Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

Scienze della terra, della vita e dell'ambiente

Ciclo XXX

Settore Concorsuale: 05/B1 - Zoologia e Antropologia

Settore Scientifico Disciplinare: BIO/08 - Antropologia

Unraveling the combined effects of demography and natural selection in shaping the genomic background of Southern Himalayan populations

Presentata da
Guido Alberto Gnecchi Ruscone

Coordinatore Dottorato
Prof. Giulio Viola

Supervisore
Dott. Marco Sazzini

Esame finale anno 2018

Table of contents

Abstract

The populations inhabiting the high-altitude Himalayan valleys and the Tibetan plateau represent an exceptional case of human adaptation to a challenging environment having evolved multifaceted physiological adjustments that allow them to cope with hypobaric hypoxia. Recently, several studies shed new light into the ancestry of populations inhabiting the northern regions of the Tibetan plateau, partially elucidating also the genetic bases of their high-altitude adaptation. Nevertheless, the polygenic nature of such an adaptive phenotype, while being previously hypothesized, has not been formally tested so far. Moreover, less attention has been devoted to the study of populations from the Southern slopes of the Himalayas and to the history of migrations, admixture and/or isolation of the many non-Tibetan trans-Himalayan Tibeto-Burman speaking populations.

In the present study, we examined genome-wide variation of previously unsurveyed Tibeto-Burman (i.e. Sherpa and Tamangs) and Indo-Aryan communities from remote Nepalese valleys along with literature data for many South/East Asian populations. Our analyses showed that most of Southern Himalayan Tibeto-Burmans derived their East Asian ancestry not from the Tibetan/Sherpa lineage, but from low-altitude ancestors who plausibly migrated across Northeast India/Myanmar, having experienced extensive admixture that reshuffled the ancestral Tibeto-Burman gene pool. These demographic inferences were also confirmed by the absence in Tamangs of the classical Tibetan/Sherpa "hard sweep" signatures at the high-altitude associated *EPAS1* and *EGLN1* genes.

Finally, by generating a 20X whole genome sequence of a Sherpa individual and by merging it with published whole genome sequence data from Sherpa and Tibetan subjects, we applied an innovative gene network-based pipeline for the detection of signatures of positive selection. This approach enabled us to identify possible signals of polygenic adaptation occurred at the level of gene subnetworks belonging to functional pathways involved in controlling angiogenesis, thus expanding the knowledge about the genetic determinants underlying the complex Tibetan/Sherpa adaptive phenotype.

1. Introduction

1.1 Molecular Anthropology and the study of human biodiversity

Molecular Anthropology, as the name suggests, is a discipline that conjugates anthropological approaches to the theoretical and methodological frameworks proper of Molecular Biology and, especially, Population Genetics/Genomics to study the natural history and evolution of modern humans (*Homo sapiens*) and their relative extinct species (e.g. *Homo neanderthalensis*) (reviewed in Destro-Bisol et al. 2010).

In the 1960s, Cavalli-Sforza and colleagues were among the first scholars who proposed that human evolution could be studied by depicting the biological diversity of extant human populations, thus planting the seeds for this research field (reviewed in Destro-Bisol et al. 2010).

Accordingly, the firsts Molecular Anthropology studies were based on the analysis of the so-called "classical markers" (i.e. proteins), such as the ABO, Rh, and Duffy antigens that determine blood groups, as well as the iso-enzymes of the erythrocytes (reviewed in Cavalli-Sforza et al. 1994).

1.1.1 Studies based on uniparental genetic markers

The development of modern, DNA-based, Molecular Anthropology approaches was then made possible by the introduction of Polymerase Chain Reaction (PCR) in the 1980s, which indeed revolutionized the whole field of Molecular Biology.

In fact, direct analysis of DNA sequences on one side allowed to increase the number of independent genetic markers that could be studied simultaneously, on the other allowed to discriminate between neutral evolving (e.g. mainly non-coding non-regulatory DNA) and potentially non-neutral evolving loci (e.g. coding or regulatory DNA). The latter category of genomic regions was essential to distinguish for the first time patterns of genetic variability caused by random allele frequency changes (i.e. genetic drift) due to specific population demographic events from those affected by the action of natural selection.

Accordingly, hereafter as "genetic marker" we will adopt the definition reported by Jobling et al. 2014, which defines it as a "region of the genome characterized by known patterns of variability and that presents two or more allelic variants whose frequencies are different in diverse human populations".

In detail, the first modern Molecular Anthropology studies focused on the direct analysis of DNA sequences were based only on a limited fraction of the human genome: the uniparentally-inherited

genetic systems (i.e. the non-recombining region of the Y-chromosome, NYR and the mitochondrial DNA, mtDNA). Markers located on these genomic regions were used to reconstruct their phylogeny and to trace the routes of migrations and admixture that characterized the dispersal of human populations all over the world. Although they represent only a small portion of the whole human genome, these sequences have the great advantage of not being subjected to recombination during meiosis. This allows researchers to reconstruct accurate phylogenies of genetic lineages that are inherited intact from the father to sons (Y-chromosome) or from the mother to the entire offspring (mtDNA), also identifying the occasional variations caused by mutations, which represent one of the main sources of genetic diversity (Fig. 1.1).

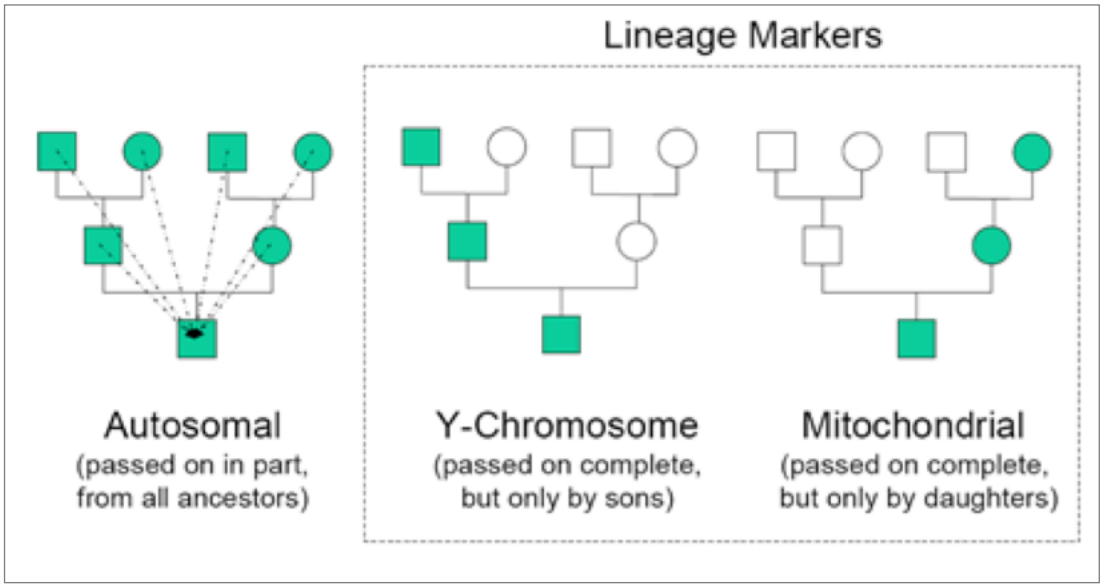


Figure 1.1. Mechanism of inheritance of autosomal and uniparental genetic markers (Butler 2005).

Studies based on uniparental markers were essential to reconstruct human evolution and to gain insights into many aspect of human history. One of the most outstanding discoveries that have been made by studying uniparental markers is that *Homo Sapiens* is actually a recent species in terms of evolutionary times, who originated ~200 thousand years ago in East Africa, and that since its spread out of Africa gene flow between neighboring groups or even more extensive admixture events between distant populations have often occurred throughout its history.

For these reasons, the overall intra-specific genetic diversity of our species is relatively low when compared to that reported for other mammals. For example, our most closely related living species, the apes, show a higher genetic diversity between different sub-groups living in different regions of the same continent than that observable between all human populations from all over the world (Gagneux et al. 1999) (Fig. 1.2).

Therefore, it is biologically impossible to identify human sub-species (or even more incorrectly human races). Instead, there is only one wide-spread human species that presents regional patterns of genetic variability, mostly due to genetic drift or isolation experienced by the different populations or by the occurrence of local adaptations (Jeong & Di Rienzo 2014; Fan et al. 2017).

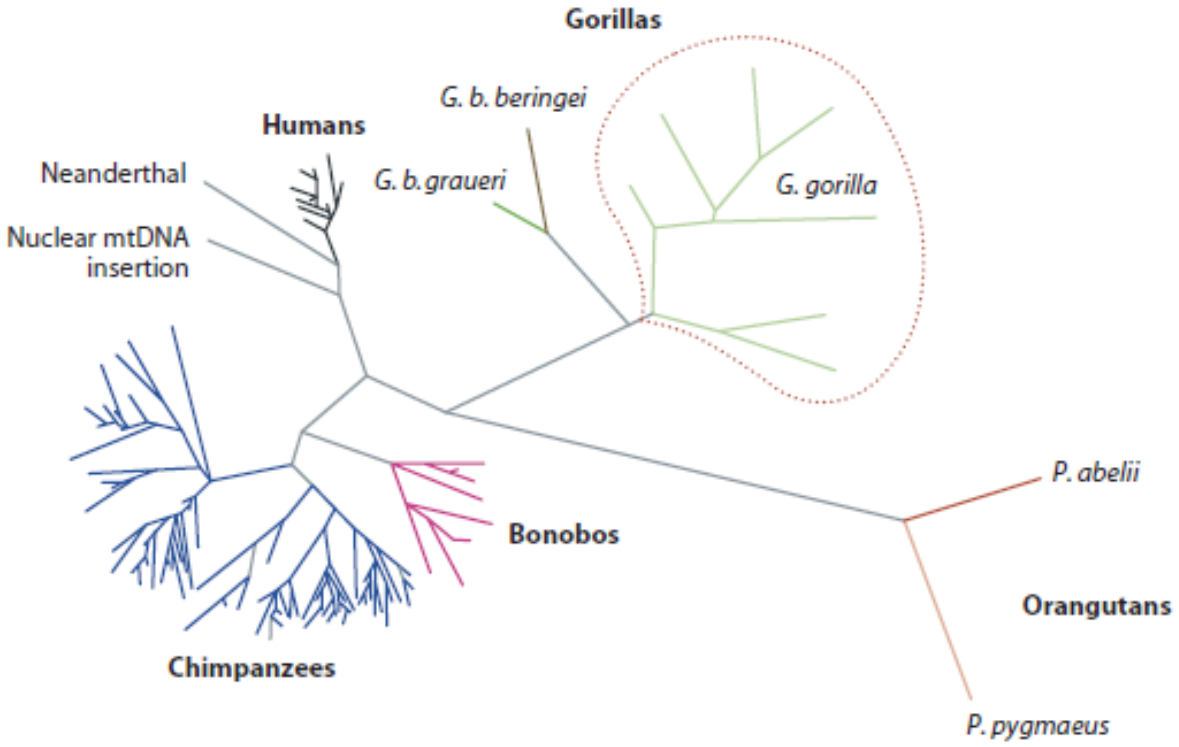


Figure 1.2. mtDNA phylogeny reconstructed from HVSI sequence data of several ape species, modern humans and Neanderthals (Gagneux et al. 1999).

A few other examples of important discoveries made through the analyses of uniparental markers are the identification of the major routes of human diffusion out of Africa, those related to the diffusion of Neolithic farmers from Levant to Europe, as well as inferences about the peopling of the Americas from the Beringia land corridor during the last glacial maximum (Jobling et al. 2014). Nowadays, the studies of uniparental markers coupled with the burst of genome-wide genotyping and whole genome sequencing technologies that took place in the last decade is providing new and unexpected fuel to the study of human evolution.

1.1.2 Studies based on autosomal genome-wide markers

Molecular Anthropology studies based on the information contained in the whole human genome (i.e. both the recombining autosomes and the X chromosome) indeed present the great advantage of

providing a more complete view of the actual genetic diversity of our species than mtDNA and Y-chromosome.

Nevertheless, this improvement came with several methodological issues caused by the reshuffling of genetic markers at every generation due to recombination, which complicates significantly the study of the past relationships between the examined variants. In fact, when considering uniparental inherited markers, the entire set of variants (i.e the entire chromosome) is passed to the next generation making it possible to reconstruct intact phylogenies keeping track of the occasional mutations originating a new lineage. Instead as regards autosomal markers, recombination breaks down the chromosome at every generation causing the phylogenies to mix together. For this reason, up to date it is not possible to reconstruct deep phylogenies with autosomal variants because it is not so straightforward to interpret if two nearby polymorphisms have been inherited together because they share a common origin or because a recombination event brought them together.

However, in the last years important developments have been achieved also in the Population Genetics/Genomics frameworks that are aimed at taking advantages of the effects of recombination. In fact, new statistical methods have been fine-tuned to enable scientists to reconstruct at a level of accuracy never reached before the demographic events that have shaped present-day patterns of human genetic variation.

A second main advantage of studying high-density genomic data from an anthropological evolutionary perspective is that it has the potential to lead to the identification of the genetic bases of human adaptation (Grossman et al. 2013). In particular, whole genome sequence data may shed light on the adaptive processes occurred according to the "polygenic adaptation" model, for which many genes simultaneously represent the targets of a given selective pressure related to specific environmental conditions, such as climate, dietary habits or pathogens (Hernandez et al. 2011).

In fact, despite the recent origin of *Homo sapiens* and its low genetic intra-specific variability, modern human populations have colonized almost all regions of the world, including several extreme environments (e.g. artic regions, rain forests, deserts, high-altitude plateaus, etc.) and experienced tremendous changes in lifestyles and diets along this way (e.g. the Neolithic agriculture shift, the industrial revolution, etc.). For this reason, it can be assumed that local adaptation to the environment has been one of the major driving forces that shaped the evolution of our species and that accounts for the phenotypic differences observable among present-day human populations. Nevertheless, the genetic determinants of most of the known adaptive traits have been not elucidated so far.

1.2 The Southern route hypothesis and the first peopling of the Asian continent

Nowadays, it is widely accepted by the scientific community that anatomically modern humans (AMH) evolved in Africa from earlier forms of Homo, and that from there they spread all over the world (Jobling et al. 2014). One of the main evidence supporting the African origin of *Homo sapiens* is the highest genetic diversity observable in African populations compared to other human groups.

Interestingly, when considering people from outside Africa, the region hosting populations with the highest genetic diversity is India, so it is likely that this area was peopled during one of the earliest human migrations "Out of Africa" (Majumder 2010). This finding, along with archaeological records and genetic evidence from both uniparental markers and whole genome sequences helped to identify a likely migration route originated in the Horn of Africa and followed by AMH around 50,00-100,000 years ago through Saudi Arabia or alternatively through Egypt and the Levant (Pagani et al. 2015). This route may have led modern humans along the coasts of India, which were presumably reached no later than 50,000 years ago (Cann, 2001), to South East Asia and beyond (Stoneking & Delfin 2010).

In fact, almost all non-African mitochondrial DNAs belong to the M and N haplogroups, which derived from the African L3 one. Some M and N lineages also showed a rapid radiation from basal nodes, many of which are restricted to "refuge" populations of the Southeast Asian coasts (e.g. the Andamanese, the Semangs in Malaysia and Negritos in the Philippines), dated to 50,000-60,000 years ago (Forster et al. 2001; Macaulay et al. 2005; Thangaraj et al. 2005).

Similarly, the distribution of Y-chromosomes C-M130 and D-M174 haplogroups suggests an ancient migration along the South coast of Asia, in line with the hypothesis of a southern route. In particular, the C-M130 haplogroup is sporadically detected in India and Southeast Asia, reaching the highest frequency in Indonesia and giving rise to sub-branches in Oceania (Underhill 2004; Mona et al. 2009). The D-M174 haplogroup instead dates back to 60,000 years ago and derives from the DE-M1 (Yap +) lineage, being found at high frequencies in Andamanese, Tibetans and Japanese, while it is only occasionally present at low frequencies in other Asian regions (Shi et al. 2008).

It is noteworthy to bear in mind that inferences about migration events based on molecular data are approximate and sometimes controversial due to uncertainty about past population sizes and mutation rates. However, the hypothesis of an early southern route followed by AMH seems to be the most reasonable one according to the phylogeny of the above-mentioned haplogroups and their distribution in extant human populations (Stoneking & Delfin 2010) (Fig. 1.3).

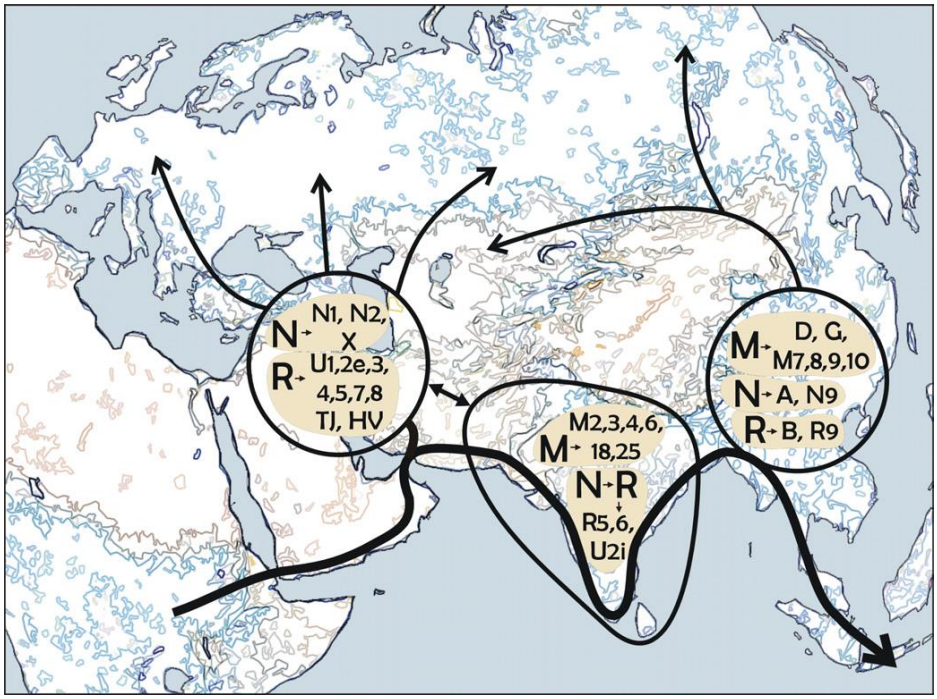


Figure 1.3. Map showing the southern route hypothesis (thicker black line) according to mtDNA haplogroup phylogeny (Metspalu et al. 2004). It is worth nothing that a recent paper based on whole genome sequence data denied the out of Arabia conduit in favor of a diffusion out of Egypt and Levant (Pagani et al. 2015).

1.3 The genetic history of South Asian populations

South Asia includes the present-day States of India, Nepal, Pakistan, Bangladesh and Sri Lanka. India is the biggest of these countries and is peopled by a great number of largely endogamous ethnic groups (Majumder 2010), which could be grouped together according to socio-cultural or linguistic characteristics.

For instance, from a socio-cultural perspective, the majority of the Indian population can be roughly divided into two groups: tribal groups and societies based on the caste system (i.e. Hindu). Tribes generally represent small independent groups of hunters-gatherers or farmers widespread throughout the Indian subcontinent, but particularly concentrated in the eastern and northeastern regions. On the contrary, the majority of the Indian population belongs to the Hindu culture, which represents for many aspects a modern globalized society and is based on an advanced agricultural system and on a written alphabet. In detail, four main castes exist in India, being ordered from higher to lower social rank: the Brahmins (priests), the Kshatriya (Warriors), the Vaisya (farmers and traders) and the Sudra (Kosambi 1964).

From a linguistic perspective, four linguistic families can be instead recognized in India. Indo-European languages are spread mainly in the North of India and Nepal, Dravidian languages are limited to the South of India, while the two remaining linguistic families are spoken exclusively by tribal groups, with the Austro-Asiatic languages being concentrated in the East of the country and the Tibeto-Burman languages being instead present in the northeastern regions and in Nepal.

Moreover, tribal groups are culturally very diverse among them. For example, all the four language families have several tribal representatives, while Hindu societies are predominantly Indo-European speaking. However, only 400 tribal groups exist in India in spite of around 4,000 caste groups (Singh 1992).

Inbreeding among different tribal groups and between tribal groups and Hindu societies is virtually absent (Majumder 2010). There is also little exchange between different castes inside the Hindu society, mainly due to cultural norms governing marriages within the caste system. However, according to genetic evidence, it has been attested that historically there have been extensive gene flow between the original Indian tribal populations and the Neolithic immigrants coming from the West. In fact, several studies reported that at a global scale the genetic structure of extant Indian populations can be modeled as a result of an admixture event between these two major gene pools (Basu et al. 2003; Cordaux et al. 2003; Cordaux et al. 2004a; Metspalu et al. 2004; Kumar et al. 2007; Reich et al. 2009; Majumder 2010; Basu et al. 2016). These studies also showed that, with the exception of Tibeto-Burmans, which contrarily to all other Indian groups present East Asian ancestries (see section 1.6), tribal groups retain most of the genetic background of the descendants of the first Paleolithic inhabitants of India.

Among them, the tribes speaking Austro-Asiatic languages exhibit the highest frequency of ancient lineages of the mtDNA M haplogroup, in particular they show the highest frequency of the M2 sub-branch (20%), which is characterized by higher HVS-I diversity (Basu et al. 2003). From a genome-wide perspective, they instead have higher proportions of an ancestral genetic component that could represent an ancient link between South Asian and East Asian gene pools (Basu et al. 2016). Accordingly, the Austro-Asiatic speaking groups have been initially proposed as the descendants of the first human groups arrived in India during the Paleolithic via a southern migration route. Studies based on Y-chromosome data seem to support this hypothesis. For instance, the O-M95 haplogroup is widespread in India and seems to have originated in the Austro-Asiatic populations about 65,000 years ago (Kumar et al. 2007). With the Neolithic transition, the spread of agriculture in India was proposed to be associated with the arrival of pro-Dravidic speaking peoples and, in later periods (about 3,500 years ago), with the arrival of Indo-European groups (Majumder 2010). However, high frequencies of ancient mtDNA haplogroups (in particular the M2 group, Kumar et al. 2008), which

are comparable to those observed in Austro-Asiatic tribes, are also found in Dravidic tribes. Therefore, it would be premature to affirm that the Austro-Asiatic groups are the only descendants of the first Indian inhabitants. In fact, ancient genetic lineages turned out to be dispersed in all Indian populations, although with higher frequencies in the Austro-Asiatic speakers.

Instead, uniparental lineages typical of Western Eurasians are present in higher proportions in the northern part of South Asia so that populations of Northern India were found to be genetically more closely related to Pakistani and Central Asian groups than to people from Southern India (Basu et al. 2003). Analyses of genome-wide autosomal markers further confirmed such a North to South gradient of genetic variation in Indian populations (Fig. 1.4, Reich et al. 2009; Basu et al. 2016). For example, Brahmins from the northern Indian provinces are found to be genetically closer to Central Asian populations than to other Brahmin groups from Southern India (Reich et al. 2009).

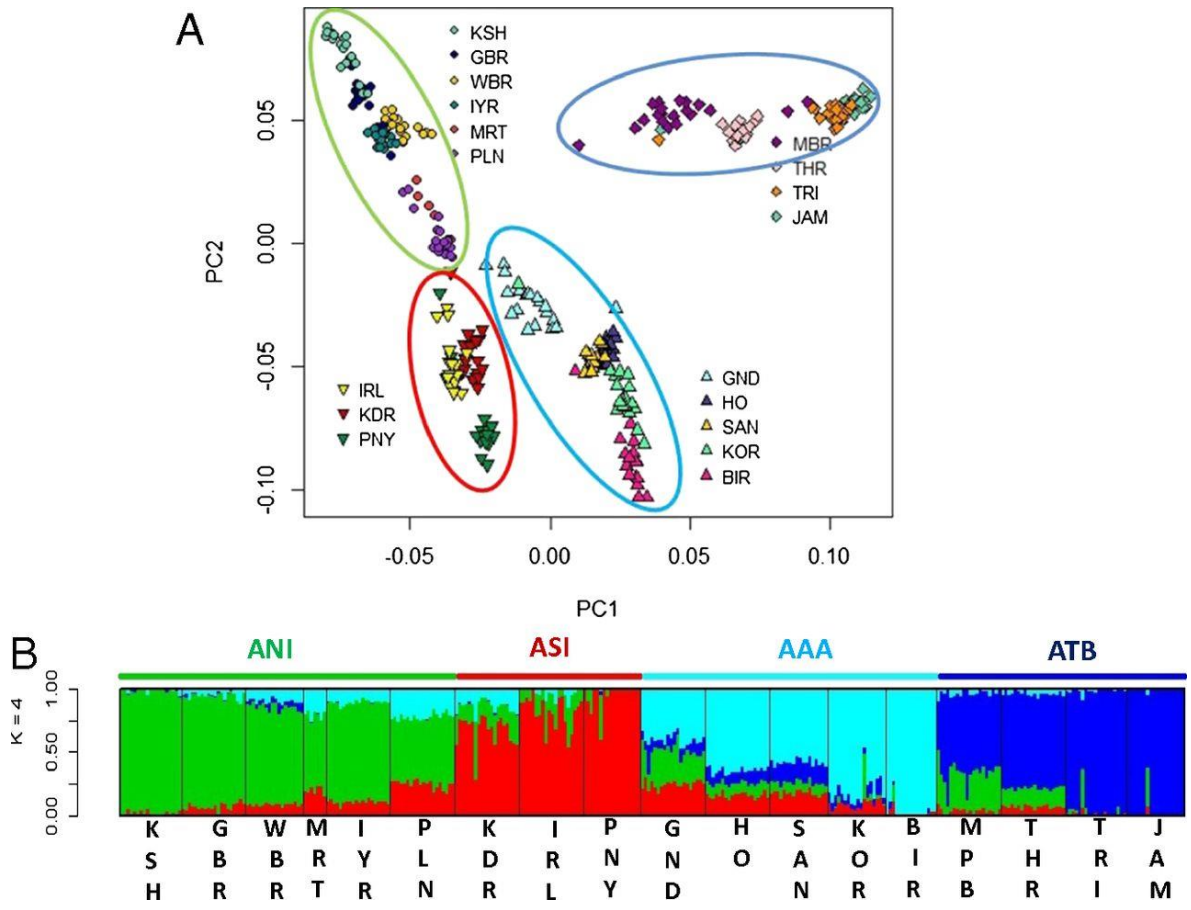


Figure 1.4. Genomic structure of Indian populations. (A) PCA based on genome-wide SNP data including 18 Indian populations and describing the four main clines and clusters of Indian genomic variation. (B) Estimates of the four main ancestry components in the same dataset of 18 Indian populations: Ancestral Northern Indian component (ANI, green), Ancestral Southern Indian component (ASI, red); Ancestral Autro-Asiatic component (AAA, cyan), Ancestral Tibeto-Burman component (ATB, blue) (Basu et al. 2016).

1.4 The peopling of East Asia: southern and northern routes

Another debate that puzzles scholars in the anthropological research field concerns the peopling of East Asia, a wide geographic area comprehending present-day China, the Indochina peninsula, Myanmar, the Korean peninsula, Mongolia, Japan and the easternmost part of Russia. In particular, two hypotheses have been formulated so far: a peopling enabled by migrations from South to North by the descendants of the first human groups that reached the Asian continent along the "southern route", as previously discussed, or by diffusions of groups coming from West Eurasia via a "northern route" (Stoneking & Delfin 2010).

In fact, according to genetic evidence, no hypothesis seems to prevail on the other. Several studies based on both uniparental and autosomal markers suggest most of extant East Asian populations to have derived from diffusions of human groups moving from South to North. Indeed, it has been observed that the most frequent haplogroups in East Asia have originated and diversified in the South part of the region and subsequently spread to the North. This is demonstrated by the presence of a South to North cline of decreasing genetic diversity in East Asian populations (Fig. 1.5, Su et al. 1999; Yao et al. 2002a; Shi et al. 2005; Wen et al. 2004a/b; Shi et al. 2008, Abdulla et al. 2009).

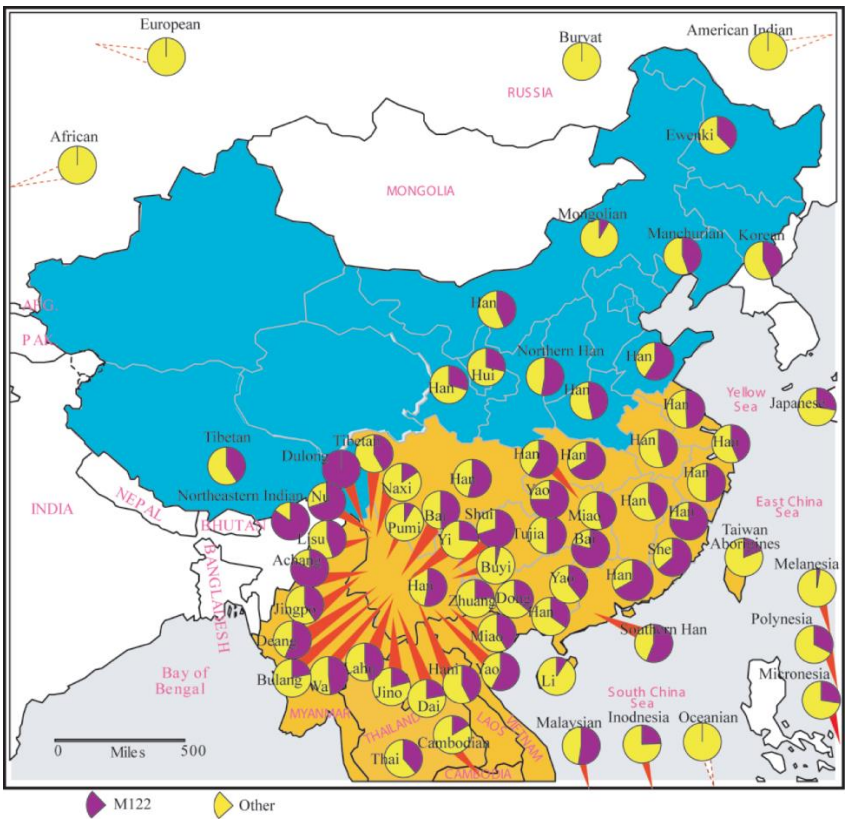


Figure 1.5. Distribution of O3-M122 Y-chromosome haplogroup frequencies in East Asian populations (Shi et al. 2005).

In particular, Su and colleagues (1999) analyzed patterns of Y-chromosome diversity and were the first to hypothesize an ancient northward diffusion of populations inhabiting the coastal areas of Southern Asia, as proven by the southern origin of the O-M175 lineage (the most frequent haplogroup in East Asia).

Also autosomal data support such a scenario by identifying a gradient of decreasing genetic diversity from South to North (The HUGO Pan-Asian SNP Consortium, Abdulla et al. 2009). Nevertheless, the analysis of the genome extracted from a 45,000 years old specimen (Ust'-Ishim) found in a Siberian cave showed that he is not more closely related to modern Andaman people, which are supposed to be the descendants of the first human groups spread along the southern route, than to present-day East Asians or Native Americans (Fu et al. 2014). Furthermore, a genome of a 24,000 years old individual from Central South Siberia (Mal'ta) resulted more closely related to Western Eurasians and ancient European hunter-gatherers than to modern East Asians (Raghavan et al. 2014).

These evidence provided by the analysis of ancient DNA thus suggest that an early peopling of northern East Asian regions must have happened at least 45,000 years ago from a group ancestral to all present-day Asians (i.e. both South and East Asians) before the diffusion and differentiation of populations along the "southern route". Unfortunately, the genetic traces of this ancient thread connecting northern West Eurasian regions to northern East Asian ones have been almost canceled by more recent events. It is worth nothing that the highest proportion of Ust'-Ishm introgression was found in present-day Tibetans (Lu et al. 2016), suggesting that these isolated Himalayan groups kept some traces of that early migration.

To sum up, both studies based on uniparental data and autosomal genetic variation show that present-day East Asian populations derive most of their genetic ancestry from the diffusion of human groups settled along the southern coast of Asia, in agreement with a southern dispersal route, occurred at least 60,000 years ago or more.

According to dating of uniparental genetic lineages, this diffusion may have taken place around 25,000-30,000 years ago. However, analyses on ancient DNA samples instead showed a more complex picture, adding to such a scenario an ancient peopling of northern East Asian regions via an unknown "northern route" occurred at least as ancient as 30,000-45,000 years ago (Fig. 1.6).

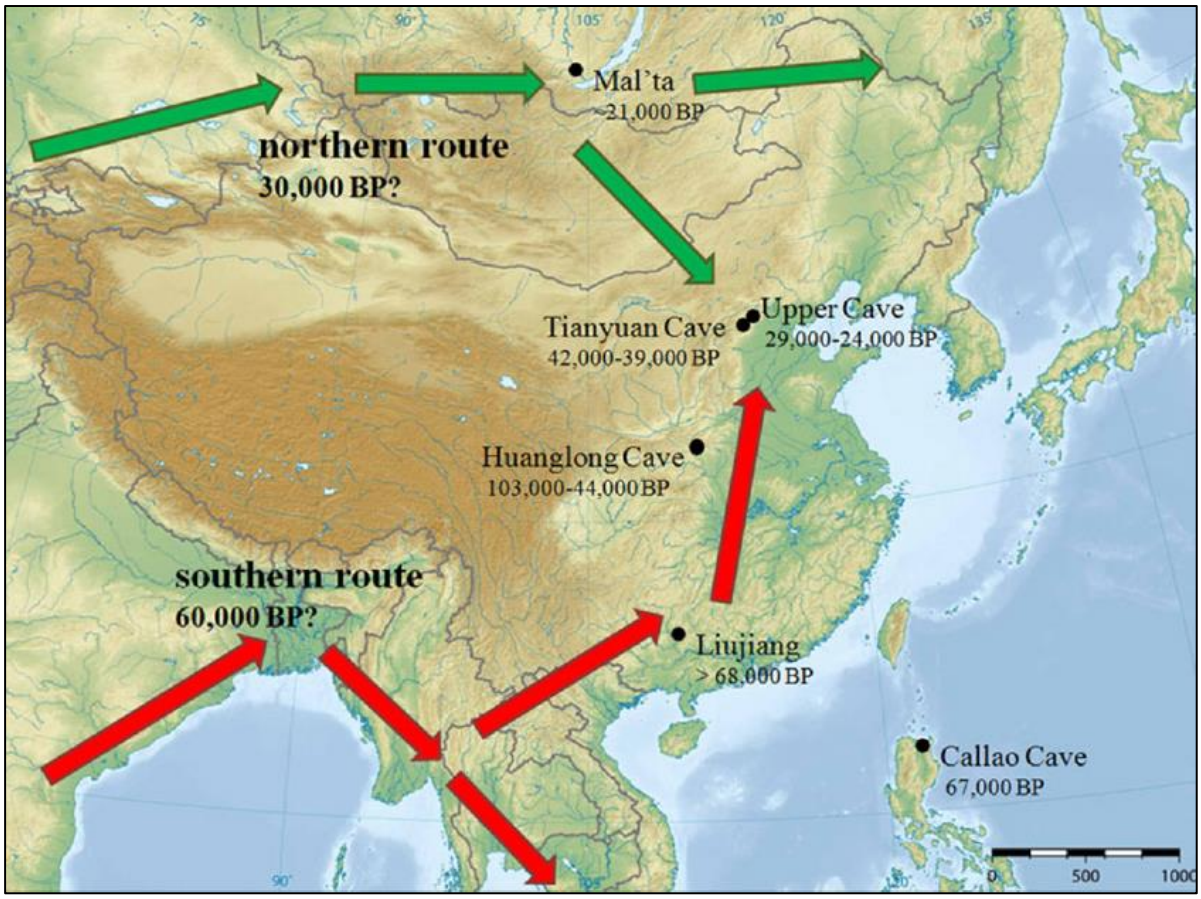


Fig. 1.6. Hypothesized northern (green) and southern (red) routes of human population diffusions in East Asia (Sanchez-Mazas et al. 2011).

1.5 Historical migrations in East Asia

Genetic traces of the first northward expansions occurred from the southern coast to the rest of East Asia have been partially masked by multiple recent demographic events. Just to mention some of the most relevant ones we can cite:

- the diffusion of the Han ethnic group (Wen et al. 2004a);
- the spread of populations speaking Tibeto-Burman languages (Su et al. 2000);
- the migrations occurred along the silk road and that introduced western genetic components in the East Asian gene pool in more recent times (Yao et al. 2004).

Among these population movements, the expansion of the Han had a great impact in reshuffling the overall patterns of genetic diversity in East Asia.

Han people are thought to descend from the Huaxia tribes originally located in Northern China. Between the 1st century and the twelfth century BC, these tribes underwent a series of southwards expansions that finally brought them to occupy all of present-day China (Ge et al. 1997).

Nowadays, Han is indeed the largest ethnic group in the world, with 1.16 billion representatives across the country.

An early study (Yao et al. 2002a) analyzed the variability of mtDNA markers in this group and suggested that the expansion of the Hans was mainly driven by a mechanism of cultural assimilation of different ethnic groups under the dominant Han culture, rather than an actual demographic diffusion. This theory is supported by the clear South to North cline of genetic diversity observable in different Han groups. In fact, Southern Hans are found to be genetically closer to non-Hans Southeast Asian ethnic groups than to the Northern Hans. On the contrary, this latter group resulted genetically closer to Tibetans than to Southern Hans. Despite the numerous migrations documented in both historical and contemporary times, this genetic differentiation between North Hans and South Hans is still appreciable (Yao et al. 2002a).

However, a later study that analyzed both the uniparental genetic markers objected against the theory of Han cultural assimilation, supporting instead the historical record of demic expansions (Wen et al. 2004a). In fact, Southern Hans and Northern Hans were found to preserve a high degree of shared Y-chromosome lineages. Therefore, this study concludes that the ancestors of Southern Hans were both the Northern Hans and the “native” populations of Southeast Asia who mixed with northern migrants. These “native” groups were as well most likely the ancestors of the many ethnic minorities living nowadays in Southern China and speaking Daic, Hmong-Mien, and Austro-Asiatic languages.

In more detail, as regards male lineages Southern and Northern Hans share similar frequencies of many Y-chromosome haplogroups, among which the most frequent are O3-M122 and its derivative O3a5-M134. The O1b-M110, O2a1-M88 and O3d-M7 lineages, which are common in the southern non-Han groups, are instead present in Southern Hans with low frequencies (average 4%), being completely absent in Northern Hans. These findings suggest that the Y-chromosome contribution of the southern "natives" to the Han gene pool has been quite limited.

Conversely, further analyses of the distribution of mitochondrial haplogroups were in agreement with early studies and revealed a substantial differentiation between Southern and Northern Hans. The Northeast Asian A, C, D, G, M8a, Y and Z haplogroups are indeed found at high frequencies in Northern Hans (55%), being instead less represented in Southern Hans (36%). On the contrary, predominant southern haplogroups (e.g. B, F, R9a, R9b and N9a) are present with high frequencies in Southern Hans (55%), but show lower occurrence in Northern Hans (33%).

In conclusion, these findings highlighted a strong sex-biased pattern in the admixture profiles of Hans. In fact, mtDNA genetic diversity decreases gradually northward, thus confirming the Southeast Asian origin of all East Asian lineages, while no cline of Y-chromosome variation is

observed. This seems to suggest that male driven migrations occurred along with the expansion of the Han civilization, partially masking the effects of the ancient northward expansions (Wen et al. 2004a).

1.6 The spread of Tibeto-Burman populations

According to the evidence described in the previous sections, present-day East Asians are thought to have derived most of their gene pool from populations who originally settled in Asia along the southern coastal regions (via a southern migration route) and subsequently spread to North East Asia. Morevoer, analyses conducted on the DNA of 45,000 and 24,000 years old ancient individuals found respectively in Mongolia and Siberia also suggested the occurrence of a more ancient peopling of East Asia via a still debated "northern route", but whose genetic traces have been now almost completely diluted in modern populations with the exception of few isolated groups (i.e. Tibetans).

Major demographic events subsequent to these early processes have then contributed to reshuffle the genetic composition of East Asian populations. As mentioned before, one of these events is represented by the spread of the Han culture, with a male-mediated gene flow from North to South East Asia.

Another key cultural and demographic event that characterized the Neolithicization of East Asia, but that is far from being completely elucidated, was the diffusion of proto-Tibeto-Burman languages and peoples.

Tibeto-Burman is one of the two sub-branches of the Sino-Tibetan language family together with the Sinitic one, which includes all major modern Chinese idioms. At least 350 different languages have been estimated to belong to the Tibeto-Burman sub-family (Martisoff 1991; Paul et al. 2016). The populations speaking these languages are scattered in several regions in both South and East Asia including Tibet, Nepal, Bhutan, India, Bangladesh, China (e.g. Qinghai provinces, Sichuan Yunnan and Hunan), Myanmar and other countries in Indochina.

In spite of the geographical and cultural heterogeneity observed among the many Tibetan-Burman groups, historical, linguistic and uniparental genetic data seem to point to a possible common origin of all these populations (Su et al. 2000; Yao et al. 2002b; Cordaux et al. 2004a; Cordaux et al. 2004b; Wen et al. 2004b; Gayden et al. 2007; Gayden et al. 2009; Gayden et al. 2013; Stoneking & Delfin 2010; Zhao et al. 2011; Kang et al. 2013; Qi et al. 2013). In particular, the roots of this common origin might trace back to the beginning of the Neolithic.

According to archaeological studies, the first Neolithic cultures appeared in East Asia around 10,000 years ago in several different locations along the Yellow River. One of these cultures, the Yang-Shao specialized in the cultivation of millet and arose in the upper and middle part of the Yellow River basin (North and Northeast China) around 8,500 years ago. Many scholars believe that the Yang-Shao is the culture in which the proto-Sino-Tibetan languages evolved (Wang 1994) (Fig. 1.7).

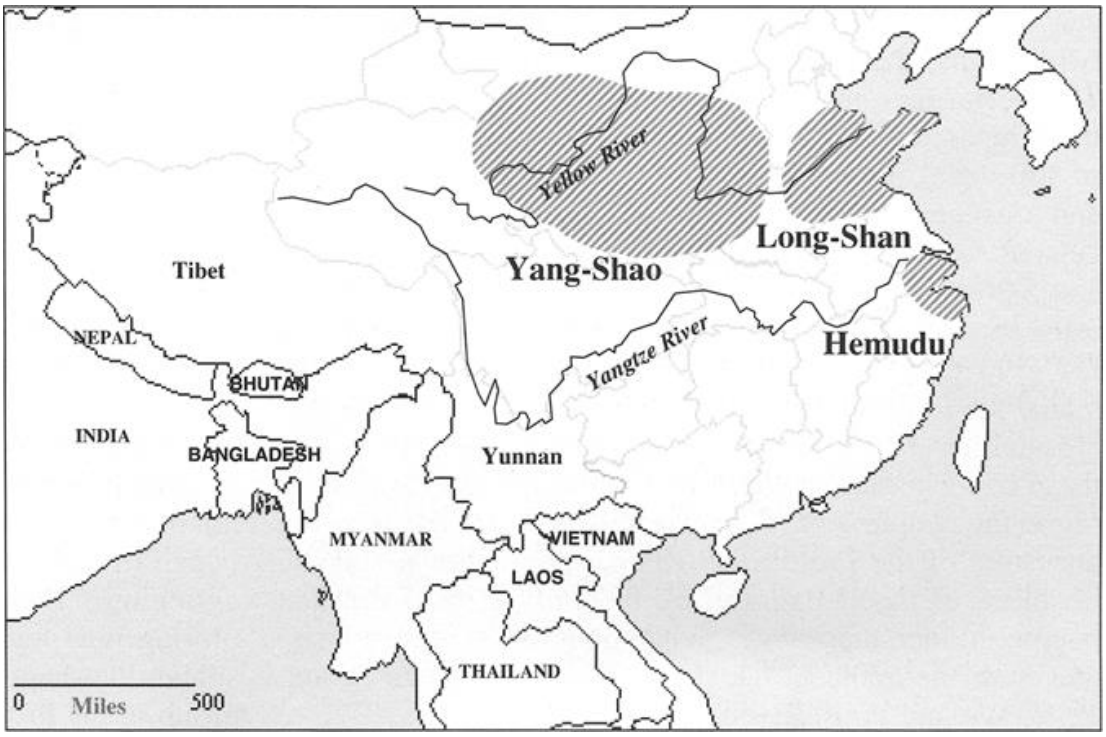


Figure 1.7. Geographic distribution of the first Neolithic cultures in East Asia (Cavalli-Sforza et al. 1994).

Plausibly during the Yang-Shao period, the Di-Qiang tribes, founders of the Ma-Jia-Yao culture, flourished. Linguistic data indicate that people speaking proto-Sino-Tibetan languages lived in the upper part of the Yellow River until around 6,000-5,000 years ago, when the divergence between the Tibeto-Burman and the Sinitic branches of the Sino-Tibetan language family probably occurred (as estimated based on a lexical data, Wang et al. 1998). In particular, historical records corroborating these linguistic findings suggest that the Di-Qiang tribes were located in the present-day Chinese provinces of Gansu, Ningxia and Qinghai and from there they spread southward. Two main migratory waves have been attested: the first one occurred around 5,000-4,000 years ago and the second one instead around 2,500-2000 years ago in conjunction with the expansion of the Qing dynasty (Wang et al. 1994).

Up to date, genetic studies conducted to shed light on the origins of Tibeto-Burman populations have been almost exclusively based on the analysis of uniparental markers. One of the very first of these studies estimated the age of divergence between Hans and Tibetans as occurred 5,900-5,200 years ago according to variability patterns of the Tibetan-specific Y-chromosome haplotypes belonging to the O3-M134 haplogroup. This represents one of the most frequent lineages in both groups, but the O3-M122 Han-specific sub-branch seems to show an older origin (20-40 thousand years ago), in agreement with the southern origin of O3-M134 (Shi et al. 2005) and of all East Asian populations. Accordingly, the O3-M134 lineage may have diversified in continental East Asia, then spreading to Tibet only after or along with the Tibeto-Burman Neolithic migrations (Su et al. 2000) (Fig. 1.8).

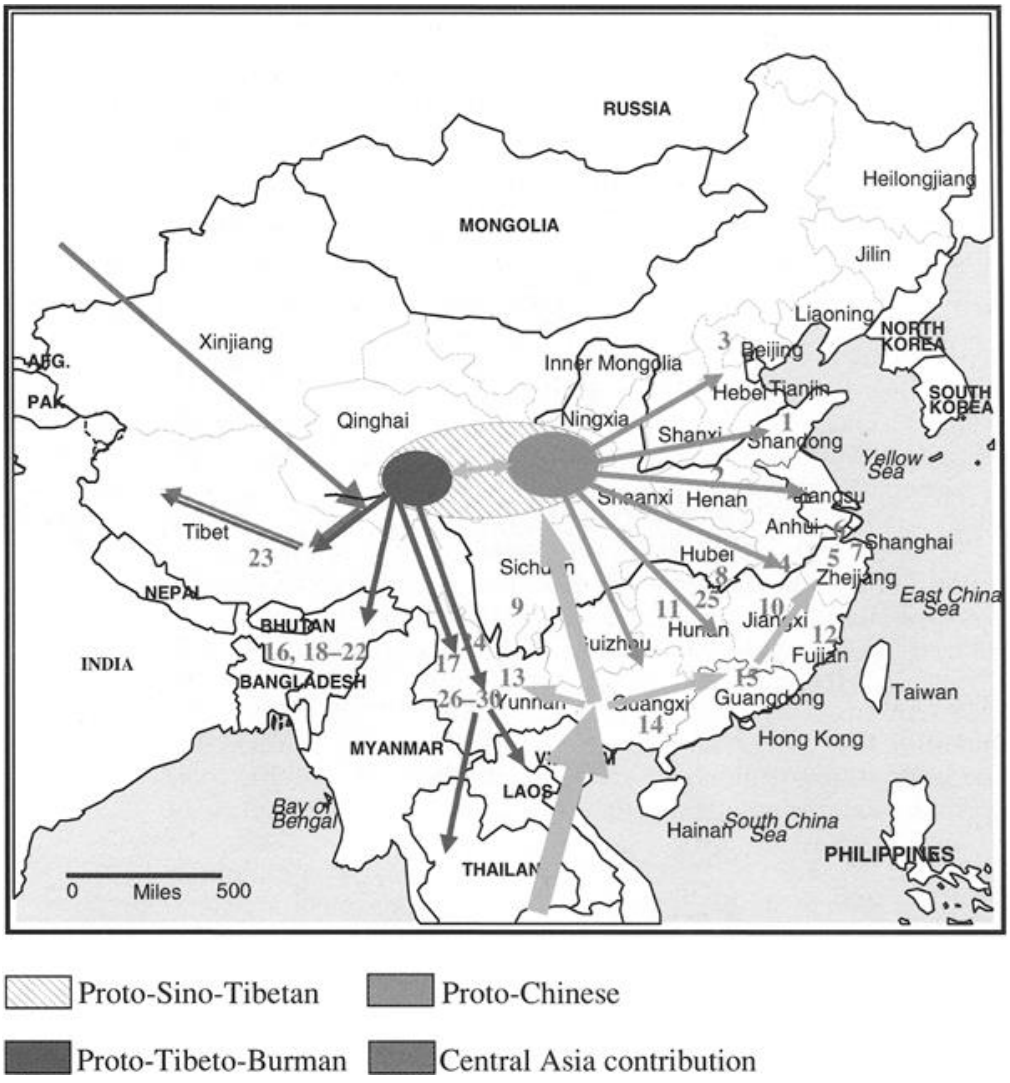


Figure 1.8. Possible diffusion routes of Neolithic Sino-Tibetan populations. The hypotized gene flow from Central Asia into the Tibetan Plateau was proved wrong by more recent studies (Su et al. 2000).

The analysis of Y-chromosome and mtDNA data generated for Tibeto-Burmans from North East India, then confirmed the common origin of these tribes, showing how they are genetically closer to East Asian populations than to other South Asian groups (Cordaux et al. 2004b). These populations thus represent the only exception of a more general genetic discontinuity between the broadly speaking South Asian and East Asian gene pools. This discontinuity is widely known and attested and would provide evidence for the fact that the Northeastern transition to India has been actually a genetic barrier limiting gene flow between the Indian subcontinent and East Asia for a period of at least 30,000 years (Cordaux et al. 2003; Cordaux et al. 2004b).

Another study based on the analysis of uniparental markers and conducted on Tibeto-Burman populations resident in the Yunnan Province of China showed that they may be the result of admixture between the ancient descendants of Tibeto-Burman migrants and "native" populations of Southern China (Wen et al. 2004b). Similarly to the case of the previously discussed Han diffusion (Wen et al., 2004a), a sex-bias pattern was inferred according to the observed admixture proportions. In particular, the paternal component of Tibeto-Burman migrants seems to have contributed with a slightly higher proportion (62%) than the maternal one (56%) to the gene pool of present-day Tibeto-Burman groups (Wen et al. 2004b). Ancient DNA studies on both mtDNAs and Y-chromosomes extracted from remains found in archaeological sites attributed to the Di-Qiang civilization further confirmed that present-day Tibeto-Burman populations plausibly originated from these ancient tribes and showed that Di-Qiang people also contributed to a fraction of the ancestry of modern Hans (Zhao et al. 2011).

More recently, a study on Y-chromosome variability conducted on the Tibeto-Burman Adi and Deng groups located east of the Himalayas, as well as on other Tibeto-Burmans, reinforced the claim of their common origin from the Di-Qiang arguing that the current Qiang ethnic group could in fact be the direct descendant of those ancient tribes (Kang et al. 2011). This study also pointed out that two possible migratory routes could be hypothesized based on the slight, but significant difference between groups distributed in the West (e.g. Tibet, Nepal and Northeast India) and those located in the East (e.g. China, Myanmar and Indochina) as regards the spatial variability of O3 derived haplotypes.

In conclusion, genetic data based on the analyses of uniparental lineages seem to support a common origin of Tibeto-Burman peoples, which are supposed to be the descendants of an ancient population who lived in the North East region of present-day China around 5,000 years ago. Nonetheless, all these studies are based exclusively on inferences drawn from uniparentally inherited markers, which provide only a limited resolution for the reconstruction of the historical processes involved in the Tibeto-Burman diffusion.

Unfortunately, up to date only a couple of studies have generated genome-wide data for Tibeto-Burman populations, especially for those residing South of the Himalayan arc, and they were not aimed at disentangling the overall genetic history of such a heterogeneous population group (Chaubey et al. 2011, Basu et al. 2016).

1.7 The peopling of the Tibetan Plateau and its adaptive implications

The Himalayan mountain range extends for 2,500 km and represents the highest cordillera in the world. It is oriented Northwest to Southeast and it touches several countries from Pakistan to Myanmar, crossing India, Nepal, Bhutan and Tibet. The Himalayas bare several of the highest peaks in the world, for example the world's highest mountain Everest (8,848 m a.s.l.) and many others over 7,000 m a.s.l. For these characteristics, it represents a natural barrier dividing the South Asian subcontinent from the Tibetan plateau.

On the southern side of the Himalayas, the valleys of Nepal are characterized by a humid and temperate climate,while on the northern side the vast Tibetan plateau is instead dominated by a cold and arid climate. This region extends for over 2,500,000 km^2 at an average height above 4,000 m a.s.l., being flanked by the Kunlun mountains on the northern border and by the Karakoram mountain range on the western border.

Despite such an environment represents a clear obstacle to human survival, the plateau is currently inhabited by over five million people. Over time, Tibetan populations have developed various technological innovations in order to survive in their harsh environment. In addition to these cultural adaptations, Tibetans display also unique hereditary physiological compensations that facilitate life in such high altitude conditions (see section 1.9), thus representing one of the very few human populations showing such a type of biological adaptation (Beall 2000; Beall 2006; Beall 2007).

1.7.1 Archeological evidence for the peopling of the Tibetan Plateau

For how long, from where and who were the people who first settled in these remote regions is still intensively debated. Historical and archeological records attest that Tibetan populations have been stably living in the plateau for at least a 4-5 millennia (Aldenderfer & Yinong 2004). Moreover, both archeological (Aldenderfer & Yinong 2004; Aldenderfer 2011; Rhode 2016) and genetic (Su et al., 2000; Zhao et al., 2009; Qin et al., 2010) evidence do not rule out the possibility of a Paleolithic

occupation of some Tibetan regions, while supporting the establishment of stable high-altitude settlements only after the Last Glacial Maximum (LGM, Aldenderfer 2011).

In more detail, archaeological record dates the first settlement of the Tibetan Plateau back to the Upper Paleolithic (40,000-30,000 years ago), as shown in Fig. 1.9. Probably, the first inhabitants of this region were hunter-gatherers who lived in relatively low altitudes areas (< 3,000 m a.s.l.), while higher settlements appeared after the arrival of farmers around 8,200-6,000 years ago (Aldenderfer & Yinong 2004; Aldenderfer 2011).

However, the archaeological sites attesting the Paleolithic occupation of the plateau present a scarce amount of rudimentary lithic materials, which cannot be dated precisely. Instead, post LGM sites, especially the Neolithic ones, are characterized by the presence of a more classifiable lithic culture.

For this reason, it has not yet been possible to clarify precisely the period in which the Paleolithic colonization of the plateau took place, as well as the origin of the populations that first inhabited those regions (Rhode 2016).

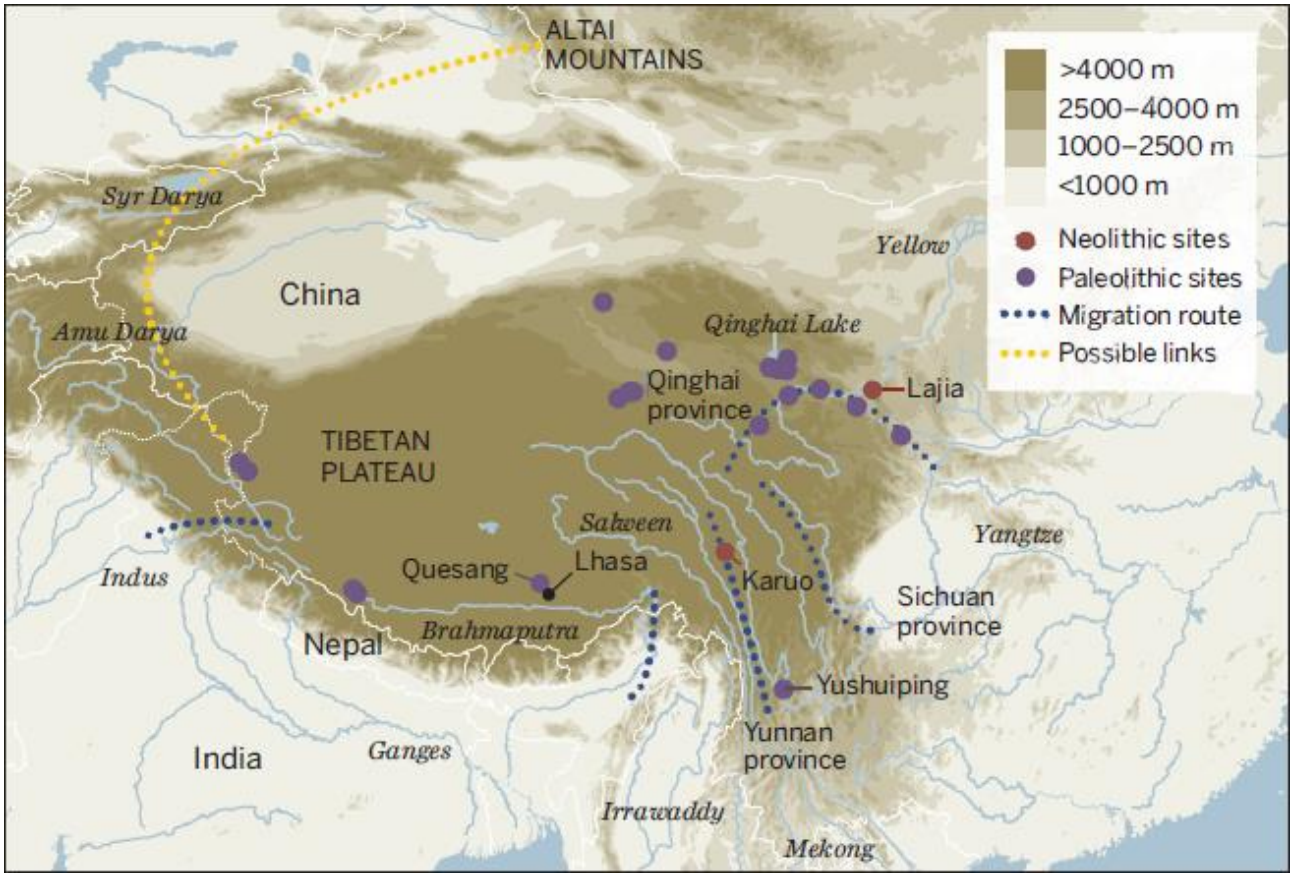


Figure. 1.9. Distribution of the Paleolithic and Neolithic archeological sites found on the Tibetan Plateau (Qiu 2015).

1.7.2 Genetic evidence for the peopling of the Tibetan Plateau

In addition to archaeological and historical studies, researches based on the analysis of the distribution of uniparental genetic lineages have pinpointed the coexistence in present-day Tibetans of both Neolithic-related and Paleolithic-related genetic components (Su et al. 2000; Zhao et al. 2009; Qin et al. 2010).

Recently, Qi and colleagues conducted an extensive survey by examining both uniparental markers and few genome-wide SNPs in a sample of 6,109 Tibetans native from 41 different geographic areas distributed on the plateau (Qi et al. 2013). Most of the Y-chromosome haplogroups observed in these populations are thus found to be of East Asian origin and among them the D-M174 lineage showed the highest frequency (54.33%), followed by O-M175 (33.47%) and N-M231 (5.65%).

In particular, sub-lineages of D-M174 have been dated to 28,000-18,000 years ago, with the exception of the most frequent lineage, D3a-P47, which was dated to around 10,000 years ago. Moreover, network analysis based on STR haplotypes of D3a-P47 showed a "star-like" structure with a "core" haplotype closely connected to several derived haplotypes. This pattern is generally assumed to be a result of strong and recent events of population expansion.

The second most common haplogroup observed in Tibetans is O3-M122 (29.82%). Interestingly, among the four East Asian O3-M122 sub-branches (i.e. O3a1, O3a2, O3a3 and O3a4) only O3a3c1-M117 is present in modern Tibetans. The estimated age for this haplogroup is 18,000 years ago, an appreciably more recent date than the 32,000 years estimated for the Han lineages. This suggests that O3a3c1 was brought in Tibet via human migrations, in accordance with the southern origin and subsequent North/East expansion of the O3-M122 haplogroup started 30,000-25,000 years ago (Shi et al. 2005). Again network analysis based on STR haplotypes of O3a3c1-M117 showed a "star-like" structure indicating a recent population expansion.

Analyses of mtDNA haplogroup distribution showed similar results to those described above as regards Y-chromosome data. Most haplogroups were indeed found to have an East Asian origin (e.g. M9a, D, A, F, G, M8 , M13, B, M62; cumulative frequency of 90.99%).

In particular, the M9a lineage turned out to be the most frequent among Tibetans (22.48%) and can be subdivided into two sub-branches, M9a1a and M9a1b1, both showing a "star-like" haplotype network and being dated 7,000 and 9,000 years, respectively.

The D haplogroup is instead the second most common one (16.53%), being followed by A (14.63%) and F (11.44%). Among the sub-branches of D, D4 is the most frequent and was dated to around 30,000 years ago, while D5 was dated to 40,000 years ago, thus confirming the presence of considerably ancient lineages in the Tibetan gene pool also as regards mtDNA.

The sub-lineage A4, which is widely spread in East Asian populations, was dated to around 15,000 years ago, while A11 was found to be Tibetan-specific and was dated to around 13,000 years ago (Zhao et al. 2009; Qin et al. 2010).

Finally, the F haplogroup was proposed to have probably originated in Southeast Asia where it presents the highest variability, being subdivided into several sub-lineages (e.g. F1*, F1-16284, F1a, F1b, F1c and F2) that were found also in Tibetan populations and for which an average age of 10,000 and 20,000 years was estimated.

Moreover, using an F_{st}-based method based on genome-wide data Qi and colleagues estimated the divergence between Tibetan and Han Chinese at around 12,000 years ago, thus in line with what obtained according to mtDNA data (Qi et al. 2013).

In conclusion, the currently available genetic and archaeological data provide evidence that support a first human settlement on the Tibetan plateau dating back to the Upper Paleolithic (about 30,000 years ago). Interestingly, paleoclimatic data seem to support such a scenario. In fact, between 40,000 and 30,000 years ago the average annual temperature on the Tibetan plateau has been estimated to be 3-4 °C higher than the actual one (Gao et al. 2008). Subsequently, during the LGM, it is reasonable to hypothesize that Paleolithic inhabitants of the plateau went through a reduction in population sizes and remained isolated in few warmer areas, but did not went extinct. After the end of the LGM, a demographic re-expansion probably occurred and after the Neolithic transition started in North East Asia, a new wave of migrants along with the newly developed agricultural technologies colonized the plateau. These new people most likely descend from the Di-Qiang tribes (as discussed in section 1.6), which have progressively admixed with the Paleolithic inhabitants of the Tibetan plateau who survived the glacial peak.

1.7.3 Adaptive implications related to the peopling of the Tibetan Plateau

Overall, understanding for how long human populations have occupied such remotes areas and which where the demographic mechanisms that brought them there (e.g. complete population replacement vs. admixture with native groups) is crucial not only for the sake of pre-historical and historical knowledge, but also to better understand the biological and evolutionary processes that have enabled humans to adapt to a new and hostile environment.

In fact, the time when natural selection began to act on the genome of Tibetan populations should coincide roughly with the time when the first human populations settled stably on the plateau. Based on this assumption, some studies have tried to estimate the age of this adaptive event, but overall obtaining discordant results. For example, one study dated the split between Tibetans and

Han Chinese to around 2,750 years ago based on exome sequencing data (Yi et al. 2010). Nevertheless, such a recent dating has been highly debated due to the strong bias introduced in the analysis by the use of exclusively coding DNA regions to infer population genetics parameters. Another study used entire sequences of the *EPAS1* gene (one of the major candidate loci supposed to be involved in the adaptation process, see section 1.9) of Tibetan individuals to estimate that the selective pressure on this gene may have started around 18,000 years ago, a period roughly corresponding to the end of the LGM (22,000-18,000 years ago) (Peng et al. 2011).

An interesting, although debated, study showed that a specific combination of allelic variants on the *EPAS1* gene is highly frequent in Tibetan populations and almost absent in all other modern human groups (Huerta-Sánchez et al. 2014). Moreover, this haplotype combination was surprisingly present in the genome of Denisovans, an archaic human species whose fossil remains were found in Siberia and dated to around 40,000 years ago (Krause et al. 2010; Reich et al. 2010). According to the authors of the study, this advantageous haplotype was introduced into the Tibetan gene pool via "adaptive introgression", that is gene flow between Tibetans and Denisovans or some unknown Denisovan-related archaic human species (Huerta-Sánchez et al. 2014). Nevertheless, the assumptions underlining this conclusion are highly questionable given how little we know about ancient human population history in East Asia and the scarce evidence available for the Denisovan species.

1.8 The Nepalese Tibeto-Burman populations

Present-day Nepal is not the only country bordering the southern slopes of the Himalayan mountain range (i.e. also Pakistan, India and Bhutan share this geographical feature), but is the one occupying a central position and extending for most of the length of the cordillera. In the previous section, we have discussed processes plausibly related to the peopling of the Tibetan Plateau, thus referring to the vast region dwelling North of the Himalayas. Nevertheless, some of the events described in section 1.7 have been not exclusively limited to the northern plateau, but to a certain extent involved also regions located South of the Himalayas.

In fact, although the Himalayas were proposed to have represented an almost insurmountable barrier to gene flow (Cordaux et al. 2004a/b; Gayden et al. 2007; Gayden et al. 2013), the presence of Tibeto-Burman populations in both Northeastern India and Nepal could indicate migrations that plausibly originated on the Tibetan Plateau and then crossed the cordillera. While most of the southern part of Nepal is inhabited by populations and ethnic groups predominantly related to the Hindu culture (see section 1.3), several Tibeto-Burman tribes reside in the northern valleys. In fact,

during the last millennia this northern region of South Asia was part of succeeding Kingdoms ruled by both Indian-related and Tibeto-Burman populations (Lewis and Shakya 1988).

One of these tribes, the Sherpas, resides in the remotest and highest valleys of North East Nepal. Their Tibetan origins have been attested according to both historical record and genetic studies. In particular, written records suggest that Sherpas first moved from Eastern to Central Tibet and then, approximately five hundred years ago, to the previously uninhabited high-altitude Nepalese valleys of Khumbu and Rolwaling (Oppitz 1974). Their close affinity to Tibetans was thus established from both uniparental and genome-wide perspectives (Jeong et al. 2014; Bhandari et al. 2015; Lu et al. 2016; Jeong et al. 2016; Zhang et al. 2017) and they are known to share with Tibetans genetic adaptation to high altitude (see section 1.9). Nevertheless, studies based on these different sets of genetic markers drew contrasting conclusions with regard to which of these groups represents the more recently derived lineage (Jeong et al. 2014; Bhandari et al. 2015).

Unfortunately, little is known about the history of other Tibeto-Burman populations residing in these regions. For example, the origins of one of the main groups inhabiting the valleys of Eastern Nepal, the Tamangs, are scarce and ambiguous (Campbell et al. 1997). Studies on uniparental genetic variation showed that they present greater Y-chromosome affinity to Tibetans lineages (Gayden et al. 2007; Gayden et al. 2009; Gayden et al. 2012) compared to mtDNA one (Gayden et al. 2013). They are thus proposed to have derived from male-biased migrations originated in a putative Tibetan homeland and subsequently followed by admixture with non-Tibetan females. Nevertheless, genome-wide data for Tamangs are still lacking and whether they share recent ancestors with the Sherpas is another question that remains to be tested from a genomic perspective. Interestingly, answering this question may provide a key for resolving the complex relationships between modern Tibeto-Burman populations.

That being so, the anthropological patchwork centered on Nepalese Sherpa and Tamang groups suggests that Tibeto-Burmans appreciably contributed to the complex network of migrations that characterized the peopling of the southern Himalayan slopes, which remain a largely understudied region (Wang et al. 2012; Jeong et al. 2016).

1.9 High altitude adaptation: the Himalayan case study

As briefly discussed in the previous sections, the Himalayan high altitude populations represent one of the few cases of recent human biological adaptation to an extreme environment. In fact, since the first settlements in such remote regions, humans have been subjected to a strong selective pressure derived from the inhospitality of this peculiar environment.

Some of the stresses imposed on Himalayan populations are common to other remote regions of the world, and the include: cold temperatures, patchy landscape, arid soil, and high UV radiation. All these features pose serious obstacles for human survival and could thus have played a role in shaping the Himalayan adaptive phenotype. However, to a certain extent they can be also overcome with technological improvements, which of course occurred and have reduced the selective pressure on genetic variants. The only stress that cannot be handled with a "cultural adaptation" and that is specific to the high altitudes, is hypobaric hypoxia, which is caused by the reduction of atmospheric pressure and consequently of oxygen partial pressure as elevation increases. For example, at an altitude of 4,000 m a.s.l., the concentration of oxygen in the atmosphere is reduced to 40% respect to the see level, thus representing a serious threat to human survival (Fig. 1.10).

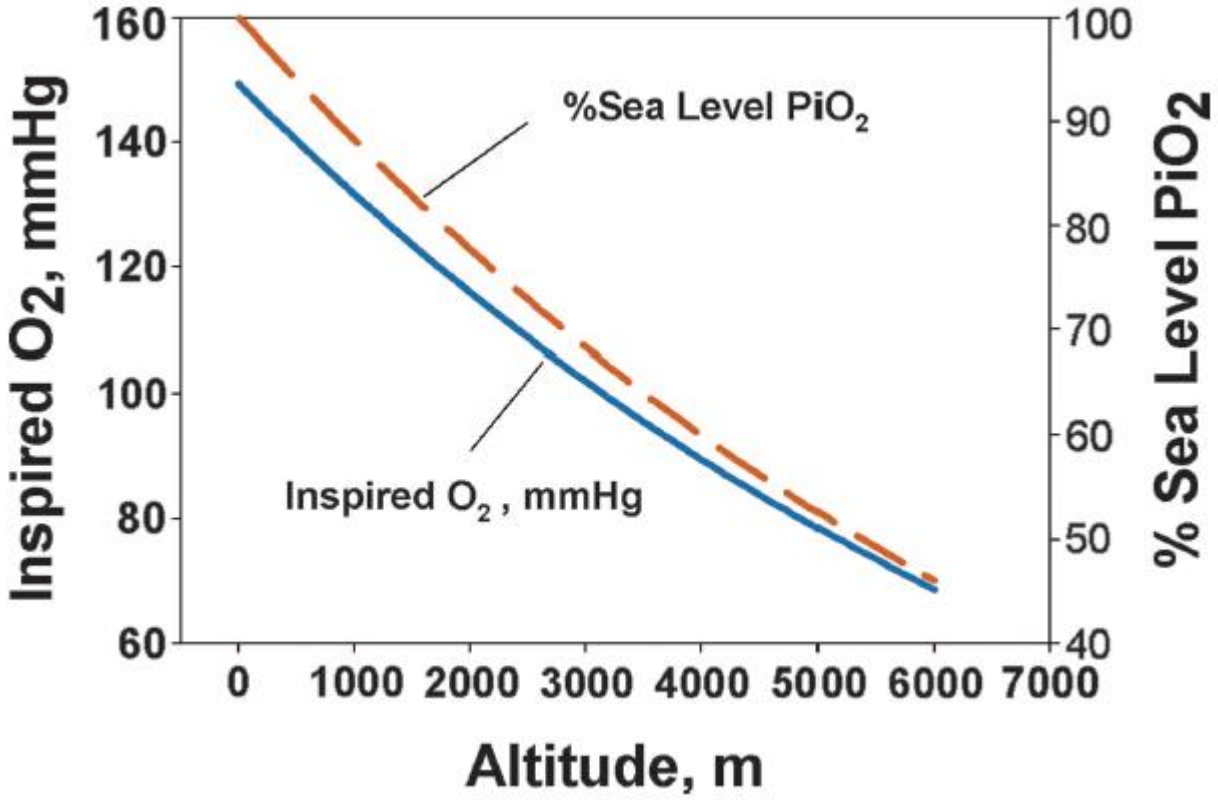


Figure 1.10. Concentration of inspired oxygen (blue line) and proportion of sea level oxygen partial pressure (red dashed line) as function of altitude expressed in meters (Beall 2007).

In fact, lowlanders travelling to high altitudes in order to compensate the reduction of oxygen availability go through several short-terms homeostatic responses which are energetically costly and in certain conditions can lead to several acute or chronic complications (Macsorley 2001). As a first response, the basal metabolic rate increases due to a regulation of heart rate, breathing and body temperature, bringing to a reduction of 20-30% in oxygen delivery (OD), which is the main cause of

acute mountain sickness. It has been observed that this metabolic response can last for 1 year of permanence in high altitude (Wu & Kayser 2006). Successively, a medium-term biological response to hypoxia consists in the increase in erythrocyte production (i.e. polycythemia) and consequently in the hemoglobin levels in bloodstream. This response enhances the amount of circulating cells carrying oxygen, which can partially balance the reduction of OD, but it also determines an increase of blood viscosity that is the main cause of insurgence of chronic mountain sickness (CMS) in lowlanders exposed for a long time to hypobaric hypoxia (Winslow & Monge 1987).

Contrarily to lowlanders, natives Himalayans, such as Tibetans and Sherpas, in extreme high altitude environments exhibit high physical performance and present physiological adaptations that have been identified and intensively studied from several decades (Beall 2013). Nevertheless, the biological parameters associated to the adaptive phenotype seem to involve mild modifications of several independent traits and it is still not clear how they work in concert to counteract the stress imposed by hypoxia (Gilbert-Kawai et al. 2014). In fact, the main process identified, which is also the most simple to be detected and measured, is not even directly related to an increase of oxygen intake (as one could expect), but instead involves a protection against the harmful effects of polycythemia (Beall 2007). In particular, Tibetans present see-level (or even lower) concentration of red blood cells, thus showing an opposite trend respect to the lowlanders residing for long periods at high altitudes (Zhuang et al. 1996). As previously mentioned, on the long run high blood viscosity caused by a high number of cells in the bloodstream is the main cause of CMS and it has been shown that native Tibetans presents a lower risk of developing such condition with respect to long-term high altitude residing lowlanders (e.g. Han Chinese inhabiting at the same altitude) (Vargas & Spielvogel 2006) . The other physiological aspects ascribable to such an adaptive trait involve several subtle shifts in biological parameters (Beall et al. 2001; Hoppeler et al. 2003; Beall et al. 2007; Erzurum et al. 2007; Vitzthum et al. 2013; Gilbert-Kawai et al. 2014), such as:

- a lower number of mitochondria in leg muscle cells, for which is still not clear what could be the benefit;
- a higher resting pulmonary ventilation, which consequentially increases oxygen absorption;
- a higher exhaled NO, which is a vasodilator and is associated to a higher blood flow in lungs;
- a denser capillary network that could enhance perfusion;
- a higher blood flow in brain during exercise;
- a higher blood flow in uterus and placenta during pregnancy, coupled with higher birth weight.

Nevertheless, it is still not clear how all these different aspects interact to balance the low arterial oxygen content observed in Tibetans and Sherpas. Recently, genome-wide studies and GWAS have identified few strong candidate genes involved in high altitude adaptation (Beall et al. 2010; Simonson et al. 2010; Yi et al. 2010; Simonson et al. 2012; Yang et al. 2017). Two of them (*EPAS1* and *EGLN1)* have been identified in almost all the different replication studies and are known to be active under hypoxic stress (Lee & Percy 2011). The identified high altitude-related variants are thus probably involved in soften the erythropoiesis cascade (although functional validation is lacking) (Fig. 1.11).

In particular, *EPAS1* encodes for the hypoxia inducible factor HIF-2α that is expressed only in some specific tissues (contrarily to the main HIF-1 gene which is a-specific) and represents the master regulator gene that becomes active in response to cellular hypoxia thus promoting the expression of several downstream genes involved in stimulating the erythropoietic cascade (Lee & Percy 2011). *EGLN1* is instead expressed as part of the main HIF-1 pathway and encodes for the PHD protein that specifically silences the HIF- α subunits, thus smoothing the erythrocyte production (Lee & Percy 2011). Tibetan specific variants on these two genes have been found to be associated with a lower hemoglobin concentration (Simonson et al. 2010; Beal et al. 2010; Yi et al. 2010; Yang et al. 2017).

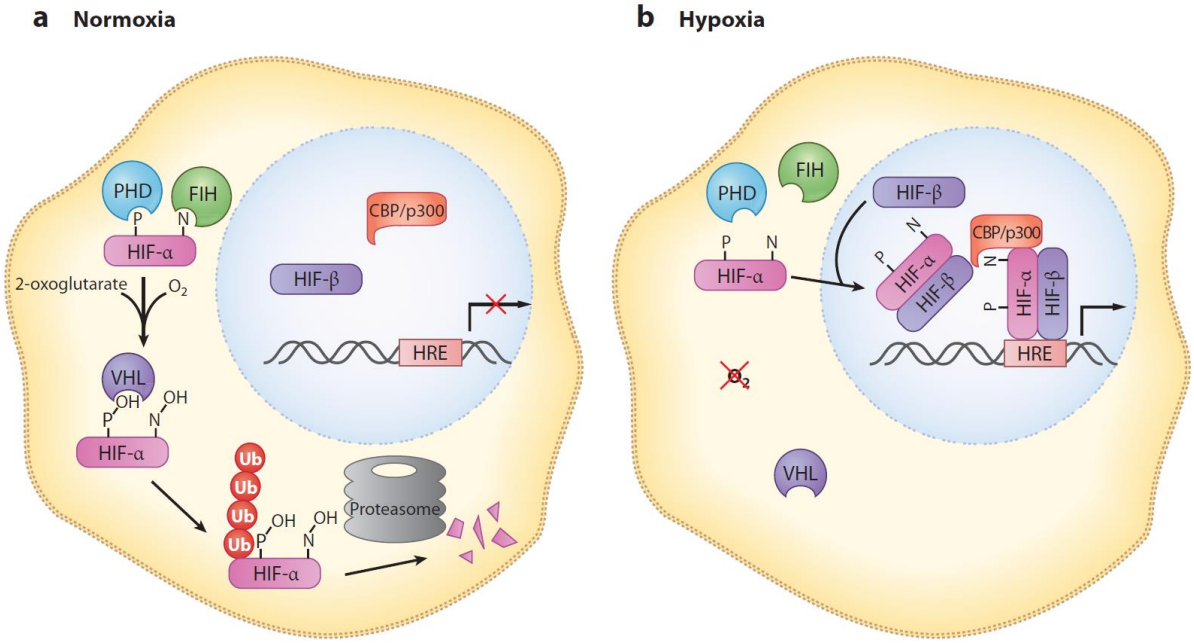


Figure 1.11. The HIF pathway under normoxia (a) and when the cell is subjected to hypoxia (b). *EPAS1* encodes for a tissue specific HIF-2α subunit, while *EGLN1* encodes for the PHD suppressor of HIF- α under normoxia (Lee & Percy 2011).

A third gene, *PPARA*, was found to be associated with a lower hemoglobin concentration in two studies (Simonson et al. 2010; Horscroft et al. 2017), but the selection signature on this gene was validated only by Simonson et al. 2010. Several others candidates have been identified so far, but most of the related studies were not able to replicate findings by using a different set of samples (e.g. from different regions or with different sizes) and statistical methods applied.

Furthermore, a mechanism that silences the erythrocyte cascade can be beneficial for CMS protection as described above, but alone cannot explain how Tibetans and Sherpas can respond to low arterial oxygen levels (Gilbert-Kawai et al. 2014). In fact, the main features of this adaptive phenotype could be controlled by several genes, each of which exerts only a little effect on the final outcome. This form of polygenic adaptation is particularly difficult to be detected with the current available statistical methods. The most serious weakness of the current approaches is that when dealing with these genomic signatures, it is extremely difficult to distinguish between those actually caused by the action of positive selection and those due to population demography and/or genetic drift (Jeong & Di Rienzo 2014).

Additionally, the adaptive phenotype itself represents a patchy composition of several biological aspects that if taken singularly can differ from one individual to another. For these reasons, it is possible that the genes involved in the adaptation may never be identified with an approach treating single genes or genetic regions independently. In the present study, we thus explored the possibility that positive selection have acted at a functional pathway (or even sub-network) level rather than on a single-gene level. This means that the genes targeted by positive selection may be many and the final outcome does not depend on one of them in particular. Accordingly, different individuals may carry a different set of beneficial variants on different genes, but all of them being involved in a wider biological function. For this purpose, we applied a new pipeline that has been developed to search for signatures of positive selection at a sub-network level (see the Materials & Methods section).

2. Aims of the study

The current study has the goal of exploring in depth the genetic structure of the southern Himalayan populations and of attempting to overcome the current conceptual and methodological limitations related to the reconstruction of the adaptive evolution of high-altitude Himalayans.

Although the performed sampling interested a restricted region, the Nepalese Gaurishankar Conservation Area, genomic characterization of the examined ethnic groups may be considered as a suitable case study to investigate demographic and biological processes representative of the population dynamics overall occurred in the vast South Asian regions bordering the southern slopes of the Himalayas. Indeed, the Gaurishankar populations are a set of distinct, but more or less related groups historically living at different altitudes and showing different degrees of isolation. In particular, at least three distinct ethnic groups reside in the considered geographical area. A broad and heterogeneous group of populations speaking Indo-Aryan languages and related to other South Asian populations lives at low altitude. The Tamangs, a Tibeto-Burman speaking group that has been never studied so far from a genomic perspective, live up to medium altitudes (< 2,500 m a.s.l.). Finally, an isolated Sherpa population related to the larger Sherpa community of the Nepalese Solukhumbu District lives at high altitude (> 3,600 m a.s.l.) in the Rolwaling valley.

Tamangs and Sherpas from the Gaurishankar Conservation Area are both Tibeto-Burman groups and represent the link between the Tibeto-Burman populations form the North of the Himalayas and those from the South of this mountain range. Accordingly, reconstructing their genetic history could be crucial to disentangle the origins of the entire Tibeto-Burman population group.

Furthermore, among the populations studied, those living at extreme altitude (i.e. Tibetans and Sherpas) represent one of the most iconic example of human adaptation to a highly challenging environment. In fact, the selective pressures imposed on high-altitude Himalayan populations have triggered a biological adaptation that allows them to cope especially with hypobaric hypoxia. While previous studies showed that Tibetans and Sherpas share (at least in part) the same adaptive traits, it is still unknown if other Tibeto-Burman populations from the same geographical area (e.g. the Tamangs) present such a biological adaptation. The present study thus moves from the belief that the outcomes of the genetic patterns driving this adaptive phenotype are strictly linked also to the demographic history of these populations rather than only to the intensity and nature of the natural selection acting on them.

To date, biological mechanisms contributing to the adaptation to high altitude have been confirmed in Tibetans and Sherpas from both physiological and genome-wide studies, the latter having recently detected signatures ascribable to strong “hard sweep” selective events occurred on two

genes involved in cellular response to hypoxia. Nevertheless, the candidate loci identified so far are found to reduce only the susceptibility of these populations to chronic mountain sickness and cannot explain alone the complex pattern of physiological adjustments displayed by Tibetans and Sherpas, which is probably multifactorial and has been thus shaped by mechanisms of polygenic adaptation.

That being so the main aims of the present study can be summarized by two consequential objectives:

a) At first, to resolve the intricate demographic history that shaped the genomic structure of human groups from the Gaurishankar Conservation Area in order to trace in a wider perspective the genetic ancestry of both low- and high-altitude Tibeto-Burman populations.

b) Then, to identify the genetic bases of Tibetan and Sherpa polygenic adaptation to high altitude by applying an innovative gene network-based approach designed to detect soft selective sweeps on both genome-wide genotype and whole genome sequence data.

3. Materials and Methods

3.1 Sampling campaigns and populations studied

Buccal swab and saliva samples analyzed in this study were collected during three scientific and humanitarian expeditions (i.e. "Earth Mater 2011", "Gaurishankar 2013", and "Extreme Malangur 2015") conducted in the Gaurishankar Conservation Area (GCA, Nepal). These expeditions were coordinated by Davide Peluzzi and Giorgio Marinelli, founding members of the ExPlora Nunaat International nonprofit organization (http://www.exploralimits.com/), in collaboration with the Laboratory of Molecular Anthropology (http://www.bioanthropologybologna.eu/) of the Department of Biological, Geological and Environmental Sciences (BiGeA) of the University of Bologna.

The GCA is located north of the Nepalese Dolakha District and borders with the Autonomous Region of Tibet (China). It is characterized by a gradual, but steep altitude increase proceeding towards the Tibetan border and culminating with the Rolwaling valley (Fig. 3.1), which owns its name from the union of the terms "*rolwa*" and "*ling*", meaning respectively "depression" and "location" in the Sherpa language. As such a name suggests, the valley is hidden between the surrounding high mountain peaks: the Melungtse II (7,023 m above sea level) on the Tibetan border and the Gaurishankar (7,184 m a.s.l.) on the Nepalese one.

The upper part of the Rolwaling valley is inhabited exclusively by a Sherpa community distributed between summer and winter villages, especially Beding (3,690 m a.s.l.) and Na (4,180 m a.s.l.) (Fig. 3.2).

In the lower part of the Rolwaling and south of it, the altitude rapidly decreases and several other villages exist (e.g. Tashinam and Simigaon), which are inhabited mostly by Tamang communities. More accessible towns are instead located at the southernmost part of the GCA, where more intercultural communities reside (e.g. in Jagat), being composed of both Tibeto-Burman and Indo-Aryan speaking groups (Fig. 3.3).

The above-mentioned expedition campaigns were focused on the collection of buccal swab and saliva samples from individuals belonging to the three main ethnic groups residing in the GCA and took place in different villages distributed along a wide altitudinal range: Jagat (1,150 m a.s.l.), Simigaon (2,000 m a.s.l.), Tashinam (2,235 m a.s.l.), Beding (3,690 m a.s.l.), and Na (4,180 m a.s.l.) (Fig. 3.1). In particular, people speaking Indo-Aryan languages (N = 23) were sampled in Jagat, Tamangs were recruited in both Simigaon (N = 26) and Tashinam (N = 11), while Sherpa samples (N = 32) were collected mainly in Beding and Na.

Each participant compiled a biodemographic questionnaire and biological samples were collected only from subjects whose grandparents belonged to the same ethnic group. Informed consent was obtained from all of them and the University of Bologna ethics committee released approval for the present study within the framework of the ERC-2011-AdG 295733 project. The research was designed and conducted in accordance with relevant guidelines and regulations and according to the ethical principles for medical research involving human subjects stated by the WMA Declaration of Helsinki.

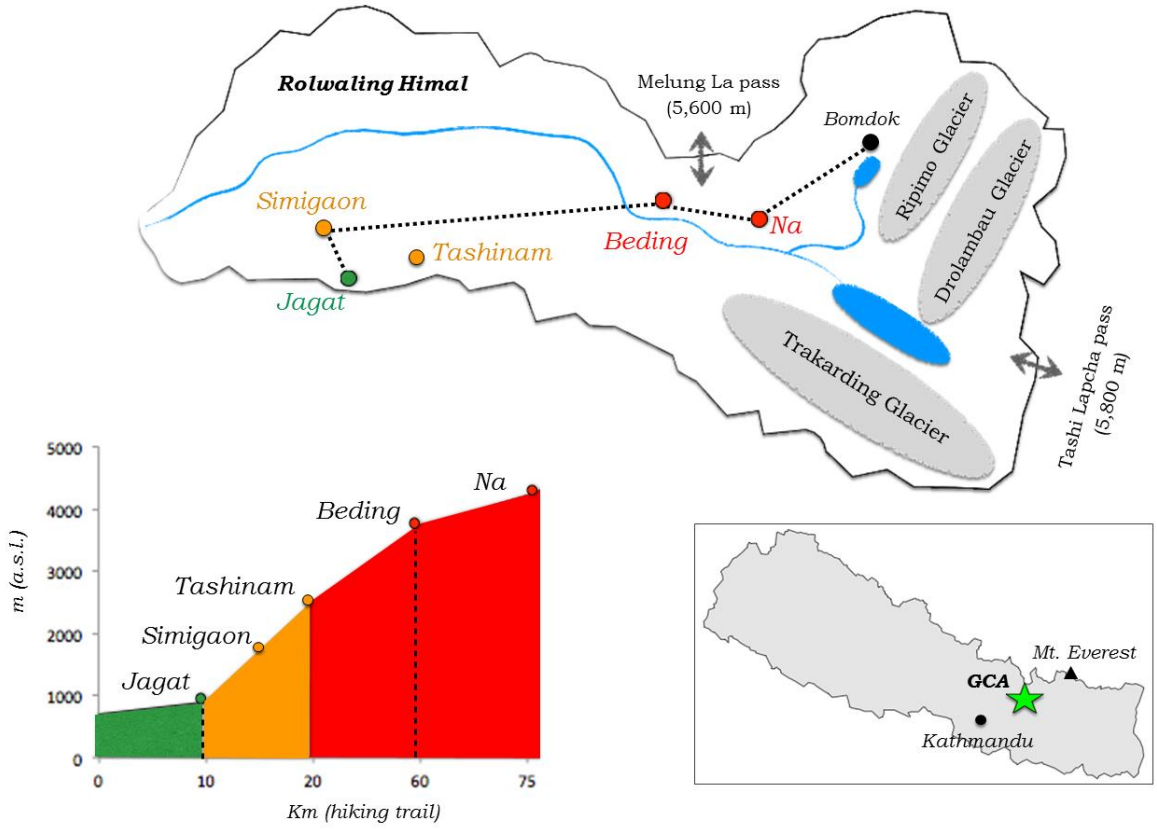


Figure 3.1. Map of the GCA and of sampling locations within and surrounding the Rolwaling Himal. On the top, the map is oriented West to East from left to right. On the bottom, the graph shows on the y-axis the elevation of the sampling locations expressed as meters above sea level (a.s.l.) and on the x-axis their distances in km of hiking trails. In the box, the location of the GCA in Nepal is specified. Bomdok is a high-altitude summer settlement at 4,900 m a.s.l. that was abandoned by the Rolwaling Sherpa community approximately 50 years ago. Both pictures are color-coded according to the ethnic groups residing in the sampling locations: green, Indo Aryan speaking groups; dark orange, Tamangs; red, Sherpas.

Figure 3.2. The Sherpa village of Beding (3,690 m a.s.l.) and a Sherpa woman with the traditional *doko* (photos by M. Sazzini).

Figure 3.3. Indo-Aryan speaking inhabitant of the village of Jagat (1,150 m a.s.l.) and Tamang woman working in the millet fields near Simigaon (2,000 m a.s.l.) (photos by M. Sazzini).

3.2 Molecular analyses

3.2.1 DNA extraction, Sanger sequencing and high-throughput genotyping

DNA was extracted from buccal swab by means of a Salting Out modified protocol and from saliva samples collected with the Oragene DNA (OG-500) kit by using the *prepIT-L2P* protocol (DNA Genotek, Ottawa, Ontario, Canada). A purifying buffer was used for lysis and protein precipitation, as well as 100% ethanol was used for DNA precipitation. The DNA was then rehydrated in 100 µl of TE (i.e. Tris-HCl 1M pH = 8.0 and 0.5M EDTA pH = 8.0) and quantified with the Quant-iT dsDNA Broad-Range Assay Kit (Invitrogen Life Technologies, Carlsbad, CA, USA).

All the 92 collected samples were sequenced at the Molecular Anthropology Lab of the University of Bologna for mtDNA hypervariable regions I and II (HVSI and HVSII) by means of Sanger sequencing and using the BigDye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific, Waltham, MA, USA). A subset of 63 male individuals were also typed at the same laboratory for 23 Y-chromosome STRs loci using the PowerPlexY23 System kit (Promega, Madison, WI, USA).

According to the results obtained by the analysis of these uniparental markers (see Results section), we then selected 75 samples representative of the Sherpa, Tamang and Indo-Aryan speaking groups, which also reached the required DNA concentration threshold of 30 ng/µl in a volume of 10 µl necessary to be genotyped with the HumanOmniExpress v 1.1 chip (Illumina, San Diego, CA, USA). This SNP-chip is designed to type simultaneously ~720,000 genome-wide SNPs distributed at an average spacing of four kb all over the genome. In particular, it includes probes for tag-SNPs, variants within 10kb of RefSeq genes, NCBI annotated non-synonymous SNPs, MHC/ADME SNPs and loci located on the sex chromosomes.

High-throughput genotyping experiments were performed at the facilities of the Human Genetics Department of the University of Chicago, by virtue of the collaboration established with Prof. Anna Di Rienzo, and at the Center for Biomedical Research & Technologies of the Italian Auxologic Institute thanks to Prof. Anna Maria Di Blasio and Dr. Davide Gentilini.

3.2.2 Whole genome library preparation and massive parallel sequencing

According to results obtained by the analysis of genome-wide SNP-chip data (see Results section), a DNA sample presenting a high DNA concentration (i.e. 48 ng/µl) and belonging to one individual from the Rolwaling Sherpa community was selected to be representative of the gene pool of such a population and to be submitted to whole genome sequencing.

We then applied a *NEBNext® Ultra™ DNA Library Prep* protocol with size selection for paired-end sequencing to construct Illumina DNA libraries for the selected sample. In detail, a preliminary DNA fragmentation step was carried out using the *NEBNext® dsDNA Fragmantase Kit* in order to obtain one to five μg of fragmented DNA. This procedure was followed by a cleaning step to cleanup enzymatic reactions and performed with the *MinElute Reaction Cleanup Kit* (QIAGEN®).
The actual library preparation protocol was finally composed of a first step of "end preparation", which consists of enzymatic blunt end repair of the DNA fragments, a second step of adapter ligation aimed at adding to the DNA fragments an Illumina unique adapter sequence, a third step of magnetic beads-based size selection in order to preserve only fragments of ~150 bp length, and a fourth step of PCR amplification that was performed by using the following cycling conditions:

- Initial denaturation at 98 °C for 30 seconds
- 15 cycles of:
    - Denaturation at 98 °C for 10 seconds
    - Annealing at 65 °C 30 for seconds
    - Extension at 72 °C for 30 for seconds
- Final Extension at 72 °C for 5 minutes

After a further cleaning step, necessary to remove waste and PCR reagents in excess, the amplified DNA fragments obtained were submitted to massive parallel sequencing with an Illumina HiSeq 4000 platform available at the facilities of the Human Genetics Department of the University of Chicago. In particular, the sequencing experiment was designed in order to be performed on a single Illumina HiSeq lane and to obtain an amount of 150 bp paired-end reads compatible with the achievement of a 20x average coverage.

3.3 Bioinformatic processing of generated raw data

3.3.1 Data curation on uniparental and genome-wide genotype data

Analyses of mtDNA and Y-chromosome haplotypes were carried out by combining the newly generated data with those publicly available for several South Asian and East Asian populations. In particular, data for 48 human groups were retrieved from literature as concerns mtDNA, while public data on 25 populations were collected to explore patterns of Y-chromosome variability (Appendix Table 1).
Nei's genetic diversity (H) was then estimated for each population included in these datasets with ARLEQUIN v.3.5.2.2 (Excoffier & Lischer 2010), while haplogroup prediction was performed

using the online HAPLOGREP2.0 (van Oven & Kayser 2009; Kloss-Brandstätter et al. 2011) and HAPLOGROUP PREDICTOR (Athey 2006) tools.

To perform quality controls (QC) on the generated genome-wide data we instead used PLINK v1.07 (Purcell et al. 2007). In particular, we retained only autosomal loci with a genotyping success rate higher than 95% and showing no significant deviations from the Hardy-Weinberg equilibrium (HWE). As regards the HWE test, we applied the Bonferroni correction for multiple testing by considering a significant threshold for p-values of 0.01 and by dividing it by the number of variants retained for the study, thus obtaining a corrected significant threshold of 1.42×10^{-8}. Moreover, one individual showing more than 5% of missing data was excluded, together with ambiguous strand SNPs characterized by A/T or G/C alleles.

The obtained high-quality dataset was then submitted to linkage disequilibrium (LD) pruning and the identified 102,980 SNPs in approximate linkage equilibrium with each other (according to a threshold of $r^2 < 0.2$) were used to estimate the degree of recent shared ancestry (IBD) for each pair of subjects by calculating genome-wide proportions of shared alleles (i.e. identity by state, IBS). According to this procedure, we removed 12 individuals showing IBD kinship coefficients higher than 0.2. It is noteworthy to mention that we used such a relaxed threshold, which in cosmopolitan populations corresponds to a value spanning between third and second degree of relativeness, since we are examining relatively isolated groups with known historical small effective population size (Jeong et al. 2014). Therefore, applying a more conservative threshold would in fact spuriously underestimate the actual degree of population inbreeding.

As a final QC step, we also applied Principal Components Analysis (PCA) on the LD-pruned dataset to further exclude outlier samples that did not cluster with the majority of subjects belonging to their self-assigned ethnic group. Accordingly, three individuals were filtered out as exceeding ± 3 standard deviations from the mean PC1 or PC2 eigenvector calculated for their respective population cluster.

After the above-mentioned QC procedures, a "GCA" dataset consisting of 59 samples successfully typed for 683,180 SNPs was generated. An "extended" dataset of 263,855 SNPs was also obtained by merging the GCA genotypes with publicly available genome-wide data retrieved from the phase3 of the 1000 Genomes Project (The 1000 Genomes Project Consortium 2015), the Human Genome Diversity Project (HGDP, Li et al. 2008), as well as many published studies focused on South Asian and/or East Asian groups (Table 3.1).

Table 3.1 Nepalese samples typed in the present study and populations included in the "extended" dataset. In bold are reported the newly generated samples of the "GCA" dataset. Acronyms correspond to: SRH = Sherpa from Rolwaling, SHT = Sherpa from Thame, SHK = Sherpa from Khumjung, TAS = Tamang from Simigaon, TAT = Tamang from Tashinam, IAR = Indo-Aryan speakers from Gaurishankar, TIB = Tibetans from Lhasa, Drav_AP = Dravidian speakers from Andhra Predesh, Drav_Kar = Dravidian speakers from Karnataka, Drav_TN = Dravidian speakers from Tamil Nadu, Gujrati_Brh = Gujarati Brahmins, IE_UTT = Indo-European speakers from Uttar Pradesh, W_Bengal_Brh = West Bengali Brahmins, N_Munda = North Munda speakers, S_Munda = South Munda speakers.

Population	Language	Country	Macro Area	Source	SNP Chip	N
Balochi	Indo-Aryan	Pakistan	SA	HGDP	Illumina 660k	21
BEB	Indo-Aryan	Bangladesh	SA	1KG	WGS	30
Birhor	Austroasiatic	India	SA	Basu 2016	Illumina 1M	9
Brahui	Dravidic	Pakistan	SA	HGDP	Illumina 660k	22
Burma	Tibeto-Burman	Myanmar	SA	Chaubey 2011	Illumina 610K	19
Burusho	Indo-Aryan	Pakistan	SA	HGDP	Illumina 660k	24
Cambodians	Austroasiatic	Cambodia	EA	HGDP	Illumina 660k	10
CDX	Tai - Kadai	China	EA	1KG	WGS	30
CEU	Indo-European	Utah (Eur origins)	Europe	1KG	WGS	30
CHB	Sinitic	China	EA	1KG	WGS	30
CHS	Sinitic	China	EA	1KG	WGS	30
Dai	Tai - Kadai	China	EA	HGDP	Illumina 660k	10
Daur	Altaic	China	EA	HGDP	Illumina 660k	7
Drav_AP	Dravidic	India	SA	Metspalu 2011	Illumina 650k	13
Drav_Kar*	Dravidic	India	SA	Behar 2010	Illumina 660k	4
Drav_Kar*	Dravidic	India	SA	Metspalu 2011	Illumina 650k	4
Drav_TN	Dravidic	India	SA	Metspalu 2011	Illumina 650k	12
GIH	Indo-Aryan	India	SA	1KG	WGS	30
Gond*	Dravidic	India	SA	Basu 2016	Illumina 1M	15
Gond*	Dravidic	India	SA	Metspalu 2011	Illumina 650k	4
Gujrati_Brh	Indo-Aryan	India	SA	Basu 2016	Illumina 1M	20
Han	Sinitic	China	EA	HGDP	Illumina 660k	43
Hezhen	Altaic	China	EA	HGDP	Illumina 660k	8
Ho	Austroasiatic	India	SA	Basu et al. 2016	Illumina 1M	18
IAR	**Indo-Aryan**	**Nepal**	**Hiamalaya**	**Present study**	**Illumina 720k**	**13**
IE_UTT	Indo-Aryan	India	SA	Metspalu 2011	Illumina 650k	85
Irula	Dravidic	India	SA	Basu 2016	Illumina 1M	18
ITU	Dravidic	India	SA	1KG	WGS	30
Iyer	Dravidic	India	SA	Basu 2016	Illumina 1M	20
Jamatia	Tibeto-Burman	India	SA	Basu 2016	Illumina 1M	18
Japanese	Altaic	Japan	EA	HGDP	Illumina 660k	27
Jarawa	Ongan	India	SA	Basu 2016	Illumina 1M	1
JPT	Altaic	Japan	EA	1KG	WGS	30
Kadar	Dravidic	India	SA	Basu 2016	Illumina 1M	19
Khasi	Austroasiatic	India	SA	Chaubey 2011	Illumina 610K	3

Khatri	Indo-Aryan	India	SA	Basu 2016	Illumina 1M	17
KHV	Austroasiatic	Vietnam	EA	1KG	WGS	30
Korwa	Austroasiatic	India	SA	Basu 2016	Illumina 1M	18
Lahu	Tibeto-Burman	China	EA	HGDP	Illumina 660k	7
Manipuri	Tibeto-Burman	India	SA	Basu 2016	Illumina 1M	19
Maratha	Indo-Aryan	India	SA	Basu 2016	Illumina 1M	7
Miaozu	Hmong-Mien	China	EA	HGDP	Illumina 660k	10
Mongola	Altaic	China	EA	HGDP	Illumina 660k	8
N_Munda	Austroasiatic	India	SA	Chaubey 2011	Illumina 610K	8
Naga	Tibeto-Burman	India	SA	Metspalu 2011	Illumina 650k	4
Naxi	Tibeto-Burman	China	EA	HGDP	Illumina 660k	6
North_Kannadi	Dravidic	India	SA	Behar 2010	Illumina 660k	3
Onge	Ongan	India	SA	Basu 2016	Illumina 1M	1
Oroqen	Altaic	China	EA	HGDP	Illumina 660k	8
Pallan	Dravidic	India	SA	Basu 2016	Illumina 1M	20
Paniya*	Dravidic	India	SA	Basu 2016	Illumina 1M	16
Paniya*	Dravidic	India	SA	Behar 2010	Illumina 660k	4
Pathan	Indo-Aryan	Pakistan	SA	HGDP	Illumina 660k	20
PJL	Indo-Aryan	Pakistan	SA	1KG	WGS	30
Pulliyar	Dravidic	India	SA	Metspalu 2011	Illumina 650k	4
S_Munda	Austroasiatic	India	SA	Chaubey 2011	Illumina 610K	12
Sakilli	Dravidic	India	SA	Behar 2010	Illumina 660k	4
Santal	Austroasiatic	India	SA	Basu 2016	Illumina 1M	20
She	Hmong-Mien	China	EA	HGDP	Illumina 660k	10
SHK	Tibeto-Burman	Nepal	Hiamalaya	Jeong 2014	Illumina 720k	21
SHT	Tibeto-Burman	Nepal	Hiamalaya	Jeong 2014	Illumina 720k	22
Sindhi	Indo-Aryan	Pakistan	SA	HGDP	Illumina 660k	22
SRH	**Tibeto-Burman**	**Nepal**	**Hiamalaya**	**Present study**	**Illumina 720k**	**30**
STU	Dravidic	India	SA	1KG	WGS	30
TAS	**Tibeto-Burman**	**Nepal**	**Hiamalaya**	**Present study**	**Illumina 720k**	**21**
TAT	**Tibeto-Burman**	**Nepal**	**Hiamalaya**	**Present study**	**Illumina 720k**	**11**
Tharu	I-A/TB	India	SA	Basu 2016	Illumina 1M	17
TIB	Tibeto-Burman	Tibet	Hiamalaya	Wang 2011	Illumina 1M	30
Tripuri	Tibeto-Burman	India	SA	Basu 2016	Illumina 1M	19
Tu	Altaic	China	EA	HGDP	Illumina 660k	9
Tujia	Sino-Tibetan	China	EA	HGDP	Illumina 660k	9
Uygur	Altaic	China	EA	HGDP	Illumina 660k	9
W_Bengal_Brh	Indo-Aryan	India	SA	Basu 2016	Illumina 1M	18
Xibo	Altaic	China	EA	HGDP	Illumina 660k	7
Yakut	Turkic	Siberia	EA	HGDP	Illumina 660k	23
Yizu	Tibeto-Burman	China	EA	HGDP	Illumina 660k	10
YRI	Niger-Congo	Nigeria	Africa	1KG	WGS	30

Genotype-based population structure and admixture analyses were then applied to 107,373 SNPs after LD-pruning ($r^2 < 0.2$) and by removing sites with a minor allele frequency (MAF) < 0.01. Such a LD-pruned dataset was also used to measure runs of homozigosity (ROH). For this purpose, the number of genomic segments in homozygosity and their mean length were calculated with PLINK v1.07 by applying default parameters settings.

On the contrary, haplotype-based population structure analyses and local ancestry inference were performed by using the unpruned "extended" dataset. In particular, haplotypes phasing was obtained by means of the statistical-based approach implemented in SHAPEIT2 v2.r790 (Delaneau et al. 2013) using default parameters and the HapMap phase3 recombination maps. This software implements a fast and accurate algorithm based on a Gibbs sampler (i.e. a variant of the Metropolis-Hastings Markov Chain Monte Carlo algorithm) for haplotype estimation and was developed to use both genotype and sequence data. This approach allows to phase individuals regardless of their degree of relatedness and, when compared to other phasing tools, it shows the great advantage that computational complexity increases linearly and not exponentially with the number of considered genetic markers.

3.3.2 Whole genome pair-end reads alignment and variant calling

Base calling and quality scoring for the sequenced whole genome library were performed by using the default Illumina softwares RTA v1.18.64.0 and CASAVA v1.8.2. In particular, the Mean Quality Score (PF) obtained for read 1 and read 2 of each pair was respectively 38.36 and 35.4 and the proportion of bases showing a quality score higher than 30 (Q30) was 92.7% and 82.2%. The sequenced reads with the respective base quality scores were then stored into two separate *fastq* files, one for reads of group 1 and the other for reads of group 2.

We mapped these reads onto the human reference sequence (hg19) using the Burrows-Wheeler Alignment Tool (BWA v.0.7; Li & Durbin 2010) by applying the BWA-MEM algorithm with the following command and parameters (below is reported a fake sample coded as "SamId" for the sake of example):

> **bwa mem -M -t** 2 **-R** "@RG\tID:SamId \tSM:SamId \tLB:SamId \tPL:ILLUMINA" hg19_all_contigs SamId_R1.fastq.gz SamId_R2.fastq.gz > SamId.sam

We then used SAMtools v1.3 (Li et al. 2009) to sort raw alignments, produce *bam* files and index them. In particular, we used bamUtil v1.0.13 to split the raw alignment file into separate files, one for each chromosome. Duplicate reads were also marked thanks to the Picard tool v1.98

(http://broadinstitute.github.io/picard/) by using the following command and parameters applied separately for each chromosome:

java -Xmx3500m -jar **picard.jar** MarkDuplicates **INPUT=** SamId _chr1.bam **OUTPUT=** SamId_rmdup_chr1.bam **METRICS_FILE=** SamId_rmdup_chr1.metric.txt **REMOVE_DUPLICATES=**false **ASSUME_SORTED=**true **MAX_FILE_HANDLES_FOR_READ_ENDS_MAP=**1000 **VALIDATION_STRINGENCY=**LENIENT **TMP_DIR=**/PATH/ > SamId_rmdup_chr1.rmdup.log

Local realignment around insertion/deletions (indels) was then performed with GATK v3.5 (DePristo et al. 2011) with the following two command lines to create target sites and to perform local realignment around the target sites, respectively:

java -Xmx4500m -jar **GenomeAnalysisTK.jar -T** RealignerTargetCreator **-R** hg19_all_contigs.fa **-I** SamId_rmdup_chr1.bam **-o** SamId_rmdup_chr1.intervals > realign1_chr1.log
java -Xmx4500m -jar **GenomeAnalysisTK.jar -T** IndelRealigner **-R** hg19_all_contigs.fa **-I** SamId_rmdup_chr1.bam -targetIntervals SamId_rmdup_chr1.intervals **-o** SamId_realigned_chr1.bam > realign2_chr1.log

GATK was also used for base quality score recalibration including the BadCigar filter and the "dbsnp_138.hg19_sorted.vcf" file to retrieve the dbSNP deposited rs codes for known variants (https://www.ncbi.nlm.nih.gov/projects/SNP/) by applying the following command line:

java -Xmx4500m -jar **GenomeAnalysisTK.jar -T** BaseRecalibrator -rf BadCigar **-R** hg19_all_contigs.fa **-I** SamId_realigned_chr1.bam **-knownSites** dbsnp_138.hg19_sorted.vcf **-o** SamId_realigned_chr1.grp > BQSR1_chr1.log

We finally used SAMtools to remove low quality pair of reads with phred-scaled mapping quality score (-q) minor than 30 and the PCR or optical duplicates previously marked with the Picard tool.
The above-mentioned alignment pipeline was applied separately also to whole genome reads of five Sherpa and six Tibetan samples presenting variable coverage levels ranging from 30x to 60x and recently published by Lu et al. (2016).
We finally called genotypes across all sites of our sequenced Sherpa sample and the 11 whole genome sequences retrieved from literature together by using the GATK UnifiedGenotyper algorithm and considering only nucleotide sites with phred-scaled quality score ≥ 30:

java -Xmx4500m **-jar GenomeAnalysisTK.jar -T** UnifiedGenotyper **-R** hg19_all_contigs.fa **-I** SamId_final_sorted.bam **-I** SamId2_final_sorted.bam **-I**… **-o** AllSamples_final_sorted.g.vcf **-stand_call_conf** 30 **--dbsnp** dbsnp_138.hg19_sorted.vcf > AllSamples_final_sorted.log

3.3.3 Data curation on whole genome sequence data

Quality controls on variants called from whole genome sequences (WGS) were performed using PLINK v1.07. In particular, we retained only single nucleotide variants (SNVs), thus removing all indel sites and obtaining a dataset made up of 9,029,495 markers. We then selected SNVs showing a genotyping success rate higher than 99% and no significant deviations from the HWE ($p > 1.5x10^{-9}$ after Bonferroni correction for multiple testing). This QC led to the generation of a high quality dataset including a total of 6,600,121 SNVs.

Two different datasets were subsequently created, a "WGS-low-density" one and "WGS-high-density" one, which were characterized by lower and higher SNV density, respectively.

The "WGS-low-density" dataset was obtained by merging WGS data with the genotypes of all the Asian populations included in the SNP-chip "extended" dataset, for a total overlap of 199,679 SNPs and by removing from the "extended" SNP-chip dataset the Sherpa individual that was also present in the WGS panel. This "WGS-low-density" dataset was used to check the consistency of the genotypes called from sequence data with those generated by genome-wide genotyping experiments conducted on the closely related populations contained in the "extended" SNP-chip dataset and to perform fine scale clustering analyses (see sections 3.4.5 and 3.4.6).

The "WGS-high-density" dataset instead consisted only of data for the 12 samples sequenced for the whole genome (i.e. five Sherpa subjects and six Tibetans from literature and the Sherpa individual sequenced in the present study), which were phased in order to reconstruct haplotypes with SHAPEIT2 v2.r790 by using default parameters and the HapMap phase3 recombination maps. Given the relatively low number of samples contained in the "WGS-high-density" dataset, whole genome sequence data for the 1000 Genomes populations (The 1000 Genomes Project Consortium 2015) were used as a reference panel of phased data in order to ensure a more robust phasing. The overlap between the 1000 Genomes and the assembled WGS data was of 4,077,599 SNVs, which is the final number of markers used for haplotypes phasing.

3.4 Population genomics analyses

3.4.1 Genotype-based population structure analyses

PCA was applied sequentially on the LD-pruned versions of both the "GCA" and "extended" SNP-chip datasets to check for the presence of potential genotypes inconsistency due to errors occurred in the merging procedure described in section 3.3.1, as well as on a subset of 75 Asian groups selected from the "extended" dataset. To compute PCA, we used the *smartpca* method implemented in the EIGENSOFT package v6.0.1 (Patterson et al. 2006).

To obtain an overall picture of ancestry proportions for each examined genome, we then applied the "extended" SNP-chip dataset the model-based clustering algorithm implemented in ADMIXTURE v1.22 (Alexander et al. 2009) by assuming K = 2 to K = 12 hypothetical ancestral populations. For this purpose, we ran fifty replicates with different random seeds for each K to monitor for convergence and only those showing the highest log-likelihood values were plotted. Concurrently, we calculated cross-validation (CV) errors for each K in order to identify the most plausible number of ancestry components characterizing the studied populations.

Once verified that the GCA samples showed no signatures of recent admixture involving ancestry sources ascribable to non-Asian populations (see the Results section), we replicated the admixture analyses on a representative subset of Asian populations by applying the same procedure to test K = 2 to K = 10 putative ancestral groups.

3.4.2 Tests aimed at providing demographic inferences

To formally test for potential admixture events occurred among the considered populations, we performed the three-population test (*f3*, Reich et al. 2009) by using the ADMIXTOOLS v3.0 package (Patterson et al. 2012). For this purpose, Z-scores were calculated via a Block Jack-knife approach to assess test significance and the identified potential source populations (i.e. those showing significant Z-scores $\leq$ -2) were submitted to further validation with ALDER v1.03 (Loh et al. 2013). This method was also used to derive the number of generations passed since the hypothesized admixture events by fitting one exponential curve (i.e. assuming a single pulse of admixture) to the data. The obtained results were finally converted in time estimates for the inferred admixture events by assuming 25 years per generation.

To test more refined demographic hypotheses (see the Results section), we used ADMIXTOOLS v3.0 to measure the shared genetic history between two populations by computing the outgroup *f3* statistics, which is less influenced than traditional approaches based on F_{ST} by the potential strong

genetic drift occurred in one of them (Raghavan et al. 2014). The same package was used also to estimate population genomic distances by computing the D-statistics (Green et al. 2010). In both cases, the Yoruba population (YRI) was used as an outgroup, being considered as a target of admixture in the *f3* test so that high positive values were interpreted as evidence for a close relationship between the two supposed source groups. According to these approaches, we tested 72 populations selected from the "extended" SNP-chip dataset to identify those most closely related to the studied GCA groups.

We further explored the phylogenetic relationships between the GCA populations, other Tibeto-Burman groups included in the "extended" SNP-chip panel, and the putative source populations pointed out by the above-mentioned analyses (see Results section), by taking into account patterns of recent admixture involving South Asian and East Asian human groups. For this purpose, we calculated a matrix of *f2* statistics between each population pair by applying the procedure implemented in MIXMAPPER v2.0 (Lipson et al. 2013) and we used it to construct a neighbor-joining tree of non-admixed Asian populations selected according to results of the *f3* test. Accordingly, the supposed admixed populations were then fitted in the scaffold tree by resolving *f2* statistics between them and each of the non-admixed groups. On the basis of the equation adopted in MIXMAPPER, it was thus possible to derive the split point of the two putative ancestral branches, the combined terminal branch length and the related mixture fraction. In particular, for each choice of the branches, the algorithm searches for the least-square solution from which it is possible to identify the combination of previous parameters presenting the smallest residual norm. Finally, to evaluate the robustness of the topology of the obtained scaffold trees and of the branch point for the test populations, we conducted 500 bootstraps replications.

Furthermore, in order to formally test whether the clade including low- and high-altitude Tibeto-Burman groups (as resulted from MIXMAPPER analysis and described in the Results section) is consistent with a single split of these two lineages from a common ancestor, we computed both *f4* and D-statistics in the form (YRI, CDX; Nagas, Tibetans/Sherpas). In detail, CDX (the Chinese Dai from the Xishuangbanna region) and the Nagas were used as proxies for low-altitude East Asian and Tibeto-Burman groups, respectively. These computations were performed by means of the functions implemented in the ADMIXTOOLS v3.0 package and by applying default parameters.

3.4.3 Haplotype-based estimates of ancestry proportions and magnitude of drift

We used CHROMOPAINTERv2 (Lawson et al. 2012) to "paint" haplotypes of single individuals as mosaics of all other individuals' haplotypes observed in the "extended" SNP-chip dataset. For each of the performed CHROMOPAINTER runs (see the Results section for a detailed description of the

tested hypothesis), we first estimated the mutation/emission and the switch rate parameters with 10 steps of the E-M algorithm on a subset of chromosomes {4,10,15,22}. We then averaged the obtained values across chromosomes, weighted by the number of markers, and then across individuals. We finally run CHROMOPAINTER on all chromosomes by using the estimated mutation/emission and switch rate parameters.

To infer individual ancestry proportions (i.e. the average proportion of the genome that each recipient population have copied from the considered donor groups), we then combined the copying length matrixes between chromosomes and we obtained an estimate of the total genomic length that each recipient individual has copied from each potential donor population. Subsequently, we performed multiple linear regression as described in Leslie et al. (2015) by applying the dedicated functions implemented in GLOBETROTTER (Hellenthal et al. 2014) and modified as described in van Dorp et al. (2015), thus allowing for "self-copy" between individuals belonging to the same group in order to account for intra-population haplotype sharing.

3.4.4 Local ancestry inference

We used the algorithm implemented in HAPMIX v1.1 (Price et al. 2009) to infer the local ancestry of Tibeto-Burman groups identified as admixed according to analyses described in section 3.4.2 and by assigning an ancestry probability score for each SNP. Chunks of contiguous ancestry in the haplotypes of the tested admixed individuals were thus identified given haploid data from two sets of reference ancestral populations. For this purpose, we selected as reference the South Asian and East Asian 1000 Genomes Project populations included in the phased "extended" SNP-chip dataset. Morevoer, we used the mean number of generations since admixture inferred with ALDER and we recovered the proportions of South Asian and East Asian admixture for individuals submitted to ADMIXTURE analysis at K = 2, finally calculating the mean proportion for each of the tested groups. We then selectively masked chunks of the genome assigned to East Asian ancestry with a probability lower than 90%.

To validate results obtained from this local ancestry inference, we performed PCA on a set of 75 Asian populations using the functions implemented in EIGENSOFT v6.0.1 to project the Tibeto-Burman East Asian ancestry-specific samples on the PCA space to overcome potential bias related to the several missing data included in the East Asian Tibeto-Burman dataset. We then replicated PCA on a subset of East Asian populations. Finally, to test how the East Asian Tibeto-Burman groups relate to the other East Asian populations, we computed outgroup *f3* statistics as described in section 3.4.2.

3.4.5 Population structure analyses based on whole genome sequence data

The global population genetic structure analyses detailed in this section were performed on the "WGS-low-density" dataset after LD-pruning of variants showing a $r^2 > 0.2$, as calculated with PLINK on sliding windows of 50 SNVs progressively shifting by five SNVs.

We then computed PCA on the subset of East Asian populations included in the "WGS-low density" dataset by using the *smartpca* method implemented in the EIGENSOFT package v6.0.1 and ADMIXTURE analyses were performed by testing K = 2 to K = 10 putative ancestral population clusters, as previously described in section 3.4.1. In particular, we ran fifty replicates with different random seeds for each K to monitor for convergence and calculated CV errors for each K to assess which was the most reliable number of clusters explaining the data. Only the run with the highest log-likelihood and the K showing the lower trend of CV-errors were finally considered.

3.4.6 Fine structure clustering analysis based on whole genome sequence data

To identify genetically homogenous population clusters at a fine scale, in order to be submitted to positive selection scans, we applied the haplotype-based methods implemented in the CHROMOPAINTER/fineSTRUCTURE pipeline.

For this purpose, we first phased the "WGS-low-density" dataset with SHAPEIT2 v2.r790 by using default parameters and the HapMap phase3 recombination maps. Then, we run CHROMOPAINTERv2 (as described in section 3.4.3) to reconstruct patterns of haplotype sharing of each individual by using all the other individuals included in the dataset as potential donors, but excluding themselves. Moreover, in the attempt to account for the differences in sample sizes between the populations typed with the genome-wide SNP-chip approach and those for which we obtained whole genome sequence data, we applied such an analysis on a subset of six randomly chosen individuals per group. This arbitrary sample size was chosen since the Tibetans were the largest group typed by WGS and was made up of exactly six individuals. We also restricted the analysis only to East Asian populations included in the "WGS-low-density" dataset by removing all Tibeto-Burman groups showing signatures of South Asian admixture according to the several analyses described in section 3.4.2.

We thus estimated the mutation/emission and recombination/switch rates using 10 steps of the Expectation maximisation (E-M) algorithm on a subset of chromosomes {4,10,15,22}. The mean values calculated across all autosomes and then across individuals, also weighted by the number of markers, were then used to run the final CHROMOPAINTER analysis on all the chromosomes by

using k = 100 to specify the number of expected chunks to define a region. Among the several CHROMOPAINTER outputs, we considered the matrix of the counts of shared haplotype chunks across all the individuals, which was summed across the 22 autosomes (i.e. "chunkcounts" matrix). This matrix was used as input for the fineSTRUCTURE version *fs2.1.1* (Lawson et al. 2012) that was run in the finestructure mode. The c-value identified by the CHROMOPAINTER algorithm was equal to 0.309719, we thus set a starting number of clusters equal to 1 and then run 1,000,000 "burn-in" iterations of MCMC, followed by another 2,000,000 iterations. The inferred clustering patterns were sampled every 10,000 runs. We finally performed 1,000,000 additional hill-climbing steps to improve the posterior probability and merge the identified clusters in a step-wise fashion.

3.5 Detection of genomic signatures of positive selection

3.5.1 Selection scans on SNP-chip genotype data

All the tests described in this section and aimed at detecting genomic signatures of positive selection were performed on a dataset obtained by merging the "GCA" data with those available from the phase3 of the 1000 Genomes Project (The 1000 Genomes Project Consortium 2015), as well as with those for Nepalese Sherpa and Tibetan individuals retrieved from Jeong et al. (2014), Jeong et al. (2016) and Jeong et al. (2017). Data from these latter studies were generated with the same HumanOmniExpress v 1.1 chip used to type the GCA samples. The final "selection SNP-chip" dataset on which we performed the analyses thus consisted of 378,174 SNPs.

We then calculated the population branch statistics (PBS, Yi et al. 2010) and the cross population extended haplotype homozygosity (XP-EHH, Sabeti et al. 2007) to detect genome-wide signatures ascribable to the action of positive selection on the studied GCA populations. These statistics represent two complementary tests based respectively on the allele frequencies differentiation and on the comparison of the extension of haplotype homozygosity between populations. Accordingly, they are particularly suited to identifying relatively ancient selective events due mostly to hard sweeps that acted just on one of the two (or three in the case of PBS) examined populations.

In particular, we used a customized Python script to calculate the PBS statistics based on the computation of three population F_{st} to measure for each SNP of the "selection SNP-chip" dataset the amount of allele frequency changes occurred after the split of two closely related populations (i.e. Sherpas from Rolwaling valley, SRH or Tamangs from Simigaon, TAS, and Han Chinese, CHB) with respect to an outgroup population of European origins (CEU, Fig. 3.4.).

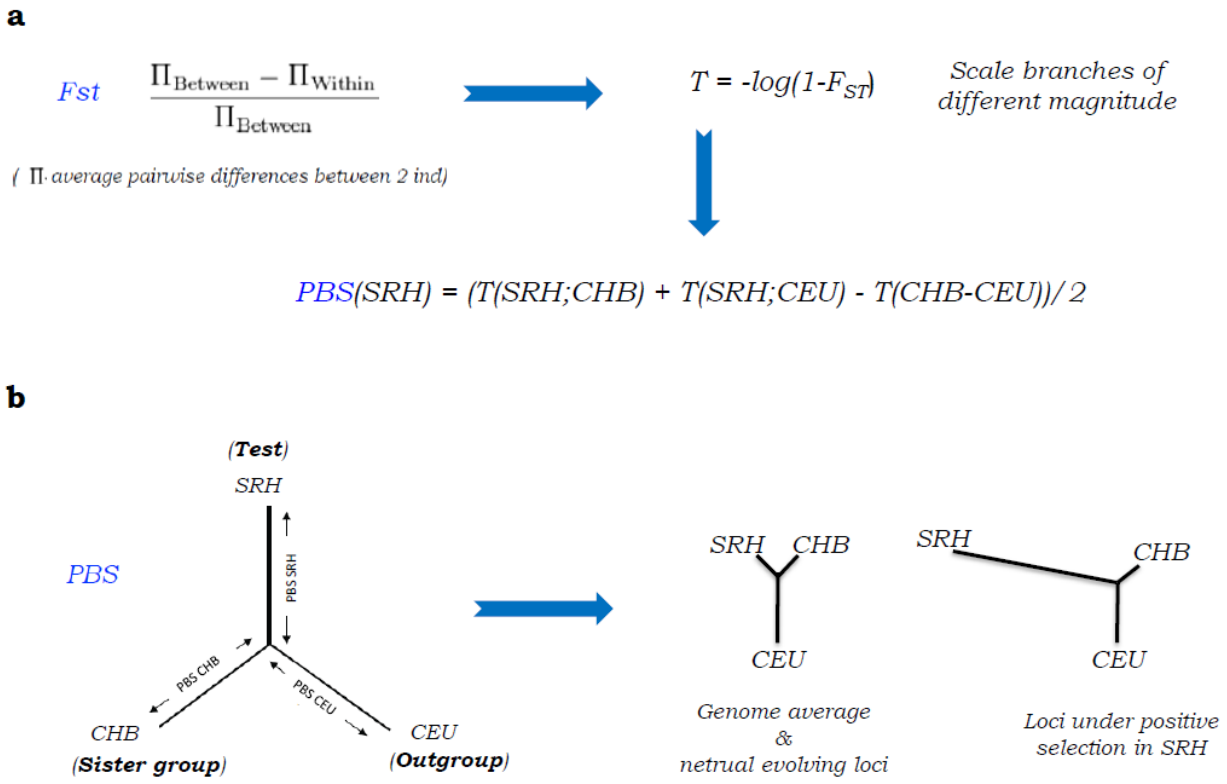


Figure 3.4. Description of the mathematical and phylogenetic characteristics of the PBS test. a) Mathematical derivation of the PBS statistics from the F_{st} calculated between the three groups of interest: SRH (test), CHB (sister group), CEU (outgroup). A logarithmic scaled F_{st} (T) is used in order to compare branches of different magnitude. b) Graphical schematic representation of the phylogenetic concepts behind the PBS test. High values of PBS indicate deviation from neutral evolution (modified from Jobling et al. 2014).

The two-population XP-EHH test was computed with the algorithms implemented in *selscan* v1.1.0b (Szpiech & Hernandez 2014) and was instead used to detect high frequency alleles associated to long-range haplotypes in Sherpa or Tamang genomes, but not in Chinese ones. This test, as others (e.g. iHS), is based on the calculation of the extended haplotype homozygosity statistics (EHH), which represents the decay of homozygosity moving away from a focal SNP (Fig. 3.5). Specifically, the iHHs, derived from the integration of its EHH, was first calculated separately for each population and for each marker. Given iHH_A and iHH_B as the iHHs of the two populations tested (e.g. A = SRH; B = CHB) then the unstandardized XP-EHH is defined as:

$$\ln\left(\frac{iHH_A}{iHH_B}\right)$$

This step was followed by genome-wide normalization so that we obtained the final XP-EHH score for each SNP as described below:

$$XPEHH = \frac{\ln\left(\frac{iHH_A}{iHH_B}\right) - E\left[\ln\left(\frac{iHH_A}{iHH_B}\right)\right]}{SD\left[\ln\left(\frac{iHH_A}{iHH_B}\right)\right]}$$

This statistical approach assumes that long haplotypes at high frequency are indicative of a fast (i.e. occurred over few generations) changes in the frequency of the core allele since recombination did not have enough time to break down its association with linked alleles. In order to perform such a computation, we retrieved information about the ancestral state of each variant from the dbSNP database (http://www.ncbi.nlm.nih.gov/SNP/) and we phased haplotypes with the statistical-based approach implemented in *SHAPEIT2* v2.r790 (Delaneau et al. 2010) by using default parameters and the HapMap phase3 recombination maps.

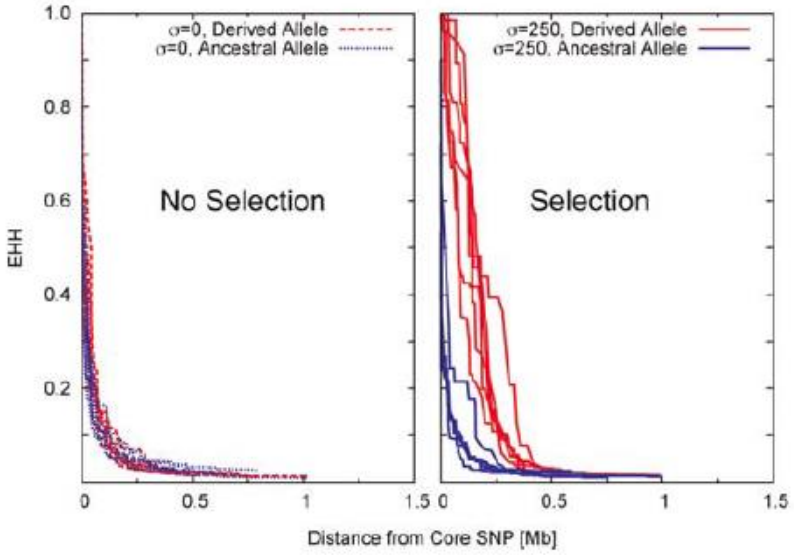


Figure 3.5. Plots showing the decay of EHH as function of the distance in Mb from the core SNP. The left panel shows a typical trend in case of no selection acting on the core SNP, while the right panel shows the deviation of EHH decay in case of selection on the derived allele (in red). The area under a curve represents the iHH (Jobling et al. 2014).

Since PBS and XP-EHH provide independent evidence of putative selected alleles having risen to or near fixation in the tested population, we combined the two statistics to reduce false positives rate by calculating a F_{cs} Fisher combined score, as described in Deschamps et al. (2016):

$$F_{cs} = -2\,[\ln(p_{psb}) + \ln(p_{xp\text{-}ehh})]$$

Where p_x are the rank p-values of the two statistics (i.e. PBS and XP-EHH) defined as the genomic rank of that statistics divided by the total number of unique values present in the related distribution

(i.e. one for each SNP presenting a value for both statistics). The F_{cs} scores obtained for each variant were then used as input to performed gene network analyses (see section 3.5.3).

As a parallel approach to the network analyses mentioned above, and to pinpoint the most plausible genomic regions targeted by selection, we divided the genome into 100 kb windows sliding by 10 kb. We then discarded windows presenting less than 10 SNPs and we grouped the remaining ones into bins according to their number of variants and with a 10 SNPs increase per bin. For each window, we also calculated the fraction of outlier SNPs (i.e. those scoring in the top 1% of overall F_{cs} distribution) with respect to the total number of SNPs in the window. Finally, for each bin, we ranked windows according to such a fraction and we considered the top 1% ones as candidate regions likely subjected to positive selection.

3.5.2 Selection scans on whole genome sequence data

As regards the previously reported approach based on low density SNP-chip data, we selected two complementary tests and merged their results in order to refine the detection of highly significant selective events that may have mainly characterized the genomic background of high altitude Himalayan groups compared to that of relatively closely related low altitude East Asians. On the contrary, the tests detailed below were chosen to take advantage of the higher density of whole genome sequence data in order to search for smoother signatures of positive selection ascribable also to soft sweeps. For this purpose, we computed two independent intra population statistics on the phased "WGS-high-density" dataset (see section 3.3.3).

3.5.2.1 nSL computation

We first calculated the nSL statistics, which stands for "number of segregating sites by length". This statistics is conceptually similar to the EHH-based iHS, but outperforms it especially as regards the fact that it does not require recombination maps (Ferrer-Admetlla et al. 2014). In fact, although human genome-wide recombination maps are publicly available, they still represent a rough approximation of the actual human recombination rates, which can instead vary between different populations (O'Reilly et al. 2008). Moreover, in the case of newly generated whole genome sequences, the recombination maps may not cover the entire spectrum of the newly genotyped loci so that the recombination rate between missing variants have to be interpolated, thus introducing potential additional bias in downstream analyses.

In addition to these considerations, we decided to compute the nSL because it provides complementary information with respect to the other tests applied on SNP-chip (see section 3.5.1)

or WGS data (see section 3.5.2.2). In particular, it was designed to refine the detection of both hard and soft sweeps. Specifically, it searches for intra-population patterns of haplotype homozigosity as a consequence of the action of relatively recent positive selection, respectively on newly arisen or already existent genetic variants (Ferrer-Admetlla et al. 2014). In fact, the so-called soft sweeps can be driven by two different events: positive selection acting on a pre-existing neutral variant (i.e. standing variation), or on a newly arisen mutation that however is not the only beneficial variant. In the former case, the neutral allele has somehow become beneficial (e.g. due to environmental changes, dietary shifts, population migration into a new environment, etc.) and thus spread rapidly in the population, but insisting on a previously existing genomic background. In the latter case, since the new beneficial variant is not the sole to shape the adaptive phenotype, the selective pressure is relaxed and the selected locus thus slowly increases in frequency over generations. Both of these soft sweeps cause a similar genomic pattern around the selected variant, which appears as intermediate between those due to a hard sweep or to neutral evolution. These patterns are therefore very difficult to be detected with methods based on extreme frequency differentiation (e.g. PBS) or aimed at identifying differences between populations (e.g. PBS and XP-EHH).

The nSL score for each variant was computed with the algorithms implemented in *selscan* v1.1.0b (Szpiech & Hernandez 2014). For each core SNV (p_k), its haplotypes were divided in two groups, those associated with the derived allele ($D(k)$) and those associated with the ancestral one ($A(k)$). For any two haplotypes i and j in each group, the length of the longest consecutive stretch of segregating sites that are identical in both haplotypes was calculated ($hap_{max}(p_k)$). All these lengths were subsequently summed and divided by the squared dimension of the group. Finally, the nSL statistics is defined as the logarithm of the value obtained by dividing the "ancestral allele computation" by the "derived allele computation" for each segregating site:

$$nSL(k) = \ln \frac{\frac{2\sum_{i<j \in A(k)} hap_{max}(p_k)}{n_A(k)(n_A(k)-1)}}{\frac{2\sum_{i<j \in D(k)} hap_{max}(p_k)}{n_D(k)(n_D(k)-1)}}$$

To perform such a computation, we set the maximum extension parameter to 4,500 SNVs, meaning that for each variant it was considered a window of 4,500 consecutive loci, or less, to calculate the nSL value. Finally, we obtained an nSL score for the each variant and we used them as input for the gene network-based approach described in section 3.5.3.

3.5.2.2 DIND computation

The derived intra-allelic nucleotide diversity (DIND) statistics was used as an independent intra-population approach that, similarly to the nSL, relies on haplotype data and applies the same logic of grouping separately the haplotypes carrying the derived and the ancestral allele. Nevertheless, contrarily to the nSL, the DIND does not use a logarithmic scale, but instead searches for the longest identical stretch of consecutive alleles by counting the number of differences between any two haplotypes belonging to the same group (Fagny et al. 2014):

$$DIND(k) = \frac{\frac{\sum_{i<j \in A(k)} hap_{diff}(p_k)}{n_A(k)^2}}{\frac{\sum_{i<j \in D(k)} hap_{diff}(p_k)}{n_D(k)^2}}$$

Moreover, the conceptual difference between DIND and nSL is that this latter test is designed to detect long common segments between haplotypes and thus, by balancing the two parameters (i.e. frequency and length of homozygosity) it can recognize genomic signatures caused by both hard and soft selective sweeps. On the contrary, the DIND focuses on the overall differences between haplotypes carrying a derived allele and it is thus more suited to recognize hard selective sweeps. We therefore chose to apply this statistics in order to rely on an independent test for hard sweeps that, with respect to many other hard sweep detectors available, has been proven to be highly robust to population demography and, especially, to variation in sequencing coverage and to low sample sizes (Fagny et al. 2014). According to these DIND characteristics and to the low number of individuals examined in the present study, we thus believed it was the most suitable test to be applied on our WGS data.

After removing variants showing a derived allele frequency (DAF) lower than 0.2, which was shown to bias DIND results (Fagny et al. 2014), we calculated DIND scores for each SNV and we used them as input for the gene network-based approach described below.

3.5.3 Gene network analyses

In order to identify possible genomic signatures ascribable to selective events occurred according to the polygenic adaptation model, which could be hardly identified by considering each locus independently, we applied the gene network-based approach implemented in the *signet* R package (Gouy et al. 2017). We performed these analyses on the entire panel of variants for which we calculated the previously described positive selection statistics. Therefore, we applied the following

pipeline separately on the two lists of F_{cs} scores obtained for the "selection SNP-chip" dataset (one for Sherpas from Rolwaling and the other for Tamangs from Simigaon), as well as on the final nSL and DIND scores calculated for the "WGS-high-density" dataset.

For each variant present in these panels, we retrieved the information of the gene or the genes located within a range of 50 kb upstream and downstream its physical position. Then, for each of the considered genes, we selected the highest score (for each statistics) among those computed for all the SNPs/SNVs associated to it as the representative score for the gene of interest. This information was finally used as input for the *signet* pipeline, which takes into consideration information available for annotated biological pathways to assign genes into their network context.

In particular, we selected the National Cancer Institute (NCI) Nature Pathway Interaction Database (Schaefer et al. 2009) to reconstruct the functional pathways our input genes belong to. We then assessed the distribution of scores within such pathways in order to test whether they are significantly shifted towards extreme values. When compared to other network analyses, the key step of the *signet* pipeline is that instead of searching for pathways where all the genes show a shift in the distribution, it focuses on the identification of outlier subnetworks of genes belonging to a wider pathway. This caveat is important since biological pathways general involve several tens of genes and it is unlikely that all the genes of a given pathway have been targeted by natural selection. On the contrary, selection acting only on a subset of genes involved in the same biological function is highly plausible. The *signet* pipeline has been developed to rely on an iterative process that we set to 10,000, in which the first step is represented by the assignment of a score to every subnetwork, combining the scores of each gene involved and normalizing them in order to allow for comparison of subnetworks of different sizes:

$$\mathbf{s_k = \sum(x_i) / \sqrt{(k)}}$$

Where s_k is the normalized combined score of a subnetwork of size k, and x_i is the score of the i-th gene.

Then, for each gene network a simulating annealing algorithm (Kirkpatrick et al. 1983) was used to identify the highest scoring subnetwork (HSS). The algorithm starts by taking a random subnetwork and modifies it progressively by adding or removing one gene at the time, and by calculating a new normalized score and repeating the process until it finds the subnetwork with the highest normalized score.

Finally, in order to assign a significance score (p-value), to test whether the identified HSS is larger than what is expected by chance, we generated null distributions of HSS for each subnetwork of a specific size. The gene scores belonging to a network were thus permuted in order to produce gene

networks with random scores and we repeated this process N times to obtain the final null distribution. In detail ,we set this iteration to 50,000 for F_{cs} scores and 20,000 for nSL and DIND (for computational complexity reasons). We then obtained the rank p-values for every observed HSS by comparing them with the HSS generated by the null distribution.

The significant subnetworks identified ($p < 0.05$) were finally plotted with Cytoscape v3.6.0 (Shannon et al. 2003).

4. Results

After applying stringent quality control filtering (see the Materials and Methods section), we generated a "GCA" dataset including 59 samples typed for 683,180 SNPs, which was used to investigate patterns of population structure and genetic diversity observable within the GCA. By merging it with publicly available genome-wide data (Table 3.1), we then assembled an "extended" dataset entailing 263,855 SNPs that was used to infer the genomic relationships of GCA groups with 72 South Asian and East Asian populations (see section 4.1).

We also merged this "extended" dataset with the genotypes called from the generated whole genome sequence thus producing a "WGS-low-density" dataset on which we performed global ancestry and haplotype-based clustering analyses (see section 4.1). Finally, selection scans were performed on the "selection SNP-chip" dataset and on the "WGS-high-density" dataset (see the Materials and Methods section and section 4.2).

4.1 Population structure analyses and demographic inferences

4.1.1 Setting GCA populations into the South/East Asian genomic landscape

PCA was applied to the "GCA" dataset after removing three individuals exceeding ± 3 standard deviations from the mean PC1 or PC2 eigenvector calculated for their respective self-assigned ethnic groups.

Accordingly, three main clusters reflecting the examined ethnic groups and/or sampling locations were identified:

- Sherpas from the Rolwaling Himal (SRH, red);
- Tamangs from Tashinam (TAT, brown);
- Tamangs from Simigaon (TAS, yellow).

Conversely, samples belonging to groups speaking Indo-Aryan languages (IAR, green) could be not considered as a homogeneous population (Fig. 4.1).

This pattern thus revealed the existence of genetically distinct populations among those sampled in the examined geographical area. However, they do not perfectly match with the broad ethnic groups recognized according to individual self-reported affiliations. In fact, while Sherpa people gathered into a single well-defined cluster, the Tamang communities occupied different positions in the PCA space. In particular, the TAS cluster was displaced toward a few IAR individuals, which were considerably scattered along PC1 (Fig. 4.1).

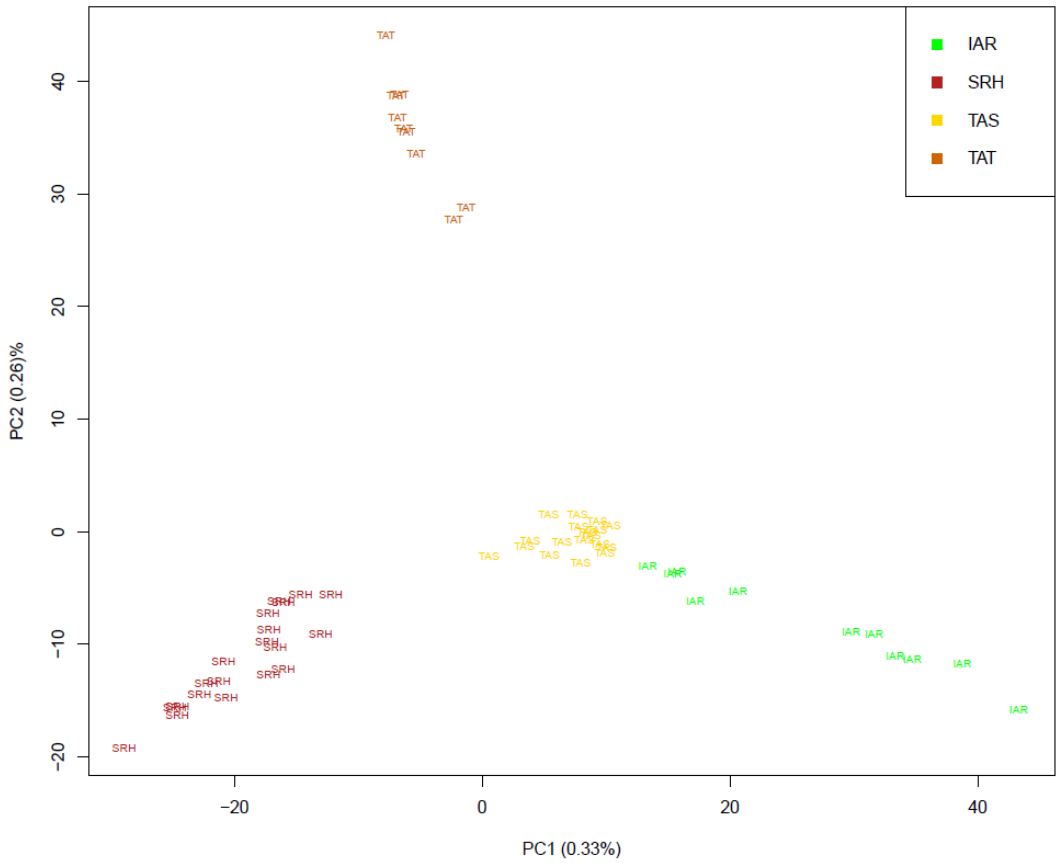


Figure 4.1. PCA performed on samples included in the "GCA" dataset.

When PCA was extended to the 75 Asian populations selected from the "extended" dataset, PC1 (accounting for 4.8% of variance) captured the known genetic differentiation between groups from the South Asian peninsula (i.e. India and Pakistan) and from East Asia (i.e. Indochina, China, Japan) (Li et al. 2008; Adbulla et al. 2009). PC2 (describing 0.6% of variance) instead reflected clines of variation observable within these two macro-geographic areas (Fig. 4.2).

In particular, the well-established North to South gradient of South Asian genetic diversity (Reich et al. 2009; Basu et al. 2016) was clearly detectable. Conversely, the East Asian cluster appeared to be more homogeneous, with the cline described by PC2 recapitulating only to a certain degree the latitudinal extension of continental East Asia. The sole exception to the North-South gradient observed along this geographic range according to PC2 was represented by the outlier position occupied by high-altitude Himalayan populations, such as Tibetans and Sherpas, as already described by several studies (Qi et al. 2013; Jeong et al. 2014; Jeong et al. 2016). In detail, SRH spread along the Sherpa cline together with samples from Khumbu above the Tibetan cluster, while TAT and, especially, TAS skew from the East Asian cline towards the South Asian cluster.

As regards IAR, they were substantially scattered in the PCA space, especially along PC1. For instance, two IAR individuals clustered within the Indian Tharu people, a group speaking an Indo-Aryan language that resides in the Terai plain and that is known to show also ancestry components typical of East Asian populations (Fornarino et al. 2009; Chaubey et al. 2011; Basu et al. 2016). Hereafter, we thus included Tharu in the bulk of Tibeto-Burman populations since the majority of South Asian groups presenting East Asian admixture actually speaks Tibeto-Burman languages.

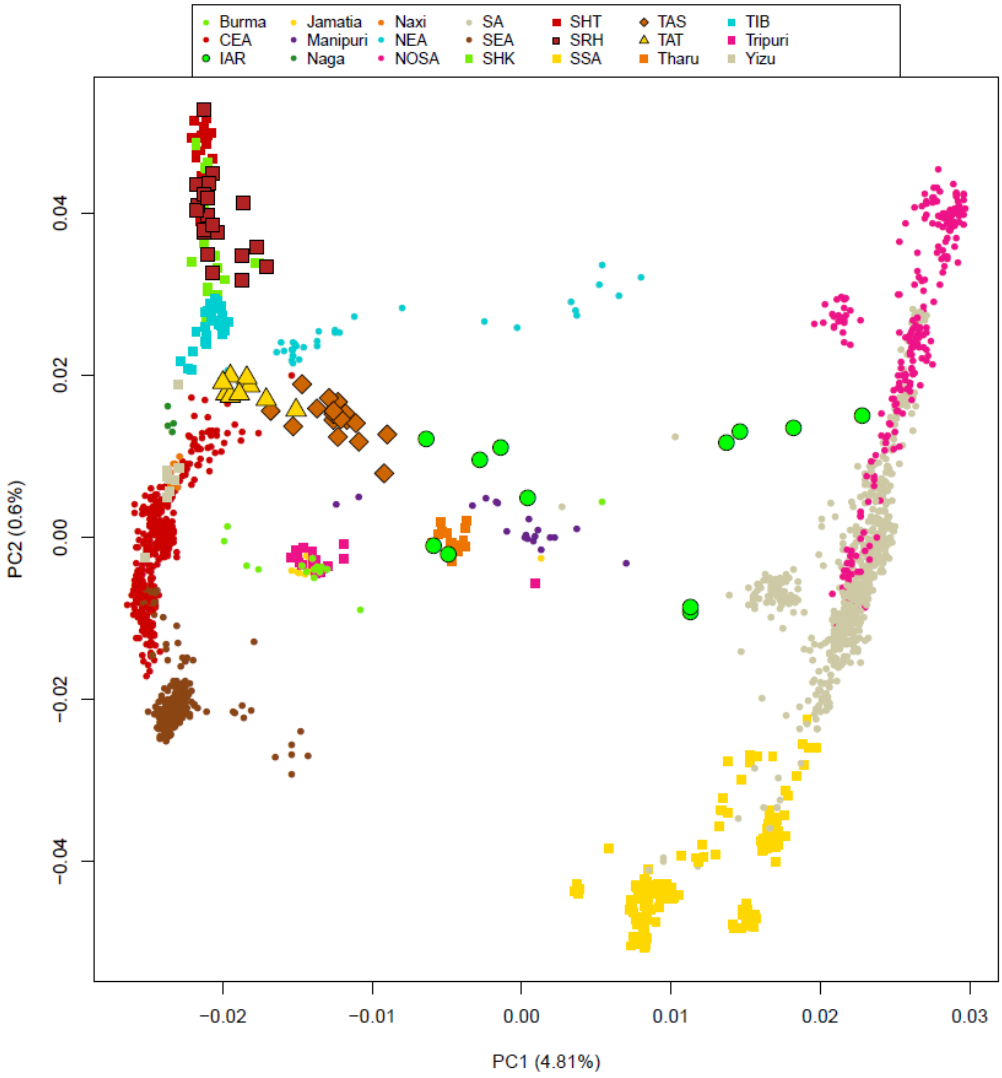


Figure 4.2. PCA performed on 75 Asian populations included in the "extended" dataset. PC1 describes genetic differentiation between East Asian (left) and South Asian (right) populations. PC2 instead shows the latitudinal gradients observable within the two macro-groups. The North to South South Asian gradient (top to bottom) and the North to South East Asian gradient (centre to bottom) are displayed together with the outlier position of Himalayan populations (Tibetans, TIB and Sherpas). Samples from the GCA populations are highlighted with bigger labels: SRH (red squares), TAT (yellow triangles), TAS (dark-orange rhombus), IAR (green circles). East Asian populations are labelled according to macro-groups: Central EA, CEA; North EA, NEA; North SA, NOSA; South SA, SSA.

4.1.2 Complex admixture patterns of GCA and Tibeto-Burman populations

ADMIXTURE analysis (Fig. 4.3) was used to infer the ancestry proportions of each GCA subject together with a large set of South Asian/East Asian samples. Overall, the results of this analysis agrees with those obtained for PCA. IAR indeed showed extremely variable levels of South Asian/East Asian admixture, with some individuals presenting a Northern Indian-like ancestry profile characterized by none or negligible East Asian ancestry and others showing instead significant proportions (up to 80%) of this genetic component. ADMIXTURE thus confirmed the pattern observed in the PCA and indicating that IAR samples do not represent a genetically homogeneous group. For this reason we excluded them from subsequent population-based analyses.
In contrast, the GCA Tibeto-Burman populations presented homogeneous ancestry profiles. Specifically, TAS were characterized by evident South Asian/East Asian admixture, with on average 23% of their ancestry being derived from South Asian genetic components and 77% from East Asian ones, through all K presenting the highest log likelihood values. The proportion of South Asian ancestry was considerably lower in TAT, with a mean value of approximately 9%. Conversely, SRH showed almost no signatures of recent South Asian admixture, in line with what observed for the Sherpas from Khumbu.
As previously reported (Jeong et al. 2014; Jeong et al. 2016; Zhang et al. 2017), the Sherpas were found to be enriched in a specific ancestry component that in our study became appreciable for K values ≥ 4 (Fig. 4.3). This component reached a proportion of 100% in almost all Sherpa individuals from Thame (SHT), a mean value of 94% in SRH and of around 80% in those from Khumjung (SHK).
At K = 6, when North East Asian and South East Asian signatures differentiated and the model achieved the best predictive accuracy (Fig. 4.4), this "Sherpa-like" component maintained relatively high proportions also in other Tibeto-Burman groups. For instance, it varied from those observed in Tibeto-Burmans from the Indian subcontinent (i.e. Manipuri, 18%; Burma, 21%; Jamatia, 23%; Tripuri, 23% and Tharu, 27%) or China (i.e. Naxi, 30%; Yizu, 32%) to those proper of Nagas (42%), Tamangs (TAS, 42%; TAT, 52%) and Tibetans (58%), being instead detected at substantially lower values in some non-Tibeto-Burman East Asian populations. Among them, only Tu people from the Chinese Qinghai-Gansu area and speaking Mongolic languages, but interlinked with Tibetan dialects (Paul et al. 2016), presented a proportion comparable to that observed in Tibeto-Burmans (26%).

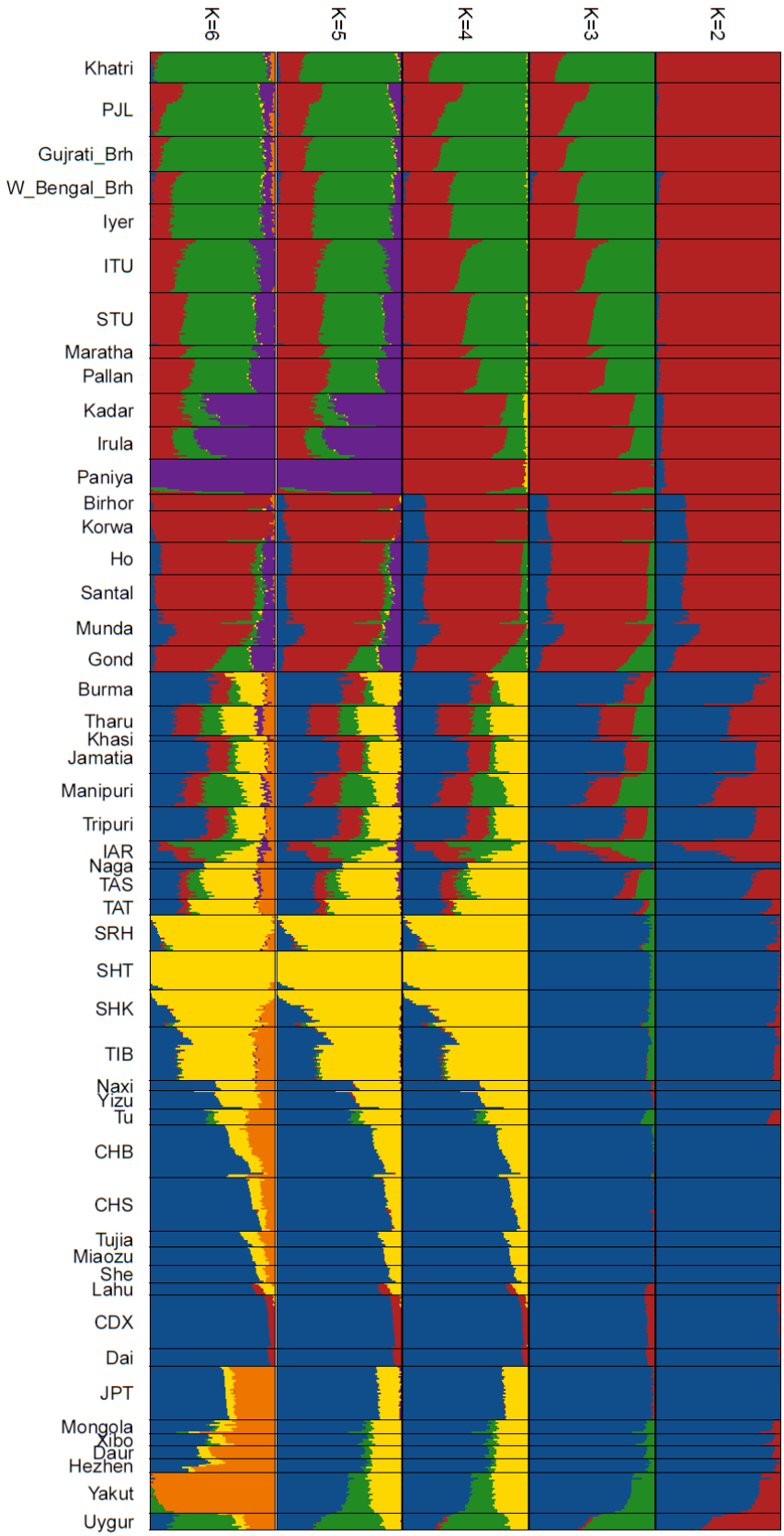


Figure 4.3. Ancestry proportions at K = 2 to K = 6 (from top to bottom) estimated with ADMIXTURE. ADMIXTURE analyses were performed on 843 individuals belonging to 50 Asian populations selected from the "extended" dataset. For each K only the runs with the highest Log-Likelihood are plotted.

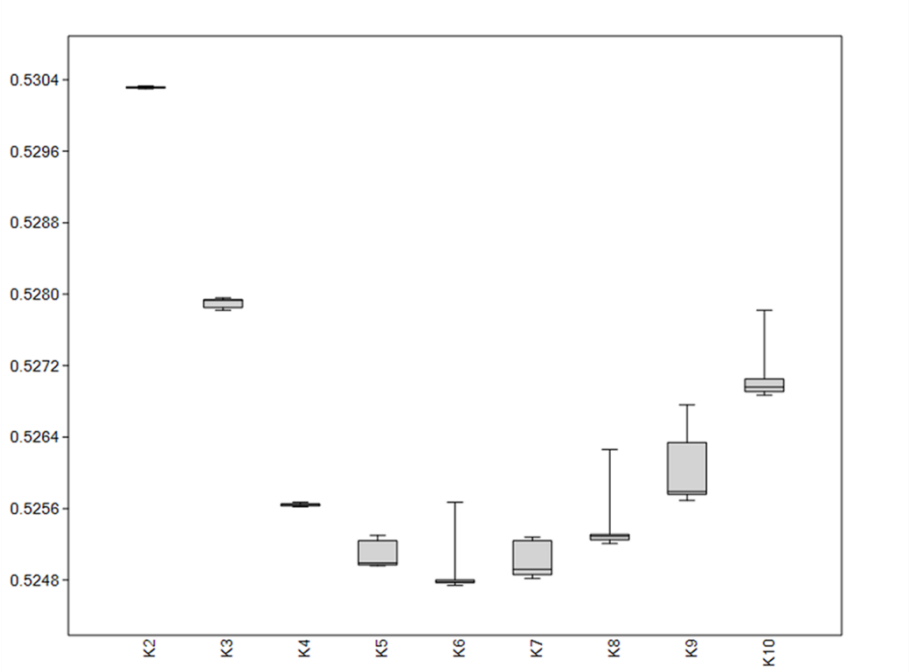


Figure 4.4. Cross-validation (CV) errors for ADMIXTURE clustering analyses. CV were computed for all the 50 runs performed for each tested K. The best predictive accuracy was achieved by the model when six ancestral components (K = 6) were tested.

We then computed the *f3* statistics by considering all South Asian/East Asian possible population pairs of the "extended" dataset as proxies for the true ancestral admixing populations (Appendix Table 2). This enabled us to validate the occurrence of admixture events involving both South Asian and East Asians ancestry components in almost all Tibeto-Burman populations residing South of the Himalayas (e.g. Burma, Tharu, Tripuri, Jamatia, Manipuri, TAS). The sole exceptions were represented by TAT, Nagas, SRH and SHT. Significantly negative *f3* values were obtained for SHK, suggesting admixture between either SHT or SRH and another East Asian or South Asian population.

Conversely, we observed admixture occurred in lowland Tibeto-Burman populations located North/East of the Himalayas, such as Yizu and Naxi, involving ancestral groups related to present-day Tibetans, Sherpas or Nagas (as one source population) and other East Asian populations. Moreover, high-altitude Tibetans also showed significant signatures of admixture involving non-admixed Sherpas (i.e. SRH and SHT) and several low-altitude East Asian populations.

These admixture patterns received further support from the LD decay-based approach implemented in the ALDER program, which also provided time estimates for the inferred admixture events (Appendix Table 3, Fig. 4.5). In particular, ALDER analysis showed that only Burma, Yizu, Naxi and Tibetans failed to be confirmed as admixed thus representing the sole exceptions together with

TAT, who turned out to be not admixed according to the computed *f3* statistics, but resulted significantly admixed with ALDER. Moreover, results obtained for the Indian Tibeto-Burman groups (i.e. Jamatia, Tripuri, Manipuri, and Tharu) were in line with those previously described by Basu et al. (2016) and achieved using a different dating method.

Finally, we considered a value of 25 years per generations to derive the time since admixture from the rate of exponential LD decay. Accordingly, Jamatia and Tharu presented the oldest time ranges, spanning from 1.6 to 0.6 thousand years ago (kya), while we identified more recent admixture events for Manipuri and Tripuri (0.8-0.2 kya). As regards TAS and TAT, the inferred dates of admixture were even more recent (0.65-0.38 kya and 0.43-0.13kya, respectively).

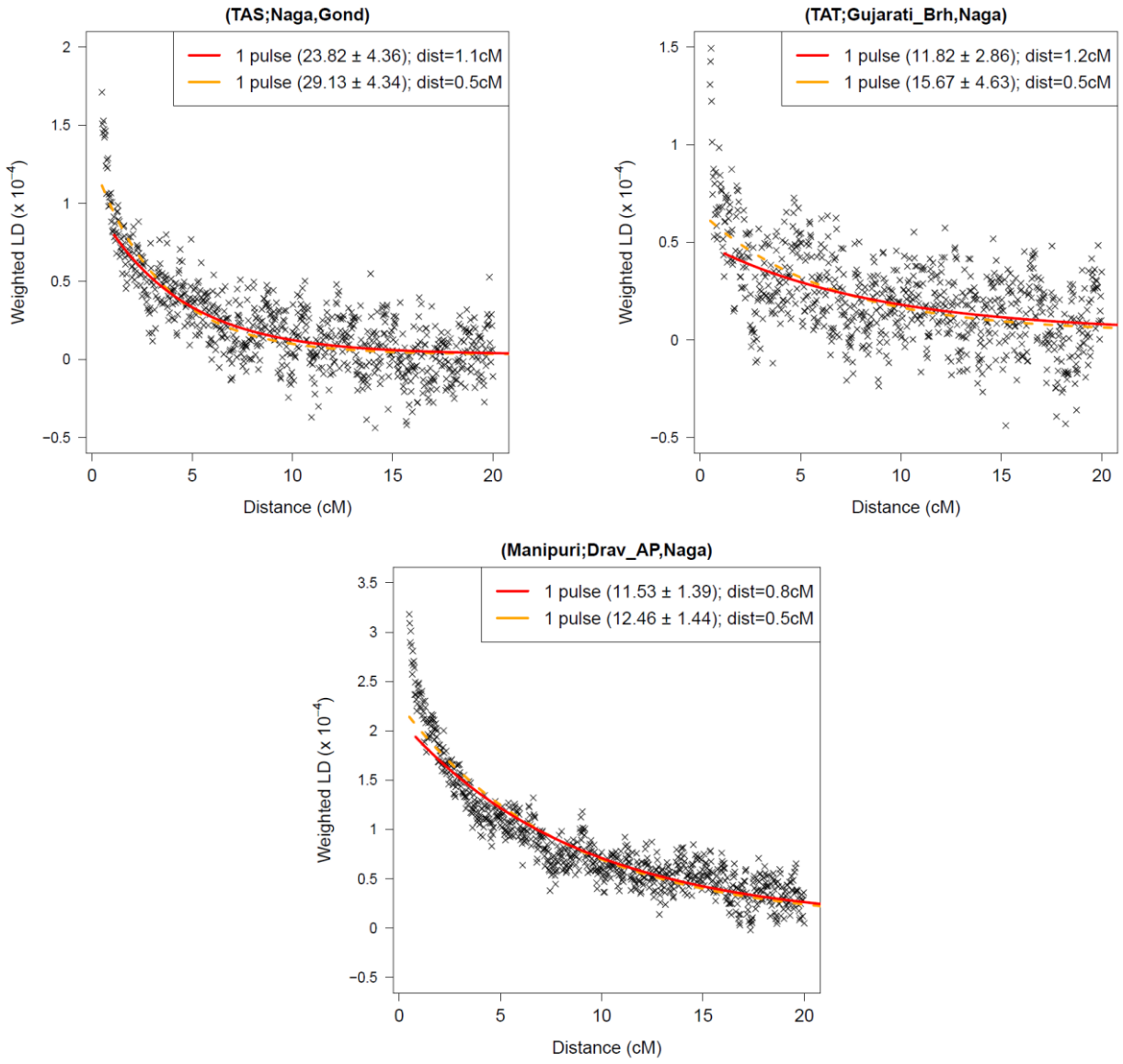


Figure 4.5. Weighted LD curves as function of genetic distance (cM) obtained with ALDER. Three LD curves are shown as examples of the several performed tests. In particular, they describe in turn TAS (top left), TAT (top right) and Manipuri (bottom) as the admixed groups, as well as Nagas as the East Asian proxy source group and a South Asian population as the other source.

4.1.3 Disentangling the impact of admixture and drift on the history of Sherpa people

By estimating haplotype diversity of mtDNA (H = 0.898) and Y-chromosome (H = 0.942) in SRH and in previously studied Sherpa communities (H = 0.906 for mtDNA) (Bhandari et al. 2015), we found reduced values with respect to the average computed for other Tibeto-Burman, South Asian and East Asian groups (H = 0.973 for mtDNA and H = 0.980 for Y-chromosome) (Appendix Table 1 and Table 4.1).

Table 4.1 mtDNA HVSI (top) and Y-chromosome STR (bottom) haplotype diversity of the GCA and Asian reference populations. Number of unique haplotypes and Nei's gene diversity indices are reported for each population.

mtDNA HVSI			
Population	**N samples**	**N haplotypes**	**Gene diversity (H)**
IAR	23	21	0.9921 +/- 0.0154
SRH	32	17	0.8986 +/- 0.0370
Tamang_GCA	24	18	0.9638 +/- 0.0253
Sherpa_Tibet	226	31	0.8686 +/- 0.0169
Sherpa_Nepal	266	41	0.9064 +/- 0.0078
Tamang	45	26	0.9576 +/- 0.0146
Newar	66	36	0.9739 +/- 0.0074
Kathmandu	77	64	0.9942 +/- 0.0034
Tibetans	156	94	0.9875 +/- 0.0033
Nakchu_Tibetan	168	100	0.9768 +/- 0.0061
Shigatse Tibetan	220	128	0.9856 +/- 0.0034
Gansu_Tibetan	83	71	0.9944 +/- 0.0034
Qinghai_Tibetan	76	62	0.9926 +/- 0.0041
Sichuan_Tibetan	62	58	0.9979 +/- 0.0034
Yunnan_Tibetan	71	17	0.8491 +/- 0.0235
Nepalese	235	135	0.9914 +/- 0.0016
Persian	82	62	0.9871 +/- 0.0054
Tajiks	44	39	0.9947 +/- 0.0061
Mongolians	47	42	0.9944 +/- 0.0060
Adi	45	35	0.9818 +/- 0.0107
Apatani	52	28	0.9585 +/- 0.0123
Naga	43	21	0.9402 +/- 0.0187
Nishi	52	33	0.9766 +/- 0.0085
Sikh Punjab	40	34	0.9910 +/- 0.0080
Taiwanese	66	59	0.9953 +/- 0.0041
Kashmir	19	14	0.9532 +/- 0.0358
Lobana	62	36	0.9778 +/- 0.0068
Tharu	37	36	0.9985 +/- 0.0067
Uttar Pradesh	73	60	0.9916 +/- 0.0047
Guangdong Han	71	62	0.9948 +/- 0.0038
Tipperah	20	19	0.9947 +/- 0.0178
Kanet	37	31	0.9880 +/- 0.0100
Punjab	109	84	0.9925 +/- 0.0031
Uighur	45	41	0.9949 +/- 0.0063
Bai	31	29	0.9957 +/- 0.0095
Dai	38	35	0.9943 +/- 0.0082
Lisu	37	20	0.9520 +/- 0.0171

Mongol	15	10	0.9429 +/- 0.0403
Nu	30	9	0.8575 +/- 0.0377
Sali	30	25	0.9885 +/- 0.0114
Yao_Tibetan	41	19	0.9354 +/- 0.0190
Tu	35	29	0.9866 +/- 0.0110
Zhuang	83	65	0.9912 +/- 0.0040
Thai	33	30	0.9943 +/- 0.0090
Wuhan_Han	42	42	1.0000 +/- 0.0052
Liaoning_Han	50	48	0.9984 +/- 0.0044
Yunnan_Han	42	40	0.9965 +/- 0.0067
Qingdao_Han	49	44	0.9949 +/- 0.0056
Xinjiang_Han	46	42	0.9952 +/- 0.0061
Guangdong	30	28	0.9954 +/- 0.0100
Karachi	99	76	0.9922 +/- 0.0032
Y-chromosome STR			
Population	**N samples**	**N haplotypes**	**Gene diversity (H)**
IAR	22	20	0.9913 +/- 0.0165
SRH	19	13	0.9415 +/- 0.0375
Tamang_GCA	22	17	0.9524 +/- 0.0365
Tamang	45	26	0.9495 +/- 0.0176
Newar	66	29	0.9399 +/- 0.0157
Kathmandu	77	70	0.9969 +/- 0.0029
Tibet	156	142	0.9988 +/- 0.0009
Tibetan_Lhasa_a	132	119	0.9984 +/- 0.0012
Tibetan_Lhasa_b	112	107	0.9992 +/- 0.0013
Han Chinese	131	129	0.9998 +/- 0.0010
Japanese Osaka	131	125	0.9992 +/- 0.0011
Taiwanese	200	185	0.9987 +/- 0.0009
Brahmin Punjab	60	56	0.9977 +/- 0.0035
Brahmin Himachal Pradesh	61	57	0.9973 +/- 0.0038
Brahmin Rajasthan	58	39	0.9607 +/- 0.0152
Bangali	216	211	0.9998 +/- 0.0005
Tibetan Qinghai	167	162	0.9996 +/- 0.0007
Sakaldwipi Brahmin	65	57	0.9957 +/- 0.0037
Kanyakubja Brahmin	78	71	0.9973 +/- 0.0027
Konkanastha Brahmin	71	66	0.9976 +/- 0.0030
Gond	75	56	0.9917 +/- 0.0036
Iyengar	66	62	0.9967 +/- 0.0040
Kurumans	67	52	0.9860 +/- 0.0068
Tripuri	65	62	0.9986 +/- 0.0030
Riang	67	53	0.9882 +/- 0.0063
Munda	68	50	0.9842 +/- 0.0074
Balmiki	62	33	0.8503 +/- 0.0456
Mahadev Koli	64	28	0.9330 +/- 0.0162

According to these analyses, the proportions of haplotypes shared among individuals belonging to the same population were substantially higher in Sherpa communities (SHR, 54% and 32% for mtDNA and Y-chromosome respectively; SHT, 85% for mtDNA) when compared to the mean values calculated for the assembled reference datasets (i.e. 27% for mtDNA and 18% for Y-chromosome). A pattern of remarkable intra-population homogeneity emerged also from the analysis of SRH haplogroup composition, which highlighted a limited number of lineages

accounting for the great majority of the examined mtDNAs (i.e. cumulative percentage of A, C4a3b and M9a1a of 56%) and Y-chromosomes (O3, 85%) (Table 4.2).

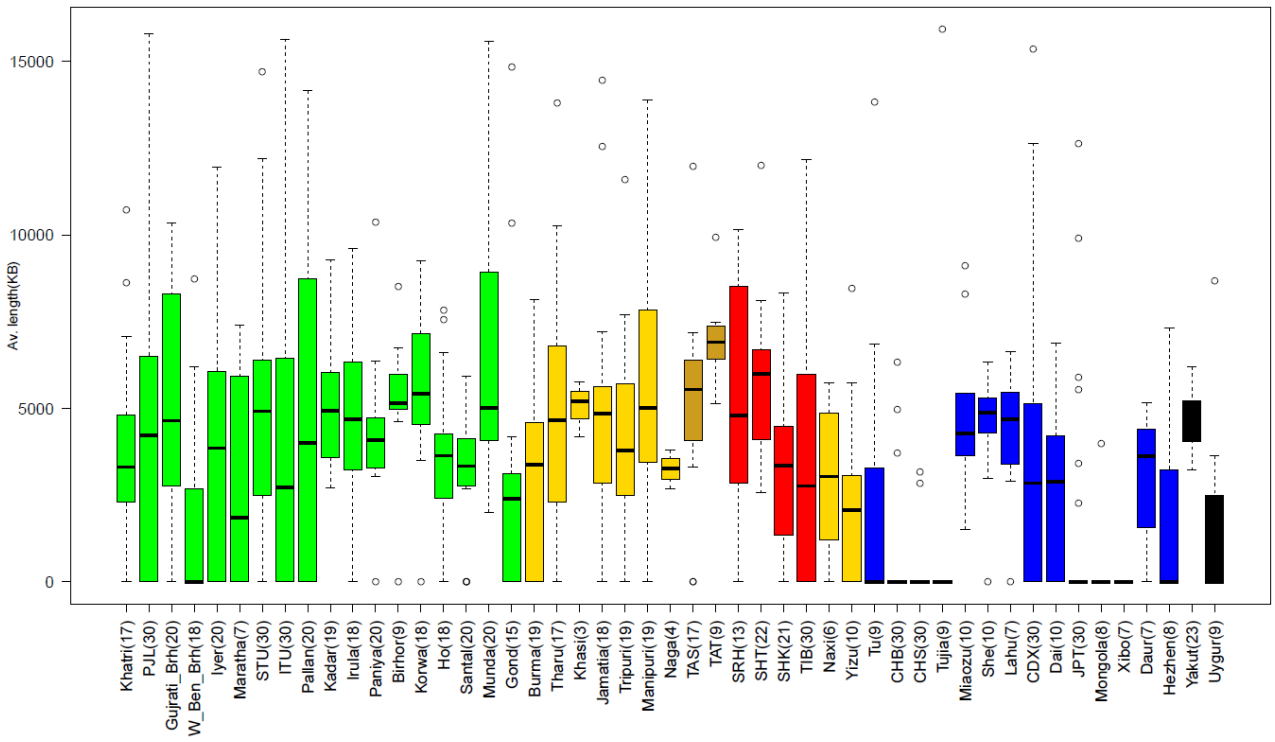
Table 4.2 mtDNA (top) and Y-chromosome (bottom) haplogroup frequencies observed in the three GCA groups.

mtDNA	SRH	Tamang_GCA	IAR	Tot.
Haplogroups/N.	**34**	**21**	**24**	**79**
A	20.60%	4.80%	0.00%	10.10%
A11	0.00%	14.30%	0.00%	3.80%
C4a1	0.00%	0.00%	4.20%	1.30%
C4a3b	20.60%	0.00%	4.20%	10.10%
C4c1b	2.90%	0.00%	0.00%	1.30%
D4m2	5.90%	4.80%	0.00%	3.80%
D5a2a1	2.90%	0.00%	4.20%	2.50%
D5a3	0.00%	4.80%	0.00%	1.30%
F1	8.80%	0.00%	12.50%	7.60%
F1a1	8.80%	0.00%	12.50%	3.80%
F1a'c'f	0.00%	4.80%	0.00%	1.30%
F1b1c	0.00%	4.80%	4.20%	2.50%
F1c1a1	0.00%	0.00%	4.20%	1.30%
G1a2	0.00%	4.80%	0.00%	1.30%
I1d	0.00%	0.00%	4.20%	1.30%
J	0.00%	4.80%	4.20%	2.50%
L3b1a	2.90%	0.00%	0.00%	1.30%
L3e1	0.00%	4.80%	4.20%	2.50%
M18	2.90%	0.00%	0.00%	1.30%
M33a1a	0.00%	0.00%	4.20%	1.30%
M3d	0.00%	0.00%	4.20%	1.30%
M6	2.90%	0.00%	4.20%	2.50%
M70	2.90%	0.00%	0.00%	1.30%
M9a1a	14.70%	4.80%	4.20%	8.90%
M9a1a2	0.00%	0.00%	8.30%	2.50%
M9a1b1	0.00%	0.00%	12.50%	3.80%
N1a1b1	0.00%	0.00%	4.20%	1.30%
T2b2b1	0.00%	4.80%	0.00%	1.30%
U2	0.00%	9.50%	0.00%	2.50%
U2b1	0.00%	19.00%	0.00%	5.10%
U7	0.00%	0.00%	4.20%	1.30%
Z	2.90%	9.50%	8.30%	6.30%
Y-chromosome	**Sherpa**	**Tamang_GCA**	**IAR**	**Tot.**
Haplogroups/N.	**15**	**20**	**16**	**51**
C3	0.00%	5.00%	0.00%	2.00%
H	6.70%	15.00%	18.80%	13.70%
J2a4	0.00%	0.00%	6.30%	2.00%
O3	86.70%	75.00%	37.50%	66.70%
Q	0.00%	0.00%	6.30%	2.00%
R1a	6.70%	5.00%	25.00%	11.80%
R2	0.00%	0.00%	6.30%	2.00%

These findings were in agreement with those described by Bhandari et al. (2015) and corroborated the hypothesis that historical isolation and subsequent drift have actually influenced patterns of uniparental variation in Sherpa populations.

A pattern of remarkable SRH intra-population homogeneity emerged also from the computation of runs of homozygosity (ROH). This approach provided a first clue supporting an appreciable impact of these demographic events also on the autosomal genome of Sherpa people, especially as regards SHT who presented remarkable values for both the average length and the number of DNA segments in homozygosity (Fig. 4.6).

Estimates obtained for these parameters were comparable to those calculated for other Sherpa groups and were observed also for different populations, especially Indian tribes, which are known to be small and isolated communities showing a higher level of inbreeding with respect to "open" populations (Basu et al. 2016). Also Tamangs, especially TAT, presented high values of average length and number of DNA segments in homozygosity, with this latter group showing even higher values with respect to SHT.



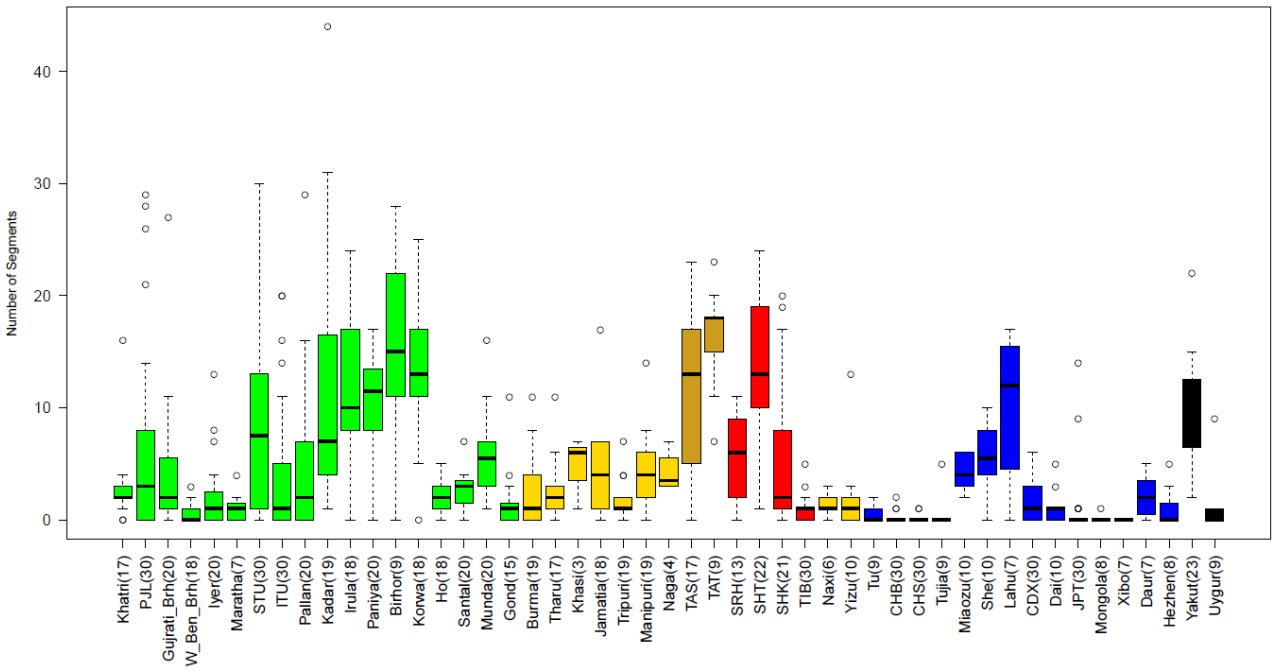


Figure 4.6. ROH calculated on a subset of Asian populations included in the "extended" dataset. The top panel shows the average length of segments in homozygosity, the bottom panel shows the total number of segments in homozygosity. In brackets, sample sizes for each population. Bar plots are color-coded as follows: green (South Asian), black (North East Asian), blue (East Asian), yellow (Tibeto-Burmans), dark-yellow (Tamangs), red (Tibetans and Sherpas).

It is noteworthy to mention that this issue should be taken into account in the light of previous studies (van Dorp et al. 2015; Falush et al. 2016) which showed that model-based clustering analyses are generally biased in assigning private ancestry components to highly drifted populations, as the Sherpas appears to be (Jeong et al. 2014).

That being so, we aimed at testing whether the strong genetic drift in the Sherpas resulted in an artificial pattern of ancestry mixture observed in the other Tibeto-Burman populations submitted to ADMIXTURE analysis, while assigning a single "Sherpa-like" ancestry component to the Sherpas. For this purpose, we applied a pipeline based on CHROMOPAINTER "chunk-lengths" outputs as described in van Dorp et al. (2015) and in the Materials and Methods section. We thus selected a subset of East Asian populations included in the "extended" dataset, from which we excluded SHK (since they appeared to be admixed with Tibetans and other South Asian/East Asian populations, (Appendix Table 2) and TAT (due to their low sample size).

We then conducted on them two separate CHROMOPAINTER runs:

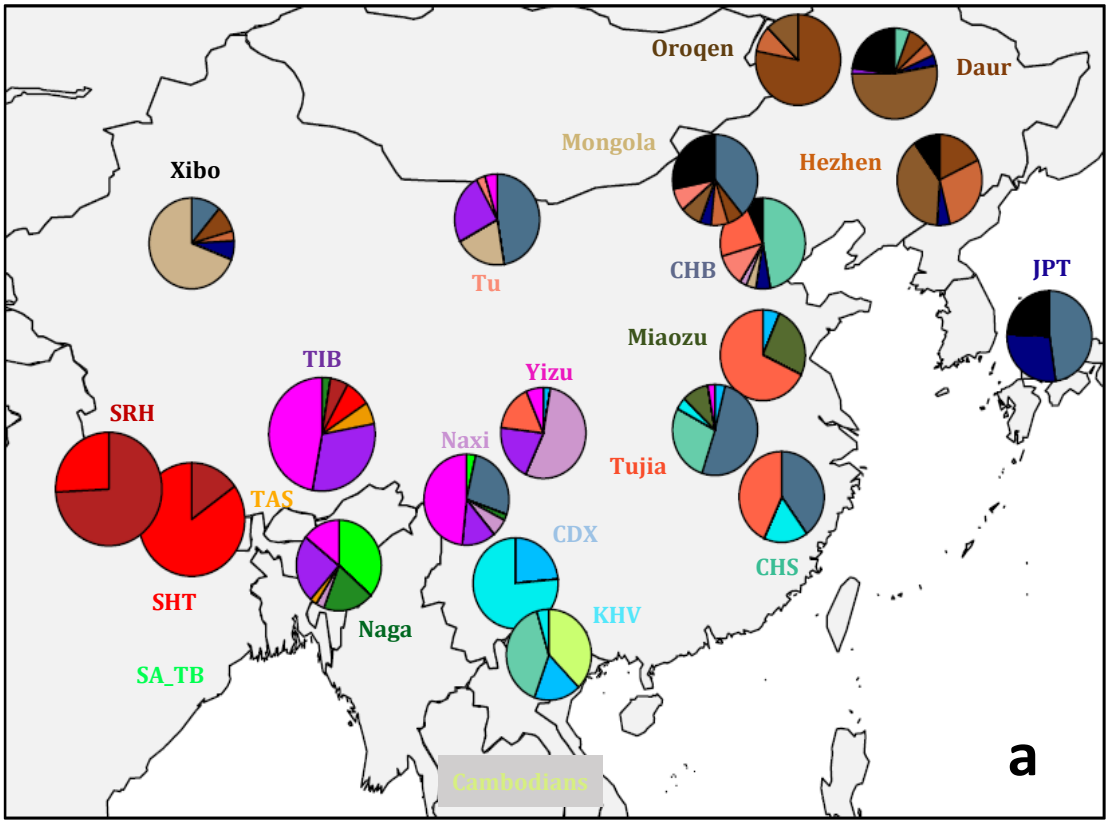
a) We searched the best matching DNA segments from individuals of each population ("recipients") in every other individual from all the considered populations ("donors") and by allowing for "self-copy" (see Materials and Methods section). The rationale beyond this

approach was that we expected the supposed differential drift levels between Himalayan groups (i.e. Tibetans, SRH and SHT) having made their painting profiles substantially different.

b) We instead excluded these populations from the donor groups in order to remove the hypothesized drift effect and to estimate haplotype sharing of Himalayan populations with respect to all the other considered groups.

Accordingly, we found that in case a) SRH and SHT derived most of their haplotype segments from themselves (74% and 85%, respectively), while the remaining ones were derived from each other (26% and 15%, respectively). On the contrary, Tibetans presented only 31% of haplotype segments ascribable to "self-copy", with the vast majority of haplotype segments being instead shared with neighbouring East Asian populations (i.e. 47% from Yizu, 8% from SHT, 6% from TAS, 5% from SRH and 3% from Nagas) (Fig. 4.7.a).

In case b) the three Himalayan groups instead showed extremely similar painting profiles (Fig. 4.7.b), thus indicating that genetic drift plausibly played a substantial role in determining the results observed in case a).



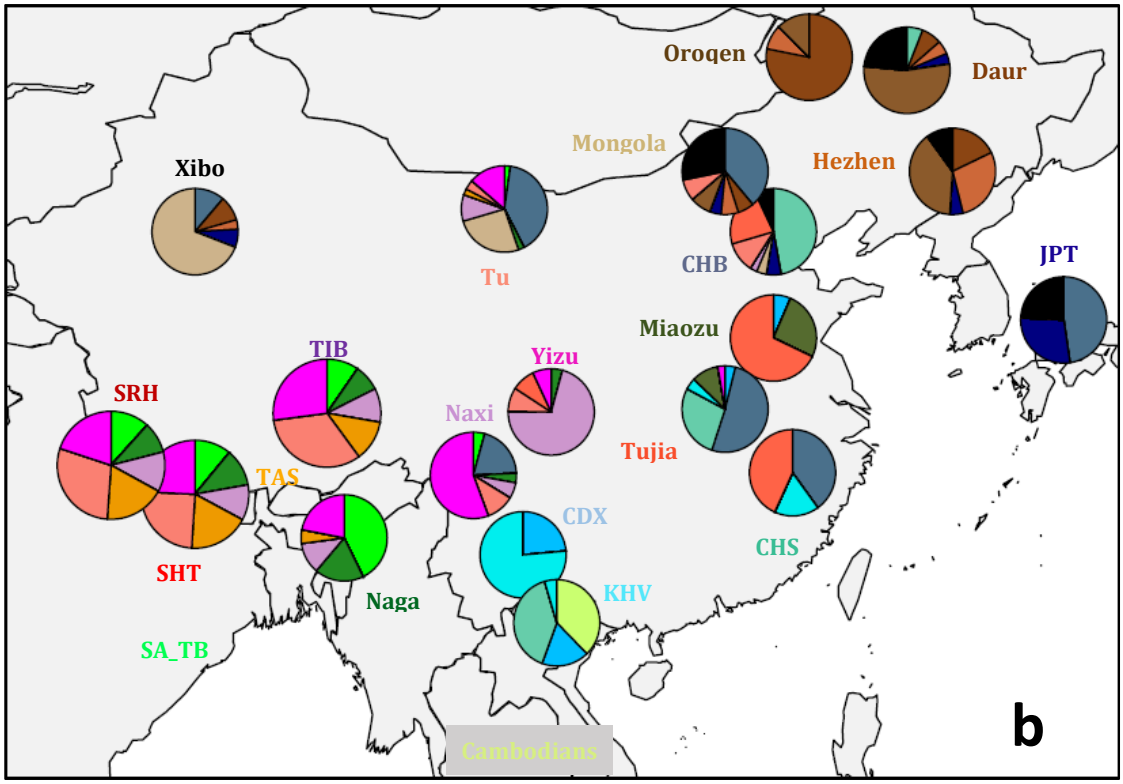


Figure 4.7. Ancestry proportions of high-altitude Himalayan groups inferred by CHROMOPAINTER analyses. a) Map showing results of case *a)* obtained by searching the best matching DNA segments from "recipients" in every other "donor". b) Map showing results of case *b)*, in which we excluded Tibetans, SRH and SHT from the donors to remove the hypothesized drift effect. In both cases, "self-copy" was allowed for all East Asian populations, while in case *b)* Himalayan populations could not "self-copy" or copy from each other since they were excluded from the "donors" in the upstream CHROMOPAINTER run. Pies charts representing inferred ancestry proportions for all East Asian populations are shown, with the exception of TAS, Cambodians and South Asian_Tibeto-Burmans

This pattern was further confirmed when the average length of shared haplotypes inferred by CHROMOPAINTER in run a) was calculated, showing that the Sherpa groups presented the highest (SHT) or the third highest (SRH) values among all the other East Asian populations included in the "extended" dataset (Table 4.3).

Table 4.3 Average lengths of haplotype segments shared between individuals belonging to the same population inferred by CHROMOPAINTER analysis.

Population (N)	Average length (cM)	Population (N)	Average length (cM)
SHT (22)	4189.297	Naga (4)	515.492
Paniya (20)	4039.213	Gond (15)	477.706
Irula (18)	3478.249	Iyer (20)	469.536
Korwa (18)	3322.204	ITU (30)	427.003
Yakut (23)	3056.981	Khatri (17)	420.149
SRH (13)	3014.960	CHS (30)	379.678
Birhor (9)	2719.409	Yizu (10)	373.138
Kadar (19)	2337.993	Daur (7)	373.055
Jamatia (18)	1967.338	STU (30)	371.407
Lahu (7)	1878.945	PJL (30)	364.312
TAS (17)	1831.986	Pallan (20)	354.627
Tharu (17)	1561.094	Burma (19)	339.553
Manipuri (19)	1389.595	Munda (20)	331.887
She (10)	1368.605	CHB (30)	302.876
Santal (20)	1356.592	Naxi (6)	282.782
TIB (30)	1032.586	Gujrati_Brh (20)	259.850
Ho (18)	949.875	W_Ben_Brh (18)	233.946
CDX (30)	838.225	Tu (9)	217.217
JPT (30)	783.775	Xibo (7)	178.653
Hezhen (8)	706.890	Tujia (9)	147.613
Miaozu (10)	703.357	Mongola (8)	118.956
Tripuri (19)	645.885	Uygur (9)	94.280
Dai (10)	528.830	Maratha (7)	73.098

4.1.4 Genomic relationships between GCA and South/East Asian populations

We used the outgroup *f3* statistics to measure the shared genetic drift between GCA samples and a large set of South Asian/East Asian groups in the attempt to pinpoint a plausible proxy for the ancestral population that could have introduced the East Asian and "Sherpa-like" genetic components in Tibeto-Burmans residing South of the Himalayas.

As expected, SRH showed the highest *f3* values when compared respectively with the Sherpas from Khumbu and Tibetans (Fig. 4.8, top left).

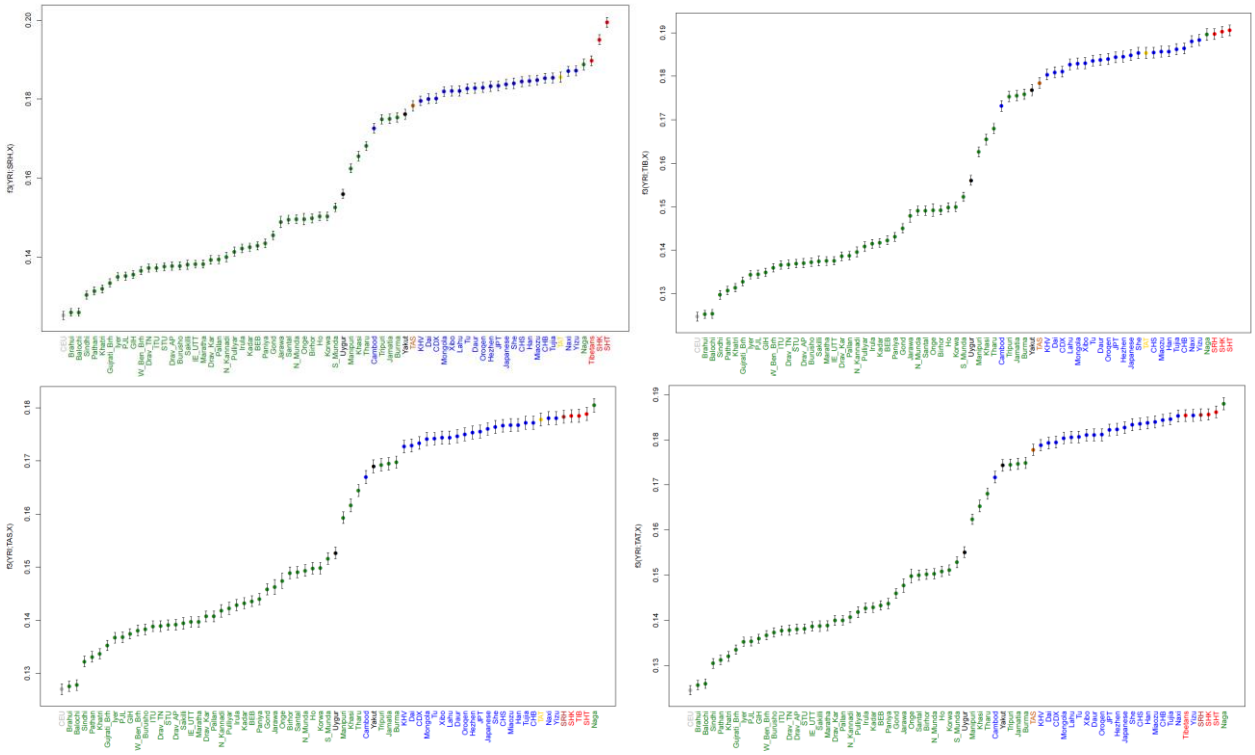


Figure 4.8. Outgroup *f3* statistics measuring the shared genetic drift between the three GCA groups and Tibetans (four panels) with a large set of South Aian/East Asian groups (X populations). The analyses were performed with the *qp3Pop* function implemented in ADMIXTOOLS on the 75 Asian populations included in the "extended" dataset. Yoruba (YRI) were used as outgroup population for every test. X populations reported on the x-axes and the dots in the plot are color-coded according to the following groups: grey (people of European origin, CEU), green (South Asian, SA), blue (East Asian), black (North East Asian), yellow (TAS), dark-orange (TAT), red (Tibetans and Sherpas). Top left panel reports the *f3*(YRI;X,SRH) test. Top right panel reports the *f3*(YRI;X,TIB) test. Bottom left panel reports the *f3*(YRI;X,TAS) test. Bottom right panel reports the *f3*(YRI;X,TAT) test.

Moreover, the *f3* score obtained when they were tested against Tibetans was comparable to that found when the Tibeto-Burman Nagas from North East India were considered. Reciprocally, the same pattern was observed when Tibetans were contrasted to SHR and Nagas (Fig. 4.8, top right). Interestingly, the Nagas were also responsible for peak *f3* scores when both TAS and TAT were tested (Figure 4.8, bottom panels).

To further disentangle the genomic relationships between these closely related populations, we calculated a series of *ad hoc* D-statistics by testing separately the three Tibeto-Burman GCA groups and each East Asian population and by considering a four-population phylogeny in the form: (East Asian, Tibetans; GCA, YRI). In particular, by forcing the GCA groups to branch out before the split between Tibetans and the rest of East Asian populations, we aimed at testing whether the poor fit of

GCA samples in such an artificial phylogeny deviates towards their closer genetic connection with Tibetans or with other East Asian groups. Accordingly, negative D values indicated high affinity with Tibetans, whereas positive values suggested close relationship with the tested East Asian population.

Overall these analyses confirmed that TAS and TAT were more closely related to the Nagas than to Tibetans, while the Sherpas showed closer affinity to Tibetans than to any other East Asian population (Fig. 4.9).

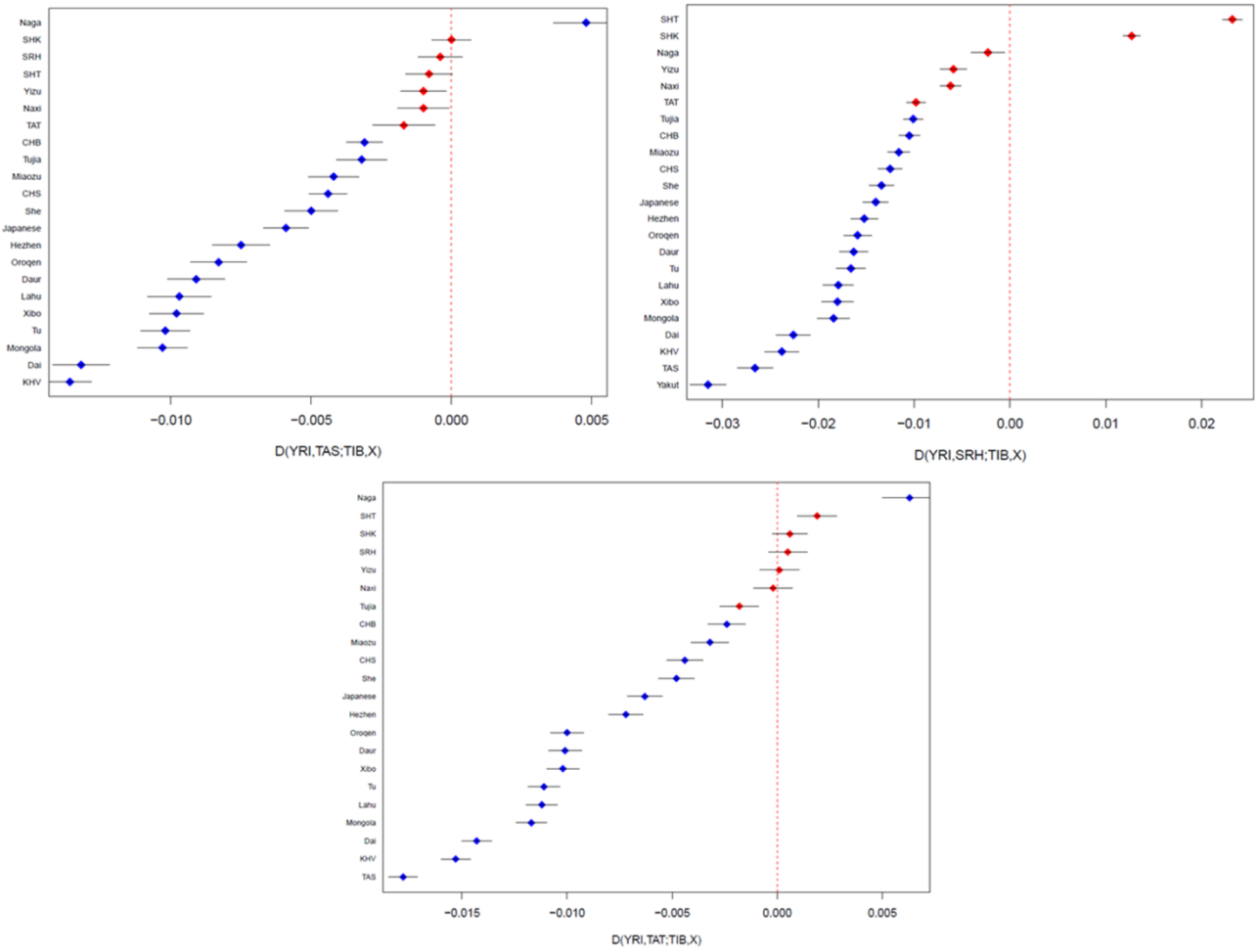


Figure 4.9. D-statistics computed for the GCA groups. The analyses were performed with the *qpDstat* function implemented in ADMIXTOOLS. On the y-axis the East Asian populations included in the "extended" dataset are displayed and used in turn for testing each GCA group (TAS, top left; SRH, top right; TAT, bottom). East Asian Tibeto-Burman populations are displayed in red, while all the other populations are coloured in blue.

4.1.5 The admixed phylogeny of Tibeto-Burman populations

Intrigued by the presence of the "Sherpa-like" ancestry component in many East Asian populations and by its highest prevalence in Tibeto-Burmans from both East Asia and South Asia, we attempted to reconstruct the phylogenetic relationships of these groups by taking into account their South Asian/East Asian admixture profiles.

For this purpose, we identified relatively un-admixed populations according to the previously computed *f3* statistics and we used them to build a scaffold tree by means of the MIXMAPPER algorithm. In detail, we selected the Nagas, SHT and SRH as representative Tibeto-Burman groups. We then included in the tree also Tibetans, which showed significantly negative *f3* values, but not when tested against the other low-altitude East Asian populations used to build the scaffold tree. In fact, we aimed at considering also the Tibetan sample because we were mainly interested in identifying the branching point between low- and high-altitude Tibeto-Burmans. Nevertheless, it is noteworthy to consider that our dataset included only one Tibetan population, which could be not representative of the overall genetic variation present across the plateau.

We then selected the non-admixed East Asian population of Dai as a proxy for groups with predominant Southern East Asian ancestry and the Yakuts as representative of people with Northern East Asian ancestry. Moreover, the Brahui and Birhor South Asian populations were also included to represent the two main sources of Indian ancestry (i.e. the ancestral northern and the ancestral southern ones) (Reich et al. 2009; Basu et al. 2016), together with Yoruba (YRI) that were used as outgroup.

The topology of the obtained tree was conserved for all the 500 bootstrap replications performed and revealed the existence of three main East Asian clades that mirror the East Asian ancestry components pointed out by ADMIXTURE analyses. The North East Asian-like cluster of Yakuts branched out first and then the South East Asian and the Tibeto-Burman clades diverged. Within this latter group, the Nagas branched out before differentiation of the Sherpas from Tibetans (Fig.4.10).

We then fitted the admixed Tibeto-Burman populations on the obtained scaffold tree. Accordingly, TAS appeared to result from the admixture between a South Asian population and the Nagas, being supported by 420 over 500 bootstrap replications. Both Yizu (500/500 replications) and Naxi (498/500 replications) were proposed to originate from the admixture between the Nagas and Dai. A population located on the internal node branching after the split of Birhor from Brahui was supposed to have admixed with either the Nagas or a group ancestral to the divergence between the Nagas and the Tibetan/Sherpa cluster, resulting in Manipuri (491/500 replications). The same South Asian population could have admixed again with the Nagas leading to the origin of Tharu (495/500

replications). The most frequent observation for Jamatia (314/500 replications) and Tripuri (388/500 replications) pointed them as an admixture between populations located on internal North South Asian nodes and a putative group occupying the internal node before the divergence between the Nagas and the Tibetan/Sherpa cluster. Finally, Burma presented the same source of South Asian ancestry of Jamatia and Tripuri, even though their East Asian ancestral component had nearly the same probability to derive from the internal node of Dai (258/500 replications) or of Tibeto-Burmans (237/500 replications) (Fig. 4.10 and Appendix Fig. 1).

It worth nothing that although the robustness of the reconstructed topology was strongly supported, the divergence between the Nagas and the high-altitude clade cannot be explained by a single split from a common ancestor. In fact, when this hypothesis was formally tested via computation of *f4* and D-statistics in the form (YRI, CDX; Nagas, Tibetans/Sherpas), significantly negative *f4* and D-statistics were obtained for all tests (Table 4.4), suggesting that Tibetans and Sherpas may represent a present-day proxy for a distinct branch of ancestral Tibeto-Burmans.

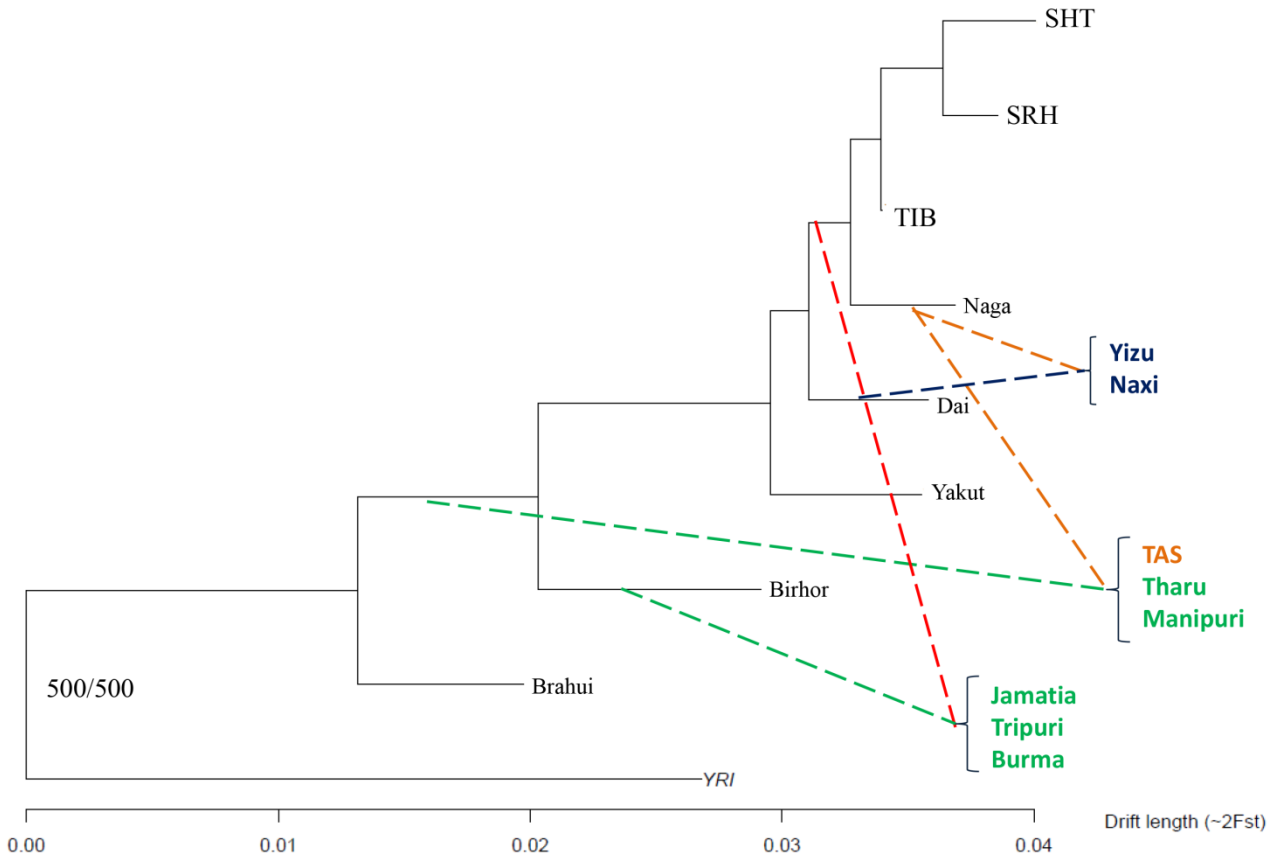


Figure 4.10. Neighbour-Joining tree reporting MIXMAPPER fitted admixture events. Scaffold phylogenetic tree of un-admixed populations with a summary of the best fits for the admixed Tibeto-Burman populations tested. Admixed populations are color-coded according to their geographic location: green (South Asia), yellow (Nepal/North India), and blue (East Asia).

Table 4.4. f4 and D-statistics in the form (YRI, CDX; Nagas, Tibetans/Sherpas). This approach was used to test if a single common origin between low- and high-altitude Tibeto-Burmans is consistent with the data.

f4						
PopW	**PopX**	**PopY**	**PopZ**	***f4***	**SD**	**Z-score**
YRI	CDX	Nagas	TIB	-0.00148	0.000151	-9.844
YRI	CDX	Nagas	SRH	-0.00177	0.000174	-10.17
YRI	CDX	Nagas	SHT	-0.00165	0.000176	-9.378
D						
PopW	**PopX**	**PopY**	**PopZ**	**Dstat**	**SD**	**Z-score**
YRI	CDX	Nagas	TIB	-0.0112	0.001141	-9.809
YRI	CDX	Nagas	SRH	-0.0133	0.001313	-10.148
YRI	CDX	Nagas	SHT	-0.0124	0.001327	-9.362

4.1.6 Isolating the East Asian ancestry of Tibeto-Burman populations

We applied the HAPMIX algorithm to infer local ancestry and to mask the South Asian ancestry chunks of Tibeto-Burman groups residing South of the Himalayas.

This resulted in a clear detection of their East Asian ancestry as tested by projecting masked haploid data in the PCA space computed on a large set of South Asian/East Asian populations. The seven examined East Asian-Tibeto-Burman populations were thus projected on the PCA space via the functions implemented in the smartpca algorithm and Tibeto-Burman populations were represented with coloured triangles (Fig. 4.11). In particular, the 68 EA populations are labelled in the figure according to the following macro-groups: Central East Asia, CEA; North East Asia, NEA; North South Asia, NOSA; South South Asia, SSA.

In this first explorative PCA, Central Siberian populations of Yakuts and Uyghurs were also identified as genetic outliers, in line with their known ancient connection with a Western Eurasian ancestry source (Raghavan et al. 2014).

To focus on the population structure observable within East Asia, we then replicated PCA only on East Asian populations (Fig. 4.12A). Interestingly, Tibeto-Burmans were found to diverge from the classical North-South cline of variation present in East Asia, forming a distinct cluster. Moreover, the Nagas, Tharu and Tamangs turned out to be located at the core of this population group from which the known Tibetan/Sherpa outlier position deviated (Fig. 4.12A). The remaining groups (i.e. Yizu, Naxi, Manipuri, Tripuri, Jamatia, and Burma) also diverged from such a core towards South East Asian populations. We can speculate that this is due to more recent gene flow between

continental East Asians and several Tibeto-Burman groups, in accordance with the results obtained by MIXMAPPER analyses.

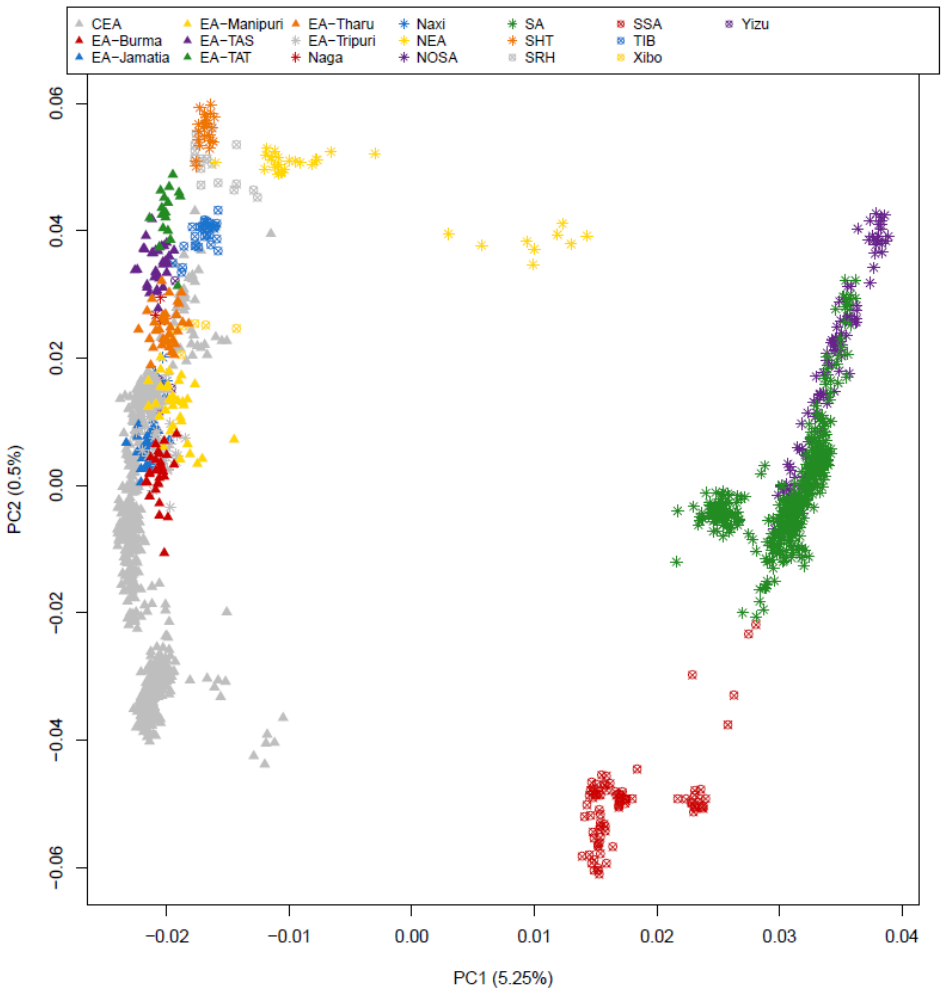


Figure 4.11. Ancestry-specific PCA performed on the 68 Asian populations included in the "extended" dataset.

Findings from the outgroup *f3* statistics computed on the masked dataset (Fig. 4.12B) were concordant with previous analyses and more recent genetic relationships between Tibeto-Burmans were also identified. In fact, Tharu from the Terai plains of Northern India and Nepal resulted more closely related to TAS (*f3* = 0.220) as well, in accordance to their geographic proximity, but again with the Nagas being responsible for the second highest *f3* peak (0.219). In addition, Jamatia and Tripuri showed closer affinity with each other (*f3* = 0.221) as they both reside in the Tripura region and speak sub-lineages of the Kok Borok branch of Tibeto-Burman languages (Paul et al. 2016), and then to the Nagas (*f3* = 0.219 and 0.218, respectively).

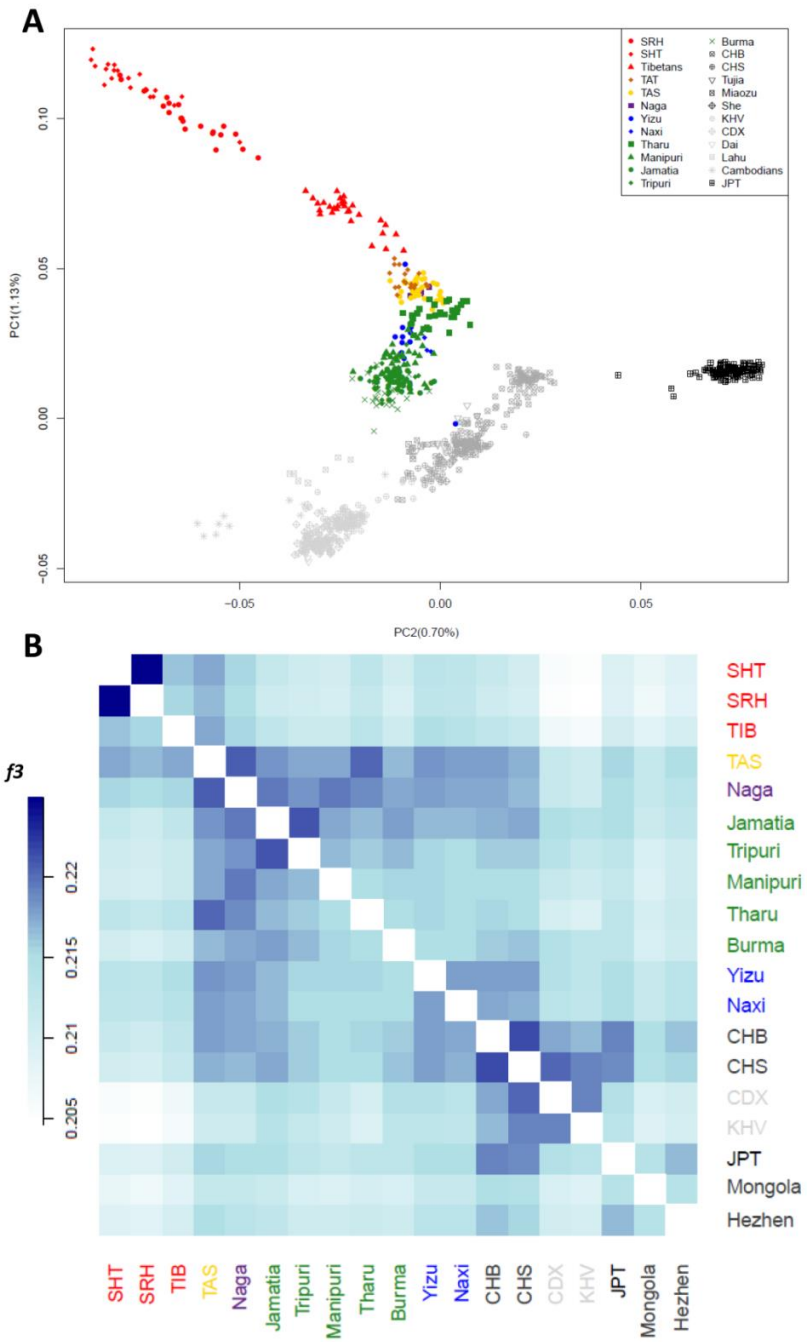


Figure 4.12. Ancestry-specific PCA and outgroup *f3* statistics computed on a subset of East Asian populations included in the "extended" dataset. A) PCA performed on the 25 East Asian populations of the "extended" dataset. Tibeto-Burmans are represented with coloured labels, whereas other East Asians are represented with grey-scale labels. **B)** Outgroup *f3* statistics performed on a subset of 18 East Asian populations selected from the masked "extended" dataset. Population labels are color-coded according to PCA plot.

Instead, Manipuri turned out to be more closely related directly to the Nagas (*f3* = 0.219), as expected according to their linguistic affiliation to the Kuki-Chin-Naga Tibeto-Burman branch (Paul et al. 2016). Finally, TAS and TAT were found to be more closely related to each other (*f3* = 0.224) and then to the the Nagas (*f3* = 0.221 and 0.220, respectively). This confirms their recent common ancestry and suggests that differences between them are probably due to recent

geographical isolation experienced by TAT, as suggested also by their high levels of homozygosity (Fig. 4.6).

4.1.7 Assessing the representativeness of Sherpa and Tibetan whole genome sequence data

In order to select the most useful samples sequenced for the whole genome to be submitted to selection scans, the generated Sherpa whole genome sequence and the Tibetan/Sherpa whole genome sequence data retrieved from literature were first set into the previously described Asian genomic landscape. This enabled to test whether they were actually representative of their populations of origin.

For this purpose, PCA was performed on the subset of East Asian and Tibeto-Burman populations included in the assembled "WGS-low-density" dataset composed of 199,679 SNPs, thus confirming the genetic structure of East Asian populations observed in previous PCA analyses based on SNP-chip data (Fig. 4.13). In particular, the East Asian populations were labelled according to the following macro-groups: Central East Asia, CEA; North East Asia, NEA; South East Asia, SEA; Tibeto-Burmans, TB.

Specifically, PC1 (accounting for 1.14% of variance) captured the main cline of variation present in East Asian populations ranging from South to North (left to right in Fig. 4.13 top). Nevertheless, PC1 alone could not discriminate between the high-altitude Tibetan/Sherpa groups and Northern East Asians, which clustered all together.

PC2 (depicting 0.66% of variance) instead highlighted the previously described South Asian admixture of Tibeto-Burman populations (which increases from bottom to top in Fig. 4.13 top). It is noteworthy to mention that the two Sherpa individuals from Tibet (SHP) occupied an intermediate position along PC2, thus suggesting some degree of South Asian admixture.

PC3 (accounting for 0.59% of variance) was then able to pinpoint the distinction between North East Asians and the outlier gradient of high-altitude Himalayan populations (top right and bottom right, respectively in Fig. 4.13 bottom).

Along this latter cline, the WGS samples turned out to be encompassed within the range of variability shown by their respective ethnic groups, with the sole exception of the two SHP individuals mentioned above.

In particular, Tibetans whole genome sequences (TBN) clustered together with the other Tibetan samples represented in the SNP-chip dataset (TIB_chip) and the rest of the SHP subjects were found to be located along the Tibetan/Sherpa drift/admixture gradient.

Finally, the generated Sherpa whole genome sequence (SRH) was instead located at the end of the cline observed among the other Nepalese Sherpas (SRH_chip, SHT, SHK).

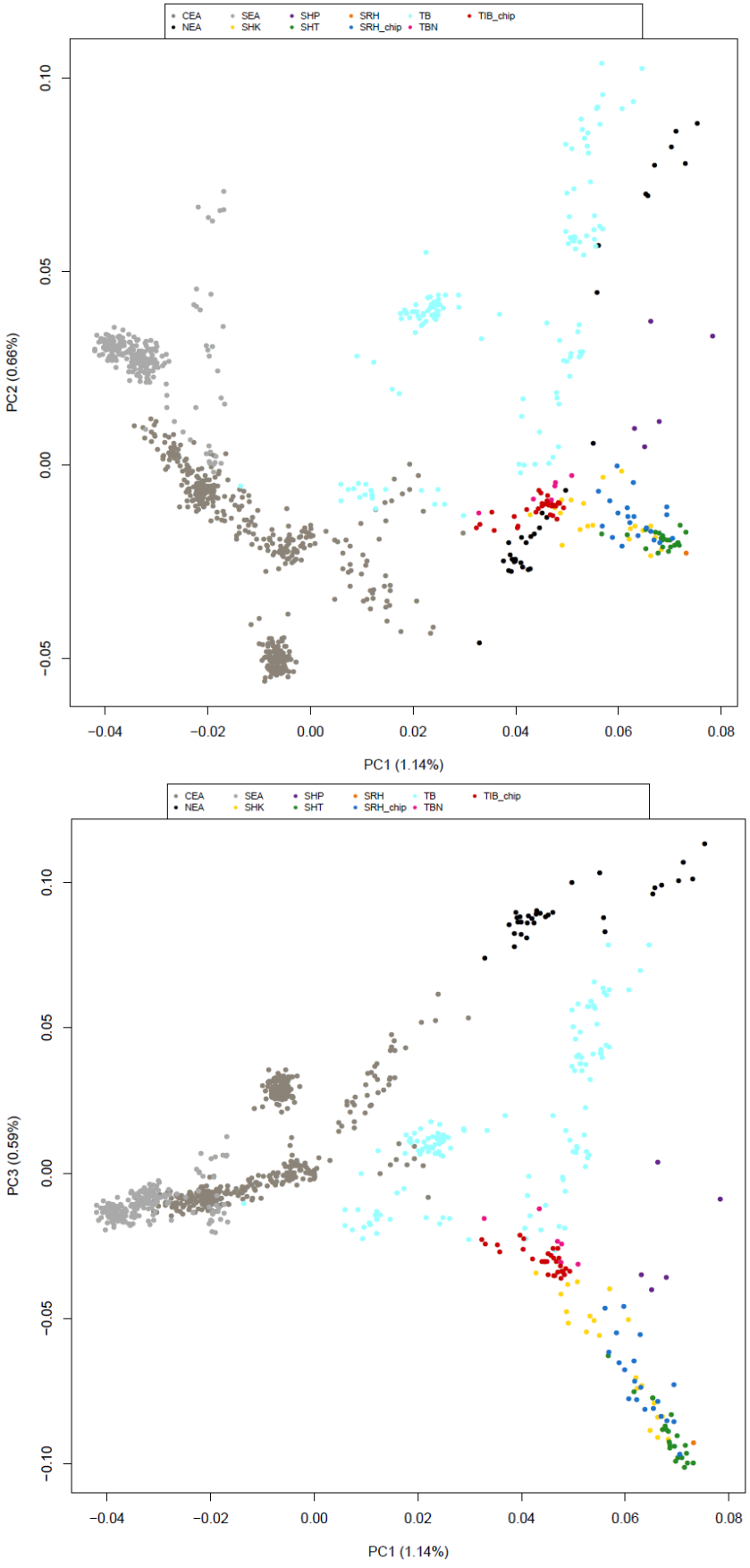


Figure 4.13. PCA performed on the East Asian groups of the "WGS-low-density" dataset. PC1 vs. PC2 (top) and PC1 vs. PC3 (bottom) are displayed.

We then performed ADMIXTURE analyses on the "WGS-low-density" dataset in order to observe the sharing of genetic components between the WGS and the SNP-chip Sherpa and Tibetan samples, as well as to infer the occurrence of potential admixture events involving South Asian ancestry components (Fig. 4.14).

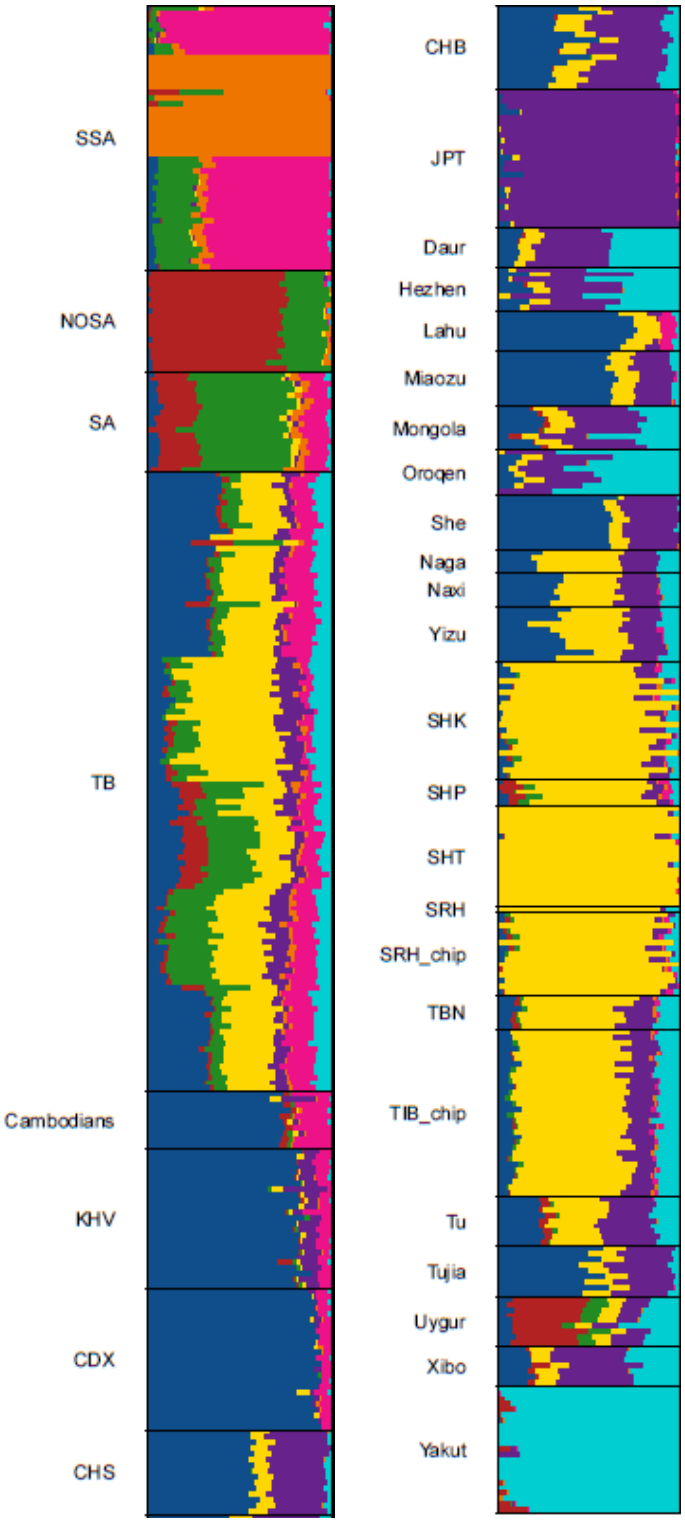


Figure 4.14. Ancestry proportions at K = 8 estimated with ADMIXTURE. ADMIXTURE analyses were performed on 540 individuals belonging to 32 Asian groups selected from the "extended" dataset.

In Fig. 4.14, the Asian populations are labelled according to the following macro-groups: South South Asia, SSA; North South Asia, NOSA; Tibeto-Burmans, TB.

Overall, distribution of South Asian and East Asian genetic components among the individuals considered in the analysis were in agreement with previous ADMIXTURE performed on the SNP-chip "extended" dataset (Fig. 4.3).

In particular, at K = 8, the one showing the lowest trend of CV-errors (Fig. 4.15), Tibetans and the SRH sample presented proportions of East Asian ancestry components comparable to those observed for the Tibetan (TIB_chip) and Sherpa (SRH_chip) subjects included in the SNP-chip dataset. On the contrary, the Sherpas from Tibet (SHP) showed overall a higher degree of South Asian admixture proportion as compared to the one displayed by the Nepalese Sherpas (SHK).

Two SHP individuals especially turned out to be outliers as regards their South Asian admixture proportions (more than 15% of South Asian ancestry), being the same two subjects that resulted outliers also in the PCA plot. Accordingly, we decided to remove these two samples form the WGS dataset used to detect potential signatures of positive selection.

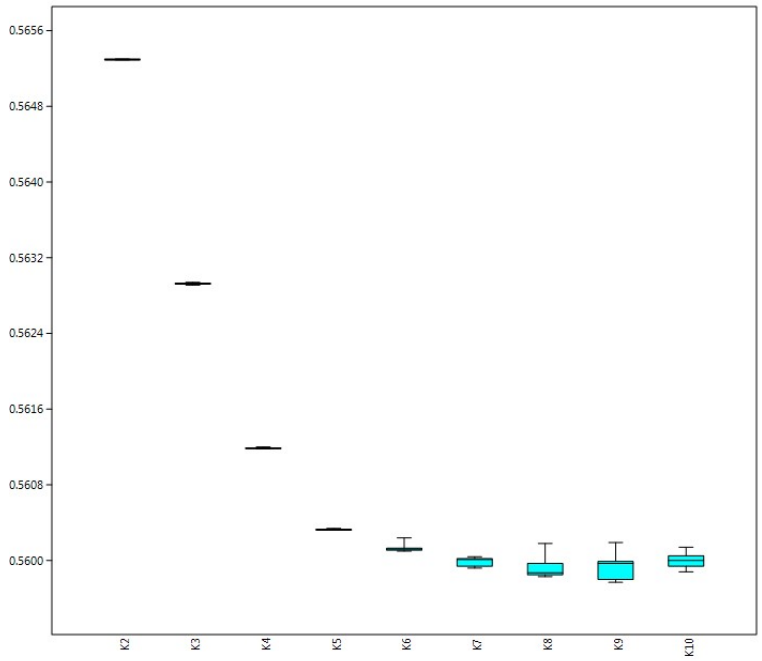


Figure 4.15. Cross-validation (CV) errors for ADMIXTURE reported in Fig. 4.14. CV were computed for all the 50 runs performed for each tested K. The best predictive accuracy was achieved by the model when eight ancestral components (K = 8) were tested.

4.1.8 Identifying homogeneous Tibetan/Sherpa genetic clusters

Previously described population structure analyses based on the SNP-chip dataset have pointed out a pattern of increasing drift and, partially, decreasing gene flow between Tibetans and the Sherpas. In particular, PCA and ADMIXTURE approaches described in section 4.1.7 have shown that overall the 12 individuals sequenced for the whole genome can be considered as good representatives of the genetic variability distributed at different locations along the Tibetan/Sherpa cline.

Nevertheless, recent studies have demonstrated that a gradient of genetic variation could in fact hide fine scale population structure, which is not identifiable with global population structure analyses (Leslie et al. 2015). According to this rationale, we thus obtained a high-resolution estimation of haplotype sharing between Tibetan and Sherpa individuals by applying the CHROMOPAINTER/fineSTRUCTURE pipeline. This enabled us to test whether along the Tibetan/Sherpa cline it is possible to identify significant subtle sub-structures (i.e. genetic clusters) or not.

Application of the algorithm implemented in fineSTRUCTURE to the "WGS-low-density" dataset thus led to the identification of 23 distinct clusters of samples (Fig. 4.16), even though it is noteworthy to mention that the dendrogram produced is the result of a hierarchical clustering function and should not be interpreted as a description of the phylogenetic relationships between the considered populations. Nevertheless, it is interesting to observe that overall East Asian populations turned out to be divided into three main groups: South East Asians, North East Asians and the Tibetan/Sherpa clade.

As regards this latter group, the algorithm actually identified a final number of seven clusters, three of which were represented by only one individual each. More interestingly, at this level of resolution it was not possible to identify clear boundaries between the two populations (i.e. Tibetans and Sherpas). In fact, although two clusters were mainly composed of Tibetans and the others exclusively of Sherpas, the 12 sequenced samples resulted spread in all of the seven identified clusters.

For this reason, we decided to consider the upper level of clustering in which all the Tibetans and Sherpas fall into a single cluster, which clearly distinguished them from the rest of East Asians.

According to these findings, the selection scans were performed on 10 WGS samples at once, thus considering them as part of an homogeneous group, after having excluded the two SHP subjects characterized by appreciable levels of South Asian admixture (as described in the previous section).

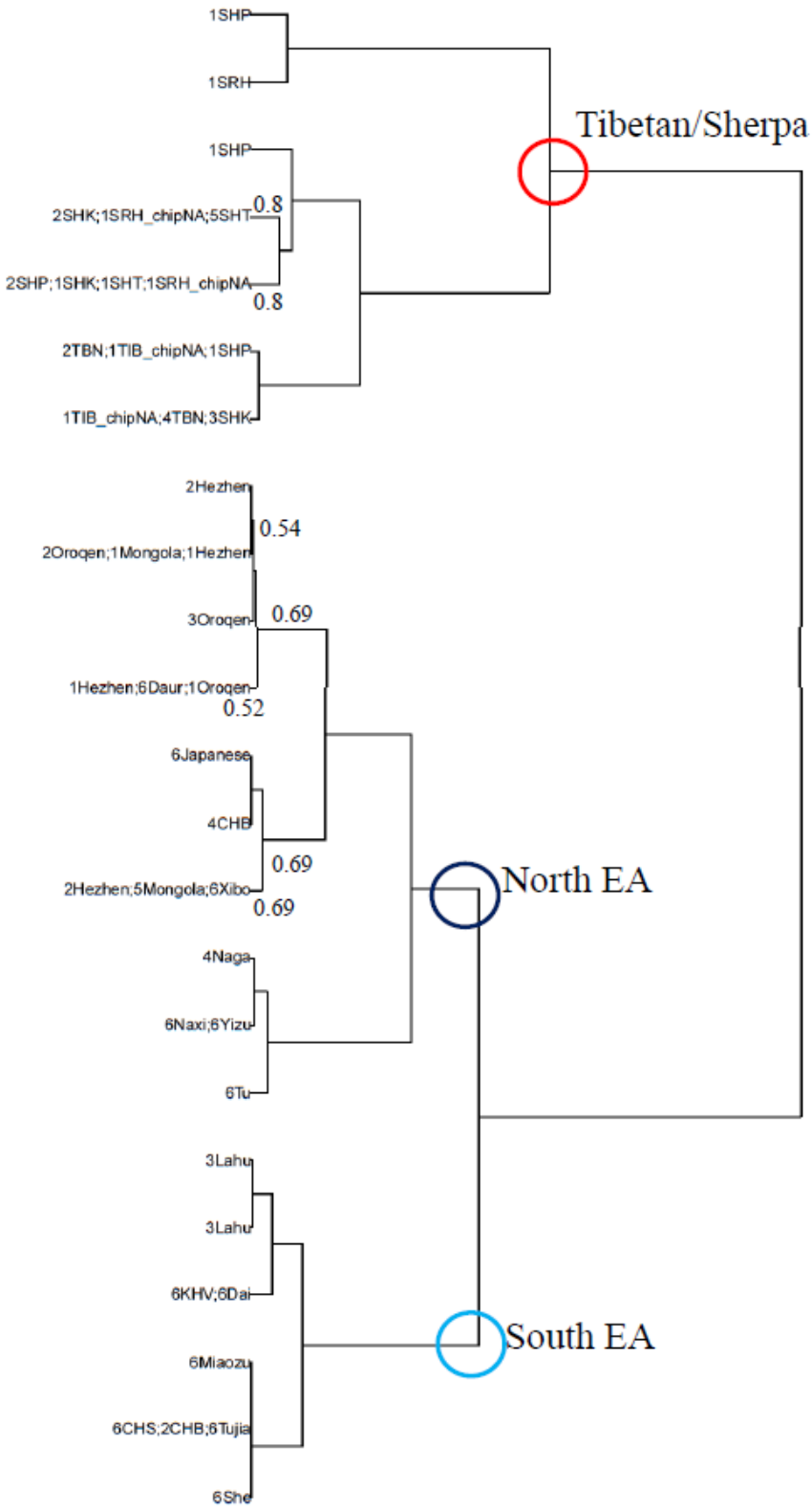


Figure 4.16. fineSTRUCTRE hierarchical clustering analysis performed on a subset of East Asian populations included in the "WGS-low-density" dataset. A total of 23 clusters were identified by the algorithm, although the upper level indicating three main clusters was considered for downstream analyses since it clearly distinguishes the Tibetan/Sherpa group from the rest of North and South East Asians. Only nodes with MCMC posterior probability lower than 1 are indicated on the dendrogram.

4.2 Dissecting the role of positive selection in shaping the Himalayan genomes

4.2.1 Genome-wide signatures of positive selection in GCA Tibeto-Burmans

The genome-wide scans performed on the "selection" SNP-chip dataset were based on statistical tests chosen in order to optimize the analysis of the generated data. In fact, with respect to the assembled WGS dataset, the SNP-chip panel contains a lower density of markers, but typed for a higher number of individuals belonging to the two Tibeto-Burman groups under consideration: a high-altitude population (SRH) and a closely related population living at intermediate altitudes (TAS). Moreover, it includes also an East Asian group living at low altitudes (Han Chinese from Northern China, CHB) and an outgroup population of European origins (CEU).

For this reason, we decided to exploit two complementary tests (i.e. PBS and XP-EHH, see the Materials and Methods section), which take advantages of the comparison between the putatively adapted group under study (in this case the two GCA Tibeto-Burman communities) and other populations that share part of their evolutionary history with it, but that have been not subjected to the same selective pressures (in our case hypobaric hypoxia).

The information provided by computation of these two statistics was finally combined in a F_{cs} Fisher score (see the Materials and Methods section) in order to reduce false positive rates. Moreover, we choose to perform these analyses separately on the high-altitude SRH group and on the intermediate altitude TAS population because for the latter, to our knowledge, no selection scans have been never conducted before. Furthermore, analyses aimed at reconstructing their demographic history (see section 4.1) showed that this group, while being closely related to both Tibetans and Sherpas, does not descend directly from this high-altitude branch of Tibeto-Burmans. Accordingly, we supposed that Tamangs do not share with the Sherpas patterns of genetic adaptation to hypobaric hypoxia and, if the selection scans would have confirmed this hypothesis, they could be thus considered as an optimal control population for studies focused on the Himalayan high-altitude groups.

As described in details in the Materials and Methods section, to get an overview of the genomic regions presenting the most robust signatures ascribable to the action of positive selection on the Sherpa/Tamang groups, we performed a windows-based filtering approach. In fact, we retained only the top 1% genomic windows showing the highest proportion of F_{cs} as plausible candidate regions that underwent positive selection. We then extracted the SNPs contained in the wider regions identified by merging consecutive top 1% windows and we produced Manhattan plots to graphically represent the obtained results (Fig. 4.17).

As regards patterns observed in the SRH population, top 1% genomic windows contained SNPs located on well-known candidate genes supposed to be involved in high-altitude adaptation, such as *EPAS1* and *EGLN1* (see the Introduction section). In particular, among these top 1% chromosomal regions, the one containing *EPAS1* was that showing the wider extension of the selection signal, with the final merged window being 320,000 kb long. Moreover, the one containing *EGLN1* was among those presenting the highest proportions of outlier SNPs (i.e. N-top/N-tot = 1), thus representing one the most significant results obtained (Appendix Table 4).

As extensively mentioned in the Introduction section, several studies identified these two genes as the most likely ones to have been subjected to strong positive selection in high-altitude Tibetans and Sherpas. In fact, variants on these loci have been found to be associated with low hemoglobin concentration (Simonson et al. 2010; Beal et al. 2010; Yi et al. 2010). Interestingly, our list of top candidates did not include the *PPARA* gene, whose variants have been as well found to be associated with low levels of hemoglobin (Horscroft et al. 2017), but showing significant signals of positive selection only in one genome-wide study (Simonson et al. 2010).

Among the top 1% windows identified, other outstanding outlier signlas were detected on a 280,000 kb region of chromosome 12 and on a 176,000 kb region of chromosome 16 (Fig. 4.17). The first one contained only one known gene (*IQSEC3*), which encodes for the IQ motif and SEC7 domain-containing protein 3 involved in guanyl-nucleotide exchange factor activities in neurons (The UniProt Consortium 2016; window_9 of chr 12, see Appendix Table 4). Furthermore, in this region are present a promoter sequence that was found active in Human Umbilical Vein Endothelial Cells (HUVEC) and an enhancer sequence that was find to be active in several cell lines including HUVEC, placenta, dermal, macrophages and lung cell lines (http://grch37.ensembl.org/index.html). Instead, the candidate region on chromosome 16 contains five protein coding genes, four of which encode for DNA binding zinc-finger proteins (*ZSCAN10*, *ZNF205*, *ZNF213*, *ZNF200*) and the fifth encodes for a Caspase (*CASP16*; The UniProt Consortium 2016). According to our window-based approach, this region resulted included in a wider window of 310,000 kb (window_1 of chr 16, see Appendix Table 4), in which other eight upstream and four downstream genes are contained. Nevertheless, these genes present low F_{cs} scores, meaning that the high significance (i.e. N-top/N-tot = 0.75) of the window was driven by the subregion described above. However, one of these loci is worth mentioning since encodes for the tumor necrosis factor receptor superfamily member 12A (TNFRSF12A), which is known to be involved in angiogenetic functions (The UniProt Consortium 2016). Anyway, none of the genes of chromosomes 12 and 16 decribed so far was present in the subnetworks analyses detailed below.

Contrarily to what observed for SRH, in the top 1% genomic regions identified for the TAS group, none of the genes belonging to the hypoxia-inducible factor (HIF) pathway were present (Appendix Table 5). The absence of significant genomic signatures including the two main candidate genes involved in high-altitude adaptation (i.e. *EPAS1* and *EGLN1*), thus supported results from demographic inferences indicating that Tamangs do not descend directly from the Tibetan/Sherpa lineage of Tibeto-Burmans.

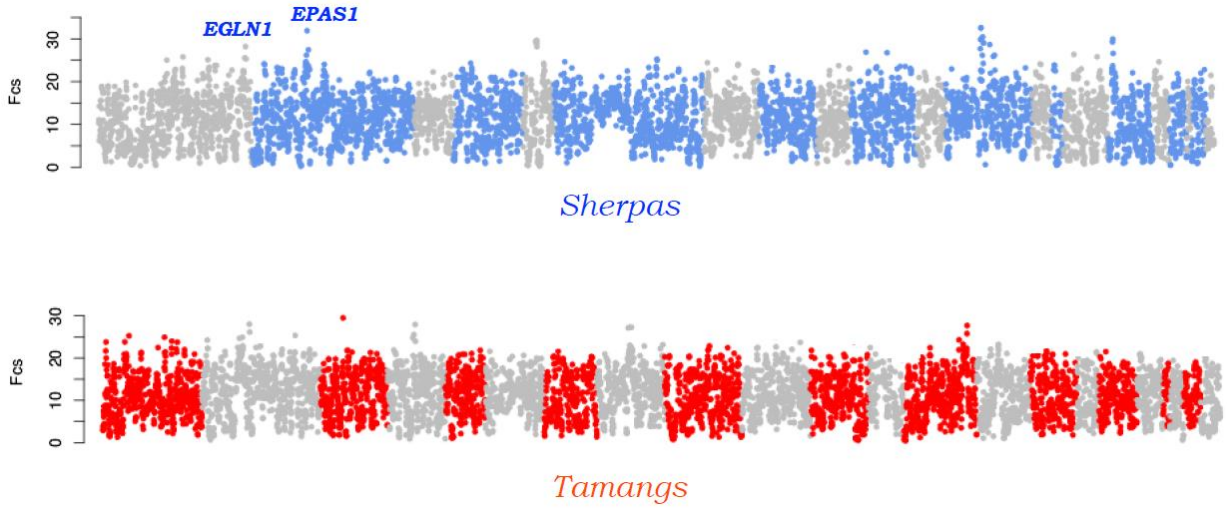


Figure 4.17. Manhattan plots of F_{cs} scores for SNPs contained in the top 1% genomic regions identified as targets of positive selection. Top and bottom plots refer to the F_{cs} scores obtained for the SRH and TAS group, respectively. Among the regions identified in SRH are present the known candidate genes that have been targeted by positive selection in Sherpas/Tibetans. Instead, TAS do not present signals around the *EGLN1* and *EPAS1* loci.

4.2.2 Fine-mapping of functional pathways involved in high-altitude adaptation

The results described in section 4.2.1 were mostly intended to serve as replications of the signatures of positive selection already observed in other Sherpa populations and in Tibetans, as well as to verify the presence or absence of such signals in Tamangs.

Moving from these preliminary results, we then attempted to identify new functional pathways having played a role in shaping the complex adaptive traits under study. In fact, the traditional approach described in section 4.2.1 considered genomic regions independently and was thus able to pinpoint only major selective events occurred according to the hard sweep model (i.e. underling selection acting on genes involved in a part of the adaptation process, for instance low hemoglobin concentration in bloodstream that protects from CMS (Buroker et al. 2012). Nevertheless, which are the genes responsible for the main processes allowing high-altitude populations to respond to hypoxic conditions is still unclear (Beall, 2007).

Given the complexity of the physiological traits observed in the Himalayan highlanders, it is highly plausible that positive selection has acted at a polygenic level, thus targeting groups of genes involved in the same pathway rather than one or two genes independently (Jeong & Di Rienzo 2014). Unfortunately, genomic signatures related to this type of selection can hardly be identified by most of the statistical methods developed so far (Pritchard & Di Rienzo 2010).

For this reason, we applied separately on the SNP-chip and WGS datasets the gene network-based approach implemented in the *signet* R package (Gouy et al. 2017), as described in detail in the Materials and Methods section. In particular, for all the following described tests, we considered as significant the highest scoring subnetworks showing a p-value < 0.05 (Table 4.5).

Table 4.5 Significant subnetworks identified by the *signet* analyses for all the selection tests performed. In bold are reported the integrin pathways, which were identified as significant in all the tests performed on high-altitude populations.

SRH *Fcs* (SNP-chip)					
Pathway	**Pat. size**	**Subnetwork size**	**HSS**	**p-value**	**Genes**
Direct p53 effectors	124	16	7.717325185	0.003416591	APC FAS BCL2 BDKRB2 VCAN MSH2 PMS2 PRKAB1 SPP1 TAP1 TP53 TP73 VDR TP63 PERP BCL2L14
Beta1 integrin cell surface interactions	55	8	7.519118343	0.004541307	ITGA6 ITGA1 ITGA2 ITGA3 LAMA3 LAMB3 LAMC1 LAMC2
p73 transcription factor network	72	11	7.000933873	0.009528256	FAS CHEK1 EP300 IL1RAP TP73 TP63 BCL2L11 UBE4B NSG1 WWOX RNF43
Validated transcriptional targets of TAp63 isoforms	52	5	6.71858042	0.014005899	FAS DST EP300 ITGA3 TP63
HIF-1-alpha transcription factor network	63	16	6.280723155	0.025953356	ARNT HK2 HMOX1 SMAD3 FURIN PGM1 ABCB1 RORA CXCL12 TERT VEGFA ABCG2 CITED2 HDAC7 EGLN1 EGLN3
Alpha6 beta4 integrin-ligand interactions	11	4	6.128575324	0.032107463	ITGA6 LAMA3 LAMC1 LAMC2
HIF-2-alpha transcription factor network	26	4	6.045622129	0.035948475	EPAS1 TWIST1 VEGFA ABCG2
Glucocorticoid receptor regulatory network	73	9	5.985263325	0.038558666	CDK5 NR3C1 GSK3B IL13 POMC PRL SGK1 TP53 NCOA2
IL1-mediated signaling events	29	7	5.947788633	0.040829319	IL1R1 IL1RAP IL1RN TRAF6 UBE2V1 ERC1 IRAK4
TAS *Fcs* (SNP-chip)					
Pathway	**Pat. size**	**Subnetwork size**	**HSS**	**p-value**	**Genes**
Direct p53 effectors	124	18	7.773539815	0.001758586	APC FAS BAX BDKRB2 PRDM1 CASP10 ARID3A MDM2 MSH2 PML TAP1 TGFA TP53 TP73 RPS27L C12orf5 NLRC4 AIFM2
p73 transcription factor network	72	3	6.888385209	0.00699197	FAS JAG2 TP73
TCR signaling in naïve CD4+ T cells	55	8	6.053346662	0.02527703	CD4 CSK FYB FYN ITK PTPN6 STIM1 ORAI1

PDGFR-beta signaling pathway	99	2	6.046901389	0.025510096	CSK SRC
Signaling events mediated by PTP1B	44	2	6.046901389	0.025510096	CSK SRC
Regulation of nuclear beta catenin signaling and target gene transcription	64	10	5.838777803	0.033858084	APC CCND2 CDX1 CTNNB1 EP300 CACNA1G CBY1 MDFIC CHD8 KCNIP4
ATF-2 transcription factor network	57	11	5.694949776	0.040998369	ARG1 ATF2 EP300 IL6 INS PRKCA SELE TH SOCS3 DUSP10 JDP2
Netrin-mediated signaling events	29	7	5.639028789	0.043985847	DAPK1 DCC DOCK1 FYN UNC5C UNC5A UNC5B
Signaling events regulated by Ret tyrosine kinase	37	4	5.638760161	0.043985847	GFRA1 PRKCA RAP1A SRC
Signaling events mediated by VEGFR1 and VEGFR2	63	11	5.633786236	0.044303663	CTNNB1 DNM2 FLT1 FYN KDR PLCG1 PTPN2 PTPN6 SRC VEGFA MYOF
TCR signaling in naïve CD8+ T cells	44	2	5.5854302	0.046613132	CSK HLA-ABC

nSL (WGS)

Pathway	Pat. size	Subnetwork size	HSS	p-value	Genes
Beta1 integrin cell surface interactions	57	7	7.245702605	0.001163775	COL6A1 ITGA6 ITGA1 ITGA2 LAMC1 LAMC2 THBS2
Alpha6 beta4 integrin-ligand interactions	11	4	6.510302058	0.004020313	ITGA6 LAMC1 LAMC2 LAMA1
Integrins in angiogenesis	57	7	6.395638798	0.005448582	CASP8 COL3A1 COL5A2 COL6A1 COL11A1 COL11A2 ITGB3
C-MYB transcription factor network	74	19	5.738368102	0.015605163	ANPEP CD34 CREBBP ELANE HES1 LECT2 KITLG MYB MYC MYF6 PAX5 SLC25A3 PRTN3 SMARCA2 SPI1 MAD1L1 IQGAP1 SND1 NLK
Stabilization and expansion of the E-cadherin adherents junction	39	6	5.282638253	0.031633517	ACTN1 CTNNA1 EFNA1 EPHA2 NCK1 LIMA1
Validated targets of C-MYC transcriptional activation	74	14	5.179334645	0.037611088	BCAT1 CCND2 HMGA1 MYC ODC1 SERPINI1 SLC2A1 TAF4B TERT TFRC SUPT3H RUVBL1 SUPT7L CDCA7
Direct p53 effectors	123	14	5.152697245	0.039462548	BAK1 BID PRDM1 BNIP3L ARID3A EPHA2 MLH1 TP53 CCNK CARM1 SESN1 DDIT4 STEAP3 TIGAR

DIND (WGS)

Pathway	Pat. size	Subnetwork size	HSS	p-value	Genes
CDC42 signaling events	62	5	10.3963464	0.00446476	CDC42 CTNNA1 CDC42BPA ARHGEF7 DIAPH3
Validated nuclear estrogen receptor alpha network	63	3	9.393763964	0.011108749	ESR1 NCOA1 UBA3
Integrins in angiogenesis	58	4	9.232414005	0.012969066	COL4A3 COL11A2 ITGAV ROCK1
Plasma membrane estrogen receptor signaling	24	2	8.548259895	0.022376953	ESR1 SOS1
Nephrin/Neph1 signaling in the kidney podocyte	20	3	8.382480017	0.02561922	TRPC6 NPHS2 CD2AP
Beta1 integrin cell surface interactions	58	2	8.369084907	0.025725524	COL11A1 COL11A2
Syndecan-1-mediated signaling events	31	2	8.369084907	0.025725524	COL11A1 COL11A2
p73 transcription factor network	76	4	8.25108448	0.027745296	CCNA2 TP73 WT1 CLCA2
FOXA1 transcription factor network	33	3	8.108776149	0.030190284	CYP2C18 ESR1 FOXA1

Signet analyses conducted on the SRH population included in the "selection SNP-chip" dataset first confirmed the results obtained by single-gene approaches (i.e. PBS and XP-EHH) pointing to the identification of significant subnetworks belonging to the HIF-1α and HIF-2α pathways and containing respectively *EGLN1* and *EPAS1* among those highlighted by the HSS algorithm. These two subnetworks are partially overlapping and are composed of a total of 18 genes (Fig. 4.18).

Among the other significant subnetworks identified for SHR, three were associated to the wide pathway of the P53 protein family of transcription factors and tumor suppressors (i.e. P53, P63 and P73 subnetworks, for a total of 28 genes). These master regulators control dozens of functions related to tissue and organism development, cell growth and apoptosis, being also involved in several pathologies (Levrero et al. 2000).

Interestingly, other two significant SHR subnetworks were highly nested and overlapping as both are part of the integrin pathways (i.e. integrin β-1 and integrin α6-β4) comprehending a total of eight genes encoding for integrin subunits, integrin ligands and effectors. Integrins are known to be involved in regulation of cell adhesion and motility in many tissues (Tulla et al. 2001). Moreover, *ITGA6* (encoding for integrin α6-β4), which was part of both the integrin networks identified, is an important receptor on platelets and is proved to play a role in angiogenesis (Avraamides et al. 2008).

Finally, other subnetworks identified by such an analysis include the glucocorticoid receptor regulatory network for a total of nine genes and the interleukin-1 signaling events for a total of seven genes (Kandasamy et al. 2010).

According to these findings, with the exclusion of the P73 pathway, the SRH and TAS populations did not present any common pathway among the identified significant subnetworks.

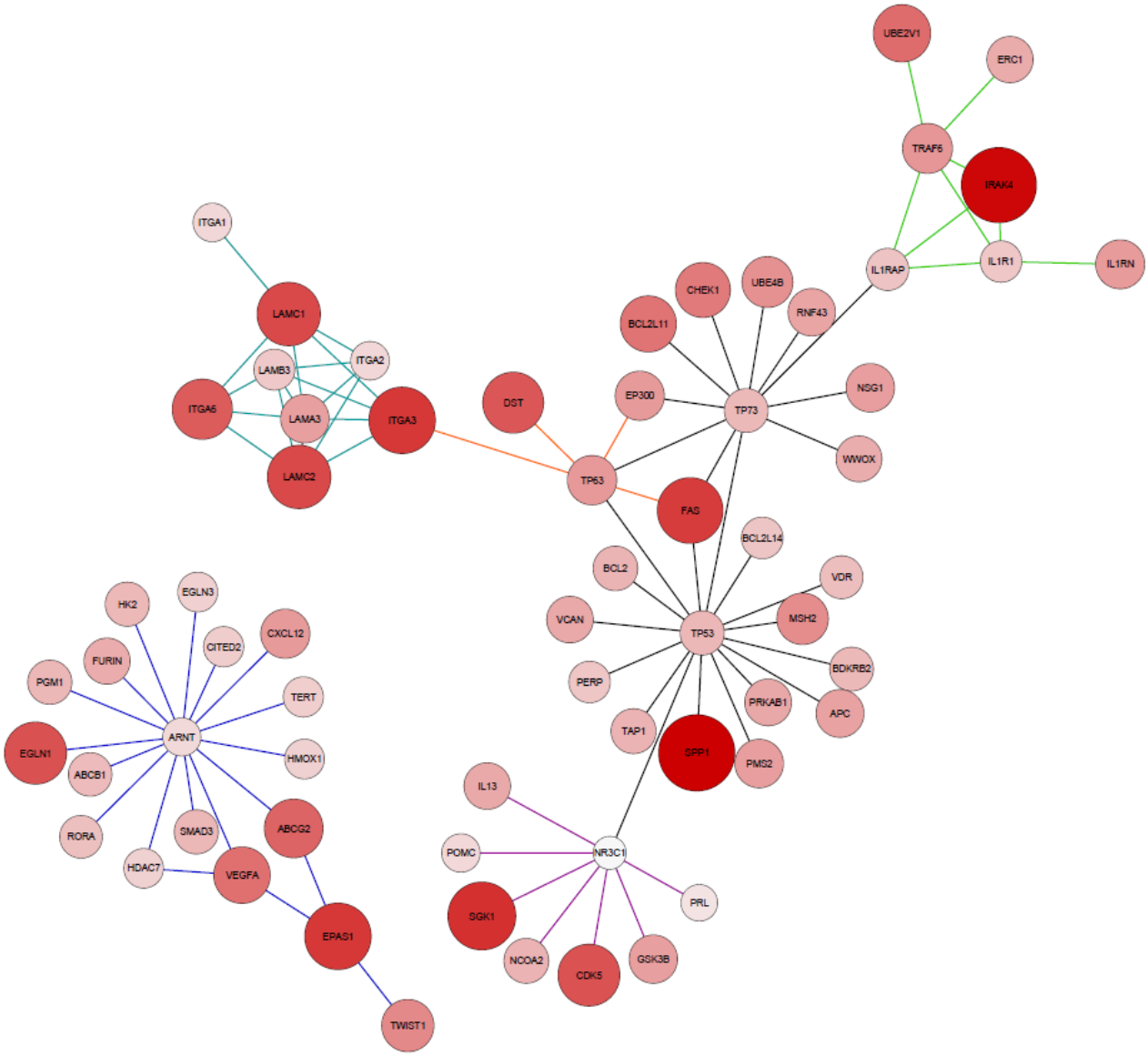


Figure 4.18. Significant subnetworks identified for the SRH population included in the "selection SNP-chip" dataset. Genes belonging to the HIF-1α and HIF-2α pathways are connected by blue lines. Integrin pathways (β-1 and α6-β4) are highly nested and connected by aquamarine lines. The glucocorticoid receptor regulatory network and the interleukin-1 signaling network are connected by purple and green lines, respectively. The P63 pathway is connected by orange lines, while the P53 and P73 are connected by black lines. The color intensity and size of the circles around the genes are proportional to the gene F_{cs} score.

As regards the "WGS-high-density" dataset, we performed the network analysis separately according to the calculated nSL and DIND scores, given the theoretical differences between the signatures of selection identifiable by the two statistics (see the Material and Methods section).

In particular, when the nSL scores were considered, seven subnetworks were identified as significant. Interestingly, the top three (i.e. presenting the lowest p-values) belonged to integrin-

associated pathways highly overlapping and nested between them, for a total of 14 genes (Fig. 4.19).

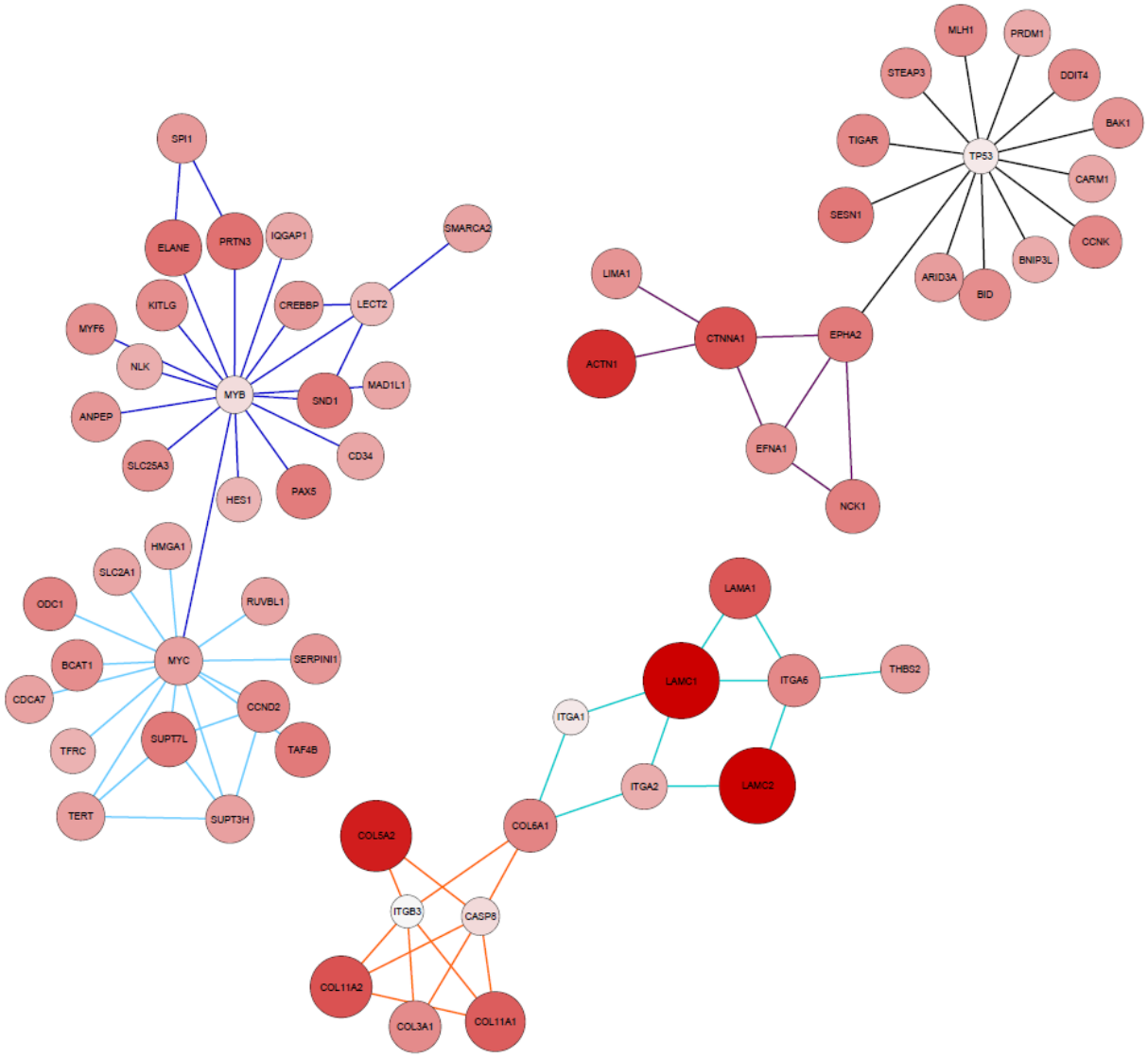


Figure 4.19. Significant subnetworks identified according to WGS data using the calculated nSL scores. Genes belonging to the integrin pathways (β-1 and α6-β4) are connected by aquamarine lines, while the integrin involved in angiogenesis pathway is connected by orange lines. Dark and light blue lines are connecting genes belonging the C-MYB and C-MYC pathways, respectively. The E-cadherin adherents junction pathway is indicated by purple lines and the P53 pathway is connected by black lines. The color intensity and size of the circles around genes are proportional to the gene nSL score.

Of these subnetworks, two related to the same integrin β-1 and integrin α6-β4 pathways identified in the SRH group according to the SNP-chip data (with *LAMC1*, *LAMC2*, *ITGA6*, *ITGA1* and *ITGA2* as common genes). The third one instead belonged to the pathway of integrin directly involved in angiogenesis (Avraamides et al. 2008).

Other two linked subnetworks belonged to the C-MYB and C-MYC transcription factor pathways, with a total of 32 genes identified. These oncogene transcription factors are known to be involved in several functions, mostly promoting the proliferation and differentiation of hematopoietic progenitor cells and sprouting angiogenesis, respectively (The UniProt Consortium 2016).

Among the last two significant subnetworks, one belonged to the stabilization and expansion of the E-cadherin adherents junction pathway (six genes) and the other to the P53 pathway (14 genes).

Finally, network analysis was performed according to results obtained from the DIND test, leading to the identification of eight significant subnetworks (Fig. 4.20).

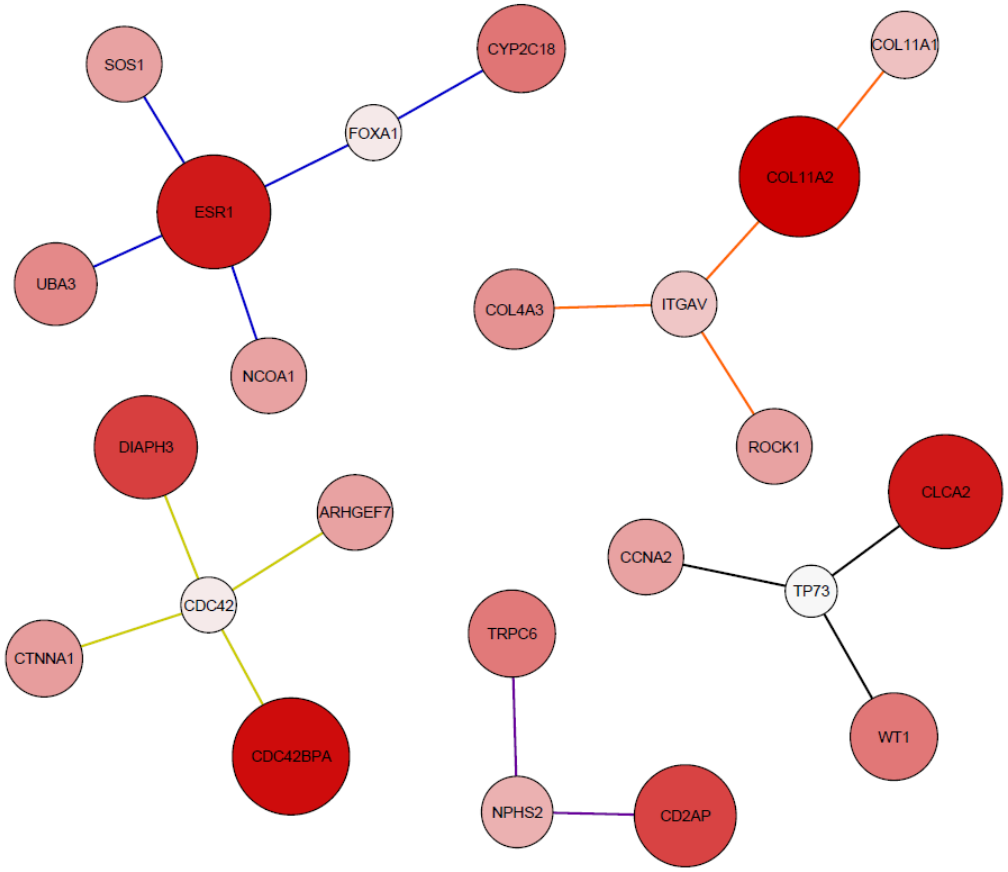


Figure 4.20. Significant subnetworks identified according to WGS data using the calculated DIND scores. Genes belonging to the integrin pathways (β-1 and integrin in angiogenesis) are connected by orange lines. Dark blue lines are connecting genes belonging different pathways associated to the *ESR1* gene. The CDC42 pathway is indicated by pale green lines. The Nephrin/Neph1 patway is connected by purple lines and the P53 pathway is connected by black lines. The color intensity and size of the circles around genes are proportional to the gene DIND score.

Interestingly among these subnetworks, two were identified also according to nSL scores and are the ones belonging to the integrin β-1 and integrin in angiogenesis pathways, being represented by a total of five genes (with *COL11A1* and *COL11A2* as common genes between the nSL-based and DIND-based computations).

The other subnetworks included one belonging to the CDC42 signaling cascade (with five representative genes), which encodes for a GTPase binding to several different effectors involved in many cellular functions (The UniProt Consortium 2016).

Another subnetwork identified represented by three genes instead belonged to the Nephrin/Neph1 signaling pathway, which is involved in controlling the glomerular permeability (Liu et al. 2003).

Three subnetworks were then composed by a few genes each (two or three), but they were all linked by the common gene *ESR1* that encodes for an estrogen receptor. Interestingly, isoform 3 of the protein encoded by *ESR1* is known to be involved in activation of NOS3 and endothelial nitric oxide production (The UniProt Consortium 2016), which was found among the top candidate regions targeted by positive selection in the SRH population (Appendix Table 4).

The last subnetwork identified belonged to the wide P73 pathway already present among the significant pathways pointed out by several tests described above.

5. Discussion

Up to date, many researches have focused on the Asian populations dwelling immediately North of the Himalayas in the attempt to investigate the demographic, biological, and cultural processes responsible for the successful diffusion of *H. sapiens* on the Tibetan plateau (Su et al. 2000; Aldenderfer & Yinong 2004; Zhao et al. 2009; Qin et al. 2010; Aldenderfer 2011; Rhode 2016).

In such a context, the Himalayan arc was supposed to have played a crucial role in shaping the history of human populations and the related patterns of genetic variation in continental Asia. In particular, it has been proposed to have acted as a geographic barrier that contributed to maintain the major cultural and genetic differences observable between South Asian and East Asian populations (Cordaux et al. 2004; Gayden et al. 2007; Gayden et al. 2013).

Nevertheless, the regions bordering the southern Himalayan slopes, together with the high-altitude transverse valleys that connect them to Tibet, seem to have witnessed a few migrations of groups speaking Tibeto-Burman languages, which have plausibly originated on the Tibetan plateau and that have then crossed the cordillera (Gayden et al. 2007; Bhandari et al. 2015; Jeong et al. 2016). Therefore, the composite anthropological picture of these southern Himalayan areas represents an excellent example of a cultural and biological melting pot where groups speaking Indo-Aryan languages and related to the South Asian Hindu culture coexist with Tibeto-Burman groups of East Asian origins, such as the Sherpas and Tamangs.

That being so, by generating genome-wide and uniparental data for previously unsurveyed Sherpa and Tamang communities from the Nepalese GCA, we first aimed at setting these populations into the context of the overall Tibeto-Burman genomic landscape. This succeeded to improve the reconstruction of the complex history of such a heterogeneous population group and to trace the most plausible sources of East Asian ancestry in South Asian Tibeto-Burmans. The obtained results provided the first genomic description of the above-mentioned GCA anthropological mosaic and shed new light on the colonization processes occurred on the southern slopes of the Himalayas.

Furthermore, some of these southern Himalayan populations share the extreme high-altitude environment with the East Asian groups living beyond the Northern slopes of the Himalayas. For a few of these populations (i.e. the Sherpas from the Nepalese Khumbu district), biological adaptation to hypobaric hypoxia has been attested from a physiological perspective and also results from selection scans have provided genomic hints about it. Conversely, as regards other Nepalese Sherpa communities (e.g. the Sherpas from the Rolwaling valley) and other ethnic groups (e.g. the Tamangs), no genome-wide studies have been ever conducted to further dissect this aspect.

More importantly, to date the genetic bases of the Tibetan/Sherpa adaptation to high-altitude have been only partially elucidated. In fact, several studies replicated the findings pointing to a strong hard selective sweep occurred in the Tibetan/Sherpa genomic regions encompassing the *EPAS1* and *EGLN1* genes. This is proved to be involved in smoothing the erythropoietic cascade that represents the ancestral response (i.e. present in all low-altitude human populations) to cope with hypobaric hypoxia (Simonson et al. 2012). Nevertheless, such a mechanism of protection against the long-term harmful effects of the polycythemia due to physiological acclimatization to high-altitude only partially explains the complex adaptive phenotype observed in Himalayan populations, which in facts involves several other biological adjustments allowing the Tibetans and Sherpas to live without harmful consequences in areas locate even above 4,000 m a.s.l (Gilbert-Kawai et al. 2014).
Therefore, we finally aimed at describing patterns of genome-wide signatures of positive selection in the Rolwaling Sherpas and Tamangs. Moreover, we assembled a high-density dataset composed of one newly generated 20-X whole genome sequence of a Rolwaling Sherpa individual and of publicly available high coverage whole genome sequences of Tibetans and Sherpas from Tibet (Lu et al. 2016). According to this approach, and by exploiting a recently developed gene network-based method (Gouy et al. 2017), we thus aimed at identifying subtle signals ascribable to the occurrence of polygenic adaptation to hypobaric hypoxia in these Himalayan groups, to put a step forward the dissection of the biological mechanisms that enable human occupation of high-altitude environments.

5.1 Dissecting the GCA genomic landscape

When contextualizing the collected samples into broad patterns of Asian genomic diversity, high heterogeneity was first revealed for IAR (Fig. 4.2), mainly due to extreme inter-individual variability in their South Asian/East Asian ancestry proportions (Fig. 4.3). Interestingly, despite admixture was proposed to be less common in IAR groups due to the strict endogamous rules imposed by the caste system to Hindu populations (Moorjani et al. 2013; Basu et al. 2016), this finding suggests that Indo-Aryan speaking communities residing close to the Himalayas have experienced appreciable gene flow from populations of East Asian origins, such as the Nepalese Tibeto-Burmans.
In contrast, these latter groups appeared to have maintained remarkable internal homogeneity, with limited genetic exchange with people of South Asian ancestry, which turned out to be most pronounced in TAS. In particular, patterns of South Asian/East Asian admixture confirmed by the *f3* and ALDER approaches (Appendix Table 2 and Appendix Table 3) suggested that the harsh

Himalayan environment has played, and still plays, an important role in limiting gene flow along a low- to high-altitude cline of geographical (more than cultural) isolation.

For instance, TAS live in a village located at around 2,000 m a.s.l., but relatively well connected with low-altitude Indo Aryan settlements through a widely known trail. Accordingly, they presented more than a twofold proportion of South Asian admixture (23%) than TAT (9%), which instead reside at higher altitude in a more isolated valley. Such a reduced South Asian ancestry of TAT could explain also their lack of admixture signatures provided by the *f3* test. Despite that, the ALDER algorithm succeeded in identifying admixture events in both of the considered Tamang communities, with the estimated dates of admixture ranging from 0.65 to 0.13 thousand years ago, in line with the hypotheses of a recent historical arrival of the first Tamang tribes in Northern Nepal (Campbell 1997).

More accurate tests aimed at drawing demographic inferences were then performed by calculating the outgroup *f3* and D-statistics. This revealed that, in addition to different admixture patterns, also a recent shared history could be ruled out for the Rolwaling Sherpa and Tamang groups, although they showed a similar profile as regards Y-chromosome lineages (Table 4.2). Accordingly, we hypothesized that Tamangs reached their present-day GCA settlements along more complex migratory routes than those followed by the Sherpas and that did not simply entailed the crossing of high-altitude Tibetan/Nepalese passes. In fact, while Rolwaling Sherpas were confirmed to be genetically similar mostly to other Sherpa groups from Khumbu and then to Tibetans, Tamangs turned out be more closely related to the low-altitude Tibeto-Burman tribe of Nagas from Northeastern India than to Sherpas or Tibetans (Fig. 4.8 and Fig. 4.9).

Interestingly, when the shared genetic drift was measured between GCA samples and a large set of South Asian/East Asian groups, the Nagas were found to be responsible for top outgroup *f3* scores in almost all the performed comparisons. These findings suggested that Tamangs did not derive their East Asian ancestry components directly from high-altitude East Asians and that this could be true also for the other Tibeto-Burman groups residing in South Asia. In particular, the pattern observed in the GCA is consistent with the hypothesis that the high-altitude specific genetic component remains restricted to Tibetans (e.g. those having occupied the Nepalese Mustang region, as described by Jeong et al. 2016) and Sherpas on the southern slopes of the Himalayas. This sheds new light on an open debate regarding the most plausible source of East Asian ancestry in Nepalese Tibeto-Burmans (Gayden et al. 2007; Wang et al. 2012; Gayden et al. 2013). In fact, contrary to previous hypotheses (Gayden et al. 2007), the obtained results proved that the Himalayas in most cases served as a barrier to gene flow even from North to South.

Unlike the Tamangs, Rolwaling Sherpas living up to 4,180 m a.s.l. did not present any significant signature of South Asian admixture according to both the *f3* and ALDER analyses. Moreover, they showed reduced mtDNA and Y-chromosome diversity with respect to a large set of South Asian/East Asian groups (Table 4.1), together with a limited number of predominant uniparental lineages (Table 4.2). This evidence was in line with those reported for the Sherpa communities of Khumbu (Bhandari et al. 2015), as well as with remarkable autosomal homozygosity (Fig. 4.6) and high values of average length of intra-population shared haplotypes (Table 4.3). Taken together these findings seem to corroborate the idea of prolonged isolation previously proposed for this ethnic group by historical and sociocultural studies (Oppitz 1974; Baumgartner 2015).

Moreover, demographic inferences based on whole genome sequences of two Sherpa individuals from Khumbu (Jeong et al. 2014) showed that this group did not experience the exponential growth in effective population size having characterized the last 30 thousands years-history of low-altitude East Asian populations, but instead went through a recent bottleneck. The CHROMOPAINTER analyses performed in the present study supported such a scenario by showing that genetic differences observed between Tibetans, Rolwaling Sherpas and Sherpas from Khumbu are also ascribable to the strong drift experienced by the Sherpas because of long-term isolation and reduced population size (Fig. 4.7).

In addition to this, a recent analysis of Tibetan cohorts from many sampling locations across the plateau, including the Tibetan sample in this study (from near Lhasa), also proved the existence of a continuum of East Asian/Tibetan gene flow along an East-West axis (Jeong et al. 2017), which further contributed to the genetic differentiation between the Tibetans and Sherpas.

5.2 Unraveling the genetic legacy of Tibeto-Burman populations

ADMIXTURE analyses conducted in the present study (Fig. 4.3) and in previous ones (Jeong et al. 2014) identified a "Sherpa-like" genetic component that reached 100% only in Sherpa individuals (especially in those from the Nepalese Khumbu region). However, such an ancestry fraction turned out to be not exclusive to the Himalayan populations, being instead shared among several other East Asian and South Asian groups. Interestingly, most populations showing significant proportions of this component speak Tibeto-Burman languages, with the highest values being observed for the Nagas and Tamangs (42%-52%), setting aside high-altitude people.

Therefore, we argued that these findings indicate a shared ancestry among all Tibeto-Burman speaking groups, which is currently in the highest proportion in the Sherpas and more diluted in

other populations due to gene flow with nearby non-Tibeto-Burman gene pools, as previously shown for Tibetans (Jeong et al. 2017) and for other Tibeto-Burmans in this study (Fig. 4.10).

Nevertheless, disentangling the thread of this genetic trace is challenging as both ADMIXTURE, f3 and ALDER tests have shown that most South Asian Tibeto-Burmans have experienced relatively recent and extensive gene flow from neighboring populations with South Asian ancestry (Appendix Table 2 and Appendix Table 3). Similarly, also East Asian Tibeto-Burmans (e.g. Yizu and Naxi) have considerably admixed with neighboring East Asian populations.

To properly consider these patterns of gene flow when reconstructing the genetic history of the examined Tibeto-Burman populations, we performed a series of *ad hoc* analyses specifically aimed at taking into account their admixed ancestry. In particular, we tested whether the bulk of southern Himalayan Tibeto-Burmans, in addition to Tamangs, derived their East Asian genetic components from low-altitude populations instead that from the high-altitude Tibetan/Sherpa lineage.

For this purpose, we first applied MIXMAPPER phylogenetic analyses that provided further evidence of a common origin of Tibeto-Burmans somewhere North of the Himalayas. Moreover, these tests showed that almost all admixed Tibeto-Burmans (from both South Asia and East Asia) were well fitted as results of admixture between the Nagas (or an internal node connecting them to the Tibetan/Sherpa clade) and other populations (Fig. 4.10). Accordingly, we propose that the Tibetan/Sherpa and the Nagas-related branches of Tibeto-Burmans split early in their history, and that, following this split, the ancestors of Nagas substantially contributed to the gene pools of many present-day populations, such as Tamangs, Naxi and Yizu. Moreover, it is noteworthy to mention that a single split from a common ancestor cannot explain the divergence between the Tibetan/Sherpa and the Nagas-related clades, as formally tested by computing *ad hoc* f4 and D-statistics (Table 4.4). Conversely, this suggests that an early gene flow with non-Tibeto-Burman East Asians may have substantially contributed to such a population divergence (Fig. 5.1).

Nevertheless, ancestry specific PCA conducted on the East Asian genome portions of Tibeto-Burmans identified by HAPMIX confirmed that all these groups form a cluster that departs from the known latitudinal gradient of variation characterizing continental East Asia (Fig. 4.12). These findings were corroborated also by the outgroup f3 statistics computed on the masked dataset (Fig. 4.12). In fact, they supported a close genetic similarity of all Tibeto-Burman populations, in line with mtDNA, Y-chromosome and linguistic evidence (Yao et al. 2002; Wen et al. 2004; Gayden et al. 2009; Stoneking & Delfin 2010), suggesting that they harbor a peculiar genetic makeup that was not described so far in East Asian populations.

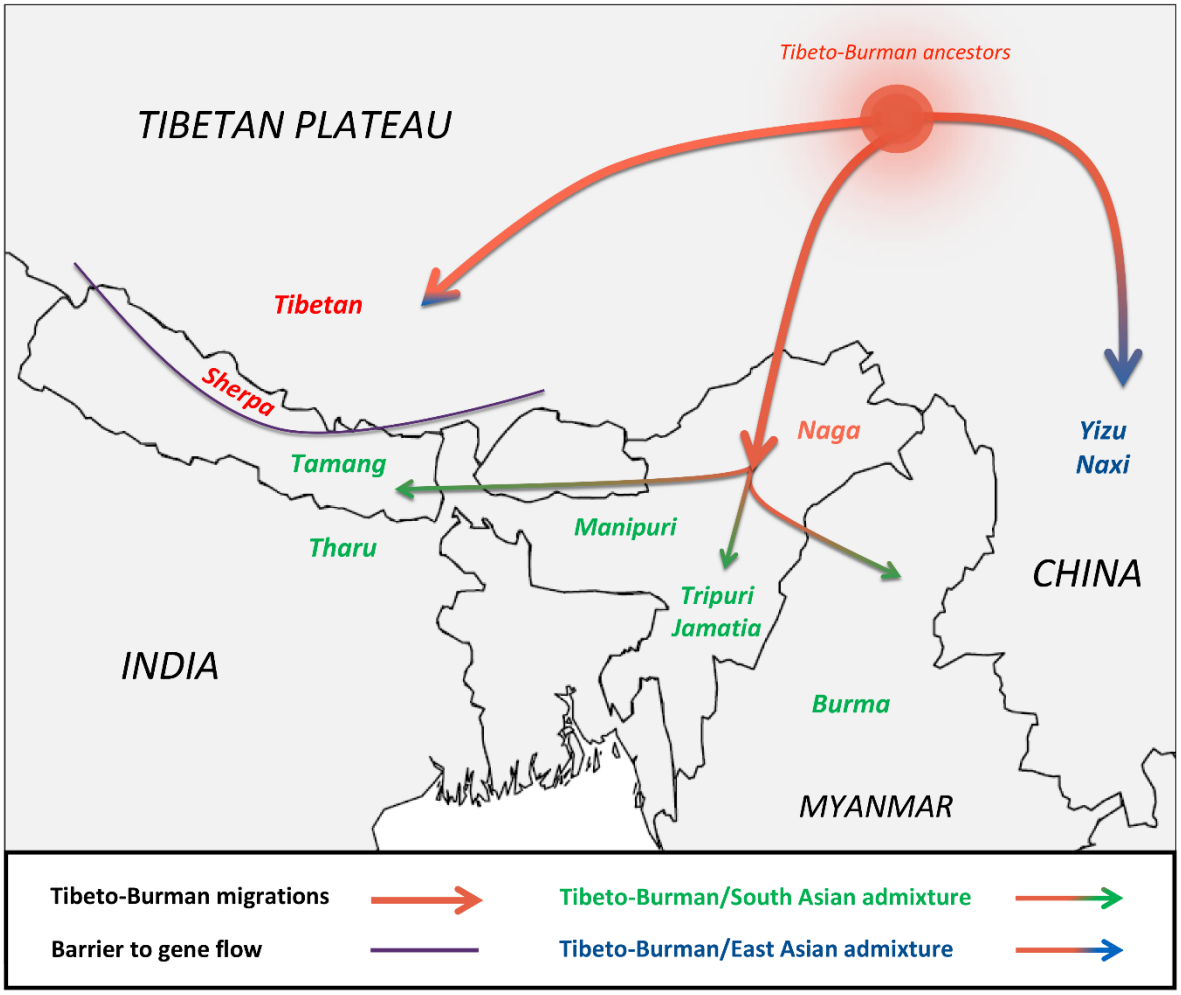


Figure 5.1. Map showing the hypothesized Tibeto-Burman migratory and admixture events as inferred by MIXMAPPER phylogenetic analyses and by isolation of the East Asian genome portions of low-altitude Tibeto-Burmans via HAPMIX. The peculiar genetic makeup of Tibeto-Burmans is proposed to have derived from an ancestral gene pool plausibly originated North of the Himalayas and progressively reshaped by historical admixture with East/South Asian non-Tibeto-Burman populations along the migratory routes that diffused Tibeto-Burman ancestry components from China to Nepal across India and/or Myanmar.

5.3 Divergent adaptive evolution of Sherpa and Tamang populations

Demographic inferences discussed in the previous section mainly pointed out that the Tibetan/Sherpa and the Nagas-related clades represent distinct branches of the Tibeto-Burman phylogeny that diverged early. In particular, we succeeded to prove that the low-altitude branch subsequently contributed to the gene pool of many present-day Tibeto-Burman populations, among which the Tamangs (Fig. 5.1).

To further validate such an evidence from a different perspective, we compared the adaptive history of the Sherpa and Tamang groups by performing haplotype-based neutrality tests (i.e. PBS and XP-EHH) aimed at identifying genomic signatures ascribable to the occurrence of hard selective sweeps.

Accordingly, we confirmed that positive selection acted on the genomic regions encompassing two well-established candidate genes, *EPAS1* and *EGLN1*, which have been previously proposed to have played a major role in high-altitude adaptation of the Himalayan populations (Fig. 4.17). In particular, adaptive variants at these loci were found to be associated with low levels of hemoglobin concentration according to several studies performed on different cohorts of Tibetan and Sherpa individuals (Beall et al. 2010; Simonson et al. 2010; Yang et al. 2017).

Interestingly, the same pipeline of analysis applied on the Tamang population sample did not result in the identification of the top candidate genomic regions putatively targeted by natural selection in the Sherpas. This indicates that these two groups diverged plausibly several thousand years ago (i.e. at the beginning of the natural history of Tibeto-Burmans) due to the long time necessary for the evolution of adaptive events according the investigated hard sweep model. Especially, the *EPAS1* and *EGLN1* genes, along with other loci potentially involved in the physiological response to hypobaric hypoxia, did not result to have been subjected to positive selection in Tamangs. This, provided an additional evidence that such an ethnic group, despite its close relationship with Tibetans and Sherpas, does not descend directly from the ancestors of the present-day high-altitude Tibeto-Burmans and could be also considered as an ideal "negative control group" in studies aimed at inferring patterns of Himalayan high-altitude adaptation.

5.4 Fine-mapping of functional pathways involved in high-altitude adaptation

Results obtained from the genome-wide selection scans described in the previous section confirmed that the Sherpa population under consideration represents a reliable case study for deepening the dissection of the adaptive mechanisms that enabled human occupation of high-altitude Himalayan regions.

However, the hypothesized beneficial outcome of the adaptive variants located at the well-established *EPAS1* and *EGLN1* candidate genes (i.e. diminished susceptibility to mountain sickness according to reduced hemoglobin levels) alone cannot explain the complex pattern of physiological responses to high altitude presented by the Tibetan/Sherpa people (Gilbert-Kawai et al. 2014). In fact, this pattern is known to involve modifications at several levels along the oxygen transport cascade with respect to what observed in native low-altitude individuals (Beall et al. 2001;

Hoppeler et al. 2003; Beall et al. 2007; Erzurum et al. 2007; Vitzthum et al. 2013; Gilbert-Kawai et al. 2014). These changes include:

- physiological macroscopic functions (e.g. increased lung ventilation);
- cellular level adjustments (e.g. reduced number of mitochondria in muscle cells);
- a series of modifications occurring at different levels of the circulatory system (e.g. augmented blood flow intensity and increased capillary distribution in several tissues).

Given the nature of this complex of adaptive traits and the failure of traditional genome-wide selection scans and/or genotype-phenotype association studies in identifying other strong candidate loci plausibly involved in adaptation to hypobaric hypoxia a part of *EPAS1* and *EGLN1*, we hypothesized that natural selection have acted on high-altitude Himalayan populations via polygenic mechanisms, rather than via the occurrence of hard selective sweeps. Accordingly, positive selection may have slightly affected many genes and variants involved in the same biological pathway and may have thus resulted in the occurrence of soft selective sweeps around the targeted loci.

In the last few years, this kind of adaptive processes has been increasingly proposed to have played a major role in the recent evolutionary history of our species, especially during the diffusion of human populations in different environmental contexts (Hernandez et al. 2011; Jeong & Di Rienzo 2014; Schrider & Kern 2017). In particular, soft selective sweeps are supposed to be driven by positive selection acting on a pre-existing neutral variant (i.e. standing variation) and/or on newly arisen mutations that however are not the sole beneficial variants. That being so, adaptive alleles would be located on a previously existing genomic background or would have experienced a slow and limited frequency increase over generations. This causes genomic patterns around the targets of natural selection that are generally intermediate between those proper of neutrally evolving genomic regions and those related to hard selective sweeps. Therefore, most of the traditional statistical methods developed to identify genomic signatures of positive selection, being based on the hard sweep model, were proved to be inadequate to draw inferences about the occurrence of polygenic adaptation. In the attempt to overcome the above-mentioned limitations of classical genome-wide selection scans, we thus applied a double approach based on: a) the implementation of a gene-network analysis and b) the calculation of a statistics (i.e. nSL) explicitly developed to detect also soft sweeps.

Accordingly, we first submitted results obtained by the combined computation (F_{cs}) of traditional neutrality tests (i.e. PBS and XP-EHH), which were chosen as the most suited to be applied on the generated genome-wide genotyping data, to the gene-network analysis implemented in a recently developed pipeline (Gouy et al. 2017). This enabled us to do not focus simply on the genomic

regions presenting outlier values of a given selection test independently, but to search for significant functional correlations among the identified candidate genes (i.e. belonging of these loci to the same biological pathway) (see the Materials and Method section).

In parallel, we aimed at overcoming the following methodological issues. The ascertainment bias proper of genotyping data, which prevents to rely on an exhaustive description of the full spectrum of genetic variation observable in a population, and the low power of PBS/XP-EHH neutrality tests (used to optimize analyses on this kind of data) to detect soft sweeps. For this purpose, we exploited the information provided by whole genome sequences to perform the nSL test, which was specifically designed to search for patterns of haplotype homozigosity due to positive selection on newly arisen and already existent genetic variants (Ferrer-Admetlla et al. 2014).

Before applying the above-mentioned approaches, population structure analyses were replicated after merging the assembled whole genome sequence data to the previously described SNP-chip "extended" dataset in order to drive the selection of the most useful sequences to be submitted to the nSL test. Accordingly, the whole genome sequence generated for the Rolwaling Sherpa individual, as well as most of those of the Tibetan Sherpas and all of those of the Tibetans retrieved from literature resulted to be reliable representatives of the full spectrum of genetic diversity observed for high-altitude Himalayans (Fig. 4.13 and Fig. 4.14).

Moreover, fine-scale haplotype-based clustering analysis performed using the fineSTRUCTURE algorithm confirmed this pattern. In fact, the 12 examined whole genome sequences turned out to be represented in all the Tibetan/Sherpa clusters identified, further suggesting that it is not possible to identify clear boundaries between these two groups (Fig. 4.16). However, it is worth nothing to mention that the Sherpa sample analyzed in the present study is quite extensive (i.e. covers different Sherpa groups from Khumbu, Rolwaling and Tibet). On the contrary, the Tibetan samples belonged to just one population from a given location in the Tibet autonomous region and thus are not representative of the entire genetic variation observable among the Tibetan populations. Nevertheless, the samples available in the present study enabled us to identify a homogeneous Tibetan/Sherpa cluster with respect to all other East Asian population groups. That being so, we decided to consider the assembled whole genome sequences all together (with the exception of two Sherpa samples removed from the analyses due to appreciable South Asian admixture, Fig. 4.14) when performing the selection scans.

We thus computed the nSL statistics on the Tibetan/Sherpa sequence data and we used a further independent test (i.e. DIND) to validate the obtained results, finally submitting them to the previously described gene-network analysis. Moreover, we compared findings from the gene-

network approach based on genotyping data with those pointed out by the analyses conducted on the sequence data.

Accordingly, we identified subnetworks of genes, rather than single outlier loci, as plausibly targeted by positive selection in the Tibetan and Sherpa populations. Among them, the subnetworks to which *EPAS1* and *EGLN1* belong to (i.e. HIF-2α and HIF-1α, respectively) were first identified according to SNP-chip data (Table 4.5; Fig. 4.18). This finding was not taken for granted and suggests that in addition to strong positive selection having acted on these loci, which indeed presented the highest F_{cs} scores, other genes (16) among those belonging to their two interconnected pathways could be partially responsible for the smoothing of the long-term damaging erythrocyte production generally triggered by hypoxic conditions.

Among the other candidate subnetworks identified, those of the macro P-53 protein family of transcription factors (i.e. P53, P63 and P73) resulted significant also when the analysis was performed on the "WGS-high-density" dataset made up of whole genome sequence data (Table 4.5; Fig. 4.18; Fig. 4.19; Fig. 4.20). The pathways associated to these master regulators are known to be involved in many different cellular activities and control the functions of hundreds of genes (Levrero et al. 2000). Therefore, it is likely that the candidate subnetworks related to them turned to be significant in all the tests performed because of the high number of associated genes (i.e. by chance). In fact, a P-53 subnetwork (i.e. the P73-related one) was found to be significant also when Tamangs were analyzed (Table 4.5), being the only candidate shared between them and the Sherpas and thus suggesting that even if it represents a real selection signal it may be not related to high-altitude adaptation.

More interestingly, we identified three highly inter-connected significant subnetworks belonging to different integrin pathways and supported by most of the tests performed on the "selection SNP-chip" dataset and on the "WGS-high-density" dataset (Table 4.5; Fig. 4.18; Fig. 4.19; Fig. 4.20). In particular, at least two of these pathways turned out to be putatively targeted by positive selection in Tibetans and Sherpas according to all the tests performed and none of them was present among the significant pathways observed in Tamangs. These subnetworks are directly involved in promoting angiogenesis (i.e. the "Integrin in angiogenesis" pathway) or include genes encoding for integrin sub-units and/or ligands and proteins of the collagens family (i.e. COL genes), which actually exert angiogenetic functions. For example, the *COL6A1* candidate gene identified according to the nSL statistics belongs to both the "β-1 integrin" and the "Integrin in angiogenesis" pathways, while the *ITGA6* gene belonging to the "β-1 integrin" pathway and its sub-pathway "Integrin α6-β4" encodes for an integrin subunit expressed on platelets and involved in angiogenesis (Avraamides et al. 2008).

The subnetworks of genes involved in these pathways were thus found to be significant according to different tests (i.e. based on patterns of inter-population differentiation, F_{cs}, or of intra-population haplotype homozygosity, nSL/DIND) and despite the different sets of samples and types of data used. This represents a strong evidence supporting their actual involvement in shaping the studied adaptive phenotype.

Moreover, to our knowledge, up to date none of the previous studies conducted on the Tibetan/Sherpa populations have pinpointed these genes among the proposed candidate loci responsible for human high-altitude adaption. One explanation for this might be that although some of such genes presented quite high scores for the applied statistics (Table 4.5; Fig. 4.18; Fig. 4.19; Fig. 4.20), when considered independently their correlation with a specific function (e.g. angiogenesis) is not so direct, since they are involved also in dozens of other basal cellular activities. Another reason may be related to the fact that although the above-mentioned pathways resulted clearly significant in all the performed network-based analyses, the individual candidate genes composing the subnetworks are not always the same in the different tests. In fact, in this case, with a traditional sliding windows-based approach many of the genes identified by a given neutrality test would not be replicated by another one.

Accordingly, the obtained results corroborated the hypothesis that adaptive events evolved in Himalayan populations in response to hypobaric hypoxia have been mediated by natural selection via polygenic adaptation. This model indeed assumes that many genes were targeted by natural selection, but the intensity of the selective pressure on each of them was relaxed because its role in shaping the adaptive phenotype has been limited. For instance, in this evolutionary scenario an adaptive allele could be even replaced by a different variant on another related gene without compromising the final adaptive outcome, so that different individuals may carry different adaptive variants on different sets of correlated genes.

Furthermore, several of the genes composing the identified subnetworks presented intermediate or even low values for the single statistics applied and would have never resulted among the top significant loci according to a traditional window-based approach aimed at identifying hard selective sweeps. Intermediate or even low scores of most neutrality tests are indeed expected under the soft sweep and polygenic adaptation models (Pritchard & Di Rienzo, 2010). In fact, the comparison of results obtained by means of the two selection tests performed on the "WGS-high-density" dataset seems to be in line with such an assumption. In more detail, the nSL, which is designed to capture also signals related to soft selective sweeps (Ferrer-Admetlla et al. 2014), was able to identify wider subnetworks (i.e. with a higher number of genes) with respect to those

pointed out by the DIND, which has been instead developed to specifically capture hard selective sweeps (Fagny et al. 2014).

This evidence further support the hypothesis that the Tibetan/Sherpa adaptation to high altitude is also driven by positive selection acting at a polygenic level and that a test such as the nSL one is more suited than DIND to detect genomic signatures related to this type of selection. Nevertheless, also the DIND statistics identified the previously mentioned integrin pathways among the significant ones, suggesting that positive selection may have acted differently on different genes involved in the same pathway, with some loci having experienced a hard selective sweep and other a soft one.

Furthermore, although with the available data and without functional studies it is impossible to say something about the actual phenotypic trait triggered by the different variants located on these genes, their role in regulating angiogenesis may be involved in some of the previously discussed modifications occurred at several levels along the oxygen transport cascade. For example, increased blood flow and capillary distribution may be achieved by changing the regulation of angiogenetic factors, thus promoting a higher vessel perfusion in response to the hypoxic stress (Gilbert-Kawai et al. 2014).

Overall, given the convergence of the different performed tests in pinpointing the described integrin pathways and their role in regulating angiogenesis, it is reasonable to consider them as valid candidate subnetworks that could be involved in shaping the Tibetan/Sherpa adaptive phenotype. Future research may thus focus on their genes to perform functional studies aimed at elucidating the impact of the identified candidate variants on protein and cell activities, as well as their potential physiological consequences.

Finally, it is worth nothing to mention that among the significant pathways identified by means of the nSL-based network analysis, other two (i.e. those related to the C-MYB and C-MYC transcription factors) are related to biological functions potentially involved in high-altitude adaptation. The proteins encoded by these genes indeed contribute to promote both the proliferation of hematopoietic progenitor cells and angiogenesis (The UniProt Consortium 2016). Interestingly, the genes belonging to these subnetworks presented intermediate/low values of nSL, again supporting the hypothesis of adaptation mediated by soft selective sweeps. As mentioned before, the nSL was specifically designed to capture also soft sweeps, contrary to most of the neutrality tests applied in previous studies conducted on the Tibetan/Sherpa populations, and this may explain why these genes have been not identified so far as candidate adaptive loci. Anyway, further replications coupled with targeted functional studies are needed to definitely validate such findings.

5.5 Conclusions

According to the results obtained by analyses aimed at reconstructing the demographic history of Tibeto-Burman populations, the present study provided new insights into the network of migration and admixture events occurred not only on the northern slopes of the Himalayas, but also on the southern ones.

In particular, by considering a previously unsurveyed Nepalese Sherpa community that is distinct from those from the Khumbu region, we attempted to disentangle the impact of admixture and drift on the history of Himalayan populations. In fact, we found genomic evidence for their common origin followed by diverging demographic histories, variable degrees of genetic isolation and recent admixture with several sources of low-altitude East Asian ancestry. Nevertheless, it is noteworthy to mention that the differential levels of admixture with low-altitude East Asians observed in the examined Himalayan groups should caution about the fact that the genetic structure of high-altitude Tibetan/Sherpa populations could be more complex than previously thought.

Finally, we argued that Tibeto-Burman groups residing South of the Himalayas did not derive their East Asian ancestry components from the Tibetan/Sherpa lineage, which indeed originated from a branch of ancestral Tibeto-Burmans that seems to have remained confined to high-altitude regions. Instead, the peculiar East Asian genetic makeup of Southern Himalayan Tibeto-Burman populations relates to a low-altitude gene pool for which the Nagas represent the present-day group that underwent the least number of demographic changes (i.e. admixture with other populations). This ancient genetic background was progressively reshaped by admixture events having involved a number of East Asian and South Asian populations and likely occurred along migratory routes that, by circumventing the Himalayas, diffused the Tibeto-Burman ancestry components from China to Nepal across India and/or Myanmar.

Therefore, this study provided a step forward into the dissection of the tangled history of recent migrations, admixture and/or geographical and cultural isolation of Tibeto-Burmans, as well as into the understanding of its role in reshuffling patterns of genetic variation in the Himalayan area.

Moreover, by elucidating the genetic legacy of Tamangs and by demonstrating that they originated from a distinct branch of Tibeto-Burmans with respect to present-day high-altitude Himalayan populations, this study also identified an ideal “negative control group” to be used in studies aimed at inferring the adaptive evolution of Tibetan and Sherpa groups.

Finally, by applying an innovative approach based on selection scans performed on both genome-wide genotyping and whole genome sequence data, and developed to detect genomic signatures ascribable to the occurrence of soft selective sweeps involved in polygenic adaptation, we

succeeded in identifying novel candidate genes/pathways potentially responsible for adaptation to hypobaric hypoxia. In particular, we pinpointed loci and gene subnetworks that act beyond the smoothed erythrocyte production ensured by the well-established *EPAS1* and *EGLN1* adaptive genes despite hypoxic conditions. In fact, a series of biological pathways regulating the proliferation of hematopoietic progenitor cells and angiogenetic processes was found to be targeted by positive selection in the Tibetan and Sherpa people, but not in Tamangs. These signatures may be likely related to significant adjustments observed at several levels along the oxygen transport cascade of native high-altitude individuals and potentially underlying some of their well-known adaptive traits evolved in response to the hypoxic stress, such as high vessel perfusion, increased blood flow and capillary distribution.

These findings thus contributed to shed new light on the genetic bases of the complex physiological responses to hypobaric hypoxia observed in high-altitude Tibetan and Sherpa populations.

Acknowledgments

I would like first of all to thank all the people of the Garurishankar Consevation Area who kindly provided their biological samples making this study possible.
Then I would like to thank Davide Peluzzi, leader of the Explora Nunaat International organization, and the other members of the team, such as Giorgio Marinelli, Luca Natali, Marco Di Marcello and Giuseppe De Angelis, who coordinated and participated to several expeditions and collected all the samples analyzed in this study.
Special thanks are deserved to the Sherpa community of the Rolwaling Himal and to Mingma G. Sherpa and Phurba T. Sherpa, members of the Mount Everest Summitter's Club, who welcomed and helped the Explora Nunaat International teams during the expeditions conducted in the Garurishankar Consevation Area. I would like to thank also the Club Alpino Italiano and all members of the Earth Mater 2011, Gaurishankar 2013 and Extreme Malangur 2015 expeditions, as well as Massimo Izzi and Maria Giustina Palmas ("L. Galvani" Interdepartmental Centre, University of Bologna) for their valuable help in organizing the sampling of biological materials. I would also like to thank Davide Gentilini and Anna Maria Di Blasio of the Centre for Biomedical Research & Technologies of the Italian Auxologic Institute IRCCS (Milan) for having allowed us to use their facilities and for having conducted most of the genome-wide genotyping experiments.
My sincere gratitude goes to all the members of the Di Rienzo Lab of the Department of Human Genetics of the University of Chicago. Especially to Anna Di Rienzo and Choongwon Jeong for hosting me in Chicago, where they supervised me carrying out crucial steps of the study and for their essential contribution in this research, from the genotyping of a part of the samples and the whole genome sequencing, to their role in the analyses and interpretation of the results.
Then I would like to thank Etienne Guichard (Dept. of Biological, Geological and Environmental Sciences, University of Bologna), Leone De Marco (University of Pavia) and Pier Massimo Zambonelli (CESIA, University of Bologna) for their help in preparing scripts for data analyses and IT assistance.
I would like also to thank Lucy van Dorp and Garrett Hellenthal for having kindly provided scripts and suggestions helpful to improve the CHROMOPAINTER analyses of haplotype sharing.
An important thank goes to all the members (present and past) of the Molecular Anthropology Lab of the Dept. of Biological, Geological and Environmental Sciences (University of Bologna), where I carried out my PhD, for their essential role in this research and beyond, for their guidance, patience and support throughout all my PhD: Michela Trancucci, Stefania Sarno, Sara De Fanti, Andrea

Quagliariello, Paolo Abondio, Luca Pagani, Eugenio Bortolini, Alessio Boattini, Donata Luiselli and Davide Pettener.

A very special thank goes to Marco Sazzini, my supervisor, who along with the Explora Nunaat International team directly coordinated and participated to the sampling campaigns and especially shared with me all the aspect of this study, which we carried out together, and finally for having followed with extreme dedication and care every aspect of my PhD and for his fundamental role as a guide and teacher during these years.

References

1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., Abecasis, G.R., 2015. A global reference for human genetic variation. Nature 526, 68–74. https://doi.org/10.1038/nature15393

Aldenderfer, M., Yinong, Z., 2004. The Prehistory of the Tibetan Plateau to the Seventh Century A.D.: Perspectives and Research from China and the West Since 1950. Journal of World Prehistory 18, 1–55.

Aldenderfer, M., 2011. Peopling the Tibetan plateau: insights from archaeology. High Alt. Med. Biol. 12, 141–147. https://doi.org/10.1089/ham.2010.1094

Alexander, D.H., Novembre, J., Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19, 1655–1664. https://doi.org/10.1101/gr.094052.109

Athey, W., T., 2006. Haplogroup prediction from Y-STR values using a Bayesian-allele-frequency approach. J. Genet. Geneal. 2, 34-39.

Avraamides, C.J., Garmy-Susini, B., Varner, J.A., 2008. Integrins in angiogenesis and lymphangiogenesis. Nat Rev Cancer 8, 604–617. https://doi.org/10.1038/nrc2353

Basu, A., Mukherjee, N., Roy, S., Sengupta, S., Banerjee, S., Chakraborty, M., Dey, B., Roy, M., Roy, B., Bhattacharyya, N.P., Roychoudhury, S., Majumder, P.P., 2003. Ethnic India: a genomic view, with special reference to peopling and structure. Genome Res. 13, 2277–2290. https://doi.org/10.1101/gr.1413403

Basu, A., Sarkar-Roy, N., Majumder, P.P., 2016. Genomic reconstruction of the history of extant populations of India reveals five distinct ancestral components and a complex structure. Proc. Natl. Acad. Sci. U.S.A. 113, 1594–1599. https://doi.org/10.1073/pnas.1513197113

Baumgartner, R, 2015. Farewell to Yak and Yeti: The Rolwaling Sherpas: Facing a Globalised World. Vajra Books.

Beall, C.M., 2013. Human adaptability studies at high altitude: research designs and major concepts during fifty years of discovery. Am. J. Hum. Biol. 25, 141–147. https://doi.org/10.1002/ajhb.22355

Beall, C.M., 2007. Two routes to functional adaptation: Tibetan and Andean high-altitude natives. Proc. Natl. Acad. Sci. U.S.A. 104 Suppl 1, 8655–8660. https://doi.org/10.1073/pnas.0701985104

Beall, C.M., 2006. Andean, Tibetan, and Ethiopian patterns of adaptation to high-altitude hypoxia. Integr. Comp. Biol. 46, 18–24. https://doi.org/10.1093/icb/icj004

Beall, C.M., 2000. Tibetan and Andean patterns of adaptation to high-altitude hypoxia. Hum. Biol. 72, 201–228.

Beall, C.M., Cavalleri, G.L., Deng, L., Elston, R.C., Gao, Y., Knight, J., Li, C., Li, J.C., Liang, Y., McCormack, M., Montgomery, H.E., Pan, H., Robbins, P.A., Shianna, K.V., Tam, S.C., Tsering, N., Veeramah, K.R., Wang, W., Wangdui, P., Weale, M.E., Xu, Y., Xu, Z., Yang, L., Zaman, M.J., Zeng, C., Zhang, L., Zhang, X., Zhaxi, P., Zheng, Y.T., 2010. Natural selection on EPAS1 (HIF2alpha) associated with low hemoglobin concentration in Tibetan highlanders. Proc. Natl. Acad. Sci. U.S.A. 107, 11459–11464. https://doi.org/10.1073/pnas.1002443107

Beall, C.M., Laskowski, D., Strohl, K.P., Soria, R., Villena, M., Vargas, E., Alarcon, A.M., Gonzales, C., Erzurum, S.C., 2001. Pulmonary nitric oxide in mountain dwellers. Nature 414, 411–412. https://doi.org/10.1038/35106641

Bhandari, S., Zhang, X., Cui, C., Bianba, null, Liao, S., Peng, Y., Zhang, H., Xiang, K., Shi, H., Ouzhuluobu, null, Baimakongzhuo, null, Gonggalanzi, null, Liu, S., Gengdeng, null, Wu, T., Qi, X., Su, B., 2015. Genetic evidence of a recent Tibetan ancestry to Sherpas in the Himalayan region. Sci Rep 5, 16249. https://doi.org/10.1038/srep16249

Buroker, N.E., Ning, X.-H., Zhou, Z.-N., Li, K., Cen, W.-J., Wu, X.-F., Zhu, W.-Z., Scott, C.R., Chen, S.-H., 2012. EPAS1 and EGLN1 associations with high altitude sickness in Han and Tibetan Chinese at the Qinghai-Tibetan Plateau. Blood Cells Mol. Dis. 49, 67–73. https://doi.org/10.1016/j.bcmd.2012.04.00

Butler, J.M., 2005. Forensic DNA Typing. 2nd Edition. ©Elsevier Science Academic Press.

Campbell B., 1997. The Heavy loads of Tamang Identity in Nationalism and Ethnicity in Nepal eds. Gellner, D., Pfaff-Czarnecka, J., Whelpton, J. 205-235. Vajra Publications.

Cann, R.L., 2001. Genetic clues to dispersal in human populations: retracing the past from the present. Science 291, 1742–1748.

Cavalli-Sforza, L. L., Menozzi, P., & Piazza, A. 1994.The history and geography of human genes. Princeton, NJ: Princeton University Press.

Chaubey, G., Metspalu, M., Choi, Y., Mägi, R., Romero, I.G., Soares, P., van Oven, M., Behar, D.M., Rootsi, S., Hudjashov, G., Mallick, C.B., Karmin, M., Nelis, M., Parik, J., Reddy, A.G., Metspalu, E., van Driem, G., Xue, Y., Tyler-Smith, C., Thangaraj, K., Singh, L., Remm, M., Richards, M.B., Lahr, M.M., Kayser, M., Villems, R., Kivisild, T., 2011. Population genetic structure in Indian Austroasiatic speakers: the role of landscape barriers and sex-specific admixture. Mol. Biol. Evol. 28, 1013–1024. https://doi.org/10.1093/molbev/msq288

Cordaux, R., Aunger, R., Bentley, G., Nasidze, I., Sirajuddin, S.M., Stoneking, M., 2004a. Independent origins of Indian caste and tribal paternal lineages. Curr. Biol. 14, 231–235. https://doi.org/10.1016/j.cub.2004.01.024

Cordaux, R., Saha, N., Bentley, G.R., Aunger, R., Sirajuddin, S.M., Stoneking, M., 2003. Mitochondrial DNA analysis reveals diverse histories of tribal populations from India. Eur. J. Hum. Genet. 11, 253–264. https://doi.org/10.1038/sj.ejhg.5200949

Cordaux, R., Weiss, G., Saha, N., Stoneking, M., 2004b. The northeast Indian passageway: a barrier or corridor for human migrations? Mol. Biol. Evol. 21, 1525–1533. https://doi.org/10.1093/molbev/msh151

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

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., McKenna, A., Fennell, T.J., Kernytsky, A.M., Sivachenko, A.Y., Cibulskis, K., Gabriel, S.B., Altshuler, D., Daly, M.J., 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43, 491–498. https://doi.org/10.1038/ng.806

Deschamps, M., Laval, G., Fagny, M., Itan, Y., Abel, L., Casanova, J.-L., 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. https://doi.org/10.1016/j.ajhg.2015.11.014

Destro-Bisol, G., Jobling, M.A., Rocha, J., Novembre, J., Richards, M.B., Mulligan, C., Batini, C., Manni, F., 2010. Molecular anthropology in the genomic era. J Anthropol Sci 88, 93–112.

Erzurum, S.C., Ghosh, S., Janocha, A.J., Xu, W., Bauer, S., Bryan, N.S., Tejero, J., Hemann, C., Hille, R., Stuehr, D.J., Feelisch, M., Beall, C.M., 2007. Higher blood flow and circulating NO products offset high-altitude hypoxia among Tibetans. Proc. Natl. Acad. Sci. U.S.A. 104, 17593–17598. https://doi.org/10.1073/pnas.0707462104

Excoffier, L., Lischer, H.E.L., 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10, 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Fagny, M., Patin, E., Enard, D., Barreiro, L.B., Quintana-Murci, L., Laval, G., 2014. Exploring the occurrence of classic selective sweeps in humans using whole-genome sequencing data sets. Mol. Biol. Evol. 31, 1850–1868. https://doi.org/10.1093/molbev/msu118

Falush, D., Dorp, L. van, Lawson, D., 2016. A tutorial on how (not) to over-interpret STRUCTURE/ADMIXTURE bar plots. bioRxiv 066431. https://doi.org/10.1101/066431

Fan, S., Hansen, M.E.B., Lo, Y., Tishkoff, S.A., 2016. Going global by adapting local: A review of recent human adaptation. Science 354, 54–59. https://doi.org/10.1126/science.aaf5098

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. https://doi.org/10.1093/molbev/msu077

Fornarino, S., Pala, M., Battaglia, V., Maranta, R., Achilli, A., Modiano, G., Torroni, A., Semino, O., Santachiara-Benerecetti, S.A., 2009. Mitochondrial and Y-chromosome diversity of the Tharus (Nepal): a reservoir of genetic variation. BMC Evol. Biol. 9, 154. https://doi.org/10.1186/1471-2148-9-154

Forster, P., Torroni, A., Renfrew, C., Röhl, A., 2001. Phylogenetic star contraction applied to Asian and Papuan mtDNA evolution. Mol. Biol. Evol. 18, 1864–1881.

Fu, Q., Li, H., Moorjani, P., Jay, F., Slepchenko, S.M., Bondarev, A.A., Johnson, P.L.F., Aximu-Petri, A., Prüfer, K., de Filippo, C., Meyer, M., Zwyns, N., Salazar-García, D.C., Kuzmin, Y.V., Keates, S.G., Kosintsev, P.A., Razhev, D.I., Richards, M.P., Peristov, N.V., Lachmann, M., Douka, K., Higham, T.F.G., Slatkin, M., Hublin, J.-J., Reich, D., Kelso, J., Viola, T.B., Pääbo, S., 2014. Genome sequence of a 45,000-year-old modern human from western Siberia. Nature 514, 445–449. https://doi.org/10.1038/nature13810

Gagneux, P., Wills, C., Gerloff, U., Tautz, D., Morin, P.A., Boesch, C., Fruth, B., Hohmann, G., Ryder, O.A., Woodruff, D.S., 1999. Mitochondrial sequences show diverse evolutionary histories of African hominoids. Proc. Natl. Acad. Sci. U.S.A. 96, 5077–5082.

Gao X., Zhou Z., Y., Guan Y., 2008. Human cultural remains and adaptation strategies in the Tibetan Plateau margin region in the late Pleistocene. Quaternary Sci. 28:969-977 (in Chinese).

Gayden, T., Bukhari, A., Chennakrishnaiah, S., Stojkovic, O., Herrera, R.J., 2012. Y-chromosomal microsatellite diversity in three culturally defined regions of historical Tibet. Forensic Sci Int Genet 6, 437–446. https://doi.org/10.1016/j.fsigen.2011.09.002

Gayden, T., Cadenas, A.M., Regueiro, M., Singh, N.B., Zhivotovsky, L.A., Underhill, P.A., Cavalli-Sforza, L.L., Herrera, R.J., 2007. The Himalayas as a directional barrier to gene flow. Am. J. Hum. Genet. 80, 884–894. https://doi.org/10.1086/516757

Gayden, T., Mirabal, S., Cadenas, A.M., Lacau, H., Simms, T.M., Morlote, D., Chennakrishnaiah, S., Herrera, R.J., 2009. Genetic insights into the origins of Tibeto-Burman populations in the Himalayas. J. Hum. Genet. 54, 216–223. https://doi.org/10.1038/jhg.2009.14

Gayden, T., Perez, A., Persad, P.J., Bukhari, A., Chennakrishnaiah, S., Simms, T., Maloney, T., Rodriguez, K., Herrera, R.J., 2013. The Himalayas: barrier and conduit for gene flow. Am. J. Phys. Anthropol. 151, 169–182. https://doi.org/10.1002/ajpa.22240

Ge JX, Wu SD, Chao SJ, 1997. Zhongguo yimin Shi [the migration history of China]. Fuzhou: Fujian People's Press [in Chinese].

Gouy, A., Daub, J.T., Excoffier, L., 2017. Detecting gene subnetworks under selection in biological pathways. Nucleic Acids Res. 45, e149. https://doi.org/10.1093/nar/gkx626

Gilbert-Kawai, E.T., Milledge, J.S., Grocott, M.P.W., Martin, D.S., 2014. King of the mountains: Tibetan and Sherpa physiological adaptations for life at high altitude. Physiology (Bethesda) 29, 388–402. https://doi.org/10.1152/physiol.00018.2014

Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H.-Y., Hansen, N.F., Durand, E.Y., Malaspinas, A.-S., Jensen, J.D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H.A., Good, J.M., Schultz, R., Aximu-Petri, A., Butthof, A., Höber, B., Höffner, B., Siegemund, M., Weihmann, A., Nusbaum, C., Lander, E.S., Russ, C., Novod, N., Affourtit, J., Egholm, M., Verna, C., Rudan, P., Brajkovic, D., Kucan, Ž., Gušic, I., Doronichev, V.B., Golovanova, L.V., Lalueza-Fox, C., de la Rasilla, M., Fortea, J., Rosas, A., Schmitz, R.W., Johnson, P.L.F., Eichler, E.E., Falush, D., Birney, E., Mullikin, J.C., Slatkin, M., Nielsen, R., Kelso, J., Lachmann, M., Reich, D., Pääbo, S., 2010. A draft sequence of the Neandertal genome. Science 328, 710–722. https://doi.org/10.1126/science.1188021

Grossman, S.R., Andersen, K.G., Shlyakhter, I., Tabrizi, S., Winnicki, S., Yen, A., Park, D.J., Griesemer, D., Karlsson, E.K., Wong, S.H., Cabili, M., Adegbola, R.A., Bamezai, R.N.K., Hill, A.V.S., Vannberg, F.O., Rinn, J.L., 1000 Genomes Project, Lander, E.S., Schaffner, S.F., Sabeti, P.C., 2013. Identifying recent adaptations in large-scale genomic data. Cell 152, 703–713.

Harpending, H., 1997. The History and Geography of Human Genes. L. Luca Cavalli-Sforza , Paolo Menozzi , Alberto Piazza. The Quarterly Review of Biology 72, 355–355. https://doi.org/10.1086/419930

Hellenthal, G., Busby, G.B.J., Band, G., Wilson, J.F., Capelli, C., Falush, D., Myers, S., 2014. A genetic atlas of human admixture history. Science 343, 747–751. https://doi.org/10.1126/science.1243518

Hernandez, R.D., Kelley, J.L., Elyashiv, E., Melton, S.C., Auton, A., McVean, G., 1000 Genomes Project, Sella, G., Przeworski, M., 2011. Classic selective sweeps were rare in recent human evolution. Science 331, 920–924. https://doi.org/10.1126/science.1198878

Hoppeler, H., Vogt, M., Weibel, E.R., Flück, M., 2003. Response of skeletal muscle mitochondria to hypoxia. Exp. Physiol. 88, 109–119.

Horscroft, J.A., Kotwica, A.O., Laner, V., West, J.A., Hennis, P.J., Levett, D.Z.H., Howard, D.J., Fernandez, B.O., Burgess, S.L., Ament, Z., Gilbert-Kawai, E.T., Vercueil, A., Landis, B.D., Mitchell, K., Mythen, M.G., Branco, C., Johnson, R.S., Feelisch, M., Montgomery, H.E., Griffin, J.L., Grocott, M.P.W., Gnaiger, E., Martin, D.S., Murray, A.J., 2017. Metabolic basis to Sherpa altitude adaptation. Proc. Natl. Acad. Sci. U.S.A. 114, 6382–6387. https://doi.org/10.1073/pnas.1700527114

Huerta-Sánchez, E., Jin, X., Asan, null, Bianba, Z., Peter, B.M., Vinckenbosch, N., Liang, Y., Yi, X., He, M., Somel, M., Ni, P., Wang, B., Ou, X., Huasang, null, Luosang, J., Cuo, Z.X.P., Li, K., Gao, G., Yin, Y., Wang, W., Zhang, X., Xu, X., Yang, H., Li, Y., Wang, J., Wang, J., Nielsen, R., 2014. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature 512, 194–197. https://doi.org/10.1038/nature13408

HUGO Pan-Asian SNP Consortium, Abdulla, M.A., Ahmed, I., Assawamakin, A., Bhak, J., Brahmachari, S.K., Calacal, G.C., Chaurasia, A., Chen, C.-H., Chen, J., Chen, Y.-T., Chu, J., Cutiongco-de la Paz, E.M.C., De Ungria, M.C.A., Delfin, F.C., Edo, J., Fuchareon, S., Ghang, H., Gojobori, T., Han, J., Ho, S.-F., Hoh, B.P., Huang, W., Inoko, H., Jha, P., Jinam, T.A., Jin, L., Jung, J., Kangwanpong, D., Kampuansai, J., Kennedy, G.C., Khurana, P., Kim, H.-L., Kim, K., Kim, S., Kim, W.-Y., Kimm, K., Kimura, R., Koike, T., Kulawonganunchai, S., Kumar, V., Lai, P.S., Lee, J.-Y., Lee, S., Liu, E.T., Majumder, P.P., Mandapati, K.K., Marzuki, S., Mitchell, W., Mukerji, M., Naritomi, K., Ngamphiw, C., Niikawa, N., Nishida, N., Oh, B., Oh, S., Ohashi, J., Oka, A., Ong, R., Padilla, C.D., Palittapongarnpim, P., Perdigon, H.B., Phipps, M.E., Png, E., Sakaki, Y., Salvador, J.M., Sandraling, Y., Scaria, V., Seielstad, M., Sidek, M.R., Sinha, A., Srikummool, M., Sudoyo, H., Sugano, S., Suryadi, H., Suzuki, Y., Tabbada, K.A., Tan, A., Tokunaga, K., Tongsima, S., Villamor, L.P., Wang, E., Wang, Y., Wang, H., Wu, J.-Y., Xiao, H., Xu, S., Yang, J.O., Shugart, Y.Y., Yoo, H.-S., Yuan, W., Zhao, G., Zilfalil, B.A., Indian Genome Variation Consortium, 2009. Mapping human genetic diversity in Asia. Science 326, 1541–1545. https://doi.org/10.1126/science.1177074

Jeong, C., Alkorta-Aranburu, G., Basnyat, B., Neupane, M., Witonsky, D.B., Pritchard, J.K., Beall, C.M., Di Rienzo, A., 2014. Admixture facilitates genetic adaptations to high altitude in Tibet. Nat Commun 5, 3281. https://doi.org/10.1038/ncomms4281

Jeong, C., Di Rienzo, A., 2014. Adaptations to local environments in modern human populations. Curr. Opin. Genet. Dev. 29, 1–8. https://doi.org/10.1016/j.gde.2014.06.011

Jeong, C., Ozga, A.T., Witonsky, D.B., Malmström, H., Edlund, H., Hofman, C.A., Hagan, R.W., Jakobsson, M., Lewis, C.M., Aldenderfer, M.S., Di Rienzo, A., Warinner, C., 2016. Long-term genetic stability and a high-altitude East Asian origin for the peoples of the high valleys of the Himalayan arc. Proc. Natl. Acad. Sci. U.S.A. 113, 7485–7490. https://doi.org/10.1073/pnas.1520844113

Jeong, C., Peter, B.M., Basnyat, B., Neupane, M., Beall, C.M., Childs, G., Craig, S.R., Novembre, J., Di Rienzo, A., 2017. A longitudinal cline characterizes the genetic structure of human populations in the Tibetan plateau. PLoS ONE 12, e0175885. https://doi.org/10.1371/journal.pone.0175885

Jobling, M., Hollox E., Hurles M., Kivisild T., Tyler-Smith C., 2014. Human Evolutionary Genetics.2nd Edition. New York: Garland Science.

Kandasamy, K., Mohan, S.S., Raju, R., Keerthikumar, S., Kumar, G.S.S., Venugopal, A.K., Telikicherla, D., Navarro, J.D., Mathivanan, S., Pecquet, C., Gollapudi, S.K., Tattikota, S.G., Mohan, S., Padhukasahasram, H., Subbannayya, Y., Goel, R., Jacob, H.K.C., Zhong, J., Sekhar, R., Nanjappa, V., Balakrishnan, L., Subbaiah, R., Ramachandra, Y.L., Rahiman, B.A., Prasad, T.S.K., Lin, J.-X., Houtman, J.C.D., Desiderio, S., Renauld, J.-C., Constantinescu, S.N., Ohara, O., Hirano, T., Kubo, M., Singh, S., Khatri, P., Draghici, S., Bader, G.D., Sander, C., Leonard, W.J., Pandey, A., 2010. NetPath: a public resource of curated signal transduction pathways. Genome Biol. 11, R3. https://doi.org/10.1186/gb-2010-11-1-r3

Kang, L., Lu, Y., Wang, C., Hu, K., Chen, F., Liu, K., Li, S., Jin, L., Li, H., Genographic Consortium, 2012. Y-chromosome O3 haplogroup diversity in Sino-Tibetan populations reveals two migration routes into the eastern Himalayas. Ann. Hum. Genet. 76, 92–99. https://doi.org/10.1111/j.1469-1809.2011.00690.x

Kang, L., Zheng, H.-X., Chen, F., Yan, S., Liu, K., Qin, Z., Liu, L., Zhao, Z., Li, L., Wang, X., He, Y., Jin, L., 2013. mtDNA lineage expansions in Sherpa population suggest adaptive evolution in Tibetan highlands. Mol. Biol. Evol. 30, 2579–2587. https://doi.org/10.1093/molbev/mst147

Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P., 1983. Optimization by simulated annealing. Science 220, 671–680. https://doi.org/10.1126/science.220.4598.671

Kloss-Brandstätter, A., Pacher, D., Schönherr, S., Weissensteiner, H., Binna, R., Specht, G., Kronenberg, F., 2011. HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. Hum. Mutat. 32, 25–32. https://doi.org/10.1002/humu.21382

Kosambi, D.D., 1965. The Culture and Civilisation of Ancient India in Historical Outline. Vikas.

Krause, J., Fu, Q., Good, J.M., Viola, B., Shunkov, M.V., Derevianko, A.P., Pääbo, S., 2010. The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. Nature 464, 894–897. https://doi.org/10.1038/nature08976

Kumar, S., Padmanabham, P.B.S.V., Ravuri, R.R., Uttaravalli, K., Koneru, P., Mukherjee, P.A., Das, B., Kotal, M., Xaviour, D., Saheb, S.Y., Rao, V.R., 2008. The earliest settlers' antiquity and evolutionary history of Indian populations: evidence from M2 mtDNA lineage. BMC Evol. Biol. 8, 230. https://doi.org/10.1186/1471-2148-8-230

Kumar, V., Reddy, A.N.S., Babu, J.P., Rao, T.N., Langstieh, B.T., Thangaraj, K., Reddy, A.G., Singh, L., Reddy, B.M., 2007. Y-chromosome evidence suggests a common paternal heritage of Austro-Asiatic populations. BMC Evol. Biol. 7, 47. https://doi.org/10.1186/1471-2148-7-47

Lachance, J., 2014. Human Evolutionary Genetics by Mark Jobling, Edward Hollox, Matthew Hurles, Toomas Kivisild, and Chris Tyler-Smith. The Quarterly Review of Biology 89, 182–183. https://doi.org/10.1086/676083

Lawson, D.J., Hellenthal, G., Myers, S., Falush, D., 2012. Inference of population structure using dense haplotype data. PLoS Genet. 8, e1002453. https://doi.org/10.1371/journal.pgen.1002453

Lee, F.S., Percy, M.J., 2011. The HIF pathway and erythrocytosis. Annu Rev Pathol 6, 165–192. https://doi.org/10.1146/annurev-pathol-011110-130321

Leslie, S., Winney, B., Hellenthal, G., Davison, D., Boumertit, A., Day, T., Hutnik, K., Royrvik, E.C., Cunliffe, B., Wellcome Trust Case Control Consortium 2, International Multiple Sclerosis Genetics Consortium, Lawson, D.J., Falush, D., Freeman, C., Pirinen, M., Myers, S., Robinson, M., Donnelly, P., Bodmer, W., 2015. The fine-scale genetic structure of the British population. Nature 519, 309–314. https://doi.org/10.1038/nature14230

Levrero, M., De Laurenzi, V., Costanzo, A., Gong, J., Wang, J.Y., Melino, G., 2000. The p53/p63/p73 family of transcription factors: overlapping and distinct functions. J. Cell. Sci. 113 (Pt 10), 1661–1670.

Lewis, T., T., Shakya, D., R., 1988. Contributions to the history of Nepal: eastern Newar diaspora settlements. Contrib Nepal Stud 15:25–65.

Li, H., Durbin, R., 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26, 589–595. https://doi.org/10.1093/bioinformatics/btp698

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., Cann, H.M., Barsh, G.S., Feldman, M., Cavalli-Sforza, L.L., Myers, R.M., 2008. Worldwide human relationships inferred from genome-wide patterns of variation. Science 319, 1100–1104. https://doi.org/10.1126/science.1153717

Lipson, M., Loh, P.-R., Levin, A., Reich, D., Patterson, N., Berger, B., 2013. Efficient moment-based inference of admixture parameters and sources of gene flow. Mol. Biol. Evol. 30, 1788–1802. https://doi.org/10.1093/molbev/mst099

Liu, G., Kaw, B., Kurfis, J., Rahmanuddin, S., Kanwar, Y.S., Chugh, S.S., 2003. Neph1 and nephrin interaction in the slit diaphragm is an important determinant of glomerular permeability. J. Clin. Invest. 112, 209–221. https://doi.org/10.1172/JCI18242

Lu, D., Lou, H., Yuan, K., Wang, X., Wang, Y., Zhang, C., Lu, Y., Yang, X., Deng, L., Zhou, Y., Feng, Q., Hu, Y., Ding, Q., Yang, Y., Li, S., Jin, L., Guan, Y., Su, B., Kang, L., Xu, S., 2016. Ancestral Origins and Genetic History of Tibetan Highlanders. Am. J. Hum. Genet. 99, 580–594. https://doi.org/10.1016/j.ajhg.2016.07.002

Macaulay, V., Hill, C., Achilli, A., Rengo, C., Clarke, D., Meehan, W., Blackburn, J., Semino, O., Scozzari, R., Cruciani, F., Taha, A., Shaari, N.K., Raja, J.M., Ismail, P., Zainuddin, Z., Goodwin, W., Bulbeck, D., Bandelt, H.-J., Oppenheimer, S., Torroni, A., Richards, M., 2005. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. Science 308, 1034–1036. https://doi.org/10.1126/science.1109792

Macsorley, F., 2001. High altitude medicine and physiology. Br J Sports Med 35, 451. https://doi.org/10.1136/bjsm.35.6.451-b

Majumder, P.P., 2010. The human genetic history of South Asia. Curr. Biol. 20, R184–187. https://doi.org/10.1016/j.cub.2009.11.053

Matisoff, J.A., 1991. Sino-Tibetan Linguistics: Present State and Future Prospects. Annual Review of Anthropology 20, 469–504.

Metspalu, M., Kivisild, T., Metspalu, E., Parik, J., Hudjashov, G., Kaldma, K., Serk, P., Karmin, M., Behar, D.M., Gilbert, M.T.P., Endicott, P., Mastana, S., Papiha, S.S., Skorecki, K., Torroni, A., Villems, R., 2004. Most of the extant mtDNA boundaries in south and southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. BMC Genet. 5, 26. https://doi.org/10.1186/1471-2156-5-26

Mona, S., Grunz, K.E., Brauer, S., Pakendorf, B., Castrì, L., Sudoyo, H., Marzuki, S., Barnes, R.H., Schmidtke, J., Stoneking, M., Kayser, M., 2009. Genetic admixture history of Eastern Indonesia as revealed by Y-chromosome and mitochondrial DNA analysis. Mol. Biol. Evol. 26, 1865–1877. https://doi.org/10.1093/molbev/msp097

Moorjani, P., Thangaraj, K., Patterson, N., Lipson, M., Loh, P.-R., Govindaraj, P., Berger, B., Reich, D., Singh, L., 2013. Genetic evidence for recent population mixture in India. Am. J. Hum. Genet. 93, 422–438. https://doi.org/10.1016/j.ajhg.2013.07.006

O'Reilly, P.F., Birney, E., Balding, D.J., 2008. Confounding between recombination and selection, and the Ped/Pop method for detecting selection. Genome Res. 18, 1304–1313. https://doi.org/10.1101/gr.067181.107

Oppitz M., 1974. Myths and facts: Reconsidering some data concerning the clan history of the Sherpas Available at: http://www.dspace.cam.ac.uk/handle/1810/227209

Pagani, L., Schiffels, S., Gurdasani, D., Danecek, P., Scally, A., Chen, Y., Xue, Y., Haber, M., Ekong, R., Oljira, T., Mekonnen, E., Luiselli, D., Bradman, N., Bekele, E., Zalloua, P., Durbin, R., Kivisild, T., Tyler-Smith, C., 2015. Tracing the Route of Modern Humans out of Africa by Using 225 Human Genome Sequences from Ethiopians and Egyptians. The American Journal of Human Genetics 96, 986–991. https://doi.org/10.1016/j.ajhg.2015.04.019

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 192, 1065–1093. https://doi.org/10.1534/genetics.112.145037

Patterson, N., Price, A.L., Reich, D., 2006. Population structure and eigenanalysis. PLoS Genet. 2, e190. https://doi.org/10.1371/journal.pgen.0020190

Paul, L. M., Simons, G. F., Fennig, C. D., 2016. Ethnologue: Languages of the World. Nineteenth edition http://www.ethnologue.com

Peng, Y., Yang, Z., Zhang, H., Cui, C., Qi, X., Luo, X., Tao, X., Wu, T., Ouzhuluobu, null, Basang, null, Ciwangsangbu, null, Danzengduojie, null, Chen, H., Shi, H., Su, B., 2011. Genetic variations in Tibetan populations and high-altitude adaptation at the Himalayas. Mol. Biol. Evol. 28, 1075–1081. https://doi.org/10.1093/molbev/msq290

Price, A.L., Tandon, A., Patterson, N., Barnes, K.C., Rafaels, N., Ruczinski, I., Beaty, T.H., Mathias, R., Reich, D., Myers, S., 2009. Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. PLoS Genet. 5, e1000519. https://doi.org/10.1371/journal.pgen.1000519

Pritchard, J.K., Di Rienzo, A., 2010. Adaptation – not by sweeps alone. Nat Rev Genet 11, 665–667. https://doi.org/10.1038/nrg2880

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575. https://doi.org/10.1086/519795

Qin, Z., Yang, Y., Kang, L., Yan, S., Cho, K., Cai, X., Lu, Y., Zheng, H., Zhu, D., Fei, D., Li, S., Jin, L., Li, H., Genographic Consortium, 2010. A mitochondrial revelation of early human migrations to the Tibetan Plateau before and after the last glacial maximum. Am. J. Phys. Anthropol. 143, 555–569. https://doi.org/10.1002/ajpa.21350

Qiu, J., 2015. Who are the Tibetans? Science 347, 708–711. https://doi.org/10.1126/science.347.6223.708

Qi, X., Cui, C., Peng, Y., Zhang, X., Yang, Z., Zhong, H., Zhang, H., Xiang, K., Cao, X., Wang, Y., Ouzhuluobu, null, Basang, null, Ciwangsangbu, null, Bianba, null, Gonggalanzi, null, Wu, T., Chen, H., Shi, H., Su, B., 2013. Genetic evidence of paleolithic colonization and neolithic expansion of modern humans on the tibetan plateau. Mol. Biol. Evol. 30, 1761–1778. https://doi.org/10.1093/molbev/mst093

Raghavan, M., Skoglund, P., Graf, K.E., Metspalu, M., Albrechtsen, A., Moltke, I., Rasmussen, S., Stafford, T.W., Orlando, L., Metspalu, E., Karmin, M., Tambets, K., Rootsi, S., Mägi, R., Campos, P.F., Balanovska, E., Balanovsky, O., Khusnutdinova, E., Litvinov, S., Osipova, L.P., Fedorova, S.A., Voevoda, M.I., DeGiorgio, M., Sicheritz-Ponten, T., Brunak, S., Demeshchenko, S., Kivisild, T., Villems, R., Nielsen, R., Jakobsson, M., Willerslev, E., 2014. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature 505, 87–91. https://doi.org/10.1038/nature12736

Reich, D., Green, R.E., Kircher, M., Krause, J., Patterson, N., Durand, E.Y., Viola, B., Briggs, A.W., Stenzel, U., Johnson, P.L.F., Maricic, T., Good, J.M., Marques-Bonet, T., Alkan, C., Fu, Q., Mallick, S., Li, H., Meyer, M., Eichler, E.E., Stoneking, M., Richards, M., Talamo, S., Shunkov, M.V., Derevianko, A.P., Hublin, J.-J., Kelso, J., Slatkin, M., Pääbo, S., 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature 468, 1053–1060. https://doi.org/10.1038/nature09710

Reich, D., Thangaraj, K., Patterson, N., Price, A.L., Singh, L., 2009. Reconstructing Indian population history. Nature 461, 489–494. https://doi.org/10.1038/nature08365

Rhode, D. A., 2016. Biogeographic perspective on early human colonization of the Tibetan Plateau. Archaeological Research in Asia 5, 33–43.

Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., McCarroll, S.A., Gaudet, R., Schaffner, S.F., Lander, E.S., International HapMap Consortium, Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M., Pasternak, S., Wheeler, D.A., Willis, T.D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., Yu, J., Zhang, B., Zhang, Q., Zhao, H., Zhao, H., Zhou, J., Gabriel, S.B., Barry, R., Blumenstiel, B., Camargo, A., Defelice, M., Faggart, M., Goyette, M., Gupta, S., Moore, J., Nguyen, H., Onofrio, R.C., Parkin, M., Roy, J., Stahl, E., Winchester, E., Ziaugra, L., Altshuler, D., Shen, Y., Yao, Z., Huang, W., Chu, X., He, Y., Jin, L., Liu, Y., Shen, Y., Sun, W., Wang, H., Wang, Y., Wang, Y., Xiong, X., Xu, L., Waye, M.M.Y., Tsui, S.K.W., Xue, H., Wong, J.T.-F., Galver, L.M., Fan, J.-B., Gunderson, K., Murray, S.S., Oliphant, A.R., Chee, M.S., Montpetit, A., Chagnon, F., Ferretti, V., Leboeuf, M., Olivier, J.-F., Phillips, M.S., Roumy, S., Sallée, C., Verner, A., Hudson, T.J., Kwok, P.-Y., Cai, D., Koboldt, D.C., Miller, R.D., Pawlikowska, L., Taillon-Miller, P., Xiao, M., Tsui, L.-C., Mak, W., Song, Y.Q., Tam, P.K.H., Nakamura, Y., Kawaguchi, T., Kitamoto, T., Morizono, T., Nagashima, A., Ohnishi, Y., Sekine, A., Tanaka, T., Tsunoda,

T., Deloukas, P., Bird, C.P., Delgado, M., Dermitzakis, E.T., Gwilliam, R., Hunt, S., Morrison, J., Powell, D., Stranger, B.E., Whittaker, P., Bentley, D.R., Daly, M.J., de Bakker, P.I.W., Barrett, J., Chretien, Y.R., Maller, J., McCarroll, S., Patterson, N., Pe'er, I., Price, A., Purcell, S., Richter, D.J., Sabeti, P., Saxena, R., Schaffner, S.F., Sham, P.C., Varilly, P., Altshuler, D., Stein, L.D., Krishnan, L., Smith, A.V., Tello-Ruiz, M.K., Thorisson, G.A., Chakravarti, A., Chen, P.E., Cutler, D.J., Kashuk, C.S., Lin, S., Abecasis, G.R., Guan, W., Li, Y., Munro, H.M., Qin, Z.S., Thomas, D.J., McVean, G., Auton, A., Bottolo, L., Cardin, N., Eyheramendy, S., Freeman, C., Marchini, J., Myers, S., Spencer, C., Stephens, M., Donnelly, P., Cardon, L.R., Clarke, G., Evans, D.M., Morris, A.P., Weir, B.S., Tsunoda, T., Johnson, T.A., Mullikin, J.C., Sherry, S.T., Feolo, M., Skol, A., Zhang, H., Zeng, C., Zhao, H., Matsuda, I., Fukushima, Y., Macer, D.R., Suda, E., Rotimi, C.N., Adebamowo, C.A., Ajayi, I., Aniagwu, T., Marshall, P.A., Nkwodimmah, C., Royal, C.D.M., Leppert, M.F., Dixon, M., Peiffer, A., Qiu, R., Kent, A., Kato, K., Niikawa, N., Adewole, I.F., Knoppers, B.M., Foster, M.W., Clayton, E.W., Watkin, J., Gibbs, R.A., Belmont, J.W., Muzny, D., Nazareth, L., Sodergren, E., Weinstock, G.M., Wheeler, D.A., Yakub, I., Gabriel, S.B., Onofrio, R.C., Richter, D.J., Ziaugra, L., Birren, B.W., Daly, M.J., Altshuler, D., Wilson, R.K., Fulton, L.L., Rogers, J., Burton, J., Carter, N.P., Clee, C.M., Griffiths, M., Jones, M.C., McLay, K., Plumb, R.W., Ross, M.T., Sims, S.K., Willey, D.L., Chen, Z., Han, H., Kang, L., Godbout, M., Wallenburg, J.C., L'Archevêque, P., Bellemare, G., Saeki, K., Wang, H., An, D., Fu, H., Li, Q., Wang, Z., Wang, R., Holden, A.L., Brooks, L.D., McEwen, J.E., Guyer, M.S., Wang, V.O., Peterson, J.L., Shi, M., Spiegel, J., Sung, L.M., Zacharia, L.F., Collins, F.S., Kennedy, K., Jamieson, R., Stewart, J., 2007. Genome-wide detection and characterization of positive selection in human populations. Nature 449, 913–918. https://doi.org/10.1038/nature06250

Sanchez-Mazas, A., Di, D., Riccio, M.E., 2011. A Genetic Focus on the Peopling History of East Asia: Critical Views. Rice 4, 159–169. https://doi.org/10.1007/s12284-011-9066-y

Schaefer, C.F., Anthony, K., Krupa, S., Buchoff, J., Day, M., Hannay, T., Buetow, K.H., 2009. PID: the Pathway Interaction Database. Nucleic Acids Res. 37, D674–679. https://doi.org/10.1093/nar/gkn653

Schrider, D.R., Kern, A.D., 2017. Soft Sweeps Are the Dominant Mode of Adaptation in the Human Genome. Mol. Biol. Evol. 34, 1863–1877. https://doi.org/10.1093/molbev/msx154

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504. https://doi.org/10.1101/gr.1239303

Shi, H., Dong, Y.-L., Wen, B., Xiao, C.-J., Underhill, P.A., Shen, P.-D., Chakraborty, R., Jin, L., Su, B., 2005. Y-chromosome evidence of southern origin of the East Asian-specific haplogroup O3-M122. Am. J. Hum. Genet. 77, 408–419. https://doi.org/10.1086/444436

Shi, H., Zhong, H., Peng, Y., Dong, Y.-L., Qi, X.-B., Zhang, F., Liu, L.-F., Tan, S.-J., Ma, R.Z., Xiao, C.-J., Wells, R.S., Jin, L., Su, B., 2008. Y chromosome evidence of earliest modern human settlement in East Asia and multiple origins of Tibetan and Japanese populations. BMC Biol. 6, 45. https://doi.org/10.1186/1741-7007-6-45

Simonson, T.S., McClain, D.A., Jorde, L.B., Prchal, J.T., 2012. Genetic determinants of Tibetan high-altitude adaptation. Hum. Genet. 131, 527–533. https://doi.org/10.1007/s00439-011-1109-3

Simonson, T.S., Yang, Y., Huff, C.D., Yun, H., Qin, G., Witherspoon, D.J., Bai, Z., Lorenzo, F.R., Xing, J., Jorde, L.B., Prchal, J.T., Ge, R., 2010. Genetic evidence for high-altitude adaptation in Tibet. Science 329, 72–75. https://doi.org/10.1126/science.1189406

Singh, K.S., 1992. People of India: An Introduction. Calcutta: Anthropological Survey of India. Oxford University Press, New Delhi.

Stoneking, M., Delfin, F., 2010. The human genetic history of East Asia: weaving a complex tapestry. Curr. Biol. 20, R188–193. https://doi.org/10.1016/j.cub.2009.11.052

Su, B., Xiao, C., Deka, R., Seielstad, M.T., Kangwanpong, D., Xiao, J., Lu, D., Underhill, P., Cavalli-Sforza, L., Chakraborty, R., Jin, L., 2000. Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. Hum. Genet. 107, 582–590.

Su, B., Xiao, J., Underhill, P., Deka, R., Zhang, W., Akey, J., Huang, W., Shen, D., Lu, D., Luo, J., Chu, J., Tan, J., Shen, P., Davis, R., Cavalli-Sforza, L., Chakraborty, R., Xiong, M., Du, R., Oefner, P., Chen, Z., Jin, L., 1999. Y-Chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age. Am. J. Hum. Genet. 65, 1718–1724. https://doi.org/10.1086/302680

Szpiech, Z.A., Hernandez, R.D., 2014. selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. Mol. Biol. Evol. 31, 2824–2827. https://doi.org/10.1093/molbev/msu211

Thangaraj, K., Chaubey, G., Kivisild, T., Reddy, A.G., Singh, V.K., Rasalkar, A.A., Singh, L., 2005. Reconstructing the origin of Andaman Islanders. Science 308, 996. https://doi.org/10.1126/science.1109987

Tulla, M., Pentikäinen, O.T., Viitasalo, T., Käpylä, J., Impola, U., Nykvist, P., Nissinen, L., Johnson, M.S., Heino, J., 2001. Selective binding of collagen subtypes by integrin alpha 1I, alpha 2I, and alpha 10I domains. J. Biol. Chem. 276, 48206–48212. https://doi.org/10.1074/jbc.M104058200

Underhill, P.A., Kivisild, T., 2007. Use of y chromosome and mitochondrial DNA population structure in tracing human migrations. Annu. Rev. Genet. 41, 539–564. https://doi.org/10.1146/annurev.genet.41.110306.130407

UniProt: the universal protein knowledgebase, 2017. Nucleic Acids Research 45, D158–D169. https://doi.org/10.1093/nar/gkw1099

van Dorp, L., Balding, D., Myers, S., Pagani, L., Tyler-Smith, C., Bekele, E., Tarekegn, A., Thomas, M.G., Bradman, N., Hellenthal, G., 2015. Evidence for a Common Origin of Blacksmiths and Cultivators in the Ethiopian Ari within the Last 4500 Years: Lessons for Clustering-Based Inference. PLoS Genet. 11, e1005397. https://doi.org/10.1371/journal.pgen.1005397

van Oven, M., Kayser, M., 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum. Mutat. 30, E386–394. https://doi.org/10.1002/humu.20921

Vargas, E., Spielvogel, H., 2006. Chronic mountain sickness, optimal hemoglobin, and heart disease. High Alt. Med. Biol. 7, 138–149. https://doi.org/10.1089/ham.2006.7.138

Vitzthum, V.J., 2013. Fifty fertile years: anthropologists' studies of reproduction in high altitude natives. Am. J. Hum. Biol. 25, 179–189. https://doi.org/10.1002/ajhb.22357

Wang, H.-W., Li, Y.-C., Sun, F., Zhao, M., Mitra, B., Chaudhuri, T.K., Regmi, P., Wu, S.-F., Kong, Q.-P., Zhang, Y.-P., 2012. Revisiting the role of the Himalayas in peopling Nepal: insights from mitochondrial genomes. J. Hum. Genet. 57, 228–234. https://doi.org/10.1038/jhg.2012.8

Wang, W.S.Y. 1998. In the Bronze Age and Early Iron Age peoples of Eastern Central Asia. University of Pennsylvania Museum Publications, pp 508-534.

Wang, Z.H., 1994. History of nationalities in China. China Social Science Press.

Wen, B., Li, H., Lu, D., Song, X., Zhang, F., He, Y., Li, F., Gao, Y., Mao, X., Zhang, L., Qian, J., Tan, J., Jin, J., Huang, W., Deka, R., Su, B., Chakraborty, R., Jin, L., 2004a. Genetic evidence supports demic diffusion of Han culture. Nature 431, 302–305. https://doi.org/10.1038/nature02878

Wen, B., Xie, X., Gao, S., Li, H., Shi, H., Song, X., Qian, T., Xiao, C., Jin, J., Su, B., Lu, D., Chakraborty, R., Jin, L., 2004b. Analyses of genetic structure of Tibeto-Burman populations reveals sex-biased admixture in southern Tibeto-Burmans. Am. J. Hum. Genet. 74, 856–865. https://doi.org/10.1086/386292

Winslow, R., M., Monge, C., C., 1987. Hypoxia, Polycythemia, and Chronic Mountain Sickness, Johns Hopkins University Press.

Wu, T., Kayser, B., 2006. High altitude adaptation in Tibetans. High Alt. Med. Biol. 7, 193–208. https://doi.org/10.1089/ham.2006.7.193

Yang, J., Jin, Z.-B., Chen, J., Huang, X.-F., Li, X.-M., Liang, Y.-B., Mao, J.-Y., Chen, X., Zheng, Z., Bakshi, A., Zheng, D.-D., Zheng, M.-Q., Wray, N.R., Visscher, P.M., Lu, F., Qu, J., 2017. Genetic signatures of high-altitude adaptation in Tibetans. Proc. Natl. Acad. Sci. U.S.A. 114, 4189–4194. https://doi.org/10.1073/pnas.1617042114

Yao, Y.-G., Kong, Q.-P., Bandelt, H.-J., Kivisild, T., Zhang, Y.-P., 2002a. Phylogeographic differentiation of mitochondrial DNA in Han Chinese. Am. J. Hum. Genet. 70, 635–651. https://doi.org/10.1086/338999

Yao, Y.-G., Kong, Q.-P., Wang, C.-Y., Zhu, C.-L., Zhang, Y.-P., 2004. Different matrilineal contributions to genetic structure of ethnic groups in the silk road region in china. Mol. Biol. Evol. 21, 2265–2280. https://doi.org/10.1093/molbev/msh238

Yao, Y.-G., Nie, L., Harpending, H., Fu, Y.-X., Yuan, Z.-G., Zhang, Y.-P., 2002b. Genetic relationship of Chinese ethnic populations revealed by mtDNA sequence diversity. Am. J. Phys. Anthropol. 118, 63–76. https://doi.org/10.1002/ajpa.10052

Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z.X.P., Pool, J.E., Xu, X., Jiang, H., Vinckenbosch, N., Korneliussen, T.S., Zheng, H., Liu, T., He, W., Li, K., Luo, R., Nie, X., Wu, H., Zhao, M., Cao, H., Zou, J., Shan, Y., Li, S., Yang, Q., Asan, null, Ni, P., Tian, G., Xu, J., Liu, X., Jiang, T., Wu, R., Zhou, G., Tang, M., Qin, J., Wang, T., Feng, S., Li, G., Huasang, null, Luosang, J., Wang, W., Chen, F., Wang, Y., Zheng, X., Li, Z., Bianba, Z., Yang, G., Wang, X., Tang, S., Gao, G., Chen, Y., Luo, Z., Gusang, L., Cao, Z., Zhang, Q., Ouyang, W., Ren, X., Liang, H., Zheng, H., Huang, Y., Li, J., Bolund, L., Kristiansen, K., Li, Y., Zhang, Y., Zhang, X., Li, R., Li, S., Yang, H., Nielsen, R., Wang, J., Wang, J., 2010. Sequencing of 50 human exomes reveals adaptation to high altitude. Science 329, 75–78. https://doi.org/10.1126/science.1190371

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 Biol. 18, 115. https://doi.org/10.1186/s13059-017-1242-y

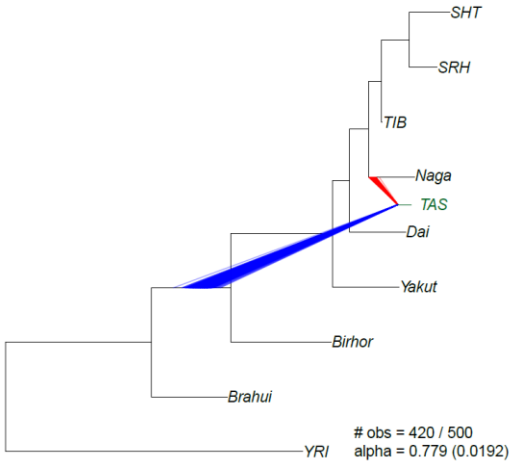
Zhao, M., Kong, Q.-P., Wang, H.-W., Peng, M.-S., Xie, X.-D., Wang, W.-Z., Jiayang, null, Duan, J.-G., Cai, M.-C., Zhao, S.-N., Cidanpingcuo, null, Tu, Y.-Q., Wu, S.-F., Yao, Y.-G., Bandelt, H.-J., Zhang, Y.-P., 2009. Mitochondrial genome evidence reveals successful Late Paleolithic settlement on the Tibetan Plateau. Proc. Natl. Acad. Sci. U.S.A. 106, 21230–21235. https://doi.org/10.1073/pnas.0907844106

Zhao, Y.-B., Li, H.-J., Li, S.-N., Yu, C.-C., Gao, S.-Z., Xu, Z., Jin, L., Zhu, H., Zhou, H., 2011. Ancient DNA evidence supports the contribution of Di-Qiang people to the han Chinese gene pool. Am. J. Phys. Anthropol. 144, 258–268. https://doi.org/10.1002/ajpa.21399

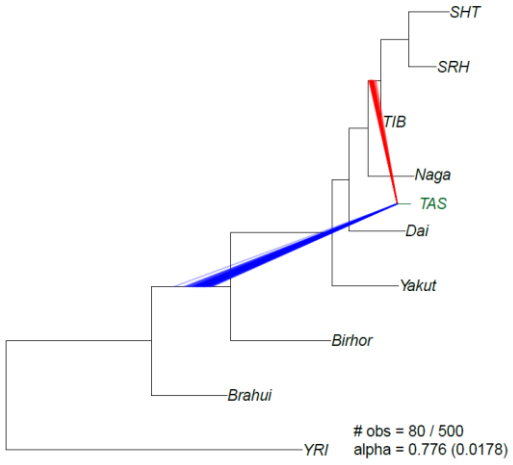
Zhuang, J., Droma, T., Sutton, J.R., Groves, B.M., McCullough, R.E., McCullough, R.G., Sun, S., Moore, L.G., 1996. Smaller alveolar-arterial O2 gradients in Tibetan than Han residents of Lhasa (3658 m). Respir Physiol 103, 75–82.

Appendix

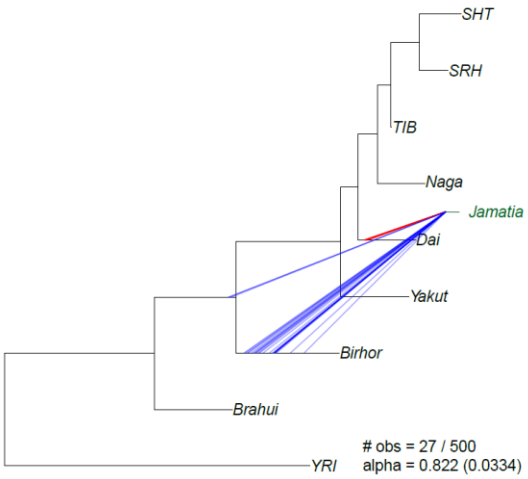
TAS : topology 1 : case 1



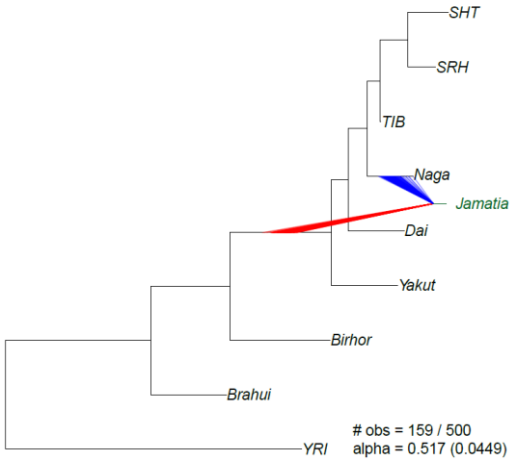
TAS : topology 1 : case 2



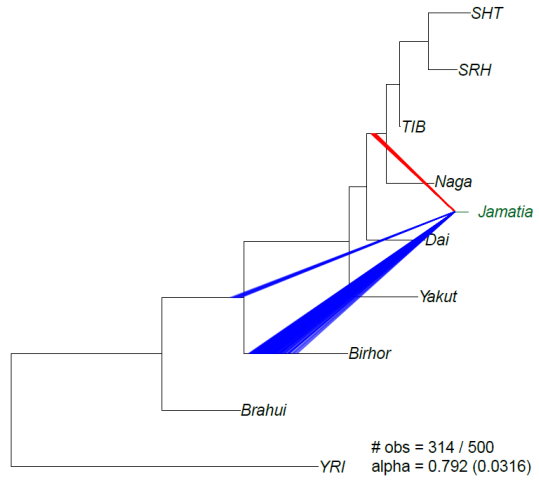
Jamatia : topology 1 : case 1



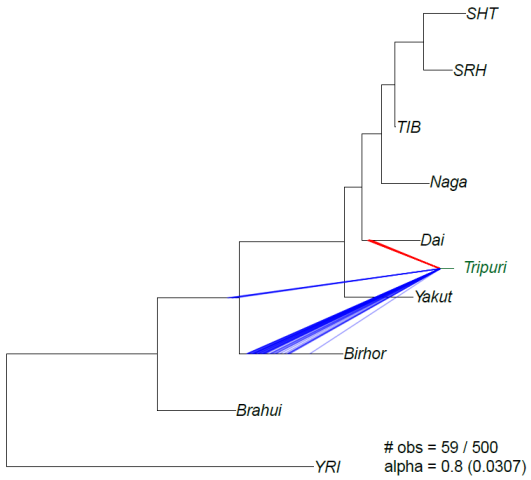
Jamatia : topology 1 : case 2



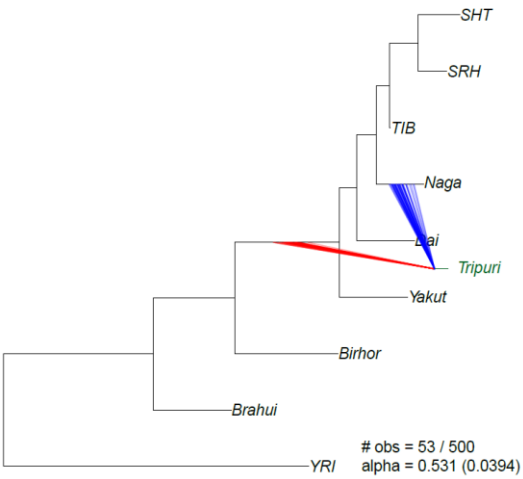
Jamatia : topology 1 : case 3



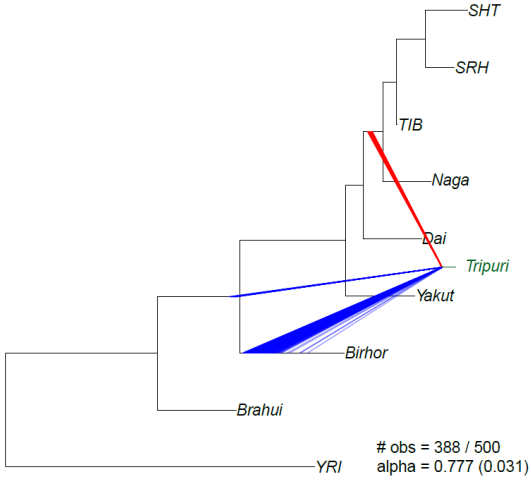
Tripuri : topology 1 : case 1



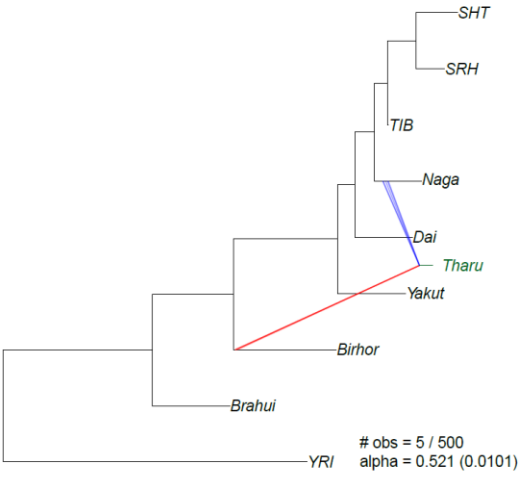
Tripuri : topology 1 : case 2



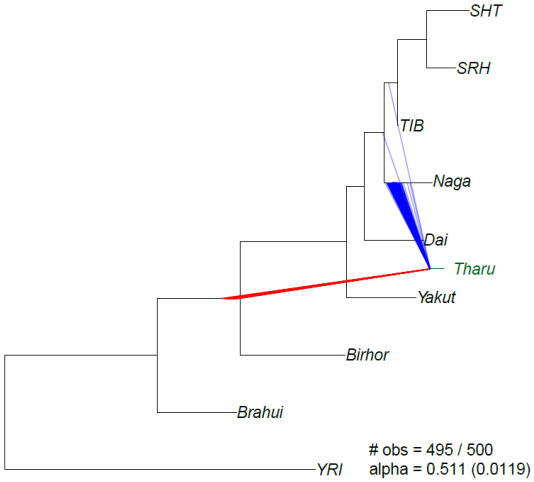
Tripuri : topology 1 : case 3



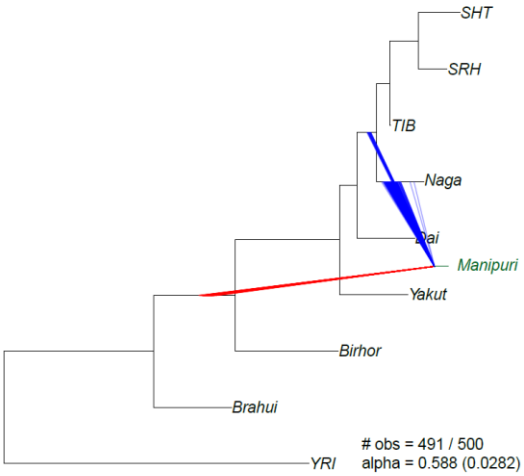
Tharu : topology 1 : case 1



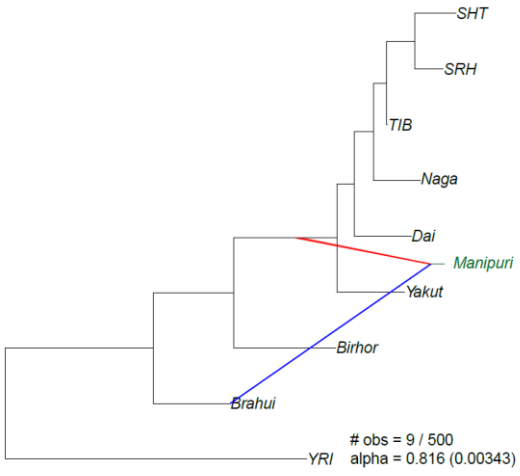
Tharu : topology 1 : case 2



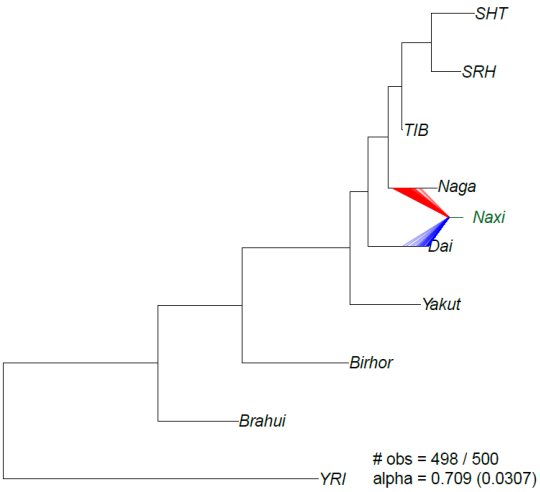
Manipuri : topology 1 : case 1



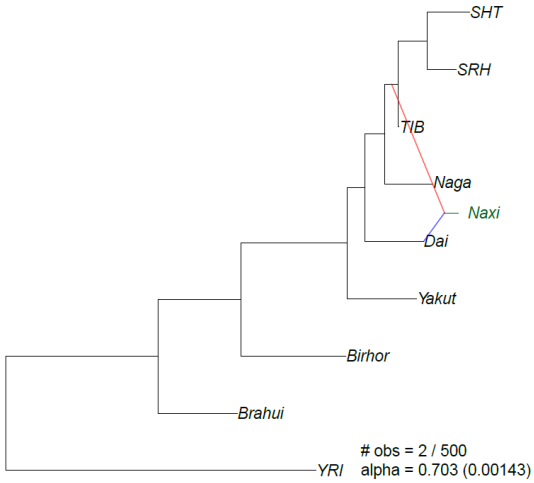
Manipuri : topology 1 : case 2



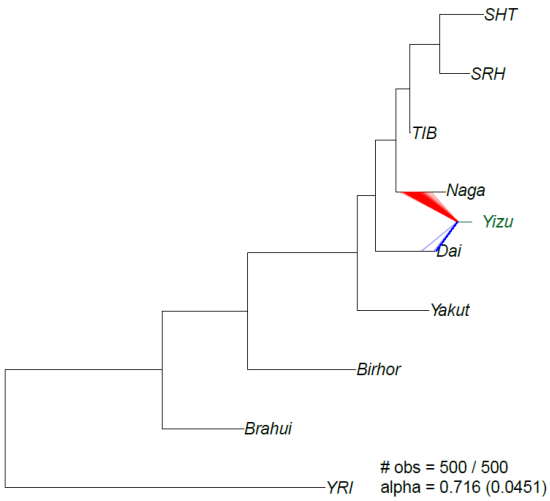
Naxi : topology 1 : case 1



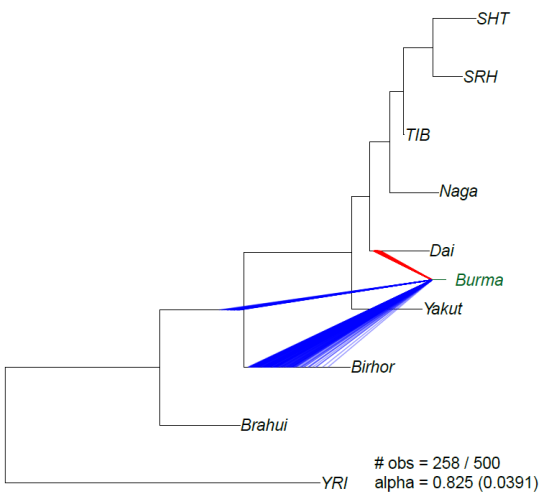
Naxi : topology 1 : case 2

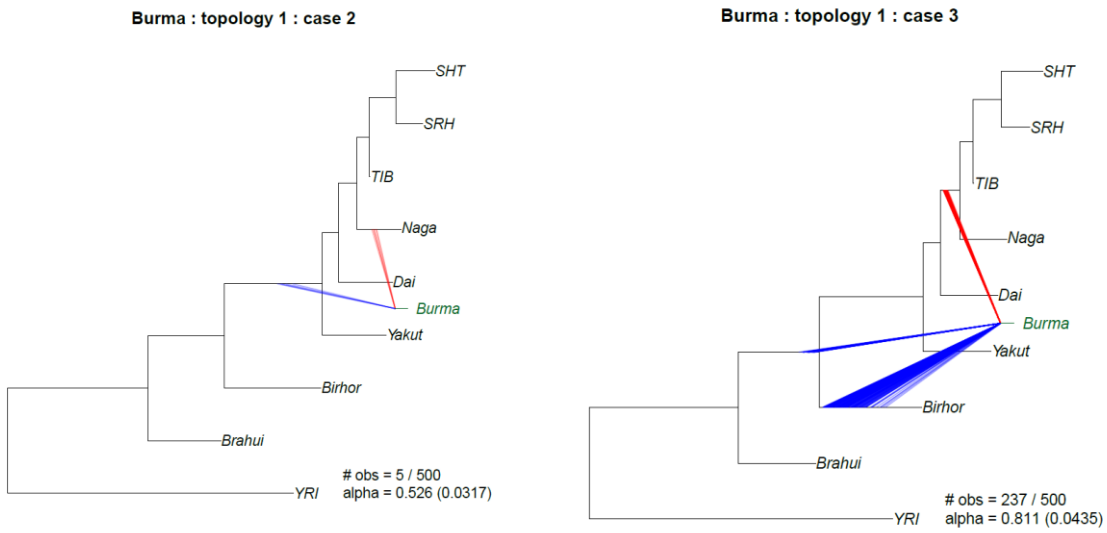


Yizu : topology 1 : case 1



Burma : topology 1 : case 1





Appendix Figure 1. Neighbour-Joining trees showing admixture events fitted with MIXMAPPER. The main ancestry sources are represented with red lines, while the secondary sources are displayed with blue lines. These lines are located in the positions of the tree presenting the best residual norm for each one of the tested cases. Only cases with at least five observations (obs) on the 500 bootstrap replications were considered. Alpha values refers to the main sources of ancestry proportions. Topology 1 was conserved for all 500 bootstrap replications.

Appendix Table 1. GCA samples typed in the present study for HVSI sequences of mtDNA and STRs of Y-chromosome data and Asian populations included in the reference panels.

mtDNA				
Population	***N***	***Language***	***Macro Area***	***Source***
SRH	32	Tibeto-Burman	Himalayas	Present study
Tamang_GCA	37	Tibeto-Burman	Himalayas	Present study
IAR	23	Indo-Aryan	Himalayas	Present study
Sherpa_Nepal	266	Tibeto-Burman	Himalayas	Bhandari 2015
Sherpa_Tibet	226	Tibeto-Burman	Himalayas	Bhandari 2016
Tamang	45	Tibeto-Burman	Himalayas	Gayden 2013
Newar	66	Tibeto-Burman	Himalayas	Gayden 2013
Kathmandu	77	Indo-Aryan	Himalayas	Gayden 2013
Tibetans	156	Tibeto-Burman	Himalayas	Gayden 2013
Nepalese	235	Non specified	Himalayas	Wang 2012
Yao_Tibetan	41	Tibeto-Burman	Himalayas	Yao 2002a,b
Nakchu_Tibetan	168	Tibeto-Burman	Himalayas	Zhao 2009
Shigatse_Tibetan	220	Tibeto-Burman	Himalayas	Zhao 2009
Qinghai_Tibetan	86	Tibeto-Burman	Himalayas	Zhao 2009
Yunnan_Tibetan	71	Tibeto-Burman	Himalayas	Zhao 2009
Sichuan_Tibetan	62	Tibeto-Burman	Himalayas	Zhao 2009
Gansu_Tibetan	83	Tibeto-Burman	Himalayas	Zhao 2009
Mongol	15	Altaic	Northeast Asia	Yao 2002a
Tu	35	Altaic	Northeast Asia	Yao 2002a
Yunnan_Han	42	Sinitic	Northeast Asia	Yao 2002b
Liaoning_Han	50	Sinitic	Northeast Asia	Yao 2002b
Qingdao_Han	49	Sinitic	Northeast Asia	Yao 2002b
Wuhan_Han	42	Sinitic	Northeast Asia	Yao 2002b
Adi	45	Tibeto-Burman	South Central Asia	Cordaux 2003
Apatani	52	Tibeto-Burman	South Central Asia	Cordaux 2003
Naga	43	Tibeto-Burman	South Central Asia	Cordaux 2003
Nishi	52	Tibeto-Burman	South Central Asia	Cordaux 2003
Sikh Punjab	40	Indo-Aryan	South Central Asia	Cordaux 2003
Kashmir	19	Indo-Aryan	South Central Asia	Kivisild 1999
Lobana	62	Indo-Aryan	South Central Asia	Kivisild 2000
Tharu	37	Indo-Aryan	South Central Asia	Kivisild 1999 & Metspalu 2004
Uttar Pradesh	73	Indo-Aryan	South Central Asia	Kivisild 1999
Tipperah	20	Tibeto-Burman	South Central Asia	Roychoudhury 2001
Kanet	37	Tibeto-Burman	South Central Asia	Metspalu 2004
Punjab	109	Indo-Aryan	South Central Asia	Metspalu 2004
Karachi	99	Indo-Aryan	South Central Asia	Quintana-Murci 2004
Taiwanese	66	Sinitic	Southeast Asia	Horai 1996
Thailand	33	Austronesian/Daic	Southeast Asia	Yao 2002a,b
Bai	31	Tibeto-Burman	Southeast Asia	Yao 2002a
Dai	38	Austronesian/Daic	Southeast Asia	Yao 2002a

Lisu	37	Tibeto-Burman	Southeast Asia	Yao 2002a
Nu	30	Tibeto-Burman	Southeast Asia	Yao 2002a
Sali	30	Tibeto-Burman	Southeast Asia	Yao 2002a
Zhuang	83	Austronesian/Daic	Southeast Asia	Yao 2002a
GuangdongHan	71	Sinitic	Southeast Asia	Kivisild 2002
Y_HanGuandong	30	Sinitic	Southeast Asia	Yao 2002b
Persian	82	Indo-Aryan	West Asia	Derenko 2007
Tajiks	44	Indo-Aryan	West Asia	Derenko 2007
Mongolians	47	Altaic	Central Asia	Derenko 2007
Y_HanXinjiang	46	Sinitic	Central Asia	Yao 2002b
Uighur	45	Altaic	Central Asia	Yao 2000
Y-chromosome				
Population	**N**	**Language**	**Macro Area**	**Source**
SRH	19	Tibeto-Burman	Himalayas	Present study
Tamang_GCA	22	Tibeto-Burman	Himalayas	Present study
IAR	22	Tibeto-Burman	Himalayas	Present study
Tamang	45	Tibeto-Burman	Himalayas	Gayden 2010
Newar	66	Tibeto-Burman	Himalayas	Gayden 2010
Kathmandu	77	Indo-Aryan	Himalayas	Gayden2010
Tibet	156	Tibeto-Burman	Himalayas	Gayden 2010
Tibetan_Lhasa_a	132	Tibeto-Burman	Himalayas	Yi 2010
Tibetan_Lhasa_b	112	Tibeto-Burman	Himalayas	Zhang 2006
Taiwanese	200	Sinitic	Northeast Asia	Huang 2008
Han Chinese	131	Sinitic	Northeast Asia	Yan 2007
Japanese Osaka	131	Altaic	Northeast Asia	Hashiyada 2008
Tibetan Qinghai	167	Tibeto-Burman	Northeast Asia	Zhu 2008
Bangali	216	Indo-European	South Central Asia	Alam 2010
Brahmin Punjab	60	Indo-European	South Central Asia	Yadav 2011
Brahmin Rajasthan	58	Indo-European	South Central Asia	Yadav 2011
Brahmin Himachal Pradesh	61	Indo-European	South Central Asia	Yadav 2011
Balmiki	62	Indo-European	South Central Asia	Ghosh 2011
Sakaldwipi Brahmin	65	Indo-European	South Central Asia	Ghosh 2011
Kanyakubja Brahmin	78	Indo-European	South Central Asia	Ghosh 2011
Konkanastha Brahmin	71	Indo-European	South Central Asia	Ghosh 2011
Mahadev Koli	64	Indo-European	South Central Asia	Ghosh 2011
Gond	75	Dravidian	South Central Asia	Ghosh 2011
Kurumans	67	Dravidian	South Central Asia	Ghosh 2011
Tripuri	65	Tibeto-Burman	South Central Asia	Ghosh 2011
Riang	67	Tibeto-Burman	South Central Asia	Ghosh 2011
Munda	68	Austro-Asiatic	South Central Asia	Ghosh 2011
Iyengar	66	Dravidian	South Central Asia	Ghosh 2011

Appendix Table 2. Significant (Z-score ≤ 2) negative f3 values.

Parental 1	Parental 2	Admixed	*f3*	Stand dev	Z-score
CHS	IE_UTT	Burma	-0.00876	0.000184	-47.744
CHS	ITU	Burma	-0.00887	0.000188	-47.269
CHS	STU	Burma	-0.0087	0.000186	-46.802
CHB	IE_UTT	Burma	-0.0084	0.00019	-44.133
CHB	ITU	Burma	-0.00856	0.000194	-44.112
Iyer	CHS	Burma	-0.00897	0.000204	-44.03
Gujrati_Brh	CHS	Burma	-0.00896	0.000204	-43.851
CHB	STU	Burma	-0.00843	0.000193	-43.768
CHB	GIH	Burma	-0.00861	0.000198	-43.446
Han	IE_UTT	Burma	-0.00836	0.000194	-43.029
CHB	PJL	Burma	-0.00837	0.000195	-42.908
Pallan	CHB	Burma	-0.0084	0.000196	-42.859
BEB	CHB	Burma	-0.00743	0.000176	-42.25
CDX	IE_UTT	Burma	-0.00787	0.000186	-42.24
W_Bengal_Brh	CHB	Burma	-0.00823	0.000197	-41.857
CHS	Drav_Kar	Burma	-0.00878	0.00021	-41.841
GIH	Han	Burma	-0.00851	0.000204	-41.642
Khatri	CHS	Burma	-0.00885	0.000213	-41.487
Iyer	CHB	Burma	-0.00858	0.000209	-41.019
CDX	GIH	Burma	-0.00807	0.000197	-40.975
Brahui	CHS	Burma	-0.00912	0.000223	-40.875
CDX	PJL	Burma	-0.00791	0.000194	-40.818
CHB	Sindhi	Burma	-0.00842	0.000207	-40.765
Balochi	CHS	Burma	-0.00893	0.000219	-40.748
CDX	STU	Burma	-0.00772	0.000191	-40.342
Burusho	CHS	Burma	-0.00762	0.000189	-40.315
W_Bengal_Brh	Han	Burma	-0.00818	0.000203	-40.262
Gujrati_Brh	CHB	Burma	-0.00847	0.000213	-39.84
CHB	Drav_Kar	Burma	-0.00856	0.000216	-39.682
ITU	Tujia	Burma	-0.00938	0.000236	-39.682
Iyer	Han	Burma	-0.0085	0.000214	-39.672
Pallan	CDX	Burma	-0.00766	0.000194	-39.455
Maratha	CHB	Burma	-0.00877	0.000223	-39.286
IE_UTT	Tujia	Burma	-0.00909	0.000232	-39.225
W_Bengal_Brh	CDX	Burma	-0.00771	0.000197	-39.07
Gujrati_Brh	Han	Burma	-0.00846	0.000217	-39.056
Gond	CHB	Burma	-0.00755	0.000194	-38.919
Irula	CHS	Burma	-0.00821	0.000211	-38.878
Gond	CHS	Burma	-0.00736	0.00019	-38.737
Iyer	CDX	Burma	-0.00796	0.000206	-38.703
BEB	CDX	Burma	-0.00669	0.000173	-38.673
GIH	Tujia	Burma	-0.00952	0.000247	-38.561
Pallan	Tujia	Burma	-0.00921	0.000239	-38.525
Han	Sindhi	Burma	-0.00841	0.000219	-38.388
Kadar	CHB	Burma	-0.0082	0.000214	-38.278
Khatri	CHB	Burma	-0.00831	0.000217	-38.202
Han	Drav_TN	Burma	-0.00846	0.000223	-38.002
ITU	Miaozu	Burma	-0.0091	0.000239	-37.987
Drav_Kar	Han	Burma	-0.00839	0.000221	-37.975
BEB	Tujia	Burma	-0.00812	0.000214	-37.891
ITU	Yizu	Burma	-0.00863	0.000228	-37.801
Maratha	Han	Burma	-0.00857	0.000227	-37.72
IE_UTT	Miaozu	Burma	-0.00894	0.000237	-37.631
GIH	Yizu	Burma	-0.0088	0.000235	-37.374
KHV	IE_UTT	Burma	-0.00672	0.00018	-37.334
Iyer	Tujia	Burma	-0.00931	0.00025	-37.159

IE_UTT	Yizu	Burma	-0.00845	0.000229	-36.966
Khatri	Han	Burma	-0.0083	0.000225	-36.951
CDX	Drav_AP	Burma	-0.00773	0.00021	-36.81
Miaozu	STU	Burma	-0.00885	0.000241	-36.758
Irula	CHB	Burma	-0.00803	0.000219	-36.727
GIH	Miaozu	Burma	-0.00923	0.000252	-36.632
Miaozu	PJL	Burma	-0.00894	0.000244	-36.597
Kadar	Han	Burma	-0.00804	0.00022	-36.496
Santal	CHS	Burma	-0.00661	0.000181	-36.431
W_Bengal_Brh	Tujia	Burma	-0.0089	0.000244	-36.421
Gujrati_Brh	Tujia	Burma	-0.00935	0.000257	-36.418
GIH	KHV	Burma	-0.00686	0.00019	-36.066
Balochi	Han	Burma	-0.00837	0.000232	-36.011
Pathan	Tujia	Burma	-0.00919	0.000255	-35.986
CDX	Drav_TN	Burma	-0.00774	0.000215	-35.961
Khatri	Tujia	Burma	-0.00936	0.00026	-35.933
Irula	Han	Burma	-0.00793	0.000221	-35.919
BEB	Yizu	Burma	-0.0075	0.000209	-35.893
KHV	STU	Burma	-0.00657	0.000183	-35.862
KHV	PJL	Burma	-0.00674	0.000188	-35.781
BEB	Miaozu	Burma	-0.00779	0.000218	-35.738
Brahui	Han	Burma	-0.00848	0.000237	-35.709
Maratha	CDX	Burma	-0.00805	0.000225	-35.705
IE_UTT	She	Burma	-0.00879	0.000246	-35.663
Burusho	CHB	Burma	-0.00694	0.000195	-35.621
Pallan	Yizu	Burma	-0.00842	0.000237	-35.453
Pallan	Miaozu	Burma	-0.0088	0.000248	-35.435
ITU	She	Burma	-0.0089	0.000252	-35.321
She	STU	Burma	-0.00868	0.000249	-34.815
Iyer	Yizu	Burma	-0.00863	0.000249	-34.704
Iyer	Miaozu	Burma	-0.00918	0.000265	-34.689
Balochi	Tujia	Burma	-0.00939	0.000271	-34.67
Gujrati_Brh	Yizu	Burma	-0.00869	0.000251	-34.627
Miaozu	Sindhi	Burma	-0.00918	0.000265	-34.574
Drav_Kar	Tujia	Burma	-0.00892	0.00026	-34.248
Ho	Han	Burma	-0.00642	0.000187	-34.246
Gujrati_Brh	KHV	Burma	-0.00695	0.000203	-34.23
Gujrati_Brh	Miaozu	Burma	-0.00908	0.000265	-34.189
N_Munda	CHB	Burma	-0.00685	0.0002	-34.174
Khatri	Yizu	Burma	-0.00872	0.000255	-34.166
W_Bengal_Brh	KHV	Burma	-0.00661	0.000194	-34.029
Iyer	KHV	Burma	-0.00682	0.000201	-34.007
CHS	Gond	Burma	-0.00798	0.000235	-33.974
BEB	KHV	Burma	-0.00566	0.000167	-33.904
BEB	She	Burma	-0.00772	0.000228	-33.892
CDX	IAR	Burma	-0.00544	0.000161	-33.851
Drav_AP	Tujia	Burma	-0.0092	0.000272	-33.784
GIH	She	Burma	-0.00885	0.000262	-33.763
Sindhi	Yizu	Burma	-0.00866	0.000257	-33.763
Irula	CDX	Burma	-0.00719	0.000214	-33.633
KHV	Sindhi	Burma	-0.00688	0.000205	-33.601
Pathan	Yizu	Burma	-0.00856	0.000256	-33.496
Dai	IE_UTT	Burma	-0.00793	0.000237	-33.47
Burusho	Tujia	Burma	-0.00789	0.000236	-33.466
Kadar	CDX	Burma	-0.0072	0.000217	-33.235
Han	Paniya	Burma	-0.00782	0.000235	-33.23
Drav_AP	Yizu	Burma	-0.00842	0.000254	-33.225
Drav_TN	Tujia	Burma	-0.00915	0.000275	-33.197
N_Munda	CHS	Burma	-0.00643	0.000194	-33.106

Maratha	Miaozu	Burma	-0.0092	0.000279	-32.984
CHB	Gond	Burma	-0.00788	0.000239	-32.964
Korwa	CHB	Burma	-0.00662	0.000201	-32.899
Burusho	KHV	Burma	-0.00617	0.000188	-32.775
CHB	Sakilli	Burma	-0.0086	0.000263	-32.768
IE_UTT	Naga	Burma	-0.01008	0.000308	-32.768
Dai	ITU	Burma	-0.00803	0.000247	-32.544
Naga	STU	Burma	-0.01017	0.000313	-32.527
Drav_AP	Miaozu	Burma	-0.0089	0.000274	-32.521
Maratha	Tujia	Burma	-0.00919	0.000282	-32.515
Balochi	Yizu	Burma	-0.00864	0.000268	-32.288
Gond	Yizu	Burma	-0.00751	0.000233	-32.27
Dai	STU	Burma	-0.00777	0.000241	-32.268
W_Bengal_Brh	She	Burma	-0.0086	0.000267	-32.236
Dai	PJL	Burma	-0.00804	0.00025	-32.232
Brahui	Miaozu	Burma	-0.00921	0.000286	-32.186
Pallan	She	Burma	-0.00864	0.000269	-32.134
Maratha	KHV	Burma	-0.00694	0.000216	-32.114
ITU	Naga	Burma	-0.01029	0.000321	-32.085
GIH	Naga	Jamatia	-0.00492	0.000364	-13.535
ITU	Naga	Jamatia	-0.00486	0.000361	-13.462
IE_UTT	Naga	Jamatia	-0.00468	0.000348	-13.449
Naga	STU	Jamatia	-0.00476	0.000362	-13.128
Naga	PJL	Jamatia	-0.00475	0.000364	-13.043
Gujrati_Brh	Naga	Jamatia	-0.00482	0.000386	-12.504
Brahui	Naga	Jamatia	-0.00502	0.000405	-12.378
W_Bengal_Brh	Naga	Jamatia	-0.00465	0.000378	-12.287
Iyer	Naga	Jamatia	-0.00475	0.000387	-12.277
Pallan	Naga	Jamatia	-0.00464	0.000388	-11.951
Khatri	Naga	Jamatia	-0.00469	0.000402	-11.652
Naga	Sindhi	Jamatia	-0.00464	0.000402	-11.537
Drav_TN	Naga	Jamatia	-0.00481	0.000421	-11.429
Naga	Pathan	Jamatia	-0.00457	0.000401	-11.391
Drav_AP	Naga	Jamatia	-0.00452	0.000398	-11.351
Gond	Naga	Jamatia	-0.00395	0.000353	-11.195
Drav_Kar	Naga	Jamatia	-0.00462	0.000415	-11.151
Balochi	Naga	Jamatia	-0.00451	0.000407	-11.082
BEB	Naga	Jamatia	-0.0036	0.000333	-10.788
Maratha	Naga	Jamatia	-0.00435	0.000406	-10.713
Santal	Naga	Jamatia	-0.0037	0.00035	-10.584
Naga	Sakilli	Jamatia	-0.00524	0.000502	-10.443
Ho	Naga	Jamatia	-0.00367	0.000354	-10.369
GIH	Tujia	Jamatia	-0.00315	0.000312	-10.11
Gond	Naga	Jamatia	-0.00471	0.000471	-10.005
CHS	GIH	Jamatia	-0.00255	0.000255	-9.988
ITU	Tujia	Jamatia	-0.00299	0.000306	-9.774
Irula	Naga	Jamatia	-0.00398	0.00041	-9.705
CHS	ITU	Jamatia	-0.00243	0.000253	-9.618
CHS	IE_UTT	Jamatia	-0.00235	0.000246	-9.558
GIH	Miaozu	Jamatia	-0.00296	0.00031	-9.527
PJL	Tujia	Jamatia	-0.00287	0.000303	-9.47
CHS	PJL	Jamatia	-0.00237	0.000251	-9.448
STU	Tujia	Jamatia	-0.00284	0.000302	-9.424
Korwa	Naga	Jamatia	-0.00347	0.000372	-9.307
Kadar	Naga	Jamatia	-0.0037	0.000399	-9.258
IE_UTT	Tujia	Jamatia	-0.00273	0.000296	-9.234
ITU	Miaozu	Jamatia	-0.00281	0.000304	-9.232
CHB	GIH	Jamatia	-0.00236	0.000257	-9.191
Iyer	CHS	Jamatia	-0.00243	0.000266	-9.131

CHS	Drav_TN	Jamatia	-0.00247	0.000271	-9.127
Pallan	Tujia	Jamatia	-0.00285	0.000315	-9.052
CHS	STU	Jamatia	-0.00227	0.000251	-9.029
IE_UTT	Miaozu	Jamatia	-0.00267	0.000296	-9.021
CHB	ITU	Jamatia	-0.00229	0.000254	-9.014
CHS	Drav_Kar	Jamatia	-0.00257	0.000285	-9.013
CHB	Drav_Kar	Jamatia	-0.00254	0.000283	-8.948
Gujrati_Brh	CHS	Jamatia	-0.00238	0.000267	-8.904
Iyer	Tujia	Jamatia	-0.00282	0.000318	-8.885
Naga	North_Kannadi	Jamatia	-0.00466	0.000528	-8.827
CHB	GIH	Manipuri	-0.01185	0.000193	-61.49
CHS	GIH	Manipuri	-0.01179	0.000194	-60.673
CHB	PJL	Manipuri	-0.01145	0.00019	-60.42
CHS	PJL	Manipuri	-0.01145	0.00019	-60.384
CHB	IE_UTT	Manipuri	-0.01078	0.000179	-60.179
CHB	ITU	Manipuri	-0.0111	0.000185	-59.971
CHS	IE_UTT	Manipuri	-0.01072	0.00018	-59.501
CHS	ITU	Manipuri	-0.01098	0.000185	-59.29
Han	PJL	Manipuri	-0.01129	0.000192	-58.665
GIH	Han	Manipuri	-0.01161	0.000199	-58.396
CHB	STU	Manipuri	-0.01068	0.000183	-58.392
Iyer	CHB	Manipuri	-0.01157	0.000201	-57.471
Iyer	CHS	Manipuri	-0.01153	0.000201	-57.417
Han	IE_UTT	Manipuri	-0.0106	0.000185	-57.232
Han	ITU	Manipuri	-0.01089	0.000191	-57.179
CHS	STU	Manipuri	-0.01053	0.000184	-57.122
GIH	Yizu	Manipuri	-0.0123	0.000217	-56.754
Gujrati_Brh	CHB	Manipuri	-0.01208	0.000215	-56.1
Gujrati_Brh	CHS	Manipuri	-0.01215	0.000217	-56.093
Han	STU	Manipuri	-0.01046	0.000187	-55.811
Khatri	CHS	Manipuri	-0.01223	0.000219	-55.723
W_Bengal_Brh	CHB	Manipuri	-0.0109	0.000196	-55.648
Khatri	CHB	Manipuri	-0.01212	0.000219	-55.325
CHS	Pathan	Manipuri	-0.01203	0.000218	-55.153
Iyer	Han	Manipuri	-0.01135	0.000207	-54.921
ITU	Yizu	Manipuri	-0.01144	0.000209	-54.752
CHB	Sindhi	Manipuri	-0.01219	0.000223	-54.697
CHB	Pathan	Manipuri	-0.01185	0.000217	-54.548
Balochi	CHB	Manipuri	-0.01273	0.000234	-54.375
W_Bengal_Brh	CHS	Manipuri	-0.01083	0.000199	-54.369
PJL	Yizu	Manipuri	-0.0119	0.000219	-54.344
CHS	Sindhi	Manipuri	-0.01223	0.000226	-54.031
Gujrati_Brh	Han	Manipuri	-0.01193	0.000221	-53.884
IE_UTT	Yizu	Manipuri	-0.0111	0.000207	-53.659
Balochi	CHS	Manipuri	-0.01284	0.00024	-53.499
Khatri	Han	Manipuri	-0.01197	0.000224	-53.442
Han	Pathan	Manipuri	-0.01183	0.000222	-53.266
CDX	GIH	Manipuri	-0.01064	0.0002	-53.254
STU	Yizu	Manipuri	-0.01097	0.000206	-53.2
W_Bengal_Brh	Han	Manipuri	-0.01071	0.000202	-53.127
CDX	PJL	Manipuri	-0.01032	0.000194	-53.068
GIH	JPT	Manipuri	-0.0105	0.000199	-52.852
GIH	Tujia	Manipuri	-0.01232	0.000234	-52.77
PJL	Tujia	Manipuri	-0.01189	0.000226	-52.641
CHB	Drav_AP	Manipuri	-0.0108	0.000206	-52.386
GIH	Naga	Manipuri	-0.01438	0.000275	-52.366
JPT	PJL	Manipuri	-0.01018	0.000195	-52.332
IE_UTT	Naga	Manipuri	-0.01327	0.000255	-52.134
Iyer	Yizu	Manipuri	-0.01189	0.000228	-52.12

Gujrati_Brh	Yizu	Manipuri	-0.01257	0.000241	-52.046
Dai	TIB	Naxi	-0.00168	0.000376	-4.457
CDX	TIB	Naxi	-0.00148	0.000339	-4.364
Dai	SHT	Naxi	-0.00187	0.000441	-4.252
Dai	SHK	Naxi	-0.00167	0.000401	-4.174
CDX	SHK	Naxi	-0.00145	0.000353	-4.11
CDX	SHT	Naxi	-0.00155	0.000387	-4.014
CDX	SRH	Naxi	-0.00144	0.000379	-3.798
Dai	SRH	Naxi	-0.00158	0.000426	-3.704
KHV	TIB	Naxi	-0.00112	0.000335	-3.342
KHV	SHT	Naxi	-0.00122	0.000384	-3.162
KHV	SHK	Naxi	-0.00111	0.000351	-3.155
KHV	SRH	Naxi	-0.00103	0.000377	-2.737
TIB	SHT	SHK	-0.00312	0.000175	-17.855
CHB	SHT	SHK	-0.00336	0.000192	-17.489
JPT	SHT	SHK	-0.00342	0.000199	-17.15
CHS	SHT	SHK	-0.00342	0.000202	-16.905
KHV	SHT	SHK	-0.00337	0.000205	-16.402
Han	SHT	SHK	-0.00332	0.000206	-16.14
CDX	SHT	SHK	-0.00336	0.000217	-15.503
Japanese	SHT	SHK	-0.00332	0.000215	-15.485
Burma	SHT	SHK	-0.00308	0.000202	-15.294
Yizu	SHT	SHK	-0.00326	0.000214	-15.252
TAS	SHT	SHK	-0.00309	0.000205	-15.106
Yakut	SHT	SHK	-0.00354	0.00024	-14.741
Tharu	SHT	SHK	-0.00312	0.000212	-14.69
Xibo	SHT	SHK	-0.00365	0.000248	-14.684
Manipuri	SHT	SHK	-0.00318	0.000218	-14.582
Tripuri	SHT	SHK	-0.00305	0.000209	-14.571
Oroqen	SHT	SHK	-0.00354	0.000244	-14.529
Miaozu	SHT	SHK	-0.00336	0.000233	-14.452
Dai	SHT	SHK	-0.00346	0.000245	-14.145
Cambodians	SHT	SHK	-0.00346	0.000246	-14.044
Tu	SHT	SHK	-0.00309	0.00022	-14.029
Naxi	SHT	SHK	-0.00326	0.000232	-14.028
She	SHT	SHK	-0.00342	0.000246	-13.942
Jamatia	SHT	SHK	-0.00298	0.000214	-13.89
Hezhen	SHT	SHK	-0.00334	0.000241	-13.887
Uygur	SHT	SHK	-0.00346	0.00025	-13.873
Lahu	SHT	SHK	-0.00341	0.000248	-13.763
Mongola	SHT	SHK	-0.00319	0.000232	-13.729
BEB	SHT	SHK	-0.00339	0.000249	-13.599
Tujia	SHT	SHK	-0.00323	0.000241	-13.423
Daur	SHT	SHK	-0.00315	0.00024	-13.1
S_Munda	SHT	SHK	-0.00322	0.000248	-12.98
STU	SHT	SHK	-0.0034	0.000263	-12.948
GIH	SHT	SHK	-0.00344	0.000269	-12.793
W_Bengal_Brh	SHT	SHK	-0.00344	0.000269	-12.77
Burusho	SHT	SHK	-0.0034	0.000267	-12.744
Iyer	SHT	SHK	-0.00347	0.000273	-12.742
Naga	SHT	SHK	-0.00344	0.00027	-12.735
PJL	SHT	SHK	-0.00342	0.000268	-12.726
IE_UTT	SHT	SHK	-0.00332	0.000264	-12.581
Gujrati_Brh	SHT	SHK	-0.00349	0.000277	-12.577
Pallan	SHT	SHK	-0.00339	0.00027	-12.536
Balochi	SHT	SHK	-0.0036	0.00029	-12.437
Sindhi	SHT	SHK	-0.00354	0.000285	-12.41
Irula	SHT	SHK	-0.00345	0.000279	-12.385
Khatri	SHT	SHK	-0.00354	0.000286	-12.373

Pathan	SHT	SHK	-0.00355	0.000287	-12.373
Kadar	SHT	SHK	-0.00342	0.000278	-12.304
ITU	SHT	SHK	-0.0033	0.000269	-12.279
Santal	SHT	SHK	-0.00319	0.000262	-12.176
Gujrati_Brh	Naga	TAS	-0.0055	0.000394	-13.966
GIH	Naga	TAS	-0.00516	0.000381	-13.523
Brahui	Naga	TAS	-0.00561	0.000418	-13.412
Naga	Sindhi	TAS	-0.00539	0.000407	-13.252
Naga	Pathan	TAS	-0.00534	0.000405	-13.177
IE_UTT	Naga	TAS	-0.00472	0.000361	-13.078
Naga	PJL	TAS	-0.00501	0.000384	-13.046
ITU	Naga	TAS	-0.0048	0.000371	-12.941
Khatri	Naga	TAS	-0.00534	0.000418	-12.791
Naga	STU	TAS	-0.00468	0.00037	-12.667
Balochi	Naga	TAS	-0.00518	0.000418	-12.391
Iyer	Naga	TAS	-0.00496	0.000402	-12.329
W_Bengal_Brh	Naga	TAS	-0.0047	0.000386	-12.187
Pallan	Naga	TAS	-0.00442	0.000388	-11.388
Drav_AP	Naga	TAS	-0.00453	0.000415	-10.93
Drav_TN	Naga	TAS	-0.00469	0.000435	-10.768
Maratha	Naga	TAS	-0.00426	0.000407	-10.464
Drav_Kar	Naga	TAS	-0.00437	0.000433	-10.092
Naga	Sakilli	TAS	-0.00479	0.000487	-9.848
Burusho	Naga	TAS	-0.00371	0.000382	-9.724
GIH	TIB	TAS	-0.00262	0.000283	-9.259
Naga	North_Kannadi	TAS	-0.00483	0.000538	-8.974
BEB	Naga	TAS	-0.00308	0.000344	-8.942
Gujrati_Brh	TIB	TAS	-0.00258	0.000292	-8.851
Sindhi	TIB	TAS	-0.00263	0.000301	-8.758
ITU	TIB	TAS	-0.00241	0.000279	-8.62
Gujrati_Brh	Yizu	TAS	-0.00292	0.000339	-8.614
Gond	Naga	TAS	-0.00377	0.000442	-8.538
GIH	Yizu	TAS	-0.00275	0.000323	-8.496
Iyer	TIB	TAS	-0.00248	0.000292	-8.487
PJL	TIB	TAS	-0.00242	0.000286	-8.477
Pathan	Yizu	TAS	-0.00291	0.000344	-8.442
Sindhi	Yizu	TAS	-0.00294	0.000349	-8.411
Balochi	TIB	TAS	-0.00255	0.000308	-8.29
Pathan	TIB	TAS	-0.00247	0.000299	-8.28
Khatri	Yizu	TAS	-0.00285	0.000348	-8.196
IE_UTT	TIB	TAS	-0.00224	0.000275	-8.138
STU	TIB	TAS	-0.00225	0.000278	-8.073
Khatri	TIB	TAS	-0.00244	0.000304	-8.033
Brahui	TIB	TAS	-0.00244	0.000307	-7.928
Irula	Naga	TAS	-0.0033	0.000419	-7.885
Balochi	Yizu	TAS	-0.00283	0.00036	-7.867
W_Bengal_Brh	TIB	TAS	-0.0022	0.000282	-7.814
PJL	Yizu	TAS	-0.00252	0.000324	-7.758
Drav_TN	TIB	TAS	-0.00229	0.000297	-7.712
Brahui	Yizu	TAS	-0.00278	0.000363	-7.652
GIH	SHT	TAS	-0.00255	0.000335	-7.595
Maratha	TIB	TAS	-0.00225	0.000302	-7.457
Drav_AP	TIB	TAS	-0.00223	0.000301	-7.408
Iyer	Yizu	TAS	-0.00243	0.000328	-7.399
ITU	Naga	Tharu	-0.00824	0.000348	-23.661
IE_UTT	Naga	Tharu	-0.00794	0.000339	-23.44
Naga	STU	Tharu	-0.00805	0.000344	-23.392
GIH	Naga	Tharu	-0.00842	0.000364	-23.118
Naga	PJL	Tharu	-0.00791	0.000363	-21.812

CHB	ITU	Tharu	-0.00559	0.000259	-21.616
Iyer	Naga	Tharu	-0.00817	0.000379	-21.556
Pallan	Naga	Tharu	-0.00802	0.000373	-21.461
ITU	Yizu	Tharu	-0.00616	0.000289	-21.338
GIH	Yizu	Tharu	-0.00646	0.000303	-21.299
CHB	IE_UTT	Tharu	-0.00533	0.000251	-21.202
CHB	STU	Tharu	-0.00538	0.000254	-21.13
CHB	GIH	Tharu	-0.00578	0.000273	-21.127
Gujrati_Brh	Naga	Tharu	-0.00819	0.00039	-20.976
IE_UTT	Yizu	Tharu	-0.00588	0.000283	-20.803
STU	Yizu	Tharu	-0.0059	0.000287	-20.525
Iyer	CHB	Tharu	-0.00555	0.000273	-20.375
Maratha	CHB	Tharu	-0.0059	0.000294	-20.085
W_Bengal_Brh	Naga	Tharu	-0.00746	0.000371	-20.083
Iyer	Yizu	Tharu	-0.0061	0.000304	-20.063
Pallan	CHB	Tharu	-0.00545	0.000274	-19.908
Naga	Sindhi	Tharu	-0.00788	0.000398	-19.821
Drav_TN	Naga	Tharu	-0.008	0.000404	-19.774
PJL	Yizu	Tharu	-0.00588	0.000298	-19.759
Drav_AP	Naga	Tharu	-0.00799	0.000405	-19.729
Drav_Kar	Naga	Tharu	-0.0079	0.000405	-19.502
CHB	Drav_Kar	Tharu	-0.00571	0.000293	-19.484
CHS	ITU	Tharu	-0.00514	0.000265	-19.414
CHB	PJL	Tharu	-0.00521	0.00027	-19.315
Khatri	Naga	Tharu	-0.00793	0.000411	-19.301
CHS	IE_UTT	Tharu	-0.00493	0.000256	-19.251
ITU	Tujia	Tharu	-0.00577	0.0003	-19.239
Gujrati_Brh	Yizu	Tharu	-0.00606	0.000315	-19.234
Maratha	Yizu	Tharu	-0.00641	0.000334	-19.18
CHS	GIH	Tharu	-0.00538	0.00028	-19.178
Drav_Kar	Yizu	Tharu	-0.0061	0.000319	-19.098
Maratha	Naga	Tharu	-0.0079	0.000415	-19.025
GIH	Tujia	Tharu	-0.00604	0.000319	-18.947
CHB	Drav_AP	Tharu	-0.00541	0.000285	-18.942
ITU	TIB	Tharu	-0.00515	0.000272	-18.895
Brahui	Naga	Tharu	-0.00794	0.000421	-18.871
Pallan	Yizu	Tharu	-0.00596	0.000317	-18.8
CHB	Drav_TN	Tharu	-0.00545	0.00029	-18.769
Gujrati_Brh	CHB	Tharu	-0.00535	0.000285	-18.76
STU	Tujia	Tharu	-0.00553	0.000295	-18.747
IE_UTT	Naxi	Tharu	-0.0056	0.000299	-18.741
CHS	STU	Tharu	-0.00489	0.000262	-18.66
Naxi	STU	Tharu	-0.00561	0.000301	-18.643
Iyer	CHS	Tharu	-0.00518	0.000278	-18.642
Sindhi	Yizu	Tharu	-0.00588	0.000316	-18.602
She	SHT	TIB	-0.00066	0.000214	-3.105
JPT	SHT	TIB	-0.00051	0.000178	-2.846
Oroqen	SHT	TIB	-0.0006	0.000222	-2.689
Yizu	SHT	TIB	-0.00051	0.000191	-2.652
CHB	SHT	TIB	-0.00041	0.00016	-2.547
CHS	SHT	TIB	-0.00042	0.000168	-2.507
Yizu	SHK	TIB	-0.00037	0.000158	-2.36
Japanese	SHT	TIB	-0.00039	0.000187	-2.083
She	SHK	TIB	-0.00037	0.000181	-2.039
Han	SHT	TIB	-0.00034	0.000167	-2.003
CHS	IE_UTT	Tripuri	-0.00723	0.000212	-34.095
CHS	ITU	Tripuri	-0.00731	0.000221	-33.044
CHS	GIH	Tripuri	-0.0074	0.000224	-33.013
CHB	IE_UTT	Tripuri	-0.00706	0.000214	-32.927

CHS	PJL	Tripuri	-0.00722	0.00022	-32.866
CHS	STU	Tripuri	-0.00714	0.000217	-32.86
CHB	ITU	Tripuri	-0.0072	0.000223	-32.308
Han	IE_UTT	Tripuri	-0.00696	0.000217	-32.112
CHB	STU	Tripuri	-0.00706	0.00022	-32.085
CHB	GIH	Tripuri	-0.00724	0.000226	-32.012
CHB	PJL	Tripuri	-0.00699	0.000222	-31.518
W_Bengal_Brh	CHS	Tripuri	-0.00685	0.000217	-31.509
CHS	Sindhi	Tripuri	-0.00734	0.000233	-31.472
Iyer	CHS	Tripuri	-0.0073	0.000233	-31.34
Han	ITU	Tripuri	-0.00708	0.000227	-31.194
CHS	Pathan	Tripuri	-0.00718	0.00023	-31.189
Han	STU	Tripuri	-0.00692	0.000223	-31.088
Han	PJL	Tripuri	-0.0069	0.000224	-30.764
GIH	Han	Tripuri	-0.00708	0.000231	-30.688
Gujrati_Brh	CHS	Tripuri	-0.00723	0.000236	-30.674
Gond	CHB	Tripuri	-0.00623	0.000204	-30.491
BEB	CHS	Tripuri	-0.00606	0.000199	-30.406
W_Bengal_Brh	CHB	Tripuri	-0.00668	0.00022	-30.375
Pallan	CHB	Tripuri	-0.00705	0.000233	-30.279
Iyer	CHB	Tripuri	-0.00711	0.000235	-30.275
Pallan	CHS	Tripuri	-0.00709	0.000234	-30.24
IE_UTT	Tujia	Tripuri	-0.00761	0.000252	-30.232
CHS	Drav_Kar	Tripuri	-0.00744	0.000248	-30.016
CHB	Sindhi	Tripuri	-0.00706	0.000235	-30.007
BEB	CHB	Tripuri	-0.00603	0.000202	-29.92
STU	Tujia	Tripuri	-0.00772	0.000258	-29.919
CHB	Drav_Kar	Tripuri	-0.00743	0.000249	-29.885
Khatri	CHS	Tripuri	-0.00715	0.00024	-29.795
Balochi	CHS	Tripuri	-0.00738	0.000248	-29.779
ITU	Tujia	Tripuri	-0.00788	0.000265	-29.723
GIH	Tujia	Tripuri	-0.008	0.000269	-29.699
CHS	Drav_TN	Tripuri	-0.00735	0.000248	-29.692
PJL	Tujia	Tripuri	-0.00771	0.000262	-29.488
Pallan	Han	Tripuri	-0.0069	0.000234	-29.472
CHB	Drav_TN	Tripuri	-0.00725	0.000247	-29.381
Maratha	CHS	Tripuri	-0.00736	0.000251	-29.364
W_Bengal_Brh	Han	Tripuri	-0.00657	0.000224	-29.304
CHB	Pathan	Tripuri	-0.00676	0.000232	-29.179
Gujrati_Brh	CHB	Tripuri	-0.00693	0.000238	-29.087
Brahui	CHS	Tripuri	-0.00731	0.000251	-29.085
BEB	Han	Tripuri	-0.00589	0.000203	-29.062
IE_UTT	Miaozu	Tripuri	-0.00748	0.000258	-28.96
GIH	Yizu	Tripuri	-0.00767	0.000266	-28.83
Han	Pathan	Tripuri	-0.00682	0.000237	-28.813
Iyer	Han	Tripuri	-0.00696	0.000242	-28.81
IE_UTT	Yizu	Tripuri	-0.00736	0.000256	-28.706
ITU	Miaozu	Tripuri	-0.00761	0.000266	-28.638
ITU	Yizu	Tripuri	-0.00751	0.000263	-28.577
IE_UTT	Naga	Tripuri	-0.00924	0.000323	-28.575
CDX	TIB	Yizu	-0.00182	0.000229	-7.943
Dai	TIB	Yizu	-0.00193	0.000264	-7.315
KHV	TIB	Yizu	-0.00152	0.00022	-6.919
CDX	SRH	Yizu	-0.0016	0.000256	-6.252
CDX	SHK	Yizu	-0.00151	0.000257	-5.891
CDX	SHT	Yizu	-0.00161	0.000281	-5.737
Dai	SHT	Yizu	-0.00185	0.000328	-5.624
Dai	SHK	Yizu	-0.00165	0.000294	-5.604
Dai	SRH	Yizu	-0.00165	0.000303	-5.438

KHV	SRH	Yizu	-0.00125	0.000248	-5.041
KHV	SHK	Yizu	-0.00123	0.000244	-5.031
KHV	SHT	Yizu	-0.00134	0.000266	-5.013
Cambodians	TIB	Yizu	-0.001	0.000272	-3.668
Lahu	TIB	Yizu	-0.0009	0.000312	-2.877
CHS	TIB	Yizu	-0.00054	0.000214	-2.511
Miaozu	TIB	Yizu	-0.00064	0.000256	-2.485

Appendix Table 3. Population trios presenting sgnificant results after ALDER analyses.

Admixed	Parental 1	Parental 2	p-value	Z-score	Decay rate	Amplitude
Jamatia	Balochi	CDX	6.20E-11	6.54	35.00 +/- 4.23	0.00009961 +/- 0.00001523
Jamatia	Balochi	CHB	3.60E-15	7.87	35.12 +/- 3.56	0.00009949 +/- 0.00001264
Jamatia	Balochi	CHS	3.50E-10	6.27	34.58 +/- 4.40	0.00010047 +/- 0.00001602
Jamatia	Balochi	Dai	2.80E-11	6.66	39.03 +/- 5.86	0.00010132 +/- 0.00001363
Jamatia	Balochi	Han	1.90E-14	7.66	33.65 +/- 4.18	0.00009865 +/- 0.00001288
Jamatia	Balochi	Japanese	4.60E-19	8.92	32.29 +/- 3.37	0.00008901 +/- 0.00000998
Jamatia	Balochi	KHV	1.50E-08	5.66	34.03 +/- 5.37	0.00008784 +/- 0.00001553
Jamatia	Balochi	She	1.90E-11	6.71	41.68 +/- 6.21	0.00011073 +/- 0.00001352
Jamatia	Balochi	Tujia	4.30E-13	7.25	41.25 +/- 5.57	0.00011618 +/- 0.00001603
Jamatia	Ho	CHB	1.10E-05	4.4	60.38 +/- 9.78	0.00007721 +/- 0.00001753
Jamatia	Ho	Han	4.80E-06	4.57	60.37 +/- 10.69	0.00007586 +/- 0.00001658
Jamatia	Ho	Tujia	1.10E-05	4.4	59.41 +/- 10.31	0.00007230 +/- 0.00001642
Jamatia	Irula	CDX	8.30E-10	6.14	45.88 +/- 7.47	0.00008165 +/- 0.00001247
Jamatia	Irula	CHB	8.80E-10	6.13	49.84 +/- 8.13	0.00009615 +/- 0.00001437
Jamatia	Irula	CHS	8.60E-06	4.45	49.08 +/- 11.03	0.00009357 +/- 0.00002103
Jamatia	Irula	Dai	2.70E-09	5.95	44.96 +/- 7.56	0.00008382 +/- 0.00001242
Jamatia	Irula	Han	5.90E-09	5.82	45.24 +/- 7.77	0.00008656 +/- 0.00001272
Jamatia	Irula	She	7.00E-09	5.79	49.23 +/- 8.50	0.00009067 +/- 0.00001438
Jamatia	Irula	Tujia	4.00E-11	6.6	50.30 +/- 7.62	0.00009545 +/- 0.00001394
Jamatia	Iyer	CDX	1.90E-06	4.76	33.43 +/- 7.02	0.00007757 +/- 0.00001617
Jamatia	Iyer	CHB	1.70E-12	7.05	38.48 +/- 4.79	0.00009459 +/- 0.00001341
Jamatia	Iyer	CHS	1.90E-09	6.01	35.13 +/- 5.85	0.00008580 +/- 0.00001357
Jamatia	Iyer	Dai	8.00E-07	4.94	38.14 +/- 6.57	0.00008848 +/- 0.00001793
Jamatia	Iyer	Han	3.10E-08	5.54	34.18 +/- 5.97	0.00008401 +/- 0.00001517
Jamatia	Iyer	Japanese	1.10E-08	5.71	32.04 +/- 5.30	0.00007221 +/- 0.00001265
Jamatia	Iyer	KHV	3.50E-07	5.09	35.61 +/- 6.64	0.00007692 +/- 0.00001510
Jamatia	Iyer	She	4.10E-07	5.07	39.33 +/- 7.76	0.00009130 +/- 0.00001651
Jamatia	Iyer	Tujia	1.30E-07	5.28	42.70 +/- 7.17	0.00010545 +/- 0.00001997
Jamatia	Kadar	CDX	5.00E-11	6.57	56.75 +/- 8.63	0.00008751 +/- 0.00001237
Jamatia	Kadar	CHB	2.90E-11	6.65	58.32 +/- 8.74	0.00010311 +/- 0.00001550
Jamatia	Kadar	CHS	3.40E-05	4.14	57.31 +/- 12.94	0.00009908 +/- 0.00002392
Jamatia	Kadar	Dai	1.30E-07	5.28	57.01 +/- 10.81	0.00009185 +/- 0.00001648
Jamatia	Kadar	Han	8.20E-10	6.14	51.83 +/- 8.44	0.00008735 +/- 0.00001322
Jamatia	Kadar	Japanese	3.10E-07	5.12	51.76 +/- 10.12	0.00007604 +/- 0.00001423
Jamatia	Kadar	She	1.90E-11	6.71	59.91 +/- 8.92	0.00009854 +/- 0.00001365
Jamatia	Kadar	Tujia	1.20E-10	6.44	58.83 +/- 8.71	0.00010359 +/- 0.00001607
Jamatia	Khatri	CDX	2.30E-17	8.48	37.45 +/- 4.24	0.00009386 +/- 0.00001107
Jamatia	Khatri	CHB	2.50E-17	8.47	38.03 +/- 3.75	0.00009744 +/- 0.00001151
Jamatia	Khatri	CHS	3.60E-15	7.87	36.83 +/- 4.44	0.00009410 +/- 0.00001196
Jamatia	Khatri	Dai	1.60E-19	9.04	44.26 +/- 4.80	0.00010928 +/- 0.00001209
Jamatia	Khatri	Han	9.30E-15	7.75	37.14 +/- 4.79	0.00009752 +/- 0.00001259
Jamatia	Khatri	Japanese	1.20E-15	8.01	34.33 +/- 4.29	0.00008509 +/- 0.00001047
Jamatia	Khatri	KHV	3.40E-11	6.63	35.42 +/- 4.60	0.00008049 +/- 0.00001214
Jamatia	Khatri	She	3.50E-12	6.95	44.87 +/- 6.45	0.00011117 +/- 0.00001409
Jamatia	Khatri	Tujia	7.10E-13	7.18	46.37 +/- 5.39	0.00012198 +/- 0.00001700
Jamatia	Korwa	CHB	2.70E-07	5.15	63.68 +/- 10.29	0.00008334 +/- 0.00001619
Jamatia	Korwa	Han	2.00E-07	5.2	59.71 +/- 10.33	0.00007462 +/- 0.00001435
Jamatia	Korwa	Tujia	1.20E-06	4.86	63.16 +/- 10.23	0.00007559 +/- 0.00001556
Jamatia	Maratha	CDX	3.40E-07	5.1	45.95 +/- 9.01	0.00009848 +/- 0.00001383
Jamatia	Maratha	CHB	1.10E-08	5.72	45.30 +/- 7.92	0.00009994 +/- 0.00001310
Jamatia	Maratha	CHS	1.50E-06	4.81	42.68 +/- 8.87	0.00009537 +/- 0.00001760
Jamatia	Maratha	Dai	2.70E-08	5.56	56.49 +/- 10.16	0.00011454 +/- 0.00001515
Jamatia	Maratha	Han	2.10E-07	5.19	44.25 +/- 8.53	0.00010004 +/- 0.00001496
Jamatia	Maratha	Japanese	9.80E-07	4.9	40.25 +/- 8.22	0.00008443 +/- 0.00001176
Jamatia	Maratha	KHV	1.80E-05	4.29	45.81 +/- 10.68	0.00008679 +/- 0.00001652
Jamatia	Maratha	She	1.70E-06	4.78	50.40 +/- 10.53	0.00010651 +/- 0.00001625

Jamatia	Maratha	Tujia	2.60E-11	6.67	51.22 +/- 7.53	0.00011835 +/- 0.00001775
Jamatia	Pallan	CDX	1.60E-12	7.06	45.16 +/- 6.39	0.00008945 +/- 0.00001144
Jamatia	Pallan	CHB	2.10E-13	7.34	47.78 +/- 6.51	0.00010160 +/- 0.00001246
Jamatia	Pallan	CHS	1.60E-07	5.24	43.95 +/- 8.38	0.00009134 +/- 0.00001679
Jamatia	Pallan	Dai	1.20E-11	6.78	51.68 +/- 7.62	0.00010653 +/- 0.00001477
Jamatia	Pallan	Han	1.10E-11	6.79	43.60 +/- 6.42	0.00009386 +/- 0.00001229
Jamatia	Pallan	Japanese	9.90E-11	6.47	41.49 +/- 6.41	0.00007845 +/- 0.00001085
Jamatia	Pallan	KHV	4.20E-07	5.06	45.38 +/- 8.97	0.00008209 +/- 0.00001615
Jamatia	Pallan	She	1.00E-10	6.46	51.73 +/- 8.01	0.00010776 +/- 0.00001406
Jamatia	Pallan	Tujia	3.90E-14	7.57	50.38 +/- 6.26	0.00011094 +/- 0.00001466
Jamatia	W_Bengal_Brh	CDX	2.30E-07	5.17	43.30 +/- 8.37	0.00009930 +/- 0.00001203
Jamatia	W_Bengal_Brh	CHB	6.80E-13	7.18	46.03 +/- 6.41	0.00010891 +/- 0.00001020
Jamatia	W_Bengal_Brh	CHS	3.40E-07	5.1	42.60 +/- 8.35	0.00010188 +/- 0.00001395
Jamatia	W_Bengal_Brh	Dai	3.00E-10	6.3	48.36 +/- 7.68	0.00010458 +/- 0.00001382
Jamatia	W_Bengal_Brh	Han	1.30E-08	5.69	43.32 +/- 7.62	0.00010437 +/- 0.00001247
Jamatia	W_Bengal_Brh	Japanese	1.20E-08	5.7	43.12 +/- 7.57	0.00009517 +/- 0.00001157
Jamatia	W_Bengal_Brh	She	2.50E-07	5.16	49.42 +/- 9.59	0.00010999 +/- 0.00001710
Jamatia	W_Bengal_Brh	Tujia	1.10E-11	6.79	47.89 +/- 7.02	0.00011281 +/- 0.00001661
Jamatia	Brahui	CDX	1.70E-15	7.96	36.53 +/- 4.59	0.00009648 +/- 0.00001147
Jamatia	Brahui	CHB	3.10E-19	8.96	37.39 +/- 3.97	0.00009930 +/- 0.00001108
Jamatia	Brahui	CHS	1.10E-13	7.43	36.56 +/- 4.92	0.00009903 +/- 0.00001234
Jamatia	Brahui	Dai	1.40E-12	7.08	41.01 +/- 5.65	0.00010185 +/- 0.00001438
Jamatia	Brahui	Han	2.80E-13	7.3	36.43 +/- 4.99	0.00010036 +/- 0.00001204
Jamatia	Brahui	Japanese	2.20E-15	7.93	35.03 +/- 4.42	0.00008950 +/- 0.00000940
Jamatia	Brahui	KHV	9.40E-10	6.12	35.28 +/- 5.49	0.00008417 +/- 0.00001376
Jamatia	Brahui	She	9.10E-11	6.48	42.79 +/- 6.60	0.00010652 +/- 0.00001433
Jamatia	Brahui	Tujia	8.30E-13	7.16	42.79 +/- 5.67	0.00011569 +/- 0.00001617
Jamatia	Burusho	CDX	1.50E-13	7.39	36.44 +/- 4.93	0.00007113 +/- 0.00000860
Jamatia	Burusho	CHB	2.20E-16	8.21	37.32 +/- 4.55	0.00007376 +/- 0.00000748
Jamatia	Burusho	CHS	4.50E-10	6.24	34.90 +/- 5.60	0.00006867 +/- 0.00000980
Jamatia	Burusho	Dai	2.00E-08	5.61	44.72 +/- 7.97	0.00008457 +/- 0.00001093
Jamatia	Burusho	Han	2.60E-11	6.67	36.10 +/- 5.41	0.00007421 +/- 0.00000873
Jamatia	Burusho	She	2.30E-05	4.23	49.56 +/- 11.70	0.00009489 +/- 0.00001725
Jamatia	Burusho	Tujia	1.50E-14	7.69	44.28 +/- 5.76	0.00009054 +/- 0.00001149
Jamatia	CDX	Drav_Kar	5.50E-13	7.21	51.79 +/- 7.18	0.00010249 +/- 0.00001146
Jamatia	CDX	GIH	8.70E-11	6.49	38.36 +/- 5.91	0.00009161 +/- 0.00001148
Jamatia	CDX	Drav_TN	4.10E-11	6.6	42.36 +/- 6.42	0.00009283 +/- 0.00001209
Jamatia	CDX	Pulliyar	0.00069	3.39	29.37 +/- 8.65	0.00005070 +/- 0.00000816
Jamatia	CDX	North_Kannadi	0.0078	2.66	66.85 +/- 12.05	0.00017463 +/- 0.00006566
Jamatia	CDX	Paniya	5.00E-06	4.57	48.75 +/- 10.68	0.00007968 +/- 0.00001596
Jamatia	CDX	Sakilli	0.0036	2.92	31.70 +/- 10.87	0.00011531 +/- 0.00003642
Jamatia	CDX	Sindhi	4.40E-08	5.47	36.71 +/- 6.71	0.00008723 +/- 0.00001311
Jamatia	CHB	Drav_Kar	3.90E-17	8.42	54.49 +/- 6.47	0.00011807 +/- 0.00001162
Jamatia	CHB	GIH	2.90E-15	7.89	39.37 +/- 4.99	0.00009858 +/- 0.00000984
Jamatia	CHB	Drav_TN	2.40E-11	6.68	43.18 +/- 6.46	0.00009814 +/- 0.00001321
Jamatia	CHB	Gond	2.70E-07	5.15	56.28 +/- 9.85	0.00011167 +/- 0.00002170
Jamatia	CHB	Pulliyar	0.00014	3.81	41.26 +/- 10.83	0.00007321 +/- 0.00001865
Jamatia	CHB	North_Kannadi	0.0087	2.62	65.45 +/- 14.00	0.00016443 +/- 0.00006264
Jamatia	CHB	Paniya	1.50E-08	5.66	51.45 +/- 9.08	0.00009530 +/- 0.00001510
Jamatia	CHB	Sakilli	0.003	2.97	34.64 +/- 11.67	0.00012212 +/- 0.00002706
Jamatia	CHB	Sindhi	3.30E-12	6.97	39.55 +/- 5.68	0.00009741 +/- 0.00001287
Jamatia	CHS	Drav_Kar	6.10E-09	5.81	54.83 +/- 9.43	0.00011985 +/- 0.00001944
Jamatia	CHS	GIH	2.20E-09	5.98	35.94 +/- 6.01	0.00008848 +/- 0.00001193
Jamatia	CHS	Drav_TN	6.10E-07	4.99	39.89 +/- 8.00	0.00008824 +/- 0.00001515
Jamatia	CHS	Gond	0.00011	3.88	55.59 +/- 12.67	0.00010458 +/- 0.00002697
Jamatia	CHS	Pulliyar	0.0054	2.78	29.40 +/- 10.56	0.00005217 +/- 0.00001477
Jamatia	CHS	North_Kannadi	0.029	2.19	63.60 +/- 15.01	0.00016521 +/- 0.00007556
Jamatia	CHS	Paniya	8.10E-05	3.94	48.23 +/- 12.24	0.00008379 +/- 0.00002046
Jamatia	CHS	Sakilli	0.005	2.81	33.48 +/- 8.94	0.00012536 +/- 0.00004461

Jamatia	CHS	Sindhi	1.70E-08	5.64	37.65 +/- 6.68	0.00009352 +/- 0.00001440
Jamatia	Dai	Drav_Kar	3.10E-15	7.89	55.53 +/- 7.04	0.00011479 +/- 0.00001268
Jamatia	Dai	GIH	9.80E-11	6.47	43.35 +/- 6.70	0.00010448 +/- 0.00001457
Jamatia	Dai	Drav_TN	7.10E-11	6.52	52.33 +/- 8.03	0.00011474 +/- 0.00001737
Jamatia	Dai	Gond	2.70E-13	7.31	61.63 +/- 7.65	0.00011969 +/- 0.00001638
Jamatia	Dai	Pulliyar	0.039	2.06	50.08 +/- 24.30	0.00007610 +/- 0.00002699
Jamatia	Dai	North_Kannadi	1.90E-07	5.21	65.12 +/- 9.25	0.00015164 +/- 0.00002913
Jamatia	Dai	Paniya	0.00013	3.82	53.13 +/- 11.87	0.00009931 +/- 0.00002598
Jamatia	Dai	Sakilli	0.014	2.47	34.79 +/- 14.11	0.00010279 +/- 0.00000989
Jamatia	Dai	Sindhi	2.60E-09	5.95	42.88 +/- 7.12	0.00009783 +/- 0.00001643
Jamatia	Drav_Kar	Han	8.30E-13	7.16	48.26 +/- 6.74	0.00010083 +/- 0.00001232
Jamatia	Drav_Kar	Japanese	2.20E-14	7.64	47.27 +/- 5.97	0.00008847 +/- 0.00001158
Jamatia	Drav_Kar	KHV	1.40E-08	5.67	58.79 +/- 9.87	0.00011226 +/- 0.00001978
Jamatia	Drav_Kar	She	6.30E-09	5.81	55.49 +/- 9.55	0.00011088 +/- 0.00001516
Jamatia	Drav_Kar	Tujia	5.70E-14	7.51	53.16 +/- 7.07	0.00011407 +/- 0.00001386
Jamatia	GIH	Han	1.40E-11	6.76	37.42 +/- 5.54	0.00009617 +/- 0.00001167
Jamatia	GIH	Japanese	1.30E-11	6.77	36.19 +/- 5.35	0.00008507 +/- 0.00001095
Jamatia	GIH	KHV	5.10E-08	5.45	36.11 +/- 6.63	0.00007889 +/- 0.00001282
Jamatia	GIH	She	2.90E-08	5.55	44.55 +/- 8.03	0.00010673 +/- 0.00001527
Jamatia	GIH	Tujia	6.90E-14	7.49	45.16 +/- 6.03	0.00011549 +/- 0.00001411
Jamatia	Han	Drav_TN	1.30E-09	6.06	41.02 +/- 6.77	0.00009600 +/- 0.00001379
Jamatia	Han	Gond	5.40E-08	5.44	56.26 +/- 9.74	0.00011284 +/- 0.00002076
Jamatia	Han	Pulliyar	0.00016	3.77	27.89 +/- 7.39	0.00005188 +/- 0.00001029
Jamatia	Han	North_Kannadi	0.0011	3.26	62.73 +/- 10.21	0.00016247 +/- 0.00004977
Jamatia	Han	Paniya	8.60E-05	3.93	45.33 +/- 11.54	0.00008162 +/- 0.00001884
Jamatia	Han	Sakilli	0.0016	3.16	31.48 +/- 9.95	0.00011398 +/- 0.00002740
Jamatia	Han	Sindhi	5.40E-10	6.21	36.48 +/- 5.88	0.00009171 +/- 0.00001281
Jamatia	Japanese	Drav_TN	2.40E-09	5.97	39.92 +/- 6.69	0.00008422 +/- 0.00001287
Jamatia	Japanese	Pulliyar	0.0001	3.89	26.35 +/- 6.77	0.00004612 +/- 0.00000884
Jamatia	Japanese	North_Kannadi	0.0022	3.06	63.92 +/- 10.81	0.00015704 +/- 0.00005136
Jamatia	Japanese	Sakilli	0.016	2.42	32.81 +/- 13.57	0.00011638 +/- 0.00002783
Jamatia	Japanese	Sindhi	7.00E-12	6.86	33.79 +/- 4.93	0.00007970 +/- 0.00000989
Jamatia	KHV	Drav_TN	8.00E-07	4.94	39.83 +/- 8.07	0.00007840 +/- 0.00001530
Jamatia	KHV	North_Kannadi	0.038	2.07	66.18 +/- 17.84	0.00015270 +/- 0.00007371
Jamatia	KHV	Sakilli	0.0074	2.68	32.21 +/- 7.17	0.00010596 +/- 0.00003959
Jamatia	KHV	Sindhi	9.80E-06	4.42	37.00 +/- 8.37	0.00008095 +/- 0.00001649
Jamatia	Drav_TN	She	7.40E-12	6.85	50.78 +/- 7.41	0.00011153 +/- 0.00001337
Jamatia	Drav_TN	Tujia	2.30E-10	6.34	49.74 +/- 7.70	0.00011669 +/- 0.00001841
Jamatia	Gond	She	1.30E-10	6.43	60.50 +/- 8.99	0.00012622 +/- 0.00001963
Jamatia	Gond	Tujia	4.50E-11	6.59	57.12 +/- 7.25	0.00012074 +/- 0.00001833
Jamatia	Pulliyar	She	0.021	2.3	39.21 +/- 17.04	0.00006171 +/- 0.00001844
Jamatia	Pulliyar	Tujia	5.20E-05	4.05	39.57 +/- 9.78	0.00007059 +/- 0.00001534
Jamatia	North_Kannadi	She	0.0003	3.61	82.51 +/- 15.98	0.00020711 +/- 0.00005734
Jamatia	North_Kannadi	Tujia	0.00062	3.42	73.84 +/- 14.03	0.00018132 +/- 0.00005295
Jamatia	Paniya	She	2.30E-05	4.23	48.56 +/- 11.47	0.00008801 +/- 0.00001852
Jamatia	Paniya	Tujia	2.50E-06	4.71	55.41 +/- 10.11	0.00010440 +/- 0.00002217
Jamatia	She	Sindhi	1.10E-06	4.87	42.87 +/- 8.81	0.00009924 +/- 0.00001798
Jamatia	Sindhi	Tujia	9.10E-11	6.48	44.23 +/- 6.82	0.00011170 +/- 0.00001701
Manipuri	Balochi	Naga	2.00E-13	7.35	11.87 +/- 1.61	0.00024147 +/- 0.00002670
Manipuri	Gond	Naga	3.70E-15	7.87	11.40 +/- 1.45	0.00014404 +/- 0.00001380
Manipuri	Gujrati_Brh	Naga	2.70E-12	6.99	11.34 +/- 1.62	0.00021500 +/- 0.00002266
Manipuri	Ho	Naga	2.30E-10	6.34	12.64 +/- 1.99	0.00012030 +/- 0.00001348
Manipuri	Irula	Naga	8.00E-14	7.47	10.62 +/- 1.42	0.00015933 +/- 0.00001555
Manipuri	Iyer	Naga	3.80E-14	7.57	11.42 +/- 1.51	0.00021271 +/- 0.00002161
Manipuri	Kadar	Naga	3.90E-12	6.94	11.89 +/- 1.71	0.00016534 +/- 0.00001710
Manipuri	Khatri	Naga	1.40E-16	8.26	11.53 +/- 1.40	0.00023071 +/- 0.00002351
Manipuri	Korwa	Naga	8.20E-15	7.76	12.79 +/- 1.65	0.00012258 +/- 0.00001273
Manipuri	Maratha	Naga	1.40E-14	7.69	12.29 +/- 1.60	0.00020509 +/- 0.00001987
Manipuri	Pallan	Naga	4.50E-17	8.4	11.89 +/- 1.42	0.00019792 +/- 0.00001847

Manipuri	Santal	Naga	7.30E-15	7.78	12.54 +/- 1.61	0.00013050 +/- 0.00001124
Manipuri	W_Bengal_Brh	Naga	1.10E-10	6.46	10.80 +/- 1.67	0.00019759 +/- 0.00002392
Manipuri	Brahui	Naga	2.00E-12	7.03	10.92 +/- 1.55	0.00022934 +/- 0.00002707
Manipuri	Burusho	Naga	1.10E-13	7.43	12.01 +/- 1.62	0.00019357 +/- 0.00002171
Manipuri	Drav_Kar	Naga	4.00E-18	8.68	11.72 +/- 1.35	0.00019710 +/- 0.00001646
Manipuri	GIH	Naga	4.00E-13	7.25	11.02 +/- 1.52	0.00020703 +/- 0.00002089
Manipuri	ITU	Naga	8.30E-14	7.46	10.52 +/- 1.41	0.00018748 +/- 0.00001801
Manipuri	Drav_AP	Naga	9.60E-17	8.31	11.53 +/- 1.39	0.00020670 +/- 0.00001822
Manipuri	Drav_TN	Naga	3.10E-17	8.44	11.67 +/- 1.38	0.00020378 +/- 0.00001868
Tripuri	Balochi	Cambodians	2.20E-07	5.18	27.75 +/- 5.35	0.00006492 +/- 0.00001106
Tripuri	Balochi	CHB	8.30E-07	4.93	22.91 +/- 4.65	0.00007139 +/- 0.00001038
Tripuri	Balochi	SHK	1.90E-10	6.37	21.33 +/- 3.35	0.00006514 +/- 0.00000743
Tripuri	Balochi	SHT	1.40E-07	5.27	21.04 +/- 4.00	0.00006287 +/- 0.00000786
Tripuri	Birhor	CHB	0.00015	3.79	22.63 +/- 5.98	0.00003652 +/- 0.00000730
Tripuri	Birhor	SHK	6.40E-06	4.51	21.24 +/- 4.71	0.00003253 +/- 0.00000636
Tripuri	Birhor	SHT	3.20E-07	5.11	21.20 +/- 4.15	0.00003525 +/- 0.00000626
Tripuri	Gond	CHB	1.70E-06	4.79	20.27 +/- 4.23	0.00003892 +/- 0.00000610
Tripuri	Gond	SHK	2.20E-08	5.6	20.13 +/- 3.60	0.00003753 +/- 0.00000562
Tripuri	Gond	SHT	5.80E-06	4.53	21.46 +/- 4.73	0.00003926 +/- 0.00000647
Tripuri	Gujrati_Brh	Cambodians	3.10E-06	4.66	24.10 +/- 5.17	0.00005338 +/- 0.00000957
Tripuri	Gujrati_Brh	CHB	5.00E-07	5.02	19.20 +/- 3.82	0.00005887 +/- 0.00000778
Tripuri	Gujrati_Brh	SHK	1.60E-09	6.03	17.27 +/- 2.86	0.00005276 +/- 0.00000648
Tripuri	Gujrati_Brh	SHT	4.30E-07	5.06	17.00 +/- 3.36	0.00005211 +/- 0.00000724
Tripuri	Ho	CHB	9.00E-05	3.92	20.30 +/- 5.18	0.00002960 +/- 0.00000460
Tripuri	Ho	SHK	6.20E-08	5.41	19.57 +/- 3.62	0.00002786 +/- 0.00000429
Tripuri	Ho	SHT	1.00E-06	4.88	20.45 +/- 4.19	0.00002987 +/- 0.00000598
Tripuri	Irula	Cambodians	2.70E-06	4.69	32.02 +/- 6.83	0.00004316 +/- 0.00000725
Tripuri	Irula	CHB	1.70E-06	4.78	26.67 +/- 5.58	0.00005512 +/- 0.00000801
Tripuri	Irula	SHK	1.50E-08	5.67	26.18 +/- 4.62	0.00005367 +/- 0.00000740
Tripuri	Irula	SHT	2.60E-11	6.67	28.44 +/- 4.27	0.00005998 +/- 0.00000620
Tripuri	Iyer	Cambodians	0.00035	3.58	22.01 +/- 6.15	0.00004701 +/- 0.00000903
Tripuri	Iyer	CHB	1.20E-05	4.37	20.10 +/- 4.60	0.00005872 +/- 0.00000910
Tripuri	Iyer	SHK	3.80E-08	5.5	18.50 +/- 3.36	0.00005331 +/- 0.00000672
Tripuri	Iyer	SHT	8.60E-07	4.92	19.01 +/- 3.86	0.00005451 +/- 0.00000726
Tripuri	Kadar	Cambodians	0.0028	2.99	24.02 +/- 8.03	0.00003586 +/- 0.00000904
Tripuri	Kadar	CHB	0.00025	3.66	22.19 +/- 6.06	0.00004799 +/- 0.00000764
Tripuri	Kadar	SHK	7.00E-06	4.5	22.33 +/- 4.97	0.00004618 +/- 0.00000726
Tripuri	Kadar	SHT	1.60E-07	5.24	25.40 +/- 4.85	0.00005175 +/- 0.00000780
Tripuri	Khatri	Cambodians	8.30E-08	5.36	25.73 +/- 4.80	0.00005702 +/- 0.00000901
Tripuri	Khatri	CHB	1.10E-05	4.4	20.64 +/- 4.69	0.00006258 +/- 0.00000924
Tripuri	Khatri	SHK	2.80E-07	5.14	19.29 +/- 3.76	0.00005762 +/- 0.00000714
Tripuri	Khatri	SHT	3.80E-07	5.08	19.72 +/- 3.88	0.00005791 +/- 0.00000777
Tripuri	Korwa	CHB	0.025	2.25	24.80 +/- 11.03	0.00003576 +/- 0.00001035
Tripuri	Korwa	SHK	0.0002	3.72	29.70 +/- 7.98	0.00003912 +/- 0.00000841
Tripuri	Korwa	SHT	0.00031	3.61	28.23 +/- 7.82	0.00003818 +/- 0.00000880
Tripuri	Maratha	Cambodians	6.80E-07	4.97	30.01 +/- 6.04	0.00005365 +/- 0.00001058
Tripuri	Maratha	CHB	7.80E-07	4.94	22.09 +/- 4.47	0.00005593 +/- 0.00000941
Tripuri	Maratha	SHK	5.00E-09	5.85	18.96 +/- 3.24	0.00004716 +/- 0.00000628
Tripuri	Maratha	SHT	2.40E-07	5.17	19.80 +/- 3.83	0.00004990 +/- 0.00000751
Tripuri	Pallan	Cambodians	0.00019	3.74	22.88 +/- 6.12	0.00004384 +/- 0.00000984
Tripuri	Pallan	CHB	7.40E-05	3.96	20.35 +/- 5.13	0.00005325 +/- 0.00000891
Tripuri	Pallan	SHK	7.40E-07	4.95	19.43 +/- 3.93	0.00004972 +/- 0.00000708
Tripuri	Pallan	SHT	1.30E-05	4.37	20.15 +/- 4.62	0.00005154 +/- 0.00000887
Tripuri	Santal	CHB	1.50E-05	4.33	23.51 +/- 5.43	0.00003880 +/- 0.00000694
Tripuri	Santal	SHK	1.50E-09	6.04	22.86 +/- 3.79	0.00003535 +/- 0.00000498
Tripuri	Santal	SHT	2.30E-07	5.17	26.19 +/- 5.07	0.00004195 +/- 0.00000804
Tripuri	Tharu	CHB	0.00029	3.62	30.58 +/- 8.45	0.00001468 +/- 0.00000395
Tripuri	W_Bengal_Brh	Cambodians	3.50E-06	4.64	23.10 +/- 4.98	0.00004884 +/- 0.00000809
Tripuri	W_Bengal_Brh	CHB	9.10E-07	4.91	18.24 +/- 3.71	0.00005430 +/- 0.00000701

Tripuri	W_Bengal_Brh	SHK	2.80E-09	5.94	17.76 +/- 2.99	0.00005188 +/- 0.00000597
Tripuri	W_Bengal_Brh	SHT	2.10E-06	4.74	18.01 +/- 3.80	0.00005126 +/- 0.00000683
Tripuri	BEB	CHB	3.20E-07	5.11	20.38 +/- 3.99	0.00004494 +/- 0.00000575
Tripuri	BEB	SHK	5.70E-12	6.89	19.57 +/- 2.84	0.00004220 +/- 0.00000447
Tripuri	BEB	SHT	2.90E-08	5.55	18.93 +/- 3.41	0.00004044 +/- 0.00000493
Tripuri	Brahui	Cambodians	1.90E-05	4.27	24.60 +/- 5.75	0.00006217 +/- 0.00001228
Tripuri	Brahui	CHB	7.80E-06	4.47	20.57 +/- 4.60	0.00006994 +/- 0.00001131
Tripuri	Brahui	SHK	4.40E-09	5.87	19.24 +/- 3.28	0.00006499 +/- 0.00000780
Tripuri	Brahui	SHT	6.60E-07	4.97	18.85 +/- 3.79	0.00006234 +/- 0.00000858
Tripuri	Burusho	Cambodians	0.00027	3.64	20.08 +/- 5.52	0.00003755 +/- 0.00000689
Tripuri	Burusho	CHB	8.10E-07	4.93	18.68 +/- 3.79	0.00004862 +/- 0.00000622
Tripuri	Burusho	SHK	3.30E-07	5.1	17.68 +/- 3.46	0.00004428 +/- 0.00000566
Tripuri	Burusho	SHT	6.10E-06	4.52	15.96 +/- 3.53	0.00003989 +/- 0.00000508
Tripuri	Cambodians	Drav_Kar	6.60E-05	3.99	28.39 +/- 7.12	0.00004882 +/- 0.00000974
Tripuri	Cambodians	GIH	2.20E-05	4.25	21.57 +/- 5.08	0.00004363 +/- 0.00000872
Tripuri	Cambodians	ITU	1.40E-08	5.67	25.66 +/- 4.52	0.00005091 +/- 0.00000734
Tripuri	Cambodians	Drav_AP	9.10E-05	3.91	18.48 +/- 4.72	0.00003871 +/- 0.00000680
Tripuri	Cambodians	Drav_TN	4.10E-06	4.61	23.71 +/- 5.15	0.00004585 +/- 0.00000763
Tripuri	Cambodians	IE_UTT	3.80E-09	5.89	27.01 +/- 4.58	0.00005225 +/- 0.00000781
Tripuri	Cambodians	Pulliyar	0.0017	3.14	54.52 +/- 17.35	0.00007465 +/- 0.00002298
Tripuri	Cambodians	IAR	0.0038	2.89	41.68 +/- 14.41	0.00003191 +/- 0.00001020
Tripuri	Cambodians	North_Kannadi	0.011	2.53	29.45 +/- 10.49	0.00007396 +/- 0.00002920
Tripuri	Cambodians	Paniya	0.0006	3.43	26.02 +/- 7.58	0.00003851 +/- 0.00000887
Tripuri	Cambodians	Pathan	8.20E-07	4.93	28.96 +/- 5.87	0.00006421 +/- 0.00001082
Tripuri	Cambodians	PJL	5.90E-08	5.42	25.82 +/- 4.76	0.00005303 +/- 0.00000831
Tripuri	Cambodians	Sakilli	0.00078	3.36	31.73 +/- 6.15	0.00006930 +/- 0.00002062
Tripuri	Cambodians	Sindhi	2.50E-05	4.22	23.41 +/- 5.55	0.00005125 +/- 0.00000927
Tripuri	Cambodians	STU	5.20E-09	5.84	26.60 +/- 4.56	0.00005113 +/- 0.00000765
Tripuri	Khasi	SHK	0.022	2.29	104.34 +/- 23.30	0.00010732 +/- 0.00004678
Tripuri	Khasi	SHT	0.018	2.36	119.60 +/- 22.10	0.00014902 +/- 0.00006307
Tripuri	N_Munda	SHK	0.00026	3.65	26.87 +/- 7.36	0.00004070 +/- 0.00001010
Tripuri	N_Munda	SHT	0.0026	3.01	20.90 +/- 6.95	0.00003058 +/- 0.00000888
Tripuri	S_Munda	CHB	0.0066	2.72	19.87 +/- 7.31	0.00002712 +/- 0.00000768
Tripuri	S_Munda	SHK	2.50E-07	5.16	19.83 +/- 3.84	0.00002835 +/- 0.00000526
Tripuri	S_Munda	SHT	0.0026	3.01	21.32 +/- 7.09	0.00002848 +/- 0.00000745
Tripuri	CHB	Drav_Kar	0.00037	3.56	20.34 +/- 5.72	0.00005042 +/- 0.00000903
Tripuri	CHB	GIH	1.30E-06	4.84	19.81 +/- 4.09	0.00005760 +/- 0.00000802
Tripuri	CHB	ITU	1.60E-08	5.65	20.40 +/- 3.61	0.00005703 +/- 0.00000702
Tripuri	CHB	Drav_AP	1.60E-07	5.24	16.17 +/- 3.08	0.00004784 +/- 0.00000589
Tripuri	CHB	Drav_TN	2.30E-07	5.18	20.30 +/- 3.92	0.00005594 +/- 0.00000672
Tripuri	CHB	Gond	3.00E-09	5.93	24.79 +/- 4.18	0.00005850 +/- 0.00000770
Tripuri	CHB	IE_UTT	4.40E-07	5.05	21.22 +/- 4.20	0.00005850 +/- 0.00000815
Tripuri	CHB	Pulliyar	0.0001	3.88	35.80 +/- 9.22	0.00007081 +/- 0.00001371
Tripuri	CHB	IAR	0.00018	3.75	19.40 +/- 5.18	0.00002270 +/- 0.00000426
Tripuri	CHB	North_Kannadi	0.00061	3.43	23.41 +/- 6.83	0.00007517 +/- 0.00002003
Tripuri	CHB	Paniya	1.70E-06	4.78	22.98 +/- 4.81	0.00005395 +/- 0.00000912
Tripuri	CHB	Pathan	6.10E-05	4.01	22.08 +/- 5.51	0.00006684 +/- 0.00001185
Tripuri	CHB	PJL	1.20E-06	4.86	20.79 +/- 4.28	0.00005986 +/- 0.00000861
Tripuri	CHB	Sakilli	6.30E-06	4.51	26.30 +/- 5.28	0.00007604 +/- 0.00001684
Tripuri	CHB	Sindhi	2.00E-05	4.26	19.18 +/- 4.50	0.00005860 +/- 0.00000817
Tripuri	CHB	STU	1.20E-07	5.3	21.36 +/- 4.03	0.00005857 +/- 0.00000788
Tripuri	Drav_Kar	SHK	4.40E-06	4.59	20.45 +/- 4.45	0.00004871 +/- 0.00000810
Tripuri	Drav_Kar	SHT	5.20E-05	4.05	21.99 +/- 5.43	0.00005267 +/- 0.00000988
Tripuri	GIH	SHK	5.80E-10	6.19	18.78 +/- 3.03	0.00005429 +/- 0.00000647
Tripuri	GIH	SHT	2.90E-07	5.13	18.00 +/- 3.51	0.00005120 +/- 0.00000717
Tripuri	ITU	SHK	9.50E-13	7.14	19.13 +/- 2.68	0.00005262 +/- 0.00000545
Tripuri	ITU	SHT	6.70E-10	6.17	19.97 +/- 3.23	0.00005413 +/- 0.00000619
Tripuri	Drav_AP	SHK	4.60E-11	6.58	15.75 +/- 2.39	0.00004498 +/- 0.00000481
Tripuri	Drav_AP	SHT	6.00E-08	5.42	15.57 +/- 2.87	0.00004439 +/- 0.00000514

Tripuri	Drav_TN	SHK	2.10E-11	6.7	18.30 +/- 2.73	0.00005003 +/- 0.00000558
Tripuri	Drav_TN	SHT	2.50E-07	5.16	17.74 +/- 3.44	0.00004901 +/- 0.00000660
Tripuri	IE_UTT	SHK	8.70E-11	6.49	19.74 +/- 3.04	0.00005389 +/- 0.00000659
Tripuri	IE_UTT	SHT	6.90E-08	5.39	19.57 +/- 3.63	0.00005339 +/- 0.00000725
Tripuri	Pulliyar	SHK	9.80E-06	4.42	33.36 +/- 7.54	0.00006673 +/- 0.00001357
Tripuri	Pulliyar	SHT	1.70E-06	4.79	32.64 +/- 6.81	0.00006600 +/- 0.00001379
Tripuri	North_Kannadi	SHK	0.00035	3.57	21.29 +/- 5.32	0.00006639 +/- 0.00001858
Tripuri	North_Kannadi	SHT	2.70E-06	4.69	17.52 +/- 3.74	0.00005408 +/- 0.00000979
Tripuri	Paniya	SHK	5.90E-09	5.82	23.87 +/- 4.10	0.00005349 +/- 0.00000829
Tripuri	Paniya	SHT	3.40E-08	5.52	24.59 +/- 3.91	0.00005641 +/- 0.00001022
Tripuri	Pathan	SHK	1.10E-09	6.1	20.32 +/- 3.33	0.00006075 +/- 0.00000777
Tripuri	Pathan	SHT	5.70E-08	5.43	20.32 +/- 3.74	0.00006080 +/- 0.00000804
Tripuri	PJL	SHK	2.20E-10	6.35	19.39 +/- 3.06	0.00005486 +/- 0.00000669
Tripuri	PJL	SHT	6.40E-09	5.81	19.91 +/- 3.43	0.00005678 +/- 0.00000721
Tripuri	Sakilli	SHK	5.60E-06	4.54	24.57 +/- 3.72	0.00007077 +/- 0.00001558
Tripuri	Sakilli	SHT	1.50E-14	7.69	24.27 +/- 3.16	0.00006580 +/- 0.00000853
Tripuri	Sindhi	SHK	3.60E-09	5.9	17.42 +/- 2.95	0.00005268 +/- 0.00000588
Tripuri	Sindhi	SHT	4.00E-06	4.61	16.41 +/- 3.56	0.00005047 +/- 0.00000674
Tripuri	STU	SHK	2.70E-12	6.99	20.44 +/- 2.92	0.00005463 +/- 0.00000608
Tripuri	STU	SHT	9.30E-10	6.12	21.00 +/- 3.43	0.00005565 +/- 0.00000684
Tharu	Balochi	CDX	2.50E-05	4.22	33.03 +/- 7.83	0.00010533 +/- 0.00002349
Tharu	Balochi	CHB	0.0025	3.02	29.57 +/- 9.78	0.00010314 +/- 0.00003023
Tharu	Balochi	Han	2.40E-05	4.22	33.60 +/- 7.96	0.00011697 +/- 0.00002406
Tharu	Balochi	JPT	0.00035	3.58	33.98 +/- 9.50	0.00011538 +/- 0.00002754
Tharu	Balochi	KHV	4.10E-06	4.6	34.80 +/- 7.56	0.00010876 +/- 0.00002033
Tharu	Balochi	Lahu	1.00E-06	4.89	36.83 +/- 7.53	0.00012965 +/- 0.00002553
Tharu	Balochi	Miaozu	9.90E-06	4.42	33.43 +/- 7.56	0.00011883 +/- 0.00002109
Tharu	Balochi	Naxi	1.10E-06	4.88	33.95 +/- 6.96	0.00012558 +/- 0.00002515
Tharu	Balochi	She	0.0012	3.24	31.79 +/- 9.82	0.00011801 +/- 0.00002761
Tharu	Balochi	SHK	0.0042	2.86	25.36 +/- 8.87	0.00008855 +/- 0.00002395
Tharu	Balochi	Tu	7.90E-05	3.95	34.56 +/- 8.76	0.00010857 +/- 0.00002364
Tharu	Gond	CHB	0.00026	3.66	37.39 +/- 10.22	0.00007580 +/- 0.00001979
Tharu	Gond	Han	1.40E-05	4.35	37.35 +/- 8.59	0.00007498 +/- 0.00001490
Tharu	Gond	JPT	0.00049	3.49	35.88 +/- 10.29	0.00006810 +/- 0.00001571
Tharu	Gond	Miaozu	5.30E-05	4.04	35.21 +/- 8.71	0.00007007 +/- 0.00001416
Tharu	Gond	Naxi	4.60E-06	4.58	37.94 +/- 8.14	0.00008519 +/- 0.00001859
Tharu	Gond	She	0.0071	2.69	35.25 +/- 13.09	0.00007021 +/- 0.00002210
Tharu	Gond	SHK	0.001	3.28	28.54 +/- 8.70	0.00005578 +/- 0.00001442
Tharu	Gond	Tu	0.00083	3.34	37.07 +/- 11.09	0.00006736 +/- 0.00001791
Tharu	Gujrati_Brh	CDX	0.00085	3.34	27.72 +/- 8.31	0.00008015 +/- 0.00001845
Tharu	Gujrati_Brh	CHB	0.0053	2.79	28.24 +/- 10.13	0.00009160 +/- 0.00002632
Tharu	Gujrati_Brh	Han	0.002	3.09	27.72 +/- 8.98	0.00008801 +/- 0.00002265
Tharu	Gujrati_Brh	JPT	0.0028	2.99	29.27 +/- 9.79	0.00009004 +/- 0.00002244
Tharu	Gujrati_Brh	KHV	0.0023	3.05	25.24 +/- 8.26	0.00007208 +/- 0.00001751
Tharu	Gujrati_Brh	Lahu	0.00079	3.36	25.19 +/- 7.51	0.00007729 +/- 0.00001925
Tharu	Gujrati_Brh	Miaozu	0.00069	3.39	25.82 +/- 7.61	0.00008307 +/- 0.00001802
Tharu	Gujrati_Brh	Naxi	0.00012	3.84	29.70 +/- 7.72	0.00009843 +/- 0.00002177
Tharu	Gujrati_Brh	SHK	0.0065	2.72	20.63 +/- 7.58	0.00007074 +/- 0.00001776
Tharu	Gujrati_Brh	Tu	0.0024	3.04	33.27 +/- 10.94	0.00009600 +/- 0.00002636
Tharu	Ho	CHB	0.00092	3.31	48.00 +/- 12.97	0.00006937 +/- 0.00002093
Tharu	Ho	Naxi	7.50E-06	4.48	51.86 +/- 8.96	0.00008251 +/- 0.00001843
Tharu	Ho	SHK	0.014	2.45	31.35 +/- 12.78	0.00004198 +/- 0.00001415
Tharu	Irula	CDX	0.00021	3.71	38.17 +/- 10.28	0.00006975 +/- 0.00001649
Tharu	Irula	CHB	0.0025	3.02	37.67 +/- 12.46	0.00008036 +/- 0.00002425
Tharu	Irula	Han	0.0003	3.62	40.26 +/- 11.14	0.00008583 +/- 0.00001960
Tharu	Irula	JPT	0.00055	3.46	39.42 +/- 11.40	0.00007929 +/- 0.00001702
Tharu	Irula	KHV	7.00E-05	3.98	40.64 +/- 10.22	0.00007288 +/- 0.00001364
Tharu	Irula	Lahu	6.30E-05	4	37.74 +/- 9.43	0.00007587 +/- 0.00001600
Tharu	Irula	Miaozu	0.0002	3.72	44.38 +/- 11.93	0.00009214 +/- 0.00001970

Tharu	Irula	Naxi	8.40E-06	4.45	37.52 +/- 8.42	0.00008765 +/- 0.00001804
Tharu	Irula	She	0.0071	2.69	37.35 +/- 13.87	0.00008462 +/- 0.00002440
Tharu	Irula	SHK	0.017	2.39	30.73 +/- 12.86	0.00006767 +/- 0.00002331
Tharu	Irula	Tu	0.0028	2.99	40.94 +/- 13.70	0.00008049 +/- 0.00002347
Tharu	Iyer	CDX	0.0078	2.66	25.95 +/- 9.75	0.00007356 +/- 0.00002199
Tharu	Iyer	Han	0.012	2.52	26.62 +/- 10.55	0.00008449 +/- 0.00002754
Tharu	Iyer	JPT	0.016	2.41	28.63 +/- 11.88	0.00008701 +/- 0.00002829
Tharu	Iyer	KHV	0.021	2.31	23.77 +/- 10.27	0.00006728 +/- 0.00002277
Tharu	Iyer	Lahu	0.0075	2.67	25.43 +/- 9.52	0.00007454 +/- 0.00002354
Tharu	Iyer	Miaozu	0.0059	2.76	23.14 +/- 8.40	0.00007566 +/- 0.00002123
Tharu	Iyer	Naxi	5.80E-05	4.02	31.81 +/- 7.91	0.00010293 +/- 0.00002313
Tharu	Iyer	SHK	0.046	1.99	17.37 +/- 8.72	0.00006187 +/- 0.00002281
Tharu	Iyer	Tu	0.029	2.19	28.37 +/- 12.97	0.00008314 +/- 0.00003091
Tharu	Kadar	CDX	4.00E-07	5.07	41.91 +/- 8.27	0.00007845 +/- 0.00001395
Tharu	Kadar	CHB	1.30E-05	4.37	42.20 +/- 9.66	0.00009477 +/- 0.00002161
Tharu	Kadar	Han	6.50E-07	4.97	42.54 +/- 8.55	0.00009366 +/- 0.00001623
Tharu	Kadar	JPT	3.20E-05	4.16	41.21 +/- 9.92	0.00008641 +/- 0.00001604
Tharu	Kadar	KHV	1.10E-07	5.31	42.98 +/- 8.09	0.00007874 +/- 0.00001170
Tharu	Kadar	Lahu	3.20E-06	4.66	39.56 +/- 8.49	0.00008157 +/- 0.00001439
Tharu	Kadar	Miaozu	3.40E-05	4.14	43.56 +/- 10.51	0.00009287 +/- 0.00001804
Tharu	Kadar	Naxi	7.80E-10	6.15	40.64 +/- 6.61	0.00009902 +/- 0.00001486
Tharu	Kadar	She	0.00036	3.57	40.79 +/- 11.43	0.00009248 +/- 0.00002208
Tharu	Kadar	SHK	0.0012	3.25	36.43 +/- 11.22	0.00008150 +/- 0.00002217
Tharu	Kadar	Tu	7.90E-05	3.95	45.12 +/- 11.43	0.00009369 +/- 0.00002162
Tharu	Khatri	CDX	0.0006	3.43	28.98 +/- 8.44	0.00008276 +/- 0.00002169
Tharu	Khatri	CHB	0.0024	3.03	28.18 +/- 9.30	0.00009120 +/- 0.00002572
Tharu	Khatri	Han	0.0018	3.12	27.05 +/- 8.67	0.00008558 +/- 0.00002351
Tharu	Khatri	JPT	0.0055	2.77	28.70 +/- 10.35	0.00008797 +/- 0.00002666
Tharu	Khatri	KHV	0.0013	3.21	25.75 +/- 8.01	0.00007276 +/- 0.00001942
Tharu	Khatri	Lahu	0.0003	3.62	25.11 +/- 6.94	0.00007399 +/- 0.00001907
Tharu	Khatri	Miaozu	0.00061	3.43	24.53 +/- 7.16	0.00007901 +/- 0.00001841
Tharu	Khatri	Naxi	0.00036	3.57	30.43 +/- 8.53	0.00010019 +/- 0.00002479
Tharu	Khatri	SHK	0.004	2.88	20.87 +/- 7.24	0.00007094 +/- 0.00001920
Tharu	Khatri	Tu	0.0098	2.58	28.10 +/- 10.88	0.00008207 +/- 0.00002764
Tharu	Maratha	CDX	5.70E-06	4.54	32.33 +/- 7.13	0.00007668 +/- 0.00001640
Tharu	Maratha	CHB	0.00014	3.81	31.06 +/- 8.16	0.00008489 +/- 0.00002041
Tharu	Maratha	Han	4.60E-05	4.07	30.30 +/- 7.44	0.00008092 +/- 0.00001749
Tharu	Maratha	JPT	0.00012	3.84	30.36 +/- 7.91	0.00007923 +/- 0.00001725
Tharu	Maratha	KHV	2.40E-05	4.22	29.93 +/- 7.09	0.00006899 +/- 0.00001493
Tharu	Maratha	Lahu	0.00054	3.46	31.57 +/- 7.40	0.00008136 +/- 0.00002350
Tharu	Maratha	Miaozu	1.80E-05	4.29	30.21 +/- 7.04	0.00007996 +/- 0.00001554
Tharu	Maratha	Naxi	2.60E-05	4.21	32.00 +/- 7.39	0.00009226 +/- 0.00002193
Tharu	Maratha	She	0.049	1.96	25.79 +/- 13.12	0.00007022 +/- 0.00002784
Tharu	Maratha	SHK	0.00019	3.74	23.53 +/- 6.30	0.00006567 +/- 0.00001399
Tharu	Maratha	Tu	0.0052	2.79	29.31 +/- 10.50	0.00007237 +/- 0.00002134
Tharu	Pallan	CDX	5.10E-06	4.56	35.25 +/- 7.73	0.00008014 +/- 0.00001467
Tharu	Pallan	CHB	0.0002	3.72	35.64 +/- 9.57	0.00009606 +/- 0.00002348
Tharu	Pallan	Han	1.70E-06	4.79	37.18 +/- 7.76	0.00009883 +/- 0.00001787
Tharu	Pallan	JPT	3.90E-05	4.11	35.25 +/- 8.57	0.00009101 +/- 0.00001734
Tharu	Pallan	KHV	1.40E-05	4.34	35.04 +/- 8.08	0.00007901 +/- 0.00001466
Tharu	Pallan	Lahu	7.80E-05	3.95	34.73 +/- 8.79	0.00008419 +/- 0.00001738
Tharu	Pallan	Miaozu	3.30E-05	4.15	34.71 +/- 8.36	0.00009062 +/- 0.00001748
Tharu	Pallan	Naxi	6.10E-08	5.42	35.03 +/- 6.47	0.00009823 +/- 0.00001633
Tharu	Pallan	She	0.0056	2.77	33.06 +/- 11.94	0.00009046 +/- 0.00002587
Tharu	Pallan	SHK	0.002	3.09	25.30 +/- 8.20	0.00007003 +/- 0.00001786
Tharu	Pallan	Tu	9.80E-05	3.9	36.63 +/- 9.40	0.00009083 +/- 0.00002162
Tharu	W_Bengal_Brh	CDX	5.70E-06	4.54	37.26 +/- 8.21	0.00009146 +/- 0.00001782
Tharu	W_Bengal_Brh	CHB	8.10E-05	3.94	36.33 +/- 9.22	0.00010272 +/- 0.00002344
Tharu	W_Bengal_Brh	Han	1.90E-05	4.28	35.70 +/- 8.35	0.00009778 +/- 0.00001984

Tharu	W_Bengal_Brh	JPT	9.30E-05	3.91	36.33 +/- 9.30	0.00009788 +/- 0.00002121
Tharu	W_Bengal_Brh	KHV	7.70E-05	3.95	34.08 +/- 8.62	0.00008116 +/- 0.00001747
Tharu	W_Bengal_Brh	Lahu	5.30E-06	4.55	35.90 +/- 7.89	0.00009003 +/- 0.00001933
Tharu	W_Bengal_Brh	Miaozu	0.0001	3.88	33.13 +/- 8.53	0.00009175 +/- 0.00001883
Tharu	W_Bengal_Brh	Naxi	1.00E-07	5.32	36.36 +/- 6.83	0.00010722 +/- 0.00001901
Tharu	W_Bengal_Brh	She	0.019	2.35	32.36 +/- 13.79	0.00009184 +/- 0.00003192
Tharu	W_Bengal_Brh	SHK	0.00029	3.63	27.88 +/- 7.69	0.00007901 +/- 0.00001716
Tharu	W_Bengal_Brh	Tu	4.40E-05	4.09	39.36 +/- 9.63	0.00009980 +/- 0.00002297
Tharu	BEB	CDX	8.70E-06	4.45	35.52 +/- 7.99	0.00006859 +/- 0.00001459
Tharu	BEB	CHB	4.70E-05	4.07	35.30 +/- 8.67	0.00008030 +/- 0.00001811
Tharu	BEB	Han	0.0001	3.88	33.61 +/- 8.66	0.00007392 +/- 0.00001728
Tharu	BEB	JPT	0.00027	3.64	35.31 +/- 9.70	0.00007547 +/- 0.00001792
Tharu	BEB	Lahu	8.00E-05	3.95	30.95 +/- 7.74	0.00006321 +/- 0.00001602
Tharu	BEB	Miaozu	0.00028	3.63	31.08 +/- 8.55	0.00006859 +/- 0.00001570
Tharu	BEB	Naxi	3.00E-07	5.13	36.05 +/- 6.82	0.00008637 +/- 0.00001685
Tharu	BEB	She	0.036	2.1	30.65 +/- 14.63	0.00006964 +/- 0.00002813
Tharu	BEB	SHK	0.0019	3.1	25.05 +/- 8.08	0.00005731 +/- 0.00001471
Tharu	BEB	Tu	0.0006	3.43	36.88 +/- 10.75	0.00007516 +/- 0.00002104
Tharu	Brahui	CDX	6.00E-05	4.01	30.41 +/- 7.58	0.00009782 +/- 0.00002075
Tharu	Brahui	CHB	0.0002	3.72	29.31 +/- 7.88	0.00010427 +/- 0.00002358
Tharu	Brahui	Han	0.0001	3.89	29.26 +/- 7.52	0.00010214 +/- 0.00002186
Tharu	Brahui	JPT	0.00051	3.48	30.62 +/- 8.81	0.00010372 +/- 0.00002422
Tharu	Brahui	KHV	0.00015	3.79	28.40 +/- 7.50	0.00008832 +/- 0.00001928
Tharu	Brahui	Lahu	8.50E-05	3.93	27.69 +/- 7.05	0.00009376 +/- 0.00002261
Tharu	Brahui	Miaozu	1.80E-05	4.28	26.85 +/- 6.27	0.00009650 +/- 0.00001656
Tharu	Brahui	Naxi	1.80E-06	4.78	31.79 +/- 6.66	0.00011794 +/- 0.00002222
Tharu	Brahui	She	0.0074	2.68	25.13 +/- 9.39	0.00009465 +/- 0.00002561
Tharu	Brahui	SHK	9.70E-05	3.9	26.66 +/- 6.84	0.00009495 +/- 0.00001818
Tharu	Brahui	Tu	4.20E-05	4.1	33.25 +/- 8.11	0.00010529 +/- 0.00002254
Tharu	Burusho	CDX	0.011	2.53	23.05 +/- 9.10	0.00005702 +/- 0.00001754
Tharu	Burusho	Han	0.02	2.32	22.28 +/- 9.60	0.00006161 +/- 0.00002062
Tharu	Burusho	JPT	0.021	2.31	23.39 +/- 10.12	0.00006251 +/- 0.00001959
Tharu	Burusho	KHV	0.016	2.41	21.05 +/- 8.73	0.00005146 +/- 0.00001619
Tharu	Burusho	Lahu	0.0038	2.89	22.20 +/- 7.67	0.00005791 +/- 0.00001697
Tharu	Burusho	Miaozu	0.016	2.4	19.36 +/- 8.06	0.00005613 +/- 0.00001598
Tharu	Burusho	Naxi	0.0022	3.06	24.84 +/- 8.13	0.00007059 +/- 0.00001951
Tharu	Burusho	SHK	0.032	2.14	14.35 +/- 6.71	0.00004677 +/- 0.00001450
Tharu	CDX	Drav_Kar	2.90E-05	4.18	32.06 +/- 7.66	0.00007364 +/- 0.00001453
Tharu	CDX	GIH	4.80E-05	4.07	31.76 +/- 7.81	0.00008705 +/- 0.00001882
Tharu	CDX	ITU	0.001	3.28	26.83 +/- 8.18	0.00006564 +/- 0.00001691
Tharu	CDX	Drav_AP	4.40E-05	4.08	34.23 +/- 8.38	0.00008043 +/- 0.00001633
Tharu	CDX	Gond	4.30E-07	5.05	40.52 +/- 8.02	0.00008402 +/- 0.00001581
Tharu	CDX	IE_UTT	1.20E-06	4.86	35.59 +/- 7.32	0.00008495 +/- 0.00001531
Tharu	CDX	Pulliyar	0.00017	3.77	52.50 +/- 11.56	0.00009555 +/- 0.00002537
Tharu	CDX	North_Kannadi	0.00098	3.3	56.19 +/- 13.20	0.00011181 +/- 0.00003392
Tharu	CDX	Pathan	2.30E-05	4.24	27.79 +/- 6.56	0.00008655 +/- 0.00001712
Tharu	CDX	PJL	0.01	2.57	23.28 +/- 9.07	0.00006320 +/- 0.00002008
Tharu	CDX	Sakilli	0.009	2.61	39.61 +/- 12.09	0.00011878 +/- 0.00004548
Tharu	CDX	Sindhi	4.20E-05	4.1	30.50 +/- 7.45	0.00009053 +/- 0.00001951
Tharu	CDX	STU	2.20E-06	4.74	35.02 +/- 7.39	0.00008187 +/- 0.00001509
Tharu	CHA_N_Munda	CHB	0.0094	2.6	51.40 +/- 14.43	0.00008504 +/- 0.00003273
Tharu	CHA_N_Munda	Naxi	0.0024	3.03	49.04 +/- 11.00	0.00009169 +/- 0.00003022
Tharu	CHA_N_Munda	SHK	0.022	2.29	35.71 +/- 15.62	0.00005066 +/- 0.00002186
Tharu	CHA_N_Munda	Tu	0.0047	2.83	54.02 +/- 15.13	0.00008285 +/- 0.00002931
Tharu	CHB	Drav_Kar	0.0025	3.02	31.62 +/- 10.47	0.00008467 +/- 0.00002566
Tharu	CHB	GIH	0.00027	3.64	31.87 +/- 8.75	0.00009958 +/- 0.00002348
Tharu	CHB	ITU	0.0027	3	27.63 +/- 9.22	0.00007810 +/- 0.00002196
Tharu	CHB	Drav_AP	0.00043	3.52	35.86 +/- 10.19	0.00009654 +/- 0.00002477
Tharu	CHB	Gond	0.00011	3.87	40.70 +/- 10.34	0.00009811 +/- 0.00002534

Tharu	CHB	IE_UTT	3.20E-05	4.16	35.64 +/- 8.57	0.00009811 +/- 0.00002119
Tharu	CHB	Pulliyar	0.0011	3.26	45.26 +/- 13.09	0.00009579 +/- 0.00002935
Tharu	CHB	North_Kannadi	0.0025	3.02	48.84 +/- 13.58	0.00010761 +/- 0.00003560
Tharu	CHB	Pathan	0.00028	3.63	28.06 +/- 7.73	0.00009689 +/- 0.00002195
Tharu	CHB	PJL	0.018	2.38	23.33 +/- 9.82	0.00007199 +/- 0.00002405
Tharu	CHB	Sakilli	0.011	2.54	38.07 +/- 13.04	0.00012446 +/- 0.00004893
Tharu	CHB	Sindhi	0.0033	2.94	27.18 +/- 9.24	0.00009136 +/- 0.00002624
Tharu	CHB	STU	5.60E-05	4.03	35.27 +/- 8.76	0.00009543 +/- 0.00002150
Tharu	Drav_Kar	Han	9.50E-05	3.9	32.45 +/- 8.31	0.00008706 +/- 0.00001826
Tharu	Drav_Kar	JPT	0.00025	3.66	32.86 +/- 8.97	0.00008364 +/- 0.00001832
Tharu	Drav_Kar	KHV	2.00E-05	4.26	32.29 +/- 7.57	0.00007405 +/- 0.00001433
Tharu	Drav_Kar	Lahu	4.80E-05	4.07	30.75 +/- 7.56	0.00007762 +/- 0.00001683
Tharu	Drav_Kar	Miaozu	1.70E-06	4.79	33.18 +/- 6.93	0.00008560 +/- 0.00001419
Tharu	Drav_Kar	Naxi	1.10E-05	4.39	33.47 +/- 7.62	0.00009633 +/- 0.00001888
Tharu	Drav_Kar	She	0.016	2.42	28.81 +/- 11.92	0.00007978 +/- 0.00002459
Tharu	Drav_Kar	SHK	0.0022	3.06	24.68 +/- 8.05	0.00006752 +/- 0.00001916
Tharu	Drav_Kar	Tu	0.0062	2.74	31.54 +/- 11.51	0.00007785 +/- 0.00002370
Tharu	GIH	Han	0.00018	3.74	30.71 +/- 8.20	0.00009322 +/- 0.00002162
Tharu	GIH	JPT	0.00082	3.35	31.92 +/- 9.54	0.00009475 +/- 0.00002313
Tharu	GIH	KHV	0.00034	3.58	29.35 +/- 8.19	0.00007810 +/- 0.00001878
Tharu	GIH	Lahu	0.0002	3.71	27.84 +/- 7.50	0.00007961 +/- 0.00001985
Tharu	GIH	Miaozu	0.00034	3.58	27.88 +/- 7.78	0.00008572 +/- 0.00001951
Tharu	GIH	Naxi	2.00E-06	4.75	33.20 +/- 6.98	0.00010698 +/- 0.00001980
Tharu	GIH	SHK	0.0027	3	23.46 +/- 7.81	0.00007544 +/- 0.00001911
Tharu	GIH	Tu	0.0018	3.12	32.69 +/- 10.48	0.00009278 +/- 0.00002687
Tharu	Han	ITU	0.0024	3.04	26.19 +/- 8.61	0.00007222 +/- 0.00002044
Tharu	Han	Drav_AP	0.00013	3.84	35.23 +/- 9.18	0.00009291 +/- 0.00002089
Tharu	Han	Gond	1.60E-07	5.24	42.08 +/- 8.03	0.00010112 +/- 0.00001664
Tharu	Han	IE_UTT	8.00E-06	4.47	34.28 +/- 7.68	0.00009179 +/- 0.00001791
Tharu	Han	Pulliyar	7.80E-05	3.95	47.44 +/- 12.00	0.00010073 +/- 0.00002502
Tharu	Han	North_Kannadi	0.00044	3.51	55.68 +/- 13.03	0.00012759 +/- 0.00003633
Tharu	Han	Pathan	5.10E-05	4.05	26.65 +/- 6.58	0.00009016 +/- 0.00001793
Tharu	Han	PJL	0.012	2.53	22.46 +/- 8.89	0.00006793 +/- 0.00002198
Tharu	Han	Sakilli	0.00083	3.34	39.39 +/- 9.87	0.00012881 +/- 0.00003852
Tharu	Han	Sindhi	5.50E-06	4.54	31.20 +/- 6.87	0.00010314 +/- 0.00001916
Tharu	Han	STU	8.10E-06	4.46	34.65 +/- 7.77	0.00009132 +/- 0.00001818
Tharu	ITU	JPT	0.0037	2.9	28.27 +/- 9.73	0.00007535 +/- 0.00002080
Tharu	ITU	KHV	0.0025	3.02	24.94 +/- 8.26	0.00005990 +/- 0.00001673
Tharu	ITU	Lahu	0.0015	3.17	24.12 +/- 7.62	0.00006264 +/- 0.00001747
Tharu	ITU	Miaozu	0.00089	3.32	24.70 +/- 7.44	0.00006811 +/- 0.00001687
Tharu	ITU	Naxi	0.0002	3.72	28.95 +/- 7.78	0.00008552 +/- 0.00001981
Tharu	ITU	SHK	0.0046	2.84	19.89 +/- 7.01	0.00005937 +/- 0.00001574
Tharu	ITU	Tu	0.011	2.54	26.52 +/- 10.44	0.00006909 +/- 0.00002309
Tharu	JPT	Drav_AP	0.0015	3.18	36.59 +/- 11.51	0.00009345 +/- 0.00002475
Tharu	JPT	Gond	8.00E-06	4.46	42.58 +/- 9.54	0.00009713 +/- 0.00001843
Tharu	JPT	IE_UTT	5.60E-05	4.03	35.06 +/- 8.71	0.00009125 +/- 0.00001863
Tharu	JPT	Pulliyar	0.00015	3.79	50.22 +/- 13.24	0.00009799 +/- 0.00002586
Tharu	JPT	North_Kannadi	0.0004	3.54	52.07 +/- 13.24	0.00011515 +/- 0.00003252
Tharu	JPT	Pathan	0.00035	3.57	28.70 +/- 8.03	0.00009352 +/- 0.00002063
Tharu	JPT	PJL	0.025	2.24	24.00 +/- 10.72	0.00007021 +/- 0.00002395
Tharu	JPT	Sakilli	0.0017	3.14	37.31 +/- 10.33	0.00011834 +/- 0.00003767
Tharu	JPT	Sindhi	0.00013	3.83	31.63 +/- 8.27	0.00010191 +/- 0.00002147
Tharu	JPT	STU	9.60E-05	3.9	34.80 +/- 8.92	0.00008946 +/- 0.00001891
Tharu	KHV	Drav_AP	0.00019	3.73	32.51 +/- 8.71	0.00007358 +/- 0.00001635
Tharu	KHV	IE_UTT	2.00E-05	4.27	33.34 +/- 7.81	0.00007722 +/- 0.00001553
Tharu	KHV	Pulliyar	7.70E-05	3.95	50.54 +/- 12.79	0.00008847 +/- 0.00002193
Tharu	KHV	North_Kannadi	0.00038	3.55	62.76 +/- 13.39	0.00013029 +/- 0.00003670
Tharu	KHV	Pathan	5.40E-05	4.04	25.86 +/- 6.41	0.00007748 +/- 0.00001541
Tharu	KHV	PJL	0.012	2.52	21.63 +/- 8.57	0.00005757 +/- 0.00001835

Tharu	KHV	Sakilli	0.0013	3.21	39.13 +/- 9.88	0.00011559 +/- 0.00003595
Tharu	KHV	Sindhi	2.10E-05	4.25	30.00 +/- 7.06	0.00008764 +/- 0.00001705
Tharu	KHV	STU	2.10E-05	4.26	33.02 +/- 7.75	0.00007451 +/- 0.00001532
Tharu	Lahu	Drav_AP	0.00012	3.85	32.13 +/- 8.33	0.00007794 +/- 0.00001770
Tharu	Lahu	Gond	0.0019	3.1	36.17 +/- 11.65	0.00007397 +/- 0.00002059
Tharu	Lahu	IE_UTT	2.90E-05	4.18	31.92 +/- 7.64	0.00007997 +/- 0.00001764
Tharu	Lahu	Pulliyar	0.0055	2.78	55.74 +/- 20.08	0.00011035 +/- 0.00003728
Tharu	Lahu	North_Kannadi	0.0091	2.61	74.31 +/- 19.24	0.00017977 +/- 0.00006891
Tharu	Lahu	Pathan	5.80E-05	4.02	26.05 +/- 6.41	0.00008682 +/- 0.00002159
Tharu	Lahu	PJL	0.0036	2.91	21.92 +/- 7.53	0.00006261 +/- 0.00001905
Tharu	Lahu	Sakilli	0.0076	2.67	46.56 +/- 10.14	0.00015815 +/- 0.00005928
Tharu	Lahu	Sindhi	2.70E-06	4.69	30.16 +/- 6.42	0.00009771 +/- 0.00002007
Tharu	Lahu	STU	3.30E-05	4.15	31.22 +/- 7.52	0.00007764 +/- 0.00001782
Tharu	Drav_AP	Miaozu	0.0008	3.35	34.02 +/- 10.14	0.00008797 +/- 0.00002099
Tharu	Drav_AP	Naxi	8.50E-06	4.45	35.38 +/- 7.95	0.00010240 +/- 0.00002019
Tharu	Drav_AP	She	0.032	2.14	28.85 +/- 13.46	0.00007830 +/- 0.00002816
Tharu	Drav_AP	SHK	0.0011	3.28	27.95 +/- 8.53	0.00007659 +/- 0.00001855
Tharu	Drav_AP	Tu	0.0065	2.72	37.63 +/- 13.82	0.00009390 +/- 0.00003043
Tharu	IE_UTT	Miaozu	3.10E-05	4.16	32.66 +/- 7.85	0.00008769 +/- 0.00001687
Tharu	IE_UTT	Naxi	2.40E-08	5.58	35.92 +/- 6.44	0.00010356 +/- 0.00001718
Tharu	IE_UTT	She	0.031	2.16	29.86 +/- 13.85	0.00008301 +/- 0.00003058
Tharu	IE_UTT	SHK	0.00099	3.29	26.42 +/- 8.02	0.00007404 +/- 0.00001743
Tharu	IE_UTT	Tu	0.00011	3.87	36.41 +/- 9.41	0.00009052 +/- 0.00002129
Tharu	Pulliyar	Miaozu	9.00E-06	4.44	52.65 +/- 11.86	0.00010337 +/- 0.00002125
Tharu	Pulliyar	Naxi	2.50E-06	4.7	50.04 +/- 10.64	0.00011097 +/- 0.00002310
Tharu	Pulliyar	She	0.0029	2.98	44.20 +/- 14.83	0.00009630 +/- 0.00002923
Tharu	Pulliyar	SHK	0.013	2.49	40.08 +/- 15.84	0.00008010 +/- 0.00003212
Tharu	Pulliyar	Tu	0.00075	3.37	48.70 +/- 13.67	0.00009358 +/- 0.00002775
Tharu	Miaozu	North_Kannadi	0.00036	3.57	57.80 +/- 13.98	0.00013703 +/- 0.00003842
Tharu	Miaozu	Pathan	5.40E-06	4.55	24.82 +/- 5.46	0.00008619 +/- 0.00001317
Tharu	Miaozu	PJL	0.0039	2.88	21.08 +/- 7.31	0.00006476 +/- 0.00001762
Tharu	Miaozu	Sakilli	4.00E-05	4.11	38.92 +/- 7.05	0.00012810 +/- 0.00003119
Tharu	Miaozu	Sindhi	1.00E-05	4.41	28.08 +/- 6.36	0.00009494 +/- 0.00001670
Tharu	Miaozu	STU	0.00013	3.83	32.50 +/- 8.48	0.00008505 +/- 0.00001751
Tharu	Naxi	North_Kannadi	0.0011	3.27	55.89 +/- 13.41	0.00014113 +/- 0.00004315
Tharu	Naxi	Pathan	7.40E-06	4.48	28.85 +/- 6.44	0.00010427 +/- 0.00002093
Tharu	Naxi	PJL	0.002	3.09	26.02 +/- 8.42	0.00008151 +/- 0.00002269
Tharu	Naxi	Sakilli	0.0031	2.96	41.30 +/- 10.35	0.00015085 +/- 0.00005093
Tharu	Naxi	Sindhi	7.70E-07	4.94	32.74 +/- 6.62	0.00011430 +/- 0.00002154
Tharu	Naxi	STU	1.20E-07	5.3	35.37 +/- 6.68	0.00010186 +/- 0.00001795
Tharu	North_Kannadi	She	0.00065	3.41	54.05 +/- 15.01	0.00012700 +/- 0.00003725
Tharu	North_Kannadi	Tu	0.0087	2.62	59.17 +/- 18.06	0.00011890 +/- 0.00004535
Tharu	Pathan	She	0.021	2.31	22.11 +/- 9.57	0.00008283 +/- 0.00002472
Tharu	Pathan	SHK	5.00E-05	4.06	21.76 +/- 5.36	0.00007825 +/- 0.00001315
Tharu	Pathan	Tu	0.0001	3.88	29.44 +/- 7.58	0.00009083 +/- 0.00001880
Tharu	PJL	SHK	0.017	2.38	16.77 +/- 7.03	0.00005525 +/- 0.00001674
Tharu	PJL	Tu	0.027	2.22	23.97 +/- 10.81	0.00006608 +/- 0.00002456
Tharu	Sakilli	She	0.00094	3.31	35.76 +/- 10.56	0.00012130 +/- 0.00003667
Tharu	Sakilli	SHK	0.0044	2.85	30.73 +/- 10.80	0.00009690 +/- 0.00003197
Tharu	Sakilli	Tu	0.00081	3.35	39.19 +/- 10.45	0.00011756 +/- 0.00003512
Tharu	She	Sindhi	0.025	2.23	26.49 +/- 11.86	0.00009400 +/- 0.00003111
Tharu	She	STU	0.026	2.23	29.74 +/- 13.36	0.00008108 +/- 0.00002921
Tharu	Sindhi	SHK	0.013	2.48	21.90 +/- 8.83	0.00007554 +/- 0.00002306
Tharu	Sindhi	Tu	0.00016	3.77	33.77 +/- 8.96	0.00010080 +/- 0.00002276
Tharu	STU	SHK	0.0007	3.39	26.91 +/- 7.94	0.00007410 +/- 0.00001758
Tharu	STU	Tu	6.80E-05	3.98	36.56 +/- 9.18	0.00009021 +/- 0.00002073
TAT	Gond	Naga	0.00097	3.3	8.61 +/- 2.61	0.00002486 +/- 0.00000579
TAT	Gond	SHK	0.0022	3.06	4.37 +/- 1.43	0.00001576 +/- 0.00000307
TAT	Gond	TIB	0.047	1.99	10.34 +/- 5.20	0.00002424 +/- 0.00000732

TAT	Gujrati_Brh	CHB	0.0057	2.76	12.68 +/- 4.59	0.00004406 +/- 0.00000741
TAT	Gujrati_Brh	CHS	0.0052	2.8	12.77 +/- 4.57	0.00004211 +/- 0.00000698
TAT	Gujrati_Brh	Han	0.0084	2.63	14.07 +/- 5.34	0.00004636 +/- 0.00000824
TAT	Gujrati_Brh	Japanese	0.0061	2.74	17.08 +/- 6.23	0.00005170 +/- 0.00001001
TAT	Gujrati_Brh	JPT	0.027	2.22	17.40 +/- 7.85	0.00004800 +/- 0.00001041
TAT	Gujrati_Brh	Naga	3.70E-05	4.13	11.82 +/- 2.86	0.00004685 +/- 0.00000871
TAT	Gujrati_Brh	Miaozu	0.0053	2.79	13.27 +/- 4.76	0.00004888 +/- 0.00000893
TAT	Gujrati_Brh	Naxi	2.60E-05	4.2	13.36 +/- 3.18	0.00004865 +/- 0.00000873
TAT	Gujrati_Brh	She	0.047	1.99	11.68 +/- 5.87	0.00004122 +/- 0.00001171
TAT	Gujrati_Brh	SHK	5.00E-06	4.57	7.77 +/- 1.70	0.00003464 +/- 0.00000569
TAT	Gujrati_Brh	TIB	0.00045	3.51	15.22 +/- 4.34	0.00005335 +/- 0.00000892
TAT	Gujrati_Brh	Yizu	0.00019	3.73	13.56 +/- 3.64	0.00004960 +/- 0.00000922
TAT	Kadar	Naga	0.015	2.44	11.34 +/- 4.65	0.00003451 +/- 0.00000909
TAT	Kadar	Naxi	0.034	2.12	9.44 +/- 4.44	0.00003079 +/- 0.00001025
TAT	Kadar	SHK	0.0036	2.91	4.80 +/- 1.65	0.00002195 +/- 0.00000452
TAT	Kadar	TIB	0.046	2	10.84 +/- 5.42	0.00003242 +/- 0.00000981
TAT	Kadar	Yizu	0.049	1.97	7.08 +/- 3.60	0.00002613 +/- 0.00000998
TAT	W_Bengal_Brh	CHB	0.0034	2.93	11.19 +/- 3.82	0.00003706 +/- 0.00000793
TAT	W_Bengal_Brh	CHS	0.0016	3.15	11.01 +/- 3.50	0.00003496 +/- 0.00000672
TAT	W_Bengal_Brh	Han	0.0077	2.66	12.19 +/- 4.58	0.00003761 +/- 0.00000858
TAT	W_Bengal_Brh	Naga	1.50E-05	4.33	11.79 +/- 2.72	0.00004193 +/- 0.00000905
TAT	W_Bengal_Brh	Miaozu	0.0051	2.8	12.45 +/- 4.45	0.00004479 +/- 0.00000971
TAT	W_Bengal_Brh	Naxi	8.20E-06	4.46	12.09 +/- 2.71	0.00004194 +/- 0.00000875
TAT	W_Bengal_Brh	She	0.031	2.16	9.86 +/- 4.57	0.00003402 +/- 0.00001036
TAT	W_Bengal_Brh	SHK	6.50E-06	4.51	7.98 +/- 1.77	0.00003076 +/- 0.00000581
TAT	W_Bengal_Brh	TIB	0.00072	3.38	14.26 +/- 4.22	0.00004596 +/- 0.00000918
TAT	W_Bengal_Brh	Yizu	0.0028	2.99	13.79 +/- 4.61	0.00004624 +/- 0.00001115
TAT	Burusho	CHB	0.00016	3.77	9.14 +/- 2.42	0.00003335 +/- 0.00000532
TAT	Burusho	CHS	0.00074	3.38	9.50 +/- 2.82	0.00003274 +/- 0.00000564
TAT	Burusho	Naga	5.60E-07	5	9.56 +/- 1.91	0.00003763 +/- 0.00000594
TAT	Burusho	Miaozu	0.0029	2.98	9.71 +/- 3.26	0.00003738 +/- 0.00000694
TAT	Burusho	Naxi	1.70E-05	4.3	10.80 +/- 2.51	0.00003901 +/- 0.00000723
TAT	Burusho	She	0.039	2.06	8.53 +/- 4.14	0.00003171 +/- 0.00000802
TAT	Burusho	SHK	5.70E-05	4.03	7.55 +/- 1.88	0.00003199 +/- 0.00000484
TAT	Burusho	TIB	0.00039	3.55	12.43 +/- 3.50	0.00004274 +/- 0.00000778
TAT	Burusho	Yizu	9.20E-05	3.91	9.99 +/- 2.55	0.00003640 +/- 0.00000714
TAT	CHB	Drav_Kar	0.028	2.2	8.22 +/- 3.73	0.00002973 +/- 0.00000904
TAT	CHB	Drav_TN	0.0065	2.72	8.96 +/- 3.29	0.00003327 +/- 0.00000697
TAT	CHB	PJL	0.0059	2.75	10.89 +/- 3.95	0.00003689 +/- 0.00000703
TAT	CHB	Sakilli	0.024	2.26	7.12 +/- 3.15	0.00002298 +/- 0.00000601
TAT	CHS	Drav_TN	0.0061	2.74	8.76 +/- 3.19	0.00003121 +/- 0.00000696
TAT	CHS	PJL	0.0089	2.62	10.80 +/- 4.13	0.00003458 +/- 0.00000708
TAT	CHS	Sakilli	0.016	2.41	6.22 +/- 2.58	0.00001941 +/- 0.00000463
TAT	Drav_Kar	Naga	0.00093	3.31	9.11 +/- 2.75	0.00003522 +/- 0.00000830
TAT	Drav_Kar	Naxi	0.036	2.1	12.89 +/- 6.14	0.00004023 +/- 0.00001243
TAT	Drav_Kar	SHK	0.011	2.54	4.67 +/- 1.84	0.00002246 +/- 0.00000610
TAT	Drav_Kar	TIB	0.0097	2.59	13.82 +/- 5.34	0.00004270 +/- 0.00001141
TAT	Drav_Kar	Yizu	0.042	2.03	10.31 +/- 5.07	0.00003676 +/- 0.00001349
TAT	Han	Drav_TN	0.0041	2.87	9.33 +/- 3.25	0.00003308 +/- 0.00000694
TAT	Han	PJL	0.016	2.42	11.93 +/- 4.93	0.00003781 +/- 0.00000817
TAT	Han	Sakilli	0.025	2.24	7.68 +/- 3.43	0.00002313 +/- 0.00000614
TAT	Japanese	PJL	0.0087	2.62	15.86 +/- 6.05	0.00004535 +/- 0.00001001
TAT	Drav_TN	Naga	0.00014	3.81	9.74 +/- 2.55	0.00003860 +/- 0.00000880
TAT	Drav_TN	Miaozu	0.02	2.33	10.22 +/- 4.39	0.00004004 +/- 0.00000972
TAT	Drav_TN	Naxi	0.00048	3.49	9.81 +/- 2.81	0.00003658 +/- 0.00000752
TAT	Drav_TN	SHK	0.0003	3.62	6.02 +/- 1.66	0.00002944 +/- 0.00000564
TAT	Drav_TN	TIB	0.00081	3.35	10.46 +/- 3.12	0.00003889 +/- 0.00000731
TAT	Drav_TN	Yizu	0.0018	3.13	9.25 +/- 2.96	0.00003605 +/- 0.00000926
TAT	Gond	Naxi	0.0012	3.24	12.31 +/- 3.80	0.00003315 +/- 0.00000774

TAT	Gond	Yizu	0.00012	3.84	15.71 +/- 3.92	0.00003932 +/- 0.00001024
TAT	Naga	PJL	0.00012	3.86	10.49 +/- 2.72	0.00004064 +/- 0.00000785
TAT	Naga	Sakilli	0.0012	3.23	9.67 +/- 2.99	0.00002830 +/- 0.00000625
TAT	Miaozu	PJL	0.028	2.19	13.29 +/- 6.06	0.00004659 +/- 0.00001131
TAT	Naxi	PJL	3.80E-05	4.12	12.32 +/- 2.99	0.00004226 +/- 0.00000792
TAT	Naxi	Sakilli	0.00056	3.45	10.25 +/- 2.97	0.00003262 +/- 0.00000657
TAT	PJL	SHK	4.30E-05	4.09	7.10 +/- 1.74	0.00003021 +/- 0.00000534
TAT	PJL	TIB	0.00065	3.41	14.06 +/- 4.12	0.00004716 +/- 0.00000837
TAT	PJL	Yizu	0.0018	3.11	12.97 +/- 4.17	0.00004466 +/- 0.00001026
TAT	Sakilli	SHK	0.002	3.09	5.21 +/- 1.69	0.00002203 +/- 0.00000413
TAT	Sakilli	TIB	0.0044	2.85	10.83 +/- 3.81	0.00003121 +/- 0.00000665
TAT	Sakilli	Yizu	0.02	2.33	8.94 +/- 3.84	0.00002965 +/- 0.00000858
TAS	Balochi	Naga	3.80E-07	5.08	23.34 +/- 4.60	0.00014947 +/- 0.00001866
TAS	Balochi	SRH	0.00079	3.36	16.89 +/- 5.03	0.00010725 +/- 0.00001857
TAS	Balochi	TIB	6.80E-05	3.98	16.63 +/- 4.17	0.00010142 +/- 0.00001502
TAS	Gond	Naga	1.80E-13	7.36	24.34 +/- 3.31	0.00008940 +/- 0.00001172
TAS	Gond	TIB	1.30E-11	6.76	19.33 +/- 2.86	0.00006889 +/- 0.00000860
TAS	Gujrati_Brh	Naga	8.00E-08	5.37	23.05 +/- 4.29	0.00013587 +/- 0.00001811
TAS	Gujrati_Brh	SRH	4.10E-05	4.1	17.06 +/- 4.16	0.00010146 +/- 0.00001475
TAS	Gujrati_Brh	TIB	3.10E-06	4.67	16.21 +/- 3.47	0.00009436 +/- 0.00001240
TAS	Ho	Naga	8.70E-09	5.75	25.91 +/- 4.50	0.00006530 +/- 0.00001022
TAS	Irula	Naga	1.50E-05	4.33	24.10 +/- 5.56	0.00009377 +/- 0.00001558
TAS	Irula	TIB	9.60E-05	3.9	16.86 +/- 4.32	0.00006816 +/- 0.00001228
TAS	Iyer	Naga	1.10E-06	4.88	22.22 +/- 4.56	0.00012308 +/- 0.00002148
TAS	Iyer	SRH	3.00E-05	4.17	16.92 +/- 4.05	0.00009317 +/- 0.00001301
TAS	Iyer	TIB	2.40E-06	4.71	15.76 +/- 3.34	0.00008608 +/- 0.00001148
TAS	Kadar	Naga	3.90E-07	5.07	23.23 +/- 4.58	0.00008873 +/- 0.00001500
TAS	Kadar	TIB	1.60E-08	5.65	15.05 +/- 2.66	0.00006200 +/- 0.00000801
TAS	Khatri	Naga	3.80E-07	5.08	24.37 +/- 4.80	0.00014791 +/- 0.00001979
TAS	Khatri	SRH	5.30E-06	4.55	19.15 +/- 4.20	0.00011188 +/- 0.00001400
TAS	Khatri	TIB	1.00E-07	5.33	18.98 +/- 3.56	0.00010743 +/- 0.00001236
TAS	Maratha	Naga	5.50E-09	5.83	23.56 +/- 4.04	0.00011249 +/- 0.00001547
TAS	Maratha	SRH	0.00016	3.77	17.01 +/- 4.51	0.00008495 +/- 0.00001273
TAS	Maratha	TIB	1.20E-06	4.86	16.70 +/- 3.44	0.00008078 +/- 0.00001037
TAS	Pallan	Naga	1.20E-10	6.43	24.26 +/- 3.77	0.00011600 +/- 0.00001462
TAS	Pallan	SRH	7.00E-05	3.98	17.73 +/- 4.46	0.00008978 +/- 0.00001430
TAS	Pallan	TIB	7.40E-07	4.95	16.95 +/- 3.42	0.00008210 +/- 0.00001122
TAS	Santal	Naga	7.90E-06	4.47	24.46 +/- 5.47	0.00006478 +/- 0.00001291
TAS	W_Bengal_Brh	Naga	2.60E-09	5.95	25.05 +/- 4.21	0.00012972 +/- 0.00001515
TAS	W_Bengal_Brh	SRH	3.90E-07	5.07	19.46 +/- 3.84	0.00010096 +/- 0.00001207
TAS	W_Bengal_Brh	TIB	2.20E-09	5.98	18.94 +/- 3.16	0.00009401 +/- 0.00000954
TAS	BEB	Naga	2.90E-06	4.68	22.10 +/- 4.72	0.00009146 +/- 0.00001611
TAS	BEB	TIB	1.10E-06	4.87	16.94 +/- 3.48	0.00007098 +/- 0.00000919
TAS	Brahui	Naga	1.20E-06	4.85	23.47 +/- 4.84	0.00014570 +/- 0.00002140
TAS	Brahui	SRH	0.00093	3.31	17.61 +/- 5.32	0.00010965 +/- 0.00001963
TAS	Brahui	TIB	0.00017	3.76	16.46 +/- 4.37	0.00010044 +/- 0.00001660
TAS	Burusho	Naga	0.00013	3.83	22.16 +/- 5.78	0.00010568 +/- 0.00002035
TAS	Burusho	TIB	0.00029	3.62	15.09 +/- 4.17	0.00007129 +/- 0.00001182
TAS	N_Munda	Naga	1.70E-10	6.39	26.16 +/- 4.08	0.00007159 +/- 0.00001121
TAS	Drav_Kar	Naga	1.80E-11	6.72	25.30 +/- 3.77	0.00012110 +/- 0.00001398
TAS	Drav_Kar	SRH	5.30E-05	4.04	19.91 +/- 4.92	0.00009576 +/- 0.00001565
TAS	Drav_Kar	TIB	4.50E-07	5.05	18.95 +/- 3.75	0.00008869 +/- 0.00001249
TAS	Druze	Naga	8.20E-09	5.76	25.62 +/- 4.44	0.00016312 +/- 0.00002230
TAS	Druze	SRH	4.30E-05	4.09	19.69 +/- 4.81	0.00012330 +/- 0.00002112
TAS	Druze	TIB	5.70E-06	4.54	19.34 +/- 4.26	0.00011593 +/- 0.00001810
TAS	GIH	Naga	1.40E-06	4.82	21.44 +/- 4.45	0.00012059 +/- 0.00001933
TAS	GIH	SRH	6.90E-05	3.98	17.37 +/- 4.36	0.00009922 +/- 0.00001527
TAS	GIH	TIB	9.90E-06	4.42	16.48 +/- 3.73	0.00009111 +/- 0.00001298
TAS	ITU	Naga	1.10E-06	4.88	21.22 +/- 4.35	0.00010750 +/- 0.00001820

TAS	ITU	SRH	3.80E-05	4.12	17.20 +/- 4.17	0.00009116 +/- 0.00001323
TAS	ITU	TIB	8.60E-07	4.92	16.50 +/- 3.35	0.00008497 +/- 0.00001105
TAS	Kalash	Naga	3.10E-09	5.93	25.69 +/- 4.33	0.00016203 +/- 0.00001745
TAS	Kalash	SRH	3.00E-05	4.18	18.32 +/- 4.39	0.00011097 +/- 0.00001530
TAS	Kalash	TIB	2.30E-07	5.17	17.89 +/- 3.46	0.00010650 +/- 0.00001290
TAS	Makrani	Naga	2.30E-07	5.18	24.31 +/- 4.70	0.00015359 +/- 0.00002123
TAS	Makrani	SRH	0.00025	3.66	18.41 +/- 5.02	0.00011510 +/- 0.00002047
TAS	Makrani	TIB	3.00E-05	4.18	18.36 +/- 4.40	0.00011078 +/- 0.00001781
TAS	Drav_AP	Naga	1.30E-08	5.68	23.73 +/- 4.17	0.00012067 +/- 0.00001582
TAS	Drav_AP	SRH	6.30E-05	4	18.31 +/- 4.58	0.00009175 +/- 0.00001561
TAS	Drav_AP	TIB	3.40E-06	4.64	17.01 +/- 3.66	0.00008330 +/- 0.00001283
TAS	Drav_TN	Naga	1.10E-08	5.71	23.76 +/- 4.16	0.00011883 +/- 0.00001551
TAS	Drav_TN	SRH	6.50E-05	3.99	18.17 +/- 4.55	0.00009564 +/- 0.00001589
TAS	Drav_TN	TIB	3.20E-05	4.16	16.49 +/- 3.97	0.00008357 +/- 0.00001376
TAS	Gond	Naga	4.50E-08	5.47	23.82 +/- 4.35	0.00009957 +/- 0.00001564
TAS	Gond	TIB	2.30E-05	4.24	18.88 +/- 4.46	0.00008001 +/- 0.00001527
TAS	IE_UTT	Naga	2.20E-07	5.18	23.38 +/- 4.51	0.00011550 +/- 0.00001804
TAS	IE_UTT	SRH	1.90E-05	4.27	18.96 +/- 4.44	0.00009580 +/- 0.00001459
TAS	IE_UTT	TIB	2.00E-07	5.2	18.26 +/- 3.51	0.00008941 +/- 0.00001171
TAS	Naga	Pulliyar	4.80E-07	5.04	18.95 +/- 3.76	0.00008417 +/- 0.00001153
TAS	Naga	North_Kannadi	0.00067	3.4	24.21 +/- 7.12	0.00011121 +/- 0.00002272
TAS	Naga	Pathan	6.80E-06	4.5	23.25 +/- 5.17	0.00013954 +/- 0.00002105
TAS	Naga	PJL	3.00E-06	4.67	21.40 +/- 4.58	0.00011129 +/- 0.00001866
TAS	Naga	Sakilli	2.50E-06	4.7	21.08 +/- 4.48	0.00010182 +/- 0.00001751
TAS	Naga	Sindhi	8.00E-08	5.37	22.51 +/- 4.19	0.00013647 +/- 0.00001920
TAS	Naga	STU	4.80E-06	4.57	22.87 +/- 5.00	0.00011138 +/- 0.00002130
TAS	Pulliyar	TIB	0.0011	3.27	12.96 +/- 3.96	0.00005858 +/- 0.00000948
TAS	North_Kannadi	SRH	0.025	2.24	15.91 +/- 7.12	0.00007548 +/- 0.00001718
TAS	North_Kannadi	TIB	0.0008	3.35	16.35 +/- 4.88	0.00007759 +/- 0.00001355
TAS	Pathan	SRH	0.00049	3.49	17.31 +/- 4.96	0.00010437 +/- 0.00001635
TAS	Pathan	TIB	6.50E-05	3.99	17.02 +/- 4.26	0.00009874 +/- 0.00001422
TAS	PJL	SRH	2.00E-05	4.26	17.99 +/- 4.22	0.00009827 +/- 0.00001429
TAS	PJL	TIB	4.00E-07	5.07	16.92 +/- 3.34	0.00008940 +/- 0.00001127
TAS	Sakilli	SRH	1.60E-05	4.31	16.97 +/- 3.94	0.00007957 +/- 0.00001050
TAS	Sakilli	TIB	7.10E-07	4.96	15.61 +/- 3.15	0.00007530 +/- 0.00000911
TAS	SRH	Sindhi	2.90E-05	4.18	17.35 +/- 4.15	0.00010438 +/- 0.00001439
TAS	SRH	STU	3.10E-05	4.17	18.56 +/- 4.45	0.00009303 +/- 0.00001490
TAS	Sindhi	TIB	4.10E-07	5.06	16.76 +/- 3.31	0.00009859 +/- 0.00001201
TAS	STU	TIB	1.40E-06	4.83	17.77 +/- 3.68	0.00008594 +/- 0.00001267

Appendix Table 4. Top 1% windows presenting highest values of outlier SNPs over the total number of SNPs present in the window (Ntop/Ntot) according to F_{cs} statistics calculated for SRH group present in the "selection SNP-chip" dataset.

Chr	Window name	Window start (bp)	Window end (bp)	Ntop/Ntot	Genes
1	window_2	9238384	9418384	0.32090669	H6PD SPSB1
1	window_3	10478384	10668384	0.3532075	PGD APITD1 APITD1-CORT CORT DFFA PEX14
1	window_4	13828384	13938384	0.20416667	LRRC38 PDPN
1	window_6	18448384	18648384	0.40929339	C1orf21 EDEM3 FAM129A
1	window_7	20448384	20548384	0.27777778	PLA2G2F PLA2G2C UBXN10
1	window_8	20478384	20588384	0.21780303	PLA2G2C UBXN10
1	window_9	20878384	21058384	0.24901513	FAM43B CDA PINK1 DDOST KIF17 SH2D5
1	window_10	24628384	24838384	0.41811279	GRHL3 STPG1 NIPAL3 RCAN3
1	window_11	36548384	36658384	0.30769231	TEKT2 ADPRHL2 COL8A2 TRAPPC3 MAP7D1
1	window_12	41388384	41498384	0.375	CTPS1 SLFNL1 SCMH1
1	window_13	47628384	47748384	0.33669109	PDZK1IP1 TAL1 STIL
1	window_15	54468384	54708384	0.63217756	LDLRAD1 TMEM59 TCEANC2 AL161915.1 CDCP2 RP11-446E24.4 AL357673.1 CYB5RL MRPL37 SSBP3 SSBP3-AS1
1	window_18	118438384	118578384	0.47272727	GDAP2 WDR3 SPAG17
1	window_19	152788384	152938384	0.34715772	LCE1A LCE6A SMCP IVL
1	window_20	184248384	184418384	0.39483173	IGSF21 C1orf21
1	window_21	184718384	184858384	0.52692308	EDEM3 FAM129A
1	window_22	208148384	208328384	0.32068285	PLXNA2
1	window_23	211588384	211798384	0.60829069	RD3 SLC30A1
1	window_24	220588384	220698384	0.33333333	AC096644.1
1	window_25	220778384	220958384	0.33785567	MARK1 C1orf115 MARC2
1	window_26	226148384	226338384	0.44506327	SDE2 H3F3A ACBD3
1	window_27	226768384	226958384	0.40887178	C1orf95 ITPKB
1	window_28	228788384	228998384	0.65941827	RHOU
1	window_29	231238384	231448384	0.49205257	TRIM67 C1orf131 GNPAT
1	**window_30**	**231428384**	**231618384**	**1**	**EXOC8 SPRTN EGLN1**
1	window_31	231668384	231838384	0.43445097	TSNAX DISC1
1	window_32	231778384	231968384	0.27729449	TSNAX-DISC1
2	window_1	3753588	3923588	0.23712713	DCDC2C
2	window_6	20533588	20683588	0.43157994	PUM2 RHOB PARD3B
2	window_7	31243588	31393588	0.38303571	GALNT14
2	window_8	42103588	42363588	0.52406381	C2orf91 PKDCC
2	**window_9**	**46483588**	**46803588**	**0.73649879**	**EPAS1 TMEM247 ATP6V1E2 RHOQ**
2	window_10	46783588	46883588	0.41666667	RHOQ PIGF CRIPT
2	window_11	46823588	46923588	0.2173913	PIGF CRIPT
2	window_12	46933588	47063588	0.33658009	SOCS5 LINC01118 LINC01119
2	window_22	64763588	64873588	0.24407115	AFTPH SERTAD2
2	window_23	64833588	64933588	0.33333333	SERTAD2

2	window_25	72243588	72423588	0.32473814	CYP26B1 EXOC6B
2	window_28	101403588	101593588	0.31192308	NPAS2
2	window_30	111703588	111893588	0.33184894	ACOXL BCL2L11
2	window_31	113233588	113373588	0.77272727	TTL POLR1B CHCHD5
2	window_32	113293588	113413588	0.63636364	TTL POLR1B CHCHD5 SLC20A1
2	window_39	163683588	163833588	0.94318182	KCNH7
2	window_42	179243588	179343588	0.30769231	OSBPL6 PRKRA DFNB59 FKBP7
2	window_47	189723588	189963588	0.45643325	COL3A1 COL5A2
3	window_2	25596417	25706417	0.35119048	RARB TOP2B
3	window_4	32506417	32616417	0.27272727	CMTM7 CMTM6 DYNC1LI1
3	window_5	40406417	40646417	0.40215673	ENTPD3 RPL14 ZNF619 ZNF620 ZNF621
3	window_7	111796417	111896417	0.27272727	TMPRSS7 C3orf52 GCSAM SLC9C1
3	window_8	111956417	112056417	0.27272727	SLC9C1 CD200
3	window_9	119376417	119546417	0.31255902	POPDC2 COX17 MAATS1 NR1I2 GSK3B
3	window_10	150146417	150256417	0.33333333	TSC22D2
4	window_1	7733908	7853908	0.23939233	SORCS2 AFAP1
4	window_6	37933908	38163908	0.52609302	TBC1D1 PTTG2
4	window_7	42543908	42703908	0.40348818	ATP8A1
4	window_10	75843908	75943908	0.27272727	PARM1
4	window_13	88833908	89113908	0.36271014	SPP1 PKD2 ABCG2
4	window_14	95653908	95843908	0.36941874	BMPR1B
4	window_23	165793908	165913908	0.32600733	TRIM61 FAM218A
4	window_24	165833908	165943908	0.58333333	TRIM61 FAM218A
4	window_25	165873908	165973908	0.66666667	TRIM61 FAM218A TRIM60
5	window_3	115873929	115993929	0.21013431	SEMA6A
6	window_2	16722897	16922897	0.34452733	ATXN1
6	window_3	17112897	17212897	0.27272727	STMND1
6	window_6	28172897	28362897	0.2753252	ZSCAN9 ZKSCAN4 NKAPL PGBD1 ZSCAN31 ZKSCAN3 ZSCAN12
6	window_7	29292897	29402897	0.45836044	OR5V1 OR12D3 OR12D2 OR11A1
6	window_8	29352897	29492897	0.46079197	OR5V1 OR12D2 OR11A1 OR10C1 OR2H1 MAS1L
6	window_9	32222897	32332897	0.63730242	C6orf10
6	window_10	32392897	32502897	0.33104213	HLA-DRA HLA-DRB5
6	window_11	32492897	32662897	0.5484314	HLA-DRB5 HLA-DRB1 HLA-DQA1 HLA-DQB1
6	window_12	32862897	32972897	0.24722222	HLA-DMB XXbac-BPG181M17.5 HLA-DMA BRD2 HLA-DOA
6	window_13	43722897	43822897	0.20833333	VEGFA
6	window_17	134482897	134672897	0.520007	SGK1
6	window_18	134592897	134712897	0.35533911	SGK1
6	window_19	136122897	136222897	0.30769231	PDE7B
6	window_20	136152897	136252897	0.28571429	PDE7B
6	window_21	143942897	144132897	0.21284118	RP3-468K18.5
6	window_24	150622897	150812897	0.38850435	IYD
7	window_1	20431442	20561442	0.41145833	ITGB8
7	window_3	20541442	20711442	0.48673598	ABCB5

7	window_7	34591442	34881442	0.45856307	AC005493.1 NPSR1
7	window_10	71061442	71231442	0.466496	WBSCR17
7	window_11	89791442	89891442	0.36363636	STEAP1 STEAP2 C7orf63
7	window_12	89811442	89911442	0.36363636	STEAP2 C7orf63
7	window_14	128421442	128571442	0.29690349	CCDC136 FLNC ATP6V1F
7	window_15	150231442	150361442	0.29761905	GIMAP4 GIMAP6
7	window_16	150691442	150881442	0.35615138	NOS3 ATG9B ABCB8 ASIC3 CDK5 SLC4A2 FASTK TMUB1 AGAP3 GBX1 ASB10
8	window_1	9551022	9781022	0.87532705	TNKS
8	window_2	9791022	9911022	0.30637255	MATN2 RPL30 C8orf47
8	window_3	14941022	15151022	0.62876184	SGCZ
8	window_4	16381022	16591022	0.48457281	MSR1
8	window_5	17451022	17551022	0.22857143	PDGFRL MTUS1
8	window_6	24191022	24381022	0.46011702	ADAM28 ADAMDEC1 ADAM7
8	window_8	25881022	25981022	0.23809524	EBF2
8	window_10	56331022	56491022	0.3241444	XKR4
8	window_11	98891022	99081022	0.41120029	MATN2 RPL30 C8orf47
8	window_12	132951022	133051022	0.27272727	EFR3A OC90
8	window_13	132971022	133131022	0.35409364	EFR3A OC90 HHLA1
9	window_2	79948153	80088153	0.37562174	VPS13A GNA14
9	window_3	80748153	80898153	0.46688034	CEP78
9	window_4	80828153	80938153	0.38088235	CEP78 PSAT1
9	window_5	91948153	92078153	0.3218032	SECISBP2 SEMA4D
9	window_6	92008153	92128153	0.31901432	SEMA4D
9	window_10	112858153	112988153	0.21533267	PALM2-AKAP2 AKAP2 C9orf152
9	window_11	116628153	116738153	0.34615385	ZNF618
10	window_1	5460172	5630172	0.23128053	NET1 CALML5 CALML3
10	window_4	25620172	25940172	0.51977592	GPR158
10	window_11	91170172	91290172	0.30891331	LIPA IFIT5 SLC16A12
10	window_12	92390172	92540172	0.48844211	HTR7
10	window_13	102750172	102970172	0.55780113	C10orf2 LZTS2 PDZD7 SFXN3 KAZALD1 TLX1NB HUG1 TLX1
10	window_17	127080172	127300172	0.48860175	TEX36
10	window_18	133060172	133260172	0.40969271	TCERG1L
11	window_2	63785920	63975920	0.39106963	MACROD1 FLRT1 STIP1 FERMT3
11	window_3	63915920	64055920	0.42453782	MACROD1 STIP1 FERMT3 TRPT1 NUDT22 DNAJC4 VEGFB FKBP2 PPP1R14B PLCB3 BAD GPR137
11	window_4	66155920	66395920	0.57703952	NPAS4 MRPL11 PELI3 DPP3 CTD-3074O7.11 BBS1 ZDHHC24 CTSF CCDC87 CCS RBM14 RBM4 RBM14-RBM4
11	window_6	103285920	103395920	0.2967033	DYNC2H1
11	window_8	125475920	125595920	0.33766234	STT3A CHEK1 ACRV1
11	window_9	125525920	125715920	0.36559524	CHEK1 ACRV1 PATE1 PATE2 PATE3 PATE4
12	window_1	280130	380130	0.24137931	IQSEC3 SLC6A12 SLC6A13 CCDC91 BICD1
12	window_2	300130	470130	0.42170423	SLC6A12 SLC6A13 KDM5A BICD1 PUS7L IRAK4 TWF1

12	window_6	19770130	19960130	0.36314325	AEBP2
12	window_7	20920130	21130130	0.64732042	SLCO1B3 LST3 SLCO1B7
12	window_8	28660130	28820130	0.58084772	IQSEC3 CCDC91
12	**window_9**	**28740130**	**28840130**	**1**	**IQSEC3**
12	window_11	32220130	32420130	0.36775754	SLC6A12 BICD1
12	window_12	44100130	44210130	1	PUS7L IRAK4 TWF1
12	window_13	51830130	51990130	0.35142466	SLC4A8 SCN8A
12	window_15	54480130	54690130	0.57197507	SMUG1 CBX5 HNRNPA1 NFE2
12	window_16	54640130	54760130	0.42307692	CBX5 HNRNPA1 NFE2 COPZ1 GPR84
12	window_17	55000130	55160130	0.38617166	LACRT DCD
12	window_18	72240130	72500130	0.63523779	TBC1D15 TPH2 TRHDE
12	window_19	79750130	79870130	0.36363636	SYT1
12	window_20	79890130	79990130	0.45454546	PAWR
12	window_21	79910130	80040130	0.46780303	PAWR
12	window_23	115760130	115950130	0.28116986	RP11-116D17.1
12	window_24	120090130	120250130	0.31739689	PRKAB1 CIT RP1-127H14.3
13	window_2	38166754	38346754	0.39073046	POSTN TRPC4
15	window_4	74419207	74539207	0.20555556	ISLR2 RP11-247C2.2 ISLR STRA6 CCDC33
15	window_5	74469207	74579207	0.21780303	ISLR STRA6 CCDC33
15	window_6	75409207	75559207	0.44194139	PPCDC C15orf39 GOLGA6C
15	window_7	79249207	79419207	0.21625906	RASGRF1
15	window_8	85349207	85499207	0.25486856	ZNF592 ALPK3 SLC28A1
15	window_9	85439207	85539207	0.23076923	SLC28A1 PDE8A
15	window_10	91269207	91369207	0.36363636	BLM
15	window_11	91289207	91399207	0.2967033	BLM
15	window_12	98199207	98439207	0.35638789	LINC00923
16	**window_1**	**3055795**	**3365795**	**0.75279578**	**CLDN9 CLDN6 TNFRSF12A HCFC1R1 THOC6 CCDC64B MMP25 IL32 ZSCAN10 ZNF205 ZNF213 CASP16 OR1F1 ZNF200 MEFV ZNF263 TIGD7 ZNF75A**
16	window_3	50455795	50585795	0.33016983	NKD1
16	window_4	50595795	50765795	0.44636405	NKD1 SNX20 NOD2
16	window_5	83645795	83835795	0.35712487	CDH13
16	window_6	84085795	84185795	0.24	MBTPS1 HSDL1 DNAAF1
16	window_7	84325795	84475795	0.23089534	WFDC1 ATP2C2
17	window_1	37128420	37318420	0.62771654	PLXDC1 ARL5C
17	window_2	48118420	48268420	0.36319444	ITGA3 PDK2 SAMD14 PPP1R9B SGCA COL1A1
17	window_3	56318420	56428420	0.32467533	LPO MPO BZRAP1 SUPT4H1
18	window_4	70501004	70711004	0.48742054	NETO1
19	window_1	8216451	8316451	0.36363636	CERS4
20	window_1	48614017	48784017	0.44876844	TMEM189-UBE2V1 UBE2V1 TMEM189
20	window_2	48714017	48844017	0.43891403	TMEM189-UBE2V1 UBE2V1 TMEM189 CEBPB
20	window_3	54944017	55124017	0.44605329	AURKA CSTF1 CASS4 RTFDC1 GCNT7 FAM209A FAM209B
20	window_4	55054017	55174017	0.37820513	RTFDC1 GCNT7 FAM209A FAM209B
20	window_5	61094017	61244017	0.33061383	C20orf166

Appendix Table 5. Top 1% windows presenting highest values of outlier SNPs over the total number of SNPs present in the window (Ntop/Ntot) according to F_{cs} statistics calculated for TAS group present in the "selection SNP-chip" dataset.

Chr	Window name	Window start (bp)	Window end (bp)	Ntop/Ntot	Genes
1	window_1	3539842	3699842	0.25882186	TPRG1L WRAP73 TP73 CCDC27 SMIM1 LRRC47
1	window_2	3619842	3729842	0.22380952	TP73 CCDC27 SMIM1 LRRC47 CEP104
1	window_3	19049842	19219842	0.28640991	PAX7 TAS1R2 RP13-279N23.2 ALDH4A1
1	window_5	95279842	95519842	0.58543407	SLC44A3 CNN3 ALG14
1	window_7	152249842	152349842	0.27272727	FLG FLG2
1	window_8	168059842	168209842	0.34292027	GPR161 TIPRL SFT2D2
1	window_9	168519842	168759842	0.28960023	XCL1 DPT
1	window_10	170959842	171149842	0.34728427	MROH9 FMO3 FMO6P
1	window_11	171069842	171169842	0.2	FMO3 FMO6P FMO2
1	window_16	203819842	203929842	0.34848485	ZC3H11A SNRPE
1	window_18	208169842	208379842	0.28864198	PLXNA2
1	window_20	212869842	213039842	0.33927743	BATF3 NSL1 TATDN3 C1orf227 FLVCR1
1	window_23	220979842	221079842	0.28571429	MARC1 HLX
1	window_24	229339842	229529842	0.45630495	TMEM78 RAB4A SPHAR CCSAP
2	window_1	3723588	3923588	0.40249145	ALLC DCDC2C
2	window_2	5333588	5463588	0.31832108	GPR75-ASB3 ASB3 CHAC2 ERLEC1 GPR75 PSME4 ACYP2
2	window_3	20153588	20313588	0.43648676	WDR35 MATN3 LAPTM4A
2	window_4	20863588	20983588	0.32638889	GDF7 C2orf43
2	window_6	26723588	26833588	0.24521531	OTOF C2orf70 CIB4
2	window_7	31233588	31393588	0.31693072	GALNT14
2	window_8	31523588	31723588	0.61822066	XDH
2	window_10	32383588	32483588	0.83333333	SLC30A6 NLRC4
2	window_11	32403588	32563588	0.69398459	SLC30A6 NLRC4 YIPF4
2	window_13	53953588	54183588	0.6493786	GPR75-ASB3 ASB3 CHAC2 ERLEC1 GPR75 PSME4
2	window_14	54183588	54303588	0.54292929	PSME4 ACYP2
2	window_17	60873588	60983588	0.40909091	PAPOLG
2	window_18	61133588	61273588	0.43752914	REL PUS10 PEX13
2	window_19	61323588	61423588	0.27272727	KIAA1841 C2orf74 AHSA2 USP34
2	window_21	69433588	69533588	0.46153846	ANTXR1
2	window_22	69483588	69583588	0.72727273	GFPT1
2	window_23	69653588	69793588	0.57307692	NFU1 AAK1
2	window_27	179163588	179343588	0.54258312	OSBPL6 PRKRA DFNB59 FKBP7
2	window_28	179313588	179523588	0.68899749	PRKRA DFNB59 FKBP7 PLEKHA3 TTN
2	window_29	183583588	183723588	0.28249084	DNAJC10 FRZB
2	window_30	206333588	206503588	0.51463832	PARD3B
2	window_31	206423588	206523588	0.23809524	PARD3B
2	window_33	222113588	222293588	0.30041482	EPHA4

2	window_34	240213588	240323588	0.2967033	HDAC4 AC062017.1
3	window_3	14146417	14246417	0.22727273	CHCHD4 TMEM43 RP11-434D12.1 XPC LSM3
3	window_4	14176417	14296417	0.22397892	TMEM43 RP11-434D12.1 XPC LSM3
3	window_6	58526417	58736417	0.4271675	FAM107A FAM3D C3orf67
3	window_8	61206417	61386417	0.66418905	FHIT
3	window_11	62116417	62286417	0.56968032	PTPRG
3	window_12	62256417	62396417	0.39668998	C3orf14 FEZF2 CADPS PTPRG
3	window_13	63516417	63706417	0.28052305	SYNPR SNTN
3	window_14	88006417	88126417	0.66666667	HTR1F CGGBP1 RP11-159G9.5
3	window_15	88076417	88176417	0.72727273	CGGBP1 RP11-159G9.5
3	window_16	140866417	141086417	0.61706784	SPSB4 ACPL2 ZBTB38
3	window_20	182776417	182896417	0.28669109	MCCC1 LAMP3 MCF2L2
3	window_21	185056417	185226417	0.49583333	MAP3K13 TMEM41A LIPH
4	window_8	96453908	96653908	0.3350074	UNC5C
4	window_12	140493908	140693908	0.39444905	SETD7 MGST2 MAML3
4	window_15	165783908	165913908	0.4271978	TRIM61 FAM218A
4	window_16	165833908	165943908	0.41666667	TRIM61 FAM218A
4	window_20	183863908	184033908	0.38292125	WWC2
4	window_21	184293908	184453908	0.37786132	CDKN2AIP ING2
5	window_2	16043929	16143929	0.27272727	GABRA6 MARCH11
5	window_3	16173929	16283929	0.27272727	MARCH11
5	window_5	16313929	16463929	0.35289988	ZNF622
5	window_13	145683929	145793929	0.27272727	RBM27 POU4F3
5	window_14	147543929	147643929	0.27272727	SPINK14 SPINK6
5	window_15	161113929	161223929	0.33566434	GABRA6
5	window_16	164603929	164713929	0.36363636	ZNF622
6	window_2	15620341	15770341	0.38589189	DTNBP1
6	window_4	25900341	26030341	0.31839827	SLC17A2 TRIM38 HIST1H1A HIST1H3A HIST1H4A HIST1H4B
6	window_5	25950341	26050341	0.27272727	TRIM38 HIST1H1A HIST1H3A HIST1H4A HIST1H4B HIST1H3B HIST1H2AB HIST1H2BB HIST1H3C
6	window_6	29760341	29860341	0.125	HLA-G
6	window_7	29810341	29960341	0.2942833	HLA-A
6	window_8	29910341	30030341	0.17193492	HLA-A ZNRD1
6	window_13	147790341	147930341	0.21897671	SAMD5
7	window_3	14901442	15051442	0.31107085	DGKB
7	window_4	18091442	18311442	0.64895111	HDAC9
7	window_6	21541442	21721442	0.34952496	SP4 DNAH11
7	window_7	36131442	36251442	0.24140212	EEPD1
7	window_9	133711442	133861442	0.49309764	EXOC4 LRGUK
7	window_10	133791442	133901442	0.81818182	LRGUK
7	window_12	133931442	134131442	0.58177181	LRGUK SLC35B4 AKR1B1
8	window_1	8241022	8421022	0.49750577	SGK223
8	window_2	17541022	17731022	0.37355029	MTUS1 FGL1

8	window_3	17921022	18071022	0.32311814	ASAH1 NAT1
8	window_4	18001022	18111022	0.31329849	NAT1
8	window_5	18041022	18211022	0.31738572	NAT1
8	window_6	76221022	76391022	0.58250916	HNF4G
8	window_7	76311022	76411022	1	HNF4G
8	window_8	76361022	76551022	1	HNF4G
8	window_12	101431022	101571022	0.30236928	ANKRD46
8	window_13	109711022	109821022	0.2967033	TMEM74
8	window_16	110341022	110451022	0.90909091	NUDCD1 ENY2 PKHD1L1
8	window_18	110431022	110571022	0.45424242	PKHD1L1 EBAG9
8	window_19	124841022	125021022	0.29602362	FER1L6
8	window_20	127541022	127691022	0.29119556	FAM84B
9	window_2	3487266	3597266	0.35119048	RFX3
9	window_8	11627266	11787266	0.28613292	ZNF618
9	window_10	14567266	14687266	0.25	ZDHHC21
9	window_17	116627266	116727266	0.30769231	ZNF618
9	window_18	126987266	127177266	0.35463478	NEK6 PSMB7
9	window_19	131487266	131587266	0.27272727	ZER1 TBC1D13 ENDOG C9orf114
9	window_20	133187266	133387266	0.29776383	HMCN2 AL354898.1 ASS1
9	window_21	137517266	137697266	0.32192612	COL5A1
10	window_1	1163246	1353246	0.33066077	LHPP RP11-12J10.3 FAM53B CCDC172 PNLIPRP3 RAB11FIP2 WDR37 ADARB2 CCDC3
10	window_2	1283246	1403246	0.28571429	CCDC3 FRMD4A
10	window_3	5703246	5863246	0.33385854	PCDH15 ASB13 FAM208B GDI2
10	window_7	30863246	30973246	0.27619048	LYZL2
10	window_8	31813246	31923246	0.56089744	ZEB1
10	window_10	33373246	33563246	0.31680855	NRP1
10	window_12	90743246	90863246	0.37529138	ACTA2 FAS
10	window_14	118053246	118153246	0.27272727	CCDC172
10	window_15	118073246	118173246	0.36363636	CCDC172
10	window_16	118093246	118263246	0.33944151	CCDC172 PNLIPRP3
10	window_17	119713246	119843246	0.28168498	RAB11FIP2
10	window_18	126273246	126403246	0.26704546	LHPP RP11-12J10.3 FAM53B
11	window_1	9795920	9895920	0.28571429	SBF2
11	window_2	15885920	16005920	0.28669109	SOX6
11	window_3	15935920	16075920	0.30694874	SOX6
11	window_4	35095920	35275920	0.28883089	AL356215.1 CD44 SLC1A2
11	window_5	36155920	36285920	0.31486167	LDLRAD3
11	window_6	73935920	74035920	0.33333333	PPME1 P4HA3
11	window_7	74775920	74965920	0.3217723	OR2AT4 SLCO2B1 TPBGL
11	window_9	100635920	100925920	0.36444778	ARHGAP42 TMEM133 PGR
11	window_10	117915920	118035920	0.27101662	TMPRSS4 SCN4B SCN2B
11	window_11	117965920	118075920	0.27903226	TMPRSS4 SCN4B SCN2B AMICA1
11	window_12	126295920	126425920	0.22844517	ST3GAL4
11	window_13	126825920	126985920	0.3657695	KIRREL3

12	window_1	12565828	12715828	0.34157509	LOH12CR1 DUSP16
12	window_2	12635828	12735828	0.36363636	DUSP16
12	window_6	103955828	104165828	0.54996267	STAB2 NT5DC3
13	window_2	28762387	28942387	0.42389514	PAN3 FLT1
13	window_4	36042387	36242387	0.39404745	MAB21L1
13	window_5	36212387	36332387	0.22242572	NBEA
13	window_6	36252387	36362387	0.20689655	DCLK1
13	window_7	36282387	36392387	0.23666667	DCLK1
13	window_9	37152387	37272387	0.33564214	SERTM1
13	window_12	51242387	51402387	0.29019627	DLEU1 DLEU7
13	window_13	77892387	78042387	0.27806258	MYCBP2
14	window_1	21862921	22062921	0.60311328	CHD8 RAB2B TOX4 METTL3 SALL2 OR10G3
14	window_2	36302921	36422921	0.35353535	BRMS1L
14	window_3	36342921	36442921	0.36363636	BRMS1L
14	window_6	63512921	63872921	0.61077171	KCNH5 RHOJ PPP2R5E
14	window_7	71652921	71792921	0.33912338	AC004817.1 SIPA1L1
14	window_8	77652921	77832921	0.26416963	RP11-463C8.4 TMEM63C NGB POMT2 GSTZ1 TMED8
14	window_9	78392921	78502921	0.22321429	ADCK1
14	window_12	105352921	105452921	0.33333333	CEP170B PLD4 AHNAK2 C14orf79
14	window_13	105502921	105692921	0.37702201	GPR132 JAG2 NUDT14 BRF1
15	window_1	24759207	24989207	0.57127396	NPAP1
15	window_2	29409207	29619207	0.48090278	APBA2 FAM189A1 NDNL2
15	window_3	33469207	33669207	0.32377942	FMN1 RYR3
15	window_4	63359207	63499207	0.44711539	TPM1 LACTB RPS27L RAB8B
15	window_5	74529207	74719207	0.47650814	CCDC33 CYP11A1 SEMA7A
15	window_6	75039207	75229207	0.38472243	CYP1A2 CSK LMAN1L CPLX3 ULK3 SCAMP2 MPI FAM219B COX5A
16	window_2	8917458	9057458	0.53433474	PMM2 CARHSP1 USP7
16	window_3	9007458	9127458	0.41258741	USP7
16	window_4	19977458	20177458	0.53182872	GPR139
16	window_5	30857458	31037458	0.48101528	BCL7C CTF1 FBXL19 ORAI3 SETD1A HSD3B7 AC135048.1 STX1B
16	window_6	87687458	87847458	0.25899693	JPH3 AC010536.1 KLHDC4 FLJ00104
17	window_4	40668420	40778420	0.59090909	NAGLU HSD17B1 COASY MLX PSMC3IP FAM134C TUBG1 ATP6V0A1
17	window_5	40698420	40898420	0.4691672	HSD17B1 COASY MLX PSMC3IP FAM134C TUBG1 TUBG2 PLEKHH3 CCR10 CNTNAP1 EZH1
17	window_6	43418420	43528420	0.27272727	ARHGAP27 PLEKHM1
17	window_7	47298420	47478420	0.40288415	ABI3 PHOSPHO1 ZNF652 RP11-81K2.1
17	window_8	49638420	49828420	0.27651731	CA10
17	window_9	77668420	77798420	0.46780303	ENPP7 CBX2 CBX8
17	window_10	78008420	78148420	0.25237154	TBC1D16 CCDC40 GAA EIF4A3 CARD14
17	window_11	80658420	80778420	0.38888889	FN3KRP FN3K TBCD
17	window_12	80698420	80828420	0.3856352	FN3K TBCD ZNF750

18	window_3	11801004	11901004	0.23809524	GNAL CHMP1B MPPE1
18	window_7	66641004	66811004	0.38600574	CCDC102B
19	window_1	7386451	7586451	0.48424917	CTB-133G6.1 CTD-2207O23.3 ARHGEF18 PEX11G C19orf45 CTD-2207O23.12 ZNF358
20	window_1	1027900	1197900	0.32893445	PSMF1 TMEM74B C20orf202
20	window_2	3857900	3987900	0.29345654	PANK2 RNF24
20	window_3	49407900	49507900	0.23809524	BCAS4 ADNP
20	window_4	49447900	49567900	0.33041101	BCAS4 ADNP DPM1
20	window_5	62237900	62337900	0.27272727	GMEB2 STMN3 RTEL1 RTEL1-TNFRSF6B TNFRSF6B ARFRP1
21	window_2	43565468	43705468	0.33710246	ABCG1
21	window_3	45195468	45295468	0.26666667	CSTB RRP1 AGPAT3
22	window_1	38841027	38941027	0.27272727	KCNJ4 KDELR3 DDX17 DMC1
22	window_2	44271027	44431027	0.26617787	PNPLA5 PNPLA3 SAMM50 PARVB
22	window_3	45791027	45961027	0.24995052	SMC1B RIBC2 FBLN1
22	window_4	50481027	50651027	0.38404552	TTLL8 MLC1 MOV10L1 PANX2 TRABD SELO