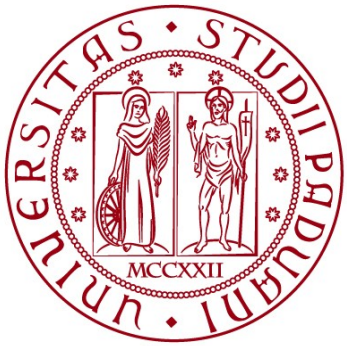
UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI BIOLOGIA

Corso di Laurea magistrale in Biologia Evoluzionistica



TESI DI LAUREA

Excavating Denisovan ancestries from the genomes of modern humans: an evaluation of the diversity of Denisovan populations

Relatore: Prof. Luca Pagani
Dipartimento di Biologia

Correlatore: Dott. Leonardo Vallini
Johannes Gutenberg-Universität Mainz

Laureanda:
Francesca Carollo

ANNO ACCADEMICO 2023/2024

Table of Contents

0. Abstract

Denisovans are a group of now extinct, archaic humans who lived in the Eastern part of Eurasia in the Middle and Late Pleistocene, and they are considered to be a sister group to Neanderthal, from which they diverged early in their evolutionary history. Even though to this date only one high coverage Denisovan genome is available (the Altai Denisova genome from Denisova Cave, Siberia), it can still give us extensive genomic data to compare to the modern human genome: analyses performed on this Denisovan genome have yielded some understanding into their population history and even highlighted the admixture of these archaic humans with Neanderthals and modern humans. The analysis of Denisovan ancestries found in the genomes of some modern human populations as introgressed genetic material left by such admixture events can provide insights into the diversity of Denisovan populations and overcome the currently limited availability of Denisova genomes. The segments of archaic ancestry I dug out of a pan-Eurasian panel of modern genomes show differing similarity to the sequenced Denisovan genome and this hints to a diversified pattern of multiple Denisovan groups, a regional mosaic of archaic humans that admixed with modern humans across a wide area in Eurasia, and gave them genetic material that can be differentiated with the proper genomics tools.

1. Introduction

1.1 History of the hominids' colonization of Eurasia

Human paleogenomics aims to reconstruct the evolutionary history of humankind, including both modern and archaic humans, from genomic data.
The investigation of our extinct closest relatives through genetic information is not only fascinating, but also serves the purpose of better understanding *who we are* and *what makes us human*, two of the relevant questions that anthropology eventually strives to answer: through comparison with others, we understand ourselves.

Our species, *Homo sapiens,* originated in Africa at least 200-150 thousand years ago. Modern humans started their route out of Africa to colonize the whole world in multiple migrations: the early diffusions likely consisted of populations on the small scale, eventually followed by a large-scale expansion at ~70–60 thousand years ago, from which all present-day non-Africans descend (Bergström et al., 2020; Freidline et al., 2023; Groucutt et al., 2018, 2021; Grün et al., 2005; Harvati et al., 2019; Hershkovitz et al., 2018).
At the time of colonization by modern humans, Eurasia was already inhabited by other archaic humans. Studying the interactions between moderns and archaics can provide insights into the population history of modern humans, shedding light on the definition of our own species through the reconstruction of the early history of non-African populations.
Given the sparse nature of archaeological findings, this type of investigation cannot transcend the use of

molecular approaches: with the advent of Molecular Anthropology, the analysis of DNA sequences retrieved from remains offers an approach that is complementary to morphology in understanding hominin relationships.

The first hominin group to leave Africa was *Homo erectus* about 1.9 million years ago, in the Late Pliocene-Early Pleistocene (Gabunia et al., 2000). After this event, archaeological as well as genetic data indicate that at least two groups of hominins left Africa during the Pleistocene: first, presumably *Homo heidelbergensis* (Hublin, 2009) at around 650 thousand years ago; and, second, anatomically modern humans about 60 thousand years ago (Krause et al., 2010; Pagani et al., 2015; Schiffels & Durbin, 2014; Soares et al., 2012).

In Western Asia and Europe, the group that had exited Africa as *H. heidelbergensis* differentiated into the hominin known as Neanderthal (Di Vincenzo & Manzi, 2023).
Neanderthals have left extensive evidence of their presence over a long span of time in West Asia and Europe: they appear in the European fossil record about 400 thousand years ago (Bischoff et al., 2007; Hublin, 2009; Stringer & Hublin, 1999) and evolved in progressively more distinctive Neanderthal forms, until their demise after 40 thousand years ago (Hajdinjak et al., 2018).
Neanderthals are the sister group to modern humans and the sequencing of their genome in 2010 was a fundamental milestone in understanding our human history: the Neanderthal contribution to the gene pool of modern humans was confirmed, and thus it was unequivocal that an admixing event with our closer relatives took place along the evolutionary path of all non-Africans.

As a matter of fact, present-day humans' genomes show that Neanderthals share significantly more derived alleles with non-Africans than with Africans, whereas they share equal amounts of derived alleles when compared either to individuals within Eurasia or to individuals within Africa. Thus, the gene flow between Neanderthals and modern humans most likely occurred before the divergence of all Eurasians, in an admixing event that presumably took place before the modern humans' further expansion into Eurasia (Green et al., 2010).

Shifting the focus geographically towards Eastern Eurasia, the debate over which hominin species were present there during the Pleistocene has yet to be settled. A series of other archaic hominins have left archeological traces of their presence in Eastern Eurasia, such as the findings attributed to *H. erectus* in China (Zhu et al., 2015) and Java Island (Baba et al., 2003; Swisher et al., 1996) and *Homo floresiensis* (P. Brown et al., 2004) in Flores Island, Indonesia. Further, many archeological findings from East Asia are contended between hominin species, according to different authors: some have emphasized the morphological affinities between Neanderthals and the specimen of Maba (Guangdong province, China), or between *H. heidelbergensis* and the Dali Man skull (Shaanxi province, China), but other authors classify these as "early *H. sapiens" (*Pope, 1992; Rightmire, 2004; Wu and Poirier, 1995). However, given the current absence of genetic data from these specimens, an ultimate agreement is yet to be achieved.

1.2 The discovery of Denisovans and their limited fossil record

In 2008, the distal manual phalanx of the fifth digit of a juvenile hominin was excavated at Denisova Cave (51°40′N; 84°68′E), a site located in the Altai Mountains in Southern Siberia (Reich et al., 2010), in Eastern Eurasia.
The excavations of the site over the past 25 years have uncovered evidence that episodic human occupation has been going on at the site since at least 125 thousand years ago.
The bone was added to the collection of other hominin fragmentary remains found in the Cave and called Denisova 3. The phalanx was found in a stratum (layer 11) dated to 48–30 thousand years ago, where an assemblage containing both Upper and Middle Palaeolithic elements has been reported (Z. Jacobs et al., 2019; Krause et al., 2010). Importantly, the bone itself was instead dated at 85 thousand years ago.
The Denisova 3 phalanx yielded both mitochondrial and nuclear DNA, which together allowed to assign it to an individual belonging to a hominin group that shares a common origin with Neanderthals (Reich et al., 2010): this archaic human population was designated as ‘Denisovans’, as they were described for the first time in relation to molecular data from the Denisova Cave, much similarly to how Neanderthals were first described based on remains found in Neander Valley, Germany (King, 1864).

Triggered by the discovery of this novel population of archaic humans, now called Denisovans, in the last decade a series of other remains have been attributed to them, according to molecular information (Table 1). The debate remains open on whether additional specimens, already present in museum collections in

the area, should be re-considered as of putative Denisovan origin.

Table 1: Denisovan remains identified based on molecular information

Specimen	Description	Origin	Age (kya)	Molecular data available	References
Denisova 2	Molar	Denisova Cave	194,4-122,7	mtDNA, 730 Mb of nuclear sequences	Slon et al., 2017
Denisova 3	Phalanx	Denisova Cave	76,2-51,6	mtDNA, high coverage nuclear genome	Bennett et al., 2019; Krause et al., 2010; Reich et al., 2010
Denisova 4	Molar	Denisova Cave	84,1-55,2	mtDNA, Y chromosome, 54.6 Mb of nuclear sequences	Reich et al., 2010, Sawyer et al., 2015
Denisova 8	Molar	Denisova Cave	136,4-105,6	mtDNA, Y chromosome, 256 Mb of nuclear sequences	Sawyer et al., 2015
Denisova 11	Long bone fragment	Denisova Cave	118,1-79,3	Ancient collagen, mtDNA (Neanderthal type)*, low coverage nuclear genome	S. Brown et al., 2016; Slon et al., 2018
Denisova 13	Fragment of skull	Denisova Cave	unknown	mtDNA	Viola et al., 2019
Denisova 19 Denisova 20 Denisova 21	Undiagnostic bone fragments	Denisova Cave	217-187	Ancient collagen, mtDNA	S. Brown et al., 2022
Xiahe 1	Right hemi-mandible with two molars	Baishiya Karst Cave	> 160	Ancient dentine peptides, mtDNA from sediment	F. Chen et al., 2019
TNH2-1	Molar	Tam Ngu Hao 2 (Cobra Cave)	164-131	Ancient enamel proteome	Demeter et al., 2022
Xiahe 2	Rib	Baishiya Karst Cave	48–32	Shotgun proteomic analysis	Xia et al., 2024

* Denisova 11 is the hybrid offspring of a Neanderthal mother and a Denisovan father (Slon *et al.*, 2018).

1.3 The genomic relationship between Denisovans, Neanderthals and modern humans

When compared to present-day human mtDNA, the Denisova 3 mitochondrial data showed twice the genetic differences than those found between Neanderthal and present-day human mtDNA. This translates to Denisova 3 and present-day humans sharing a most recent common mtDNA ancestor at about 1 million years ago, which is deeper than the most recent common mtDNA ancestor of modern humans and Neanderthals, dated at about 400 thousand years ago (Fu, Mittnik, et al., 2013; Rieux et al., 2014).
Additional archaic mitochondrial genomes, categorized in Neanderthal or Denisovan mtDNA types, suggested a sister group relationship between modern humans and Neanderthals, with Denisovans as a basal mtDNA outgroup (Meyer et al., 2012, 2016; Prüfer et al., 2014; Reich et al., 2010; Sawyer et al., 2015). Based on this deep divergence found in mitochondrial DNA, it was concluded at the time of the first analyses on the Denisova 3 mtDNA that the individual likely belonged to a previously unknown hominin, outgroup to Neanderthal and modern humans, who lived in Southern Siberia.

However, nuclear DNA data revealed a different story: the Neanderthal and Denisovan populations separated only after their common divergence from the lineage leading to modern humans, so Denisovans are the sister group to Neanderthals.
The Denisovan genome does not fall within the genetic variation of Neanderthals, so Denisovans are considered as a separated archaic population: the average number of pairwise differences between Neanderthal and Denisovan autosomal genomes is approximately four times higher than that among

Neanderthals. Yet, the average divergence of the Denisovan nuclear genome from present-day humans is similar to that of Neanderthals: Denisova and Neanderthal carry alleles in the ancestral state respectively at 11.7% and 12.2% of the positions where modern human and chimpanzee are different (Reich et al., 2010). Neanderthal and Denisova share more autosomal alleles with each other than either of them with present-day humans. This suggests that the two archaics descended from a common ancestral population that has been isolated from the ancestors of present-day humans for a long time. The dating for the population split time between the two archaic hominin groups and modern humans was estimated to 765–550 thousand years ago from nuclear DNA (Prüfer et al., 2014)

The corresponding divergence time for Neanderthal and human mtDNA dated to 400 thousand years ago was definitely in contrast with the time estimate obtained from autosomal data. Nuclear DNA sets the divergence at a deeper time in the past and puts Neanderthal and Denisova as sister groups, but Neanderthal mtDNA clusters with the modern humans'. Analyses of Middle Pleistocene hominins from the Sima de los Huesos site in northern Spain, dated at around 400 thousand years ago, cleared the understanding of the situation: the Sima de los Huesos nuclear DNA confirmed their affinity to the Neanderthal lineage.

However, in contrast to genome-wide data, the Sima de los Huesos mtDNA was found to branch off with the deeply divergent Denisovan mtDNA lineage.

These early Neanderthals from Sima de los Huesos carried mitochondrial genomes related to Denisovan mtDNA.

The phylogenetic discordancies are reconciled in a scenario in which the mtDNA of early Neanderthals

was indeed Denisovan-like, but it was later substituted by a more derived mtDNA lineage, in an introgression event from an early lineage related to modern humans (Meyer et al., 2016; Posth et al., 2017).

Similarly to mtDNA, also the Neanderthal Y-chromosome landscape seems to have been replaced: the split between Denisovan and modern human Y chromosome lineages at around 700 thousand years ago is in agreement with the split times inferred from nuclear DNA. This confirms that the differentiation of Denisovan Y chromosomes from modern humans occurred through a simple population split. On the contrary, Neanderthal and modern human Y chromosome lineages diverged from each other around 370 thousand years ago, suggesting that the Y chromosome has been replaced in Neanderthals through ancient gene flow from an early lineage closely related to modern humans (Petr et al., 2020).

1.4 Hypothesized super archaic hominin ancestry in the Denisovan genome

In some of its genomic regions, the Denisovan genome is unexpectedly divergent to modern humans and Neanderthals. Denisovans could have inherited this genetic material from an unknown deeply divergent human group, referred to as ‘super-archaic’. The inferred presence of super-archaic ancestry in the Denisovan genome has been debated and investigated with different methods, which all agree that a small contribution (0–6%) from a super-archaic hominin is at least probable (Hubisz et al., 2020; Prüfer et al., 2017).

1.5 Admixture and gene flow between modern and archaic humans

Genomic data manifest a rich legacy of interbreeding between divergent populations throughout human evolutionary history. Genomic evaluations of the relationship between modern and archaic humans have revealed the occurrence of such admixing events: admixture is defined as the interbreeding of individuals from two or more previously isolated populations. It results in gene flow, the transfer of genetic material from one population to another, that can be detected on the molecular level and used to understand the demographic history of the analyzed population.

The contribution of Neanderthals to the genomes of present-day humans has been thoroughly explored (Burbano et al., 2010; Green et al., 2010), demonstrating that all out-of-Africans share an amount of genetic material with Neanderthals. The proportion of Neanderthal-derived ancestry in people outside Africa is 1.5–2.1%. The introgression took place in a single pulse (Browning et al., 2018).

By comparing the Denisovan genome to the genomes of present-day humans, it became apparent that some ancestry segments of Denisovan origin exist in the genomes of present-day Near Oceanians, even though they are geographically far from the Altai region in which Denisova 3 was found.
The first investigation by Reich *et al.* in 2010 initially estimated that 4.5-5% of the genomes of Near Oceania inhabitants from Papua New Guinea and Bougainville Islands derive from Denisovans.
Subsequent extensive samplings of present-day human genomes from across Asia and Oceania revealed that other Indigenous Oceanian populations east of the Wallace’s line (a sea corridor that separated Asia from

Wallacea and New Guinea during the Pleistocene) as well as Indigenous populations from the Philippines, all display varying levels of Denisovan ancestry.

The estimation for Denisovan ancestry in Papuans was later refined to 3%. Further, Papuans and Indigenous Australians have virtually identical proportions of Denisovan ancestry, which indicates that the admixture event took place before the common ancestor of the two entered the Sahul continent (that is the name of the single mass of land that united New Guinea and Australia during the Pleistocene) at around 44,000 years ago (Meyer et al., 2012; Reich et al., 2011) and surely before 35,000 years ago (Malaspinas et al., 2016).

Many Philippine Negritos, such as the Agta, Ayta Magbukon and Mamanwa also carry 3–4% Denisovan ancestry ("Negritos" is the self-proclaimed label of several ethnic groups indigenous to the Andaman Islands, the Malaysian Peninsula, the Philippines and Thailand). According to Larena et al. 2021, Philippine Ayta Negritos exhibit the highest levels of Denisovan ancestry, having more Denisovan ancestry than Melanesians and Indigenous Australians (~4%)(Larena et al., 2021).

Other local populations have intermediate levels of Denisovan ancestry: present-day Taiwanese and Indonesian groups from Island Southeast Asia, possess Denisova ancestry as an indirect consequence of their history of East Asian and Oceanian admixture. Thus, their fraction of Denisovan ancestry is proportional to their fraction of Near Oceanian ancestry.

All East Asians and South Asians carry some amount of Denisovan ancestry, much less than that of Oceanians and South East Asians (Browning et al., 2018). However, some South Asians were found to carry an excess of Denisovan variants when compared to other Eurasians, a large portion of which are not shared with

East Asians or Papuans (Mallick et al., 2016; Sankararaman et al., 2016).
Siberians and Indigenous Americans also display some level of Denisovan ancestry associated with the East Asian one (G. S. Jacobs et al., 2019).
The varying levels of Denisovan-related DNA segments in present-day populations raised questions about the manner and scope of the interaction between archaic and modern humans, with noteworthy evidence for additional independent contacts with Denisovans accumulating in the past few years.

The unexpectedly higher proportion of Denisovan ancestry detected in South Asians groups hints to the fact that the ancestors of South Asians may have experienced different demographic and admixture events with archaics that gave them a high number of population-specific unique variants (Mallick et al., 2016; Sankararaman et al., 2016; Witt et al., 2022).

Browning et al. in 2018 detected in Asian genomes two components of Denisovan ancestry with differing similarity to the sequenced Denisovan genome: the component that has higher affinity to Denisova 3 from the Altai is predominantly present in East Asians and the component that is less similar to the Altai Denisovan is found in Papuans and South Asians, who on the contrary have very little of the former component. These can be interpreted as two different pulses of introgression from Denisovan lineages: some proportion of Denisovan ancestry present in contemporary populations was inherited from a second contact. The introgression event into East Asians was dated at around 48 thousand years ago (Zhang et al., 2021).

Jacobs et al. in 2019 found that the genomes of present-day Papuans comprise Denisovan ancestry from two divergent Denisovan lineages, called D1 and D2, that had been separated from each other and Denisova 3 for around 283 and 363 thousand years ago, respectively. According to their model, the introgressions from D2 and D1 into modern humans took place respectively 46 and 30 thousand years ago.
The authors also describe multiple events of introgression from the Denisovan lineage D1: after the split of an ancestral Papuan population in mainland New Guineans and Baining, a population of the Island of New Britain, subsequent events of introgression from D1 would have taken place in mainland Papuans. This would have happened after the split between said Papuan populations, that happened 16 thousand years ago. These results are compatible with a scenario in which one of the two Denisovan populations that introgressed was present East of Wallace's line, implying their technological ability to cross large bodies of water.

Dating the admixing events indicates that Denisovan populations were spanning Eastern Eurasia at least until about 50 thousand years ago. In support of this, Denisovan ancestry was detected in the genomes of a modern human from Tianyuan Cave in China, dated at 40 thousand years ago, and another modern human individual from the Salkhit Valley, in Mongolia, dated at 34 thousand years ago. For both, the admixture likely happened at least 10 thousand years before they lived (Fu, Meyer, et al., 2013; Massilani et al., 2020; Yang et al., 2017).

Some Southeast Asian Denisovan groups may have been around until as late as 16 thousand years ago (G. S. Jacobs et al., 2019). This is much later than the

extinction of Neanderthals that happened around 40 thousand years ago (Higham et al., 2014), and suggests that Denisovan-related populations may have coexisted with modern humans in New Guinea for tens of thousands of years.

1.6 Aftermath of the Denisovan introgression in modern humans

The events of admixture between archaic and modern humans introduced novel variation into modern human genomes. These archaic genetic variants were shaped by demographic and selective forces: some Denisovan introgressed fragments were favored by natural selection and nowadays are found at remarkably high frequencies in some present-day populations.

The Denisovan allele of the EPAS1 genes is perhaps the most suggestive example of adaptive introgression from an archaic into a modern human genome. The gene is involved in the physiological response to low partial pressures of oxygen in the atmosphere at higher altitudes (Beall et al., 2010; Simonson et al., 2010; Yi et al., 2010)**.**

The Denisovan haplotype of the EPAS1 gene, inherited from a population of Denisovans closely related to Denisova 3, confers Tibetans their peculiar high-altitude adaptation (Huerta-Sánchez et al., 2014; Zhang et al., 2021). The haplotype is found at high frequency (80%) in Tibetans, but is almost absent in the closely related Han Chinese. This increase in frequency was due to positive selection on standing variation: positive selection raised the frequency of an haplotype that was already present in modern humans, but became advantageous only after the colonization of a high altitude environment by the ancestors of present-day Tibetans, about 14 thousand years ago (Brantingham & Xing, 2006).

Other examples of Denisovan gene variants at high frequency in modern human populations are associated with metabolism-related genes and with pathways for the regulation of innate and adaptive immune responses (G. S. Jacobs et al., 2019; Sankararaman et al., 2016).

Apart from these exceptional examples, Denisovan ancestry much like Neanderthal ancestry has been overall deleterious to the modern human genome, which is made apparent by its depletion near genes. This implies that negative selection acted to purify the archaic variants from coding and regulatory portions of our genome. The reduction is particularly conspicuous on chromosome X and near genes expressed in testes (Petr et al., 2019; Sankararaman et al., 2016; Telis et al., 2020).
Furthermore, in the genomes of present-day humans there are some extended regions that are completely depleted from any archaic human ancestry and are hence called ‘archaic deserts’. These include modern human-specific traits that presumably were incompatible with archaic human introgressed variants that have been purged by purifying selection. The desert on chromosome 7 contains the FOXP2 gene that has been hypothesized to have a role in enabling modern human speech and language (Fisher & Scharff, 2009; Krause et al., 2007; Lai et al., 2001; Maricic et al., 2013; Vargha-Khadem et al., 1995).

1.7 Objective of the thesis

The diversity of segments of Denisovan origin in contemporary human genomes hints at the presence of geographic and, therefore, population substructure among Denisovans. In this framework, my thesis aims to retrieve such introgressed genetic material left by admixture events within a panel of pan-Eurasian

populations, to gain insights into the genetic diversity of the archaic human's populations.
The level of genetic heterogeneity of Denisovans suggests that they occupied a large portion of Eastern Eurasia, differentiating into several groups that admixed in various instances with ancient modern humans. The complex nature of the contact and the limited fossil record available so far both suggest that the archaic human occupied diverse environments, spanning from temperate steppes, to the Tibetan plateau, to tropical equatorial islands. The extent to which ecological barriers might have driven the divergence among Denisovan groups raises particularly interesting questions: Denisovans may have been capable of exploring across long distances, even surpassing remarkable geographical barriers such as Wallace's line. This seems plausible, as there is clear cut evidence of other hominins' dispersals to the area of Island SouthEast Asia, with respect to the finding of *Homo floresiensis*, discovered on the Island of Flores, Indonesia (P. Brown et al., 2004). Additionally, the peculiar excess in Denisovan ancestry observed in South Asian groups offers an interesting perspective of Denisovans being potentially capable of colonizing the secluded Indian subcontinent.
An interesting research question regarding Denisovan genomic diversity could be the following: it is known from the only sequenced Denisovan genome that Denisovans that occupied the Altai region had low levels of heterozygosity, much like Neanderthals; but what about the heterozygosity of Denisovans in other portions of the world, such as Southeast Asia? Were these Denisovans more heterozygous and thus perhaps more diverse?
Analyzing from a molecular viewpoint the complex demographic scenario of Denisovans also sheds light into our evolutionary history as humans, as together

with Neanderthals they are our closest relatives. Furthermore, the Denisovan component excavated from the genome of present-day Eurasians and Oceanians can be seen as a dye, useful to track and reconstruct the palaeolithic routes followed by these ancient human populations along their expansion out of Africa to colonize the whole world.

2. Materials and methods

2.1 Genomic data

The genomic data used for this Master Thesis is composed by high coverage (>30x) whole human genomes from the following present-day human panels: the 1000 Genomes Project (1000 Genomes Project Consortium et al., 2015), the Human Genome Diversity Panel (Bergström et al., 2020), and the GenomeAsia 100K (GenomeAsia100K Consortium, 2019).
All available individuals, with a cap to 50 for each, of 20 present-day populations were chosen, for a total of 772 pan-Eurasian samples. A Yoruban sample (NA18486 from the 1000 Genomes Project) was included as an outgroup to the Eurasian human diversity. The samples used in the genomic investigation are colored by their broad region of provenance in Table 2 and their global geographic distribution can be seen in Figure 1.
All the samples used were statistically phased by Dr. Leonardo Vallini using Shapeit 4 (Delaneau et al., 2019) to be able to discriminate the sequence of alleles of the two parental sources, obtaining a panel of 1544 statistically phased haploid samples.

Figure 1: Map of the world, showing the provenance of the populations used as samples for the modern human genomes and the location of Denisova Cave for the archaic genomes. Populations are colored in association to the broad geographic provenance: in blue Southeast Asia, in red South Asia, in green East Asia, in yellow America, in purple Africa, in black the Denisova Cave.
The map was realized with Google Maps (2024) Genome samples, 1:1000. Available at: https://www.google.com/maps/d/editmid=1pe4QqtN4F3KvEFFsKLvIjwwxJxo751U&usp=drive_link [Accessed on 23/08/2024].

Table 2: panel of modern human populations with label, panel of origin and number of samples for each. The populations have been colored according to their geographic region of provenance: in blue Southeast Asia, in red South Asia, in green East Asia, in yellow America, in purple Africa.

Population	**Label**	**Panel**	**N° of samples**
Bengali in Bangladesh	BEB	1KGP	50
Dai Chinese in Xishuangbanna, China	CDX		50
Han Chinese in Beijing, China	CHB		50
Han Chinese South	CHS		50
Colombian in Medellín, Colombia	CLM		50
Gujarati Indians in Houston, Texas, USA	GIH		50
Indian Telugu in the United Kingdom	ITU		50
Mexican Ancestry in Los Angeles, California, USA	MXL		50
Japanese in Tokyo, Japan	JPT		50
Kinh in Ho Chi Minh City, Vietnam	KHV		50
Peruvian in Lima, Peru	PEL		50

Punjabi in Lahore, Pakistan	PJL		50
Puerto Rican in Puerto Rico	PUR		50
Sri Lankan Tamil in the United Kingdom	STU		50
Yoruba in Ibadan, Nigeria	YRI		1
Papuan Highlands in New Guinea	PapuanHighlands	HGDP	9
Papuan Sepik in New Guinea	PapuanSepik		8
Papua New Guineans	Papuan	GA100K	4
Onge in Andaman Islands	Onge		11
Aeta in Philippines	Aeta		29
Paniya in India	Paniya		11

Genomic data for the archaic Altai Neanderthal and Denisova genomes was downloaded from the repository of the Max Planck Institute for Evolutionary Anthropology (http://cdna.eva.mpg.de/).
Since assemblies did not correspond between the present-day and archaic human genomes, the archaic data had to undergo the conversion of genomic coordinates from build 37 (hg19) to build 38 (GRCh38), to match the corresponding coordinates between the two editions, in a process known as liftover. The liftover was performed by Dr. Leonardo Vallini.
For both present-day and archaic humans, only data for the autosomes was included in the analyses.

Most of the computational analyses were performed on the <<University of Padova Strategic Research Infrastructure Grant 2017: "CAPRI: Calcolo ad Alte Prestazioni per la Ricerca e l'Innovazione">> cluster.

The following softwares were used for the processing of the genomic data:
Bcftools (Li, 2011) downloaded from https://github.com/samtools/bcftools
Plink (Chang et al., 2015) downloaded from https://www.cog-genomics.org/plink/1.9/
R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. downloaded from https://www.R-project.org/. - R version 4.4.0 (2024-04-24).

2.2 IBDmix

The IBDmix software (L. Chen et al., 2020) implements a probabilistic method to detect hominin introgressed material in present-day human genotype data. Unlike other similar methods, it does not require the use of an unadmixed human reference genome. IBDmix is developed based on Identity By Descent (IBD), the principle that an identical haplotype is shared by two individuals and inherited from a common ancestor. At each position, IBDmix estimates the probability of identity by descent between the archaic and modern sample based on allele frequencies and summarizes this as a LOD score. A LOD (short for “Logarithm Of the Odds”) score is a statistical estimate of the relative probability that two loci are located near each other on a chromosome and are therefore likely to be inherited together (https://www.genome.gov/genetics-glossary).
In order to identify the putatively introgressed archaic segments in the modern genome, the first step is a conversion using the generate_gt executable, to create the format-converted genotype data for the archaic and modern genomes that IBDmix requires as input. Then, the ibdmix executable applies a scanning dynamic programming algorithm to maximize the sum of LOD scores across a region above a pre-set threshold. The software proceeds site-by-site on the pair of archaic and modern human samples and sums together single-site LOD scores across a region. Variants are added consecutively to calculate the sum of the LOD scores, until the sum of the LOD scores becomes a negative value. The putative introgressed segments are the ones that maximize the sum of LOD scores above a pre-set threshold. The region with the maximized LOD score (above the preset LOD threshold) is defined as a putative introgressed segment in the modern

individual. Scanning restarts from the next variant after the putative introgressed segment. The output from IBDmix is a list of putatively introgressed segments and the probability of IBD between the archaic and modern human sample summarized as a maximized LOD score.
IBDmix was run with default parameters on the present-day genomes specified in Table 2, using the Denisovan genome as the genetic variant data of the reference archaic population.

2.3 The sstar matchrate package for S*

The S* statistic was developed to search for patterns of archaic admixture in human genomic data. It was introduced to detect patterns of variation and linkage expected in the case of introgression, by identifying which SNPs are the most likely to have mutated in a putative archaic population (Plagnol & Wall, 2006). The sstar matchrate package (Huang et al., 2022) enables efficiency and reproducibility when using S* for detecting introgression. The package allows to estimate a so-called “source match rate” for putative archaic SNPs; it computes the similarity of a certain segment of DNA from a target population (the putative Denisova segments detected in genomic data from the pan-Eurasian human panel) to a source population (Denisova or Neanderthal), using as reference an un-admixed population (the Yoruban sample). It refines the analysis by comparing significant S* haplotypes to the source genome and testing to determine whether they match it more than expected by chance.
The sstar matchrate package was run on the previously masked genomes, using as target the human populations in Table 2 and as source the Denisovan, Neanderthal and Yoruban genome, to compute respectively the source match rates to the two archaics and to the African human.

2.4 Data filtering and visualization

To filter the information, only DNA segments that met some defined criteria were chosen to be retained in the following steps: a match rate to both of the archaics over 0.2 was taken as baseline, to account for incomplete lineage sorting between archaic and modern humans.

The match rates of all target individuals to Denisova, Neanderthal and Yoruba were visualized with three dimensional scatter plots divided by population, using the scatterplot3d function from the {scatterplot3d} package in R (Figure 4a). Additionally, a rotating visualization of the 3d scatter plot for the population labelled as “Papuan” is available by scanning the QR code in Figure 4c. The QR was made with QR.io (2024). Available at: https://qr.io/ [Accessed on 24/08/2024].

To better visualize the information, correlograms for the three variables were plotted using the ggpairs function from the {GGally} package in R (Figure 4b).

2.5 Binning of the matchrate scores via pairwise correlation

Binning of the matchrate scores was performed using a perl script that takes as input sstar matchrate files and outputs a square matrix of pairwise correlation coefficients between all individuals based on binned density arrays, a measure of pairwise similarity between individuals with respect to their affinity with the reference Denisova genome. Matrices were obtained using all the combinations of the following parameters: 10 bins; match rates to Neanderthal <0.3, <0.4 and <0.5; regions matching to the Denisovan genome at least 100kb or 500kb long.

2.6 Multidimensional Scaling analysis of data from the binned matrices

The binned match rate data matrices were converted to distance matrices with the dist command from the {stats} package in R and a Multidimensional Scaling was performed using the cmdscale command from the same package. Centroids, computed as the median of x and y coordinates of all the samples from each population, were added and then color coded to their broad region of provenance.

2.7 Masking of the genomes

Once the putatively archaic ancestry has been identified, from the whole genome one can choose to reduce the size of the genetic data by only keeping the introgressed regions, as these are the target of the analysis. This process, referred to as “masking” of the genomes, creates files with genetic information only for the relevant portions of the DNA sequences, whereas the other sites are set as “missing data”. The masking was performed using a Perl script developed by Professor Luca Pagani. It requires as inputs the modern human genomic data to be masked and the regions detected as putatively archaic by IBDmix. It allows the choice of an exclusion threshold for the IBDmix LOD score, which in this case was set to 3. It also needs an outgroup, the Yoruban sample from the 1000 Genomes Project (NA18486), to keep as unmasked whole genome data for further analyses, and that the masking process will skip altogether. The script masks out the non-introgressed segments within the dataset, setting them as missing, and keeps only the putatively archaic portions. It outputs new genetic variant files,

containing only the data in the putatively archaic regions.
To individuate the haploid segments that match primarily to the Denisovan genome and to reduce the impact of the Neanderthal component as a confounder, a threshold was introduced. This quantity, for each DNA segment, is the ratio of the matchrate to Denisova divided by the sum of the matchrate to Denisova plus the matchrate to Neanderthal. The DNA segments of Denisovan ancestry that showed a value of this threshold over 0.75 were deemed as relevant. Of these, those that also displayed a matchrate to Yoruba smaller than the matchrate to Denisova for that segment were kept for the following analyses.

2.8 *f*-statistics

2.8.1 *f*3 statistic

The '*f3* statistic' is a 3-population generalization of F_{ST}. This statistic is equal to the inner product of the frequency differences between a group that is to be distinguished from two other interchangeable groups. The *f*3 is proportional to the correlated genetic drift between the considered groups and it was first introduced to test for admixture in populations (Reich et al., 2009). Raghavan *et al.*, 2013 proposed an interpretation of the *f*3, referred to as 'outgroup' *f*3 (Raghavan et al., 2013). For an unknown population P_U, they wanted to find the most closely related population from a panel of *k* extant populations. They did this by calculating $f3(P_O; P_U, P_i)$, where P_O is an outgroup population that was assumed widely diverged from P_U and all populations in the panel. This measures the shared drift (or shared branch) of P_U with the populations from the panel, and high f_3 values imply

close relatedness. The *f*3 statistic *sensu* Raghavan et al., 2013 was applied to the panel of pan-Eurasian samples for their inferred Denisova genomic segments, to investigate the evolutionary history that these individuals share, and the higher the value of the *f3,* the more evolutionary history the two individuals share as far as their Denisovan past is concerned.

*f*3(population1;population2;outgroup) were computed for all permutations of the following populations, considered of major interest: CHB, ITU, KHV, MXL, PUR, PapuanHighlands, PapuanSepik, Papuan, Onge, Aeta, Paniya. The outgroup is the Yoruban sample.

2.8.2 Multidimensional Scaling analysis from *f*3 data

Data from the *f*3 statistic was arranged in a distance matrix using the list2dist command from the {spaa} package in R. A Multidimensional Scaling was performed using the cmdscale command from the {stats} package.
Populations were represented in a plot in which they are color coded according to their broad region of provenance.

2.8.3 *f4* statistic

The '*f*4 statistic' is a generalization of the Green *et al.*, 2010 D-statistic. It is used to assess whether an unrooted phylogenetic tree of populations is consistent with the SNP allele frequency data. It tests for "treeness": given an outgroup population D, and populations A, B, and C, if the topology (A,B)(C,D) is correct, then the frequency differences between A and B should reflect genetic drift that is uncorrelated with that between C and D. Thus, the expected value of the

product of frequency differences $(p_A - p_B)(p_C - p_D)$, which is referred to as $f4(p_A,p_B,p_C,p_D)$, is zero. On the contrary, if the *f4* value is significantly different than 0, there is a deviation from the (A,B)(C,D) tree assumption. A significantly positive value of *f*4 indicates more shared genetic history between populations A,C or B,D, whereas a significantly negative value of *f*4 indicates closeness between populations B,C or A,D.

*f*4s were computed for target population (X), Neanderthal (Nea), Denisova (Den), Yoruba (YRI), Han Chinese (Han), Ancestral (Anc) in the following permutations: *f*4(X,Nea,Den,Anc); *f*4(X,Den,YRI,Anc); *f*4(Den,YRI,X,Anc); *f*4(Den,X,Nea,Anc); *f*4(Nea,Den,X,Anc); *f*4(Han,X,YRI,Anc).

*f*4s were considered as significantly different than 0 with a |Z-score|>3.

f4(Han,Deni,YRI,Anc) was computed for a modern Han Chinese genome (NA18525 from 1KGP) and used as control. It was also used for the computation of the *f*4-ratio between *f*4(X,Deni,YRI,Anc) and *f4*(Han,Deni,YRI,Anc), which consists of the ratio between the former and the latter, as a mean to infer the proportion of modern human ancestry present in the genomic material in analysis.

All f-statistics were computed with the popstats.py Python script (Skoglund et al., 2015).

2.9 Evaluation of the expected population heterozygosity

Heterozygosity is defined as the proportion of heterozygous sites within an individual’s genome, which is informative about inbreeding and effective population size of their population of origin.

Average expected heterozygosity was assessed for all populations using allele frequencies for each SNP.

Frequencies for the minor and major allele were obtained using the --freq command in the PLINK software, which allows to estimate the allele frequencies. Modern human unfiltered data was used to correct for the influence of modern human demography that is reflected in the Denisovan portions of the genomes, as an effect of the evolutionary history they shared after the admixture event.

Frequency for the filtered Denisovan portions of the genome were obtained from the PLINK estimate.

Frequency for the minor allele in human unfiltered data was obtained after correcting for the presence of Denisova alleles. The correction was performed by comouting the ration of the difference of the two products of the minor allele frequency in modern human multiplied by the number of observed chromosomes in the modern human and the minor allele frequency of Denisova multiplied by the number of observed chromosomes in Denisova, divided by the difference of number of observed chromosomes:

$$FreqH = \frac{\left(\left(MAFH * NChrObsH\right) - \left(MAFD * NChrObsD\right)\right)}{\left(NChrObsH - NChrObsD\right)}$$

Consequently, the expected heterozygosity was computed as 2*p*(1-p) for every SNP in both the modern human unfiltered and Denisovan frequency data.

The average expected heterozygosity estimates for each population were represented and compared with plots using the plot function from the {graphics} R package.

The trend of expected heterozygosity represented in orange in Figure 8 was calculated to set a baseline for

comparing the populations. The heterozygosity of the putative Denisovan material was calculated for all populations from the heterozygosity of a chosen population (Chinese Han from Beijing), known to host Denisova segments with high affinity to the Denisova 3 reference genome (Browning et al., 2018), using the following formula:

$$HetD(pop) = \frac{HetD(Han) * HetH(pop)}{HetH(Han)}$$

The trend corrects for the history that the putatively Denisovan segments shared with the modern humans, after the introgression took place: the Denisovan segments contained in human genomes undoubtedly underwent modern human demographic history. If all of the segments isolated from different populations were to follow the same trend (in this case using the Han Chinese as baseline), all the Denisovan populations that introgressed would be similar.

2.10 Correlograms

The resulting data from all previous analyses were compared in a correlogram of the total information available for all of the populations, plotted using the ggpairs function from the {GGally} package in R.

2.11 Additional analyses run during my internship

During my internship, I briefly took part in the UMFOLD project by D. Marnetto, R. Zeloni, D. Fusco, F. Pizzagalli, P. Provero (all from the University of Torino), professor L. Pagani and myself. The overall aim of the project is to assess Neanderthal's impact on present-day human brain morphology: the correlation of Neanderthal-inherited variants to specific brain phenotypes is analyzed through MRI phenotypic data of the UK BioBank database. My role in the project consisted in the extraction of the segments of putative Neanderthal ancestry, using the IBDmix software, from the samples with an MRI phenotype (about 50 thousand individuals).

3. Results

3.1 Qualitative comparison through the sstar matchrates

Using the IBDmix software, 17,989,895 putatively Denisovan regions across the 22 autosomal chromosomes were individuated in the genomes of the selected present-day populations (Table 2).
The original pan-Eurasian genomes were masked according to the output of IBDmix, to keep only the regions of putatively archaic origin.
After sifting out the segments with less than 0.2 as a match rate to both of the archaics to account for incomplete lineage sorting, the match rates of the fragments of DNA to the archaics' and to the Yoruban genome were represented with three dimensional scatter plots and corresponding correlograms. The plots are here reported only for the Papuan population, chosen as representative of all the others, which in turn are reported in the appendix (Appendix 1).
The QR code links to a rotating version of the 3d scatterplot for the Papuan population.

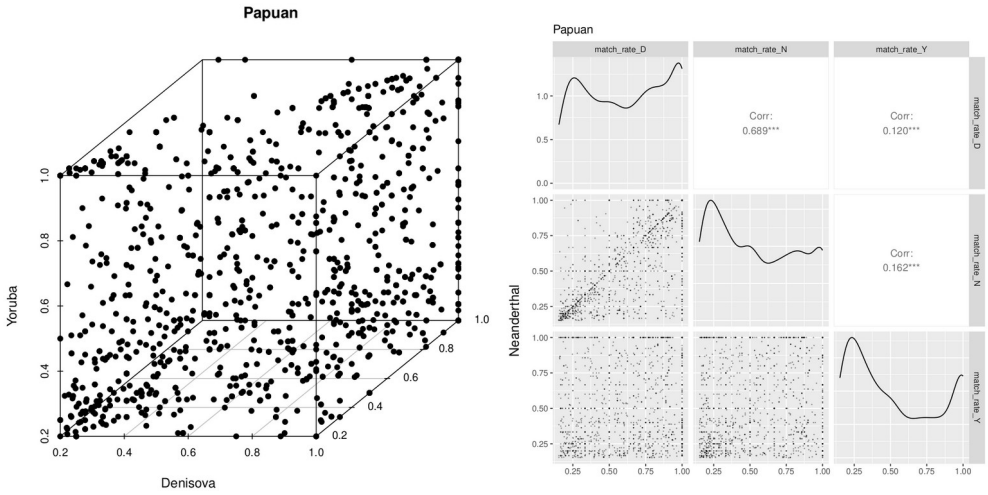


Figure 4. 3d (a) and 2d (b) scatter plots of the match rates to Denisova, Neanderthal and Yoruba for Papua New Guineans*. Panel c shows a QR code linking to the rotating 3d scatter plot for the Papuan population.*

Regarding the three dimensional scatterplot (Figure 4a), on the x, y and z axes are shown respectively the matchrate to Denisova, Neanderthal and Yoruba.
The correlogram on the right (Figure 4b) shows three sets of plots: on the left part of the figure there are scatterplots for each pair of match rates compared. It is more straightforward to analyze the correlogram, rather than the three dimensional comparison of the match rates, as these scatterplots on the left can be seen as the two-dimensional projections of the faces of the three-dimensional scatterplot cube. On the top right panels are displayed the Pearson correlations between

pairs of match rates and the plots on the diagonal represent the distribution of the match rates.

The panel that shows the comparison between the match rates to Denisova and Neanderthal can be seen as divided into two triangles by its diagonal (fourth panel of Figure 4b, counting from the left): the lower half triangle of points, in which the match rate to Denisova is higher, contains the majority of samples, showing that the extraction of Denisovan ancestry segments was successful in effectively singling out material of putative Denisovan ancestry. *Vice versa*, the upper half triangle, in which the Neanderthal contribution is higher, is mostly devoid of samples. This provides a qualitative and preliminary confirmation to the fact that the segments extracted are not of broadly speaking archaic ancestry, but for the most part belong to probable Denisovan ancestry. The subsequent filtering of the match rates data will allow to retain only the segments that have a higher probability of being of Denisovan origin.

3.2 Multidimensional Scaling analysis of data from the binned matrices

The Multidimensional Scaling analysis of binned match rate data shows different clusters of populations.

For clarity purposes, the plots generated using 10 bins, regions matching to the Denisovan genome at least 100 kilo bases long and matchrate to Neanderthal <0.3 are shown for the first three dimensions, and the plots for the other combinations of parameters are available in the appendix (Appendix 2).

East Asian populations (Japanese, Chinese South, Chinese Dai and Chinese Han, represented in green hues in the plot) form a cluster that also includes the Vietnamese Kihn from Ho Chi Minh City.
The other South East Asian populations, represented in hues of blue, are dispersed across the space of the plot. PapuanSepik and PapuanHighlands are close, but distinguished between each other, with PapuanSepik being closer to the population labelled as “Papuan” and the populations from South Asia.
South Asian populations (Bengali, Gujarati, Paniya, Punjabi, Telugu, Sri Lankan Tamil, represented in shades of red) form a cluster in an intermediate position with respect to all other populations, apart from Onge that occupies a space of the plot distinct from all other populations.
Onge and Aeta are two Negritos groups, respectively from the Andaman Islands (a South Asian population, as such represented in a shade of red) and from Luzon Island in the Philippines (a South East Asian population, represented in a shade of blue), that seem to occupy two polar opposite parts of the plot, each separated from the other clusters of populations.

Paniya are an indian population that lives in Kerala and Wayanad, India, but fall separately from all other South Asian populations.
American populations (Colombians, Mexicans, Puerto Ricans, in hues of yellow) create a group that partially overlaps with the South Asian cluster, but also extends towards the East Asian cluster with Peruvians from Lima.
Both sets of dimensions in Figure 5a and 5b show this disposition of the populations.

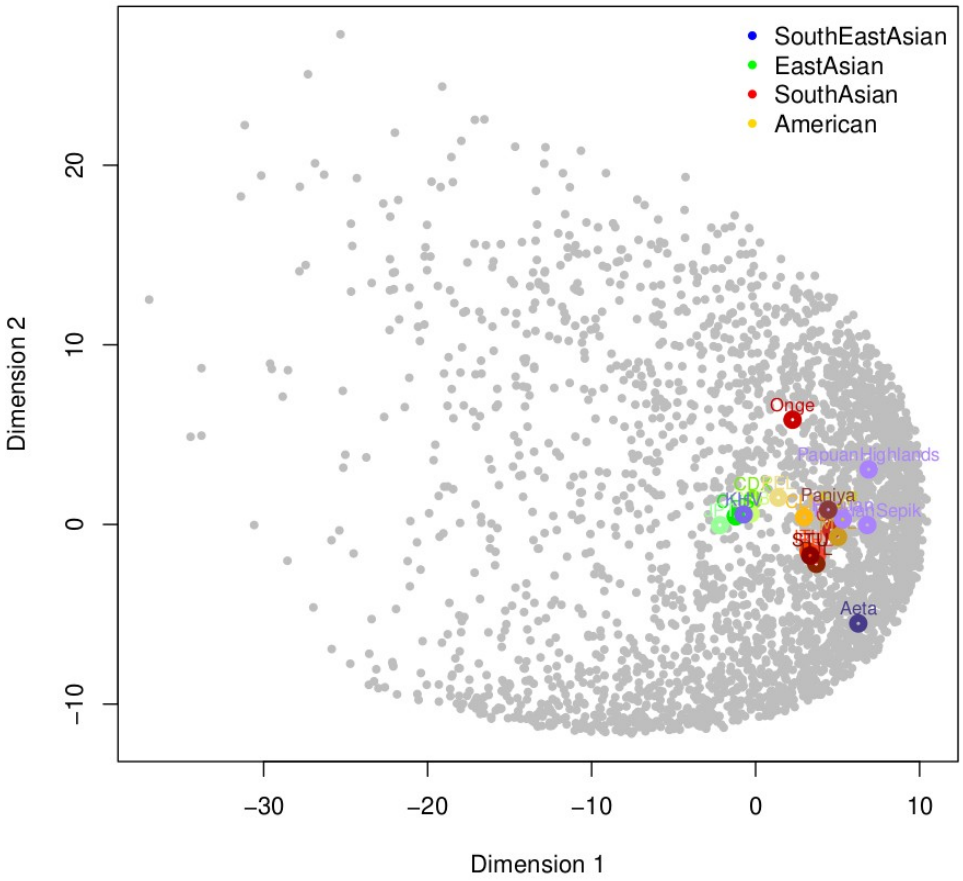


Figure 5a: Multidimensional Scatter Plots for all populations for the first and second dimensions. Populations' centroids are color coded with color hues according to the broad geographic region of provenance: blue for South East Asia, green for East Asia, red for South Asia, yellow for America.

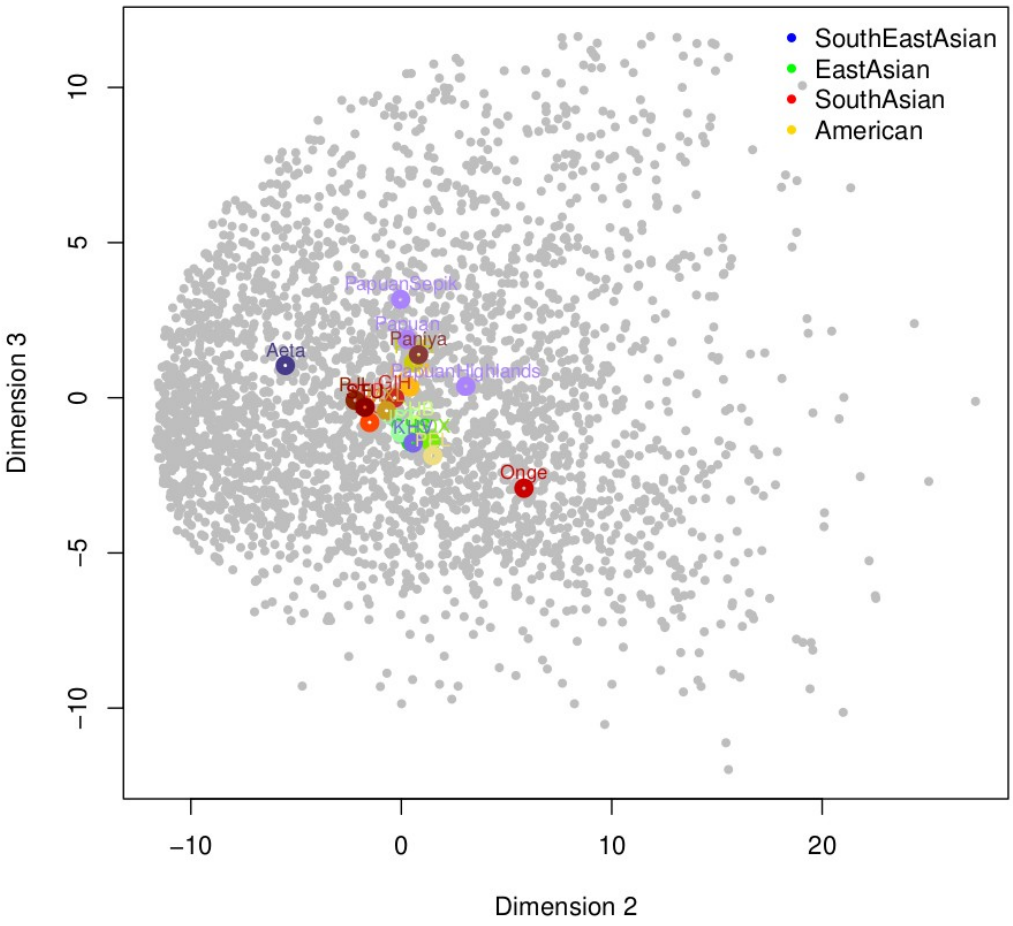


Figure 5b: Multidimensional Scatter Plots for all populations for the second and third dimensions. Populations' centroids are color coded with color hues according to the broad geographic region of provenance: blue for South East Asia, green for East Asia, red for South Asia, yellow for America.

3.3 Multidimensional scaling analysis of population f3 data

The Multidimensional Scaling analysis of the population *f3* shows different positioning of the segments of probable Denisovan origin extracted from different populations: the first dimension appears to be dominated by the distinction between the two populations ITU and KHV. The second dimension discriminates between the two populations Paniya and Onge and the aforementioned two groups. The third dimension seems to emphasize similarities between SouthEast and Oceanian groups against the rest of the populations.

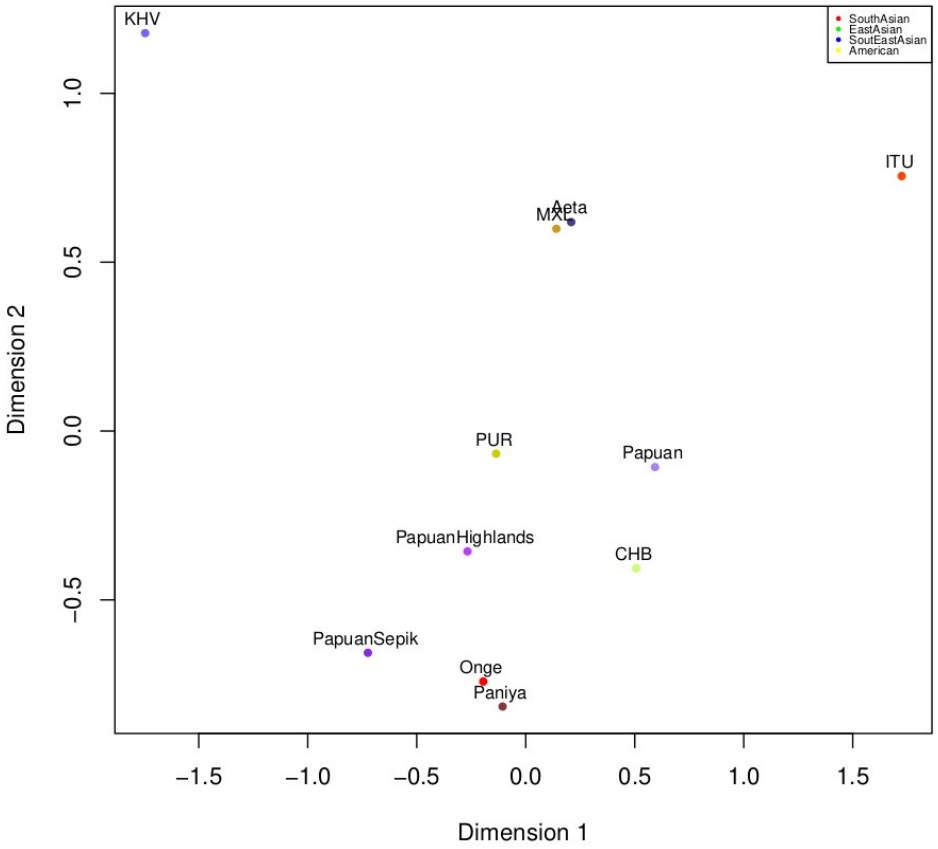


Figure 6a: Multidimensional scaling plots for a chosen subset of populations (CHB, ITU, KHV, MXL, PUR, PapuanHighlands, PapuanSepik, Papuan, Onge, Aeta, Paniya) for the first and second dimensions. Populations are color coded with color hues according to the broad geographic region of provenance: blue for South East Asia, green for East Asia, red for South Asia, yellow for America.

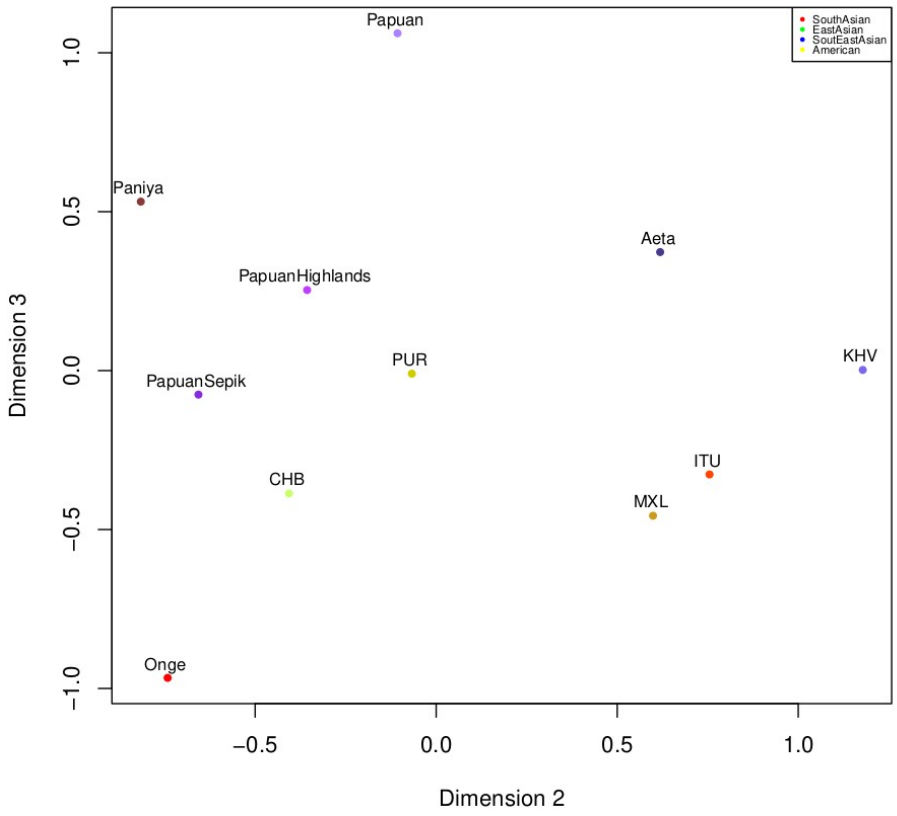


Figure 6b: Multidimensional scaling plots for a chosen subset of populations (CHB, ITU, KHV, MXL, PUR, PapuanHighlands, PapuanSepik, Papuan, Onge, Aeta, Paniya) for the second and third dimensions. Populations are color coded with color hues according to the broad geographic region of provenance: blue for South East Asia, green for East Asia, red for South Asia, yellow for America.

3.4 *f*4 statistic

In order to to understand whether the extracted DNA segments were predominantly of Denisovan origin and virtually free from Neanderthal and modern human confounders, *f*4 statistics as *f*4(p_A,p_B,p_C,p_D) were compiled for target population (X), Neanderthal (Nea), Denisova (Den), Yoruba (YRI), Han Chinese (Han), using Ancestral (Anc) as outgroup.

The *f*4s between the two archaics and the target population suggested that their genomic regions may be lacking Neanderthal confounders. From Table 3, it is noticeable how all of the segments have a propensity towards the Denisovan origin of the genomic material, suggesting probable Denisovan ancestry for these DNA segments.

Likewise, the *f*4s that show the comparison of the extracted putatively Denisovan material for each population compared to modern human material confirmed that the genomic data is scarce of modern human confounders, as can be seen in Table 4. Nevertheless, there is a certain pull towards the modern human genomic material for some populations: in particular, the putatively Denisovan material extracted from Aeta genomes appears to have some level of human confounder, as can be seen from *f*4(X,Deni,YRI,Anc).

Table 3: f4 and Z-score values for the f4 test between target population (X), Neanderthal (Nea), Denisova (Den), and Ancestral (Anc).

Population	*f4*(Deni,X,Nea,Anc)		*f4*(Nea,Deni,X,Anc)		*f4*(X,Nea,Deni,Anc)	
	f4	Z-score	*f4*	Z-score	*f4*	Z-score
BEB	0,0060	2,6915	-0,0753	-16,0997	0,0693	13,3331
CDX	0,0055	1,5816	0,0985	-10,3280	0,0929	9,5228
CHB	0,0037	0,9515	-0,0849	-12,6324	0,0811	10,1964
CHS	0,0074	2,61001	-0,0819	-12,7654	0,0744	10,6028
CLM	0,0066	3,1968	-0,0675	-13,8144	0,0608	10,7567
GIH	0,0099	3,0668	-0,0714	-11,5305	0,0614	8,5767
ITU	0,0044	1,2574	-0,0851	-9,2181	0,0806	8,4798
MXL	0,0052	3,3485	-0,0695	-22,4910	0,0643	19,0059
JPT	0,0064	2,4639	-0,0879	-14,5486	0,0814	11,8737
KHV	0,0065	1,7175	-0,0774	-8,8709	0,0708	7,2822
PEL	0,0045	2,4869	-0,0629	-15,5099	0,0584	12,9728
PJL	0,0016	0,7236	-0,0644	-11,9693	0,0628	11,6187
PUR	-0,0037	-0,6808	-0,0526	-6,0232	0,0563	6,6357
STU	0,0027	1,3132	-0,0799	-12,8525	0,0771	12,1452
PapuanHighlands	-0,0006	-0,5334	-0,0505	-18,4379	0,0512	19,4179
PapuanSepik	0,0050	6,306	-0,0526	-28,1946	0,0476	23,0480
Papuan	0,0045	5,5814	-0,0507	-23,1443	0,0461	18,9220
Onge	0,0045	4,2915	-0,0669	-22,4880	0,0624	19,2315
Aeta	0,0023	3,1774	-0,044	-29,1332	0,0418	28,3870
Paniya	0,0027	1,8103	-0,0651	-19,5148	0,0623	17,6380

Table 4: f4 and Z-score values for the f4 test between target population (X), Denisova (Den), Yoruba (YRI), Han Chinese (Han), and Ancestral (Anc).

Population	*f4*(Deni,YRI,X,Anc)		*f4*(X,Deni,YRI,Anc)		*f4*(Han,X,YRI,Anc)	
	f4	Z-score	*f4*	Z-score	*f4*	Z-score
BEB	0,0579	12,5225	-0,0017	-0,8546	0,0260	5,5582
CDX	0,0794	6,8133	0,0001	0,0312	0,0237	3,2865
CHB	0,0676	9,3767	-0,0019	-0,5124	0,0242	3,7694
CHS	0,0691	9,1996	-0,0018	-0,6445	0,0284	5,1354
CLM	0,0691	13,0476	-0,00491	-2,4406	0,0314	6,5944
GIH	0,0621	10,1405	-0,0061	-1,8898	0,0420	7,1728
ITU	0,0640	7,3268	-0,0017	-0,4917	0,0404	5,1465
MXL	0,0559	15,9807	0,0004	0,3201	0,0281	9,6979
JPT	0,0758	10,6378	-0,0014	-0,4740	0,0344	4,9231
KHV	0,0662	7,6565	-0,0058	-1,3228	0,0341	4,6232
PEL	0,0542	12,1433	-0,0018	-0,9271	0,0274	7,5503
PJL	0,0578	10,3709	-0,0017	-0,7201	0,0391	7,8160
PUR	0,0502	5,6967	-0,0029	-0,5086	0,0267	3,8516
STU	0,0636	10,1373	-0,0017	-0,7266	0,0306	4,8263
PapuanHighlands	0,0398	14,8337	0,0036	2,5430	0,0222	9,0990
PapuanSepik	0,0426	21,3004	0,0020	2,6868	0,0206	11,5211
Papuan	0,0403	17,1282	0,0042	4,6298	0,0217	10,8396
Onge	0,0592	18,2802	-0,0002	-0,2804	0,0333	10,5497
Aeta	0,0314	19,2881	0,00630	7,0502	0,0193	15,0587
Paniya	0,0569	16,40210	-0,0015	-0,9416	0,0281	9,7312

Table 5: f4 and Z-score values for the f4 test between Han Chinese (Han), Denisova (Den), Yoruba (YRI), and Ancestral (Anc).

f4(Han,Deni,YRI,Anc)	
f4	Z-score
0.01970	24.0691

The *f4*(Han,Deni,YRI,Anc) (Table 5) was used as control and did confirm that the modern human Han Chinese genomic material has an affinity to the truly human Yoruban genomic material. This data was then used in the following computation of the *f4*-ratio between *f*4(X,Deni,YRI,Anc) and *f4*(Han,Deni,YRI,Anc). The *f*4-ratio yielded the proportion of modern Eurasian material that acts as contaminant in the selected predominantly Denisovan segments (Table 6). Putative Denisovan segments that were extracted from Aeta genomes are those that appear to have the highest contribution in modern human contaminants (30% human), followed by the population labelled as Papuan (21%).

Nevertheless, it is worth noting that all of the genomic material extracted has been enriched of several-fold in probable “Denisovaness”, if compared to the overall percentage of Denisovan ancestry at the whole genome level, which for all populations does not surpass the 4% limit: these extracted segments appear to be 70% and over of Denisovan origin, which is a noticeable enrichment.

Table 6: f4-ratio between f4(X,Deni,YRI,Anc) and f4(Han,Deni,YRI,Anc) for all populations.

population	*f4*(X,Deni,YRI,Anc)/*f4*(Han,Deni,YRI,Anc)
BEB	-0,0904
CDX	0,0056
CHB	-0,0976
CHS	-0,0914
CLM	-0,2536
GIH	-0,3115
ITU	-0,0869
MXL	0,0251
JPT	-0,0735
KHV	-0,2981
PEL	-0,0929
PJL	-0,0867
PUR	-0,1486
STU	-0,0907
PapuanHighlands	0,1864
PapuanSepik	0,1052
Papuan	0,2173
Onge	-0,0149
Aeta	0,3197
Paniya	-0,0809

3.5 Evaluation of the expected population heterozygosity

Heterozygosity can be defined as the fraction of sites within an individual that differ between the chromosomes they inherit from their parents. It allows us to make inferences on genetic variation within the population of origin of the sampled individual, as it informs about inbreeding and effective population size. Heterozygosity can be evaluated at the population level as the average of the individual values of heterozygosity of the individuals of a population.

The expected heterozygosity of the populations is shown in Figure 7, split between Denisovan genomic information and modern human genomic information computed for the regions extracted as putatively Denisovan. The heterozygosity is plotted against the number of SNPs extracted for that specific population. The expected heterozygosity obtained from the data is lower for all of the Denisovan genetic material, with respect to the modern human counterpart, which always displays higher levels of expected heterozygosity. South East Asian populations such as the Papuan populations, the Aeta from the Philippines and Onge from the Andaman Islands have the lowest expected heterozygosity of all other modern human populations, according to their lower effective population size as insular, isolated populations. The other modern human populations display a higher amount of expected heterozygosity, making it so that for modern humans the data appears more dispersed along the axis of heterozygosity. The putatively Denisovan segments appear to be less dispersed along the x axis.

The three populations Aeta, PapuanHighlands and Papuan show the highest number of loci that got selected as putatively Denisovan in the extraction with the IBDmix software.

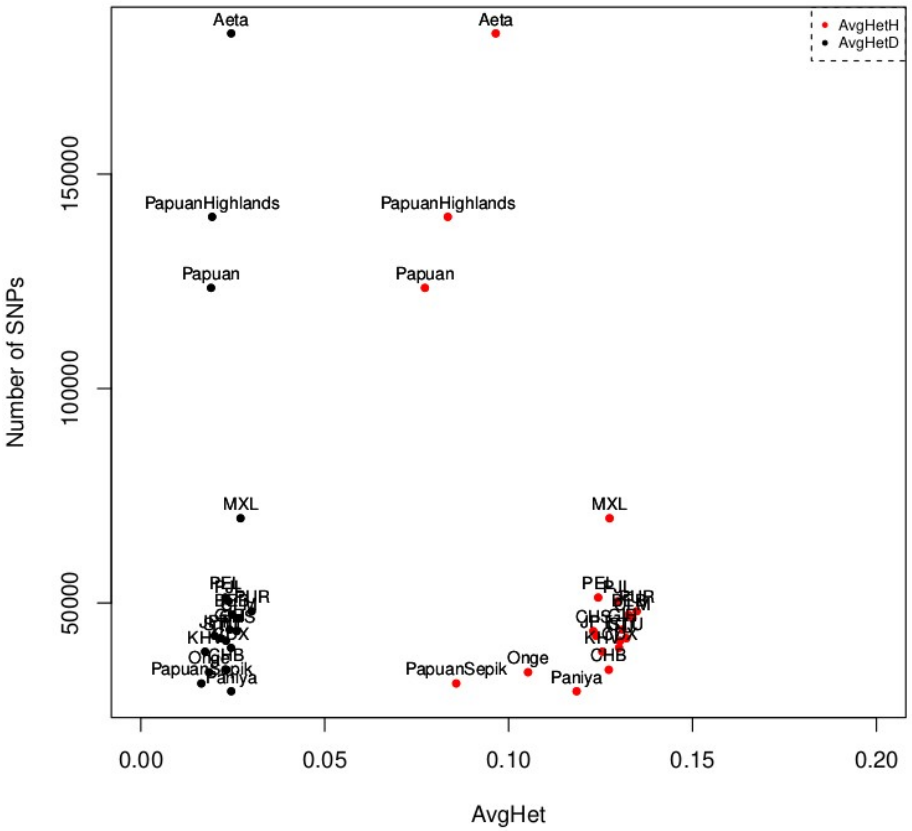


Figure 7: Plot of the average expected heterozygosity (AvgHet) for all populations represented over the number of SNPs, split between modern human (red dots) and Denisovan data (black dots).

Scattering the average expected heterozygosity of the Denisovan genomic material against the average expected heterozygosity of the modern human genomic material allows to distinguish populations on the basis of their heterozygosity (Figure 8). All populations in the axis of modern human heterozygosity fall in a much bigger range

(they are about four times more dispersed) than the Denisovan expected heterozygosity in the y axis.

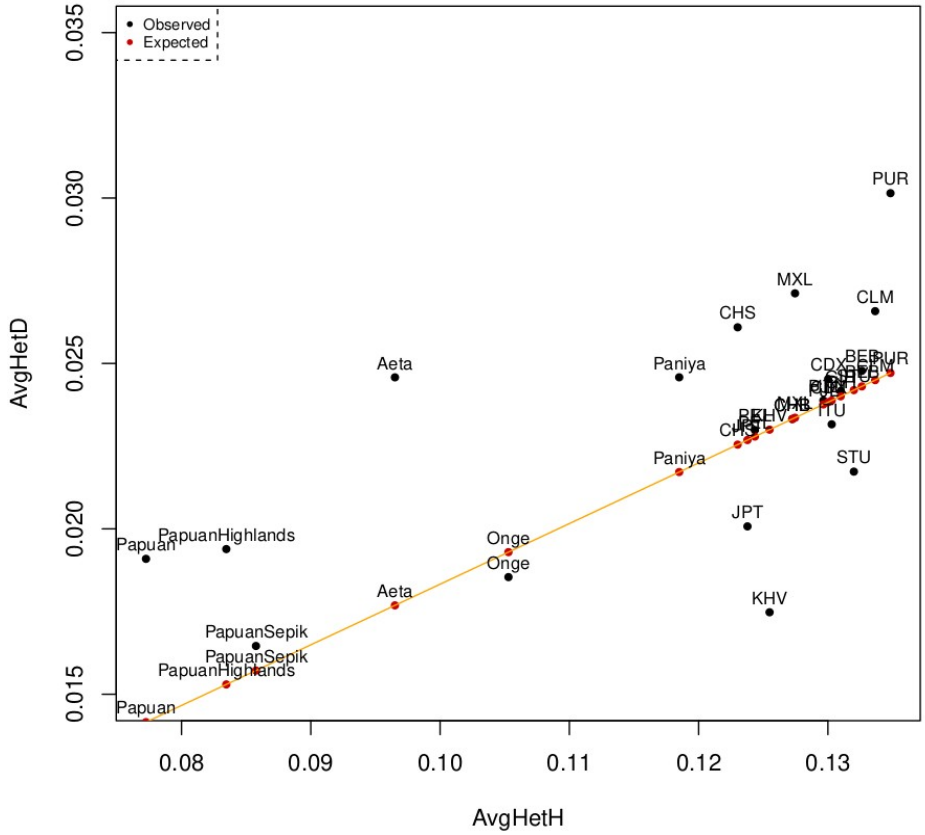


Figure 8: Plot of the average expected heterozygosity for all populations of modern human genomic material on the x axis and of Denisovan genomic material on the y axis. The black dots represent the expected heterozygosity obtained from the genomic data of this Master Thesis, whereas the red dots and the orange line mark the trend for CHB.

The trend of expected heterozygosity represented in orange in Figure 8 was calculated to set a baseline for comparing the populations to a chosen population (Chinese Han from Beijing). All putatively Denisovan segments appear to approximately follow the same trend, with a limited range of dispersal along the y axis: the

average expected heterozygosity for the putatively Denisovan segments is spread over a difference of about 0.01, whereas the expected modern human heterozygosity varies of about 0,05.

3.6 Correlograms

The correlogram in Figure 10 shows the correlation of all the various results from the previous analyses, used as variables in the plot.
The number of SNPs correlates positively with the *f*4s that control for human contaminants, *f4(Deni,YRI,X,Anc)* and *f4(X,Deni,YRI,Anc*) (Figure 10, line 1 with 9 and 11), implying that the populations that yielded the higher number of segments in the extraction are also those that contain some amount of probable human contaminant.
The first and third dimensions of the Multidimensional Scaling analysis performed on the binned data are correlated between each other and also with the *f*4s that control for modern human contaminants, whereas the second dimension does not appear to correlate with any of the previously mentioned *f*4s, but with the second dimension of the Multidimensional Scaling analysis performed on the *f*3 data. The first two dimensions of the Multidimensional Scaling analysis performed on the *f*3 data do not correlate to any other variable apart from the correlation of the second dimension from the binning with the second dimension from the *f*3 data. This can be seen as confirmation that any clustering or differentiation between populations in the first two dimensions of the MDS analysis performed on the *f*3 data and the second dimension of the MDS analysis performed on binned data is a differentiation between putative Denisovan material (and not between modern human contaminants).

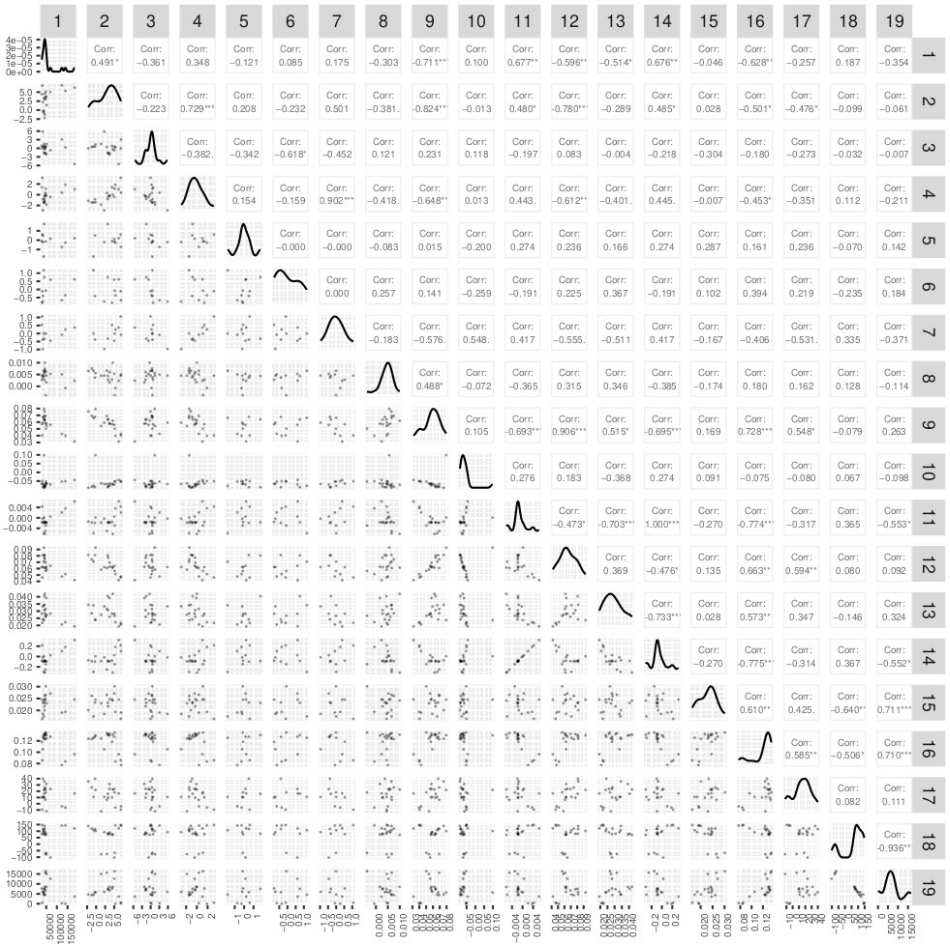


Figure 10: Correlogram of every available piece of information for all populations. The compared variables are labelled with serial numbers from 1 to 19 according to the following key: 1 - number of SNPs extracted; 2 - first dimension of the Multidimensional Scaling from binned data; 3 - second dimension of the Multidimensional Scaling from binned data; 4 - third dimension of the Multidimensional Scaling from binned data; 5 - first dimension of the Multidimensional Scaling from population f3 data; 5 - first dimension of the Multidimensional Scaling from population f3 data; 6 - second dimension of the Multidimensional Scaling from population f3 data; 7 - third dimension of the

Multidimensional Scaling from population f3 data; 8 - f4(Deni,X,Nea,Anc); 9 - f4(Deni,YRI,X,Anc); 10 - f4(Nea,Deni,X,Anc); 11 - f4(X,Deni,YRI,Anc); 12 - f4 (X,Nea,Deni,Anc); 13 - f4(Han,X,YRI,Anc); 14 - f4-ratio f4(X,Deni,YRI,Anc)/f4(Han,Deni,YRI,Anc); 15 - average expected heterozygosity for Denisovan genomic material; 16 - average expected heterozygosity for modern human genomic material; 17 - latitude; 18 - longitude; 19 - approximate distance from Papua New Guinea (-6.0620376,144.4342138).

4. Discussion

This Master Thesis evaluates the genetic diversity of the segments of Denisovan ancestry found within a panel of pan-Eurasian modern genomes as introgressed material left by admixture events. The retrieval was successful in isolating genomic material of archaic origin, manyfold enriching the samples in putatively Denisovan ancestry regions that could then be compared.

This Master Thesis is not exempt from assumptions that need to be taken into consideration when making inferences on Denisovan diversity. Firstly, the modern human genomes used for the investigation can only be proxies for the ancestors of these specific present-day humans, but are not representative of all humans that lived in the past and that first-hand experienced the events of introgression. Next, the softwares used for the extraction of the putatively archaic segments and for the calculation of the match rates both come with their own assumptions, which make it so that these can only be approximations of the reality but never a true representation. Additionally, there is the fundamental assumption that these are indeed Denisovan-related portions of the genome: as some of the extracted material appears to be still containing traces of modern human genomic material, one should use caution when making inference about Denisovan demographic history from these segments, reason for which the segments have been called as probable or putative Denisovan ancestries all throughout the Master Thesis. Finally, the genomic material sifted out comprises a limited number of SNPs, amounting to 17,989,895 putatively Denisovan regions across the 22 chromosomes.

4.1 Discussion of the key findings

In the first instance, the results of this Master Thesis show that the genomes of the modern human populations analyzed do encompass a certain amount of archaic ancestry that can be traced back to a Denisovan origin, confirming what was previously known from literature (Prüfer et al., 2014; Qin & Stoneking, 2015; Sankararaman et al., 2016; Skoglund & Jakobsson, 2011). The genomic portions do not appear to include material of Neanderthal origin, as it always appears to show a propensity towards Denisovan when the two archaics are compared in the *f*4 statistics (Table 3).
Even considering the assumption that this material could not be entirely of purely Denisovan origin, but partly encompassing some modern human genomic material, the populations can be somewhat distinguished when comparing this "probably Denisovan" enriched portion of the genome.

4.1.1 At least three different Denisovan populations

As the first two dimensions of the Multidimensional Scaling analysis on *f*3 data appear to be uncorrelated to the the variables that control for the modern human contaminants, they strike as the best proxies to be able to distinguish the Denisovan segments contained in the various populations, together with the second dimension of the Multidimensional Scaling analysis performed on the binned data. These three variables together hint to the presence of substructure between the populations of Denisovans that introgressed into the ancestors of the analyzed modern human populations: two different clusters of populations can be distinguished and isolated from the plots in Figures

5 and 6, one that includes the segments of Denisovan ancestry extracted from populations of Papua New Guinea and one from East Asian populations.
These results corroborate literature in finding that at least three distinct Denisovan populations lived at some point of human evolutionary history: one population from which the Denisova 3 from Denisova Cave comes from; another Denisovan population that left its traces into the genomes of East Asians; and finally one that introgressed into Papuan populations.
These findings consistently fit into the literature landscape set by the two previously mentioned works by (Browning et al., 2018; G. S. Jacobs et al., 2019).

4.1.2 Evaluation of expected population heterozygosity

On a negative note, the initial ambition of evaluating the genetic diversity with *f*3 statistics on individual genomes and with population *f*3 data proved itself to be far too limited, probably by the small amount of Denisovan introgressed material retrievable from the modern human genomes after applying our stringent set of filters. Even if at the individual level each sample does contain some Denisovan ancestry, as can be verified by the fact that an enrichment of 70% and over has been performed successfully on selected portions of these genomes, these regions appear to be far too sparse over the chromosomes and generally do not overlap between individuals, making it so that the software that computes the *f*3s does not perform well at the individual level (compiling *f*3 statistics for each individual was attempted, but without any success in giving interpretable results). Further, *f*3 statistics performed at the population level and analyzed with Multidimensional Scaling analysis failed to show a clear

tendency towards the formation of clear clusters for the selected populations.
This limitation is intrinsic to the type of genomic material that is under analysis in this Master Thesis and it needed to be somehow bypassed to make inference on Denisovan population diversity. Thus, the concept of heterozygosity became crucial to infer about Denisovan demography: heterozygosity can be defined as the fraction of sites within an individual that differ between the chromosomes they inherit from their parents. It allows us to make inferences on genetic variation, as it informs about inbreeding and effective population size. Heterozygosity can be evaluated at the population level as the average of the individual values of heterozygosity of the individuals of a population.

When correcting the Denisovan heterozygosity for the history they share with the modern human populations whose genomes acted as guests for these introgressed segments, most of the probable Denisovan ancestries follow the trend that correlates almost perfectly the heterozygosity of the host (modern humans) to the heterozygosity of the guest (Denisovans that introgressed in modern human populations). In Figure 8, the populations that show the highest deviation from the expected trend are the putatively Denisovan segments extracted from Aeta and Papuans, which incidentally are also the populations that show the highest contamination from modern human material in the *f*4 statistic (Tables 4 and 6). Consequently, they are not to be considered as Denisovan populations that were more heterozygous and diverse than the others: they should follow the same trend of being swept by the human shared history as the other populations, but are pulled slightly upwards in the plot because of the human contaminants they enclose.

While the increased Denisova heterozygosity displayed by Papuans and Aeta may reflect the excess of human haplotypes detected with the *f*4 rations, the similarly higher levels of Denisova heterozygosity with the respect to the underlying human heterozygosity displayed by Paniya may instead reflect a true signal of higher diversity in that local Denisovan population, as the genomic data from this population did not show evidence of modern human contamination.

Regardless, the deviation of the expected Denisovan heterozygosity on the y axis of the plot in Figure 8 is limited to a very restricted range of heterozygosity. Thus, regardless of the small differences between populations, we can infer that the diversity of Denisovan populations when they introgressed the modern human populations followed the same trend of low heterozygosity. After this, the Denisovan portions that introgressed populations that subsequently went through the Neolithic population expansion showed modest growth in their heterozygosity, as can be seen for the populations that occupy the rightmost area of the plot in Figure 8, whereas others such as Papuan populations and Aeta stayed at lower heterozygosity levels.

As for the question regarding the possible differences in heterozygosity in South East Asian Denisovan populations, this Master Thesis answers this question, as in general there does not seem to be a noticeable different level of heterozygosity in the Denisovan populations that introgressed in the area of SouthEast Asia. The particular case of Paniya suggests that there may have been a higher level of heterozygosity for the Denisovan population that introgressed the ancestors of the Indian population, but with a difference on a very limited scale.

To summarise, Denisovans were diversified in at least three populations, but all at lower heterozygosity levels, much like the population that the Altai Denisovan individual from whom the only complete genome currently at our disposal comes from (Prüfer et al., 2014; Reich et al., 2010).

4.2 Limitations of the findings

This Master Thesis both confirms literature regarding the presence of at least three distinct populations and adds insights to the level of heterozygosity that all Denisovan populations shared. Even so, the present results lacks resolution to confirm the work by Jacobs et al., 2019, in which the authors could subdivide the genomic material of Denisovan origin introgressed into Papuan populations in two distinct pulses of admixture: such a division was not evident from the resolution of our results, as all Papuan populations fell in the same cluster of similarity to the Altai Denisovan genome. Incidentally, no clear-cut evidence was found for a pronounced genetic substructure of the Denisovan segments introgressed in the ancestors of all the modern human populations tested in this Master Thesis, apart from the two major clusters of introgressed Denisovan segments. Namely, the South Asian populations that in literature appear to show an excess in Denisovan ancestry (Mallick et al., 2016; Sankararaman et al., 2016; Witt et al., 2022), do not form a separated cluster from all other Denisovan segments introgressed into Asian populations.
Considering these limitations, one should avoid to make inferences about the divergence of Denisovan populations spanning Eastern Eurasia based on the results of this Master Thesis: even if there are at least three distinct Denisovan populations, and this obviously does not put an upper limit to the possible number of

populations, it does not appear sensible to infer a further division between groups based on the results of this work, because it is not apparent from these results. What is discernible from this work is that certainly Denisovans did occupy a vast range of environments, differentiating in at least three different populations that admixed with the ancestors of modern humans in different occasions, but their demography followed the same trend of low heterozygosity. When making inference about the demographic history of Denisovans from the results of this Master Thesis, we should refrain from the unsubstantial hypothesis that any certain environment they spanned across would have brought forth an increase in heterozygosity in Denisovans.

4.3 Future prospects

In the light of these results, we can say that Denisovans were diversified in at least three populations, but all had low heterozygosity.
A recommendation for a future development of this subject is the investigation of the fact that the Aeta population does render from the extraction with IBDmix a number of putatively Denisovan regions that is much higher in number than any other of the analyzed populations: even considering that this genetic material appears to be contaminated with modern human components, there must be a reason as to why the IBDmix software confuses a great number of these regions, that should look more similar to modern human regions, with archaic, putatively Denisovan regions. It would make sense to investigate the reason as to why these regions differ for Aeta from all other populations. Still another intriguing perspective for future research is the possibility of dating the introgression of the Denisovan populations that

introgressed into the genomes of the ancestors of the modern humans analyzed in this Master Thesis.

5. Conclusion

This Master Thesis serves as an evaluation of the presence of genetic substructure among populations of the archaic human known as Denisova. The genomes of a panel of pan-Eurasian populations were enriched in Denisovan ancestry segments to gain insights into the genetic diversity of the archaic human's populations. The comparison and analysis of such introgressed genetic material yielded results that validate literature in confirming the existence of at least three distinct Denisovan populations during the course of human evolutionary history, namely a population occupying the area of the Denisova Cave and two other populations that left traces into the genomes of East Asian and Papuan populations respectively. When wondering about the level of heterozygosity, seen as proxy for inbreeding levels and effective population size to infer about the demography of Denisovan populations, this Master Thesis shows that there does not appear to be a remarkable difference for the expected population heterozygosity in the Denisovan ancestries analyzed, with only the Paniya population showing a significantly higher level of heterozygosity, and on a limited scale.

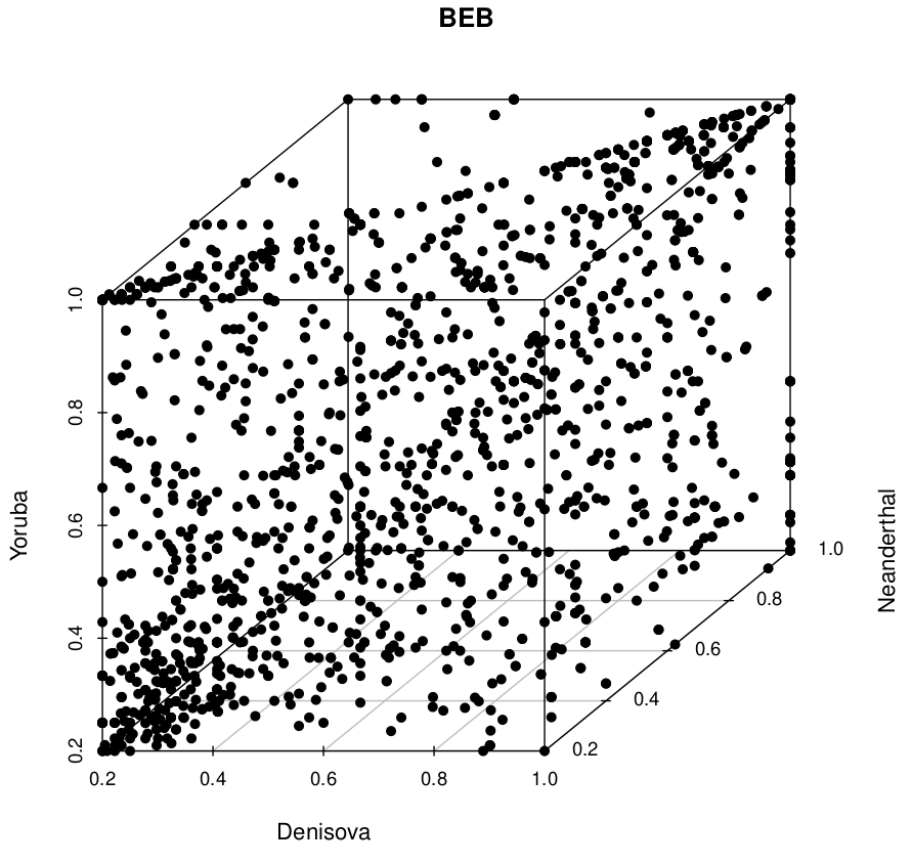
To conclude, I would like to emphasize and remark the great quantity of information that can be gained from using a Molecular approach in Anthropology, which allowed me to make inference about genetic structure and demography of the extinct hominid lineage of the Denisovans starting from modern human genomes.

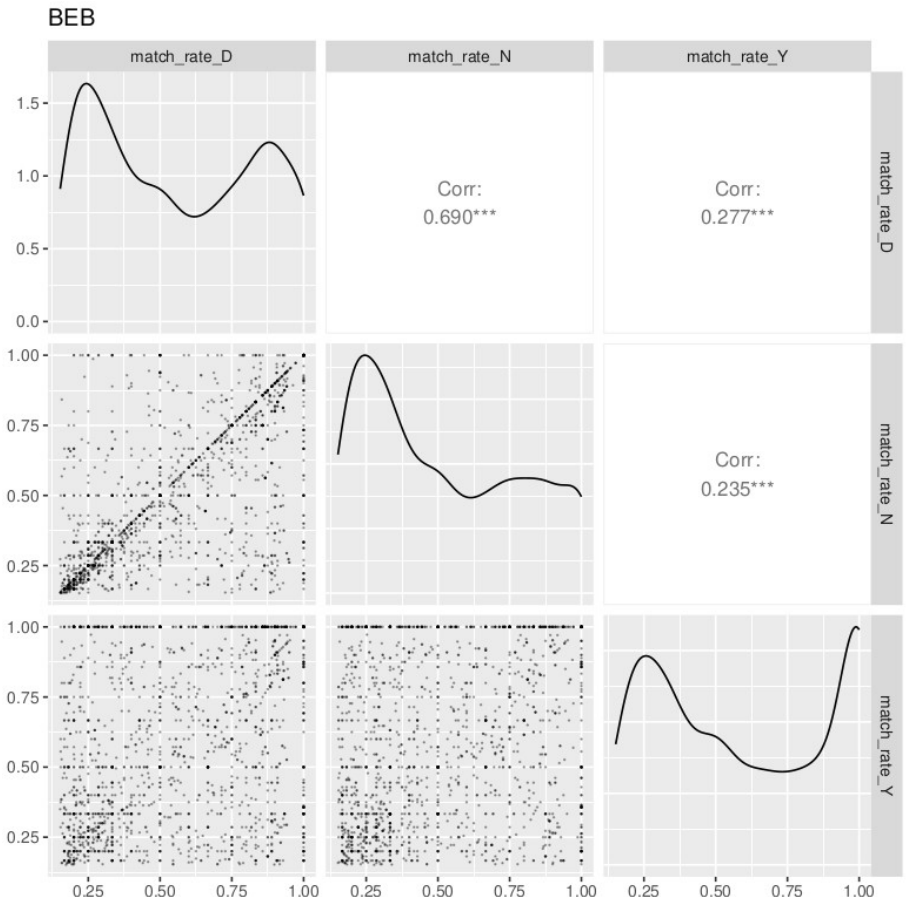
6. Appendix

6.1 Appendix 1

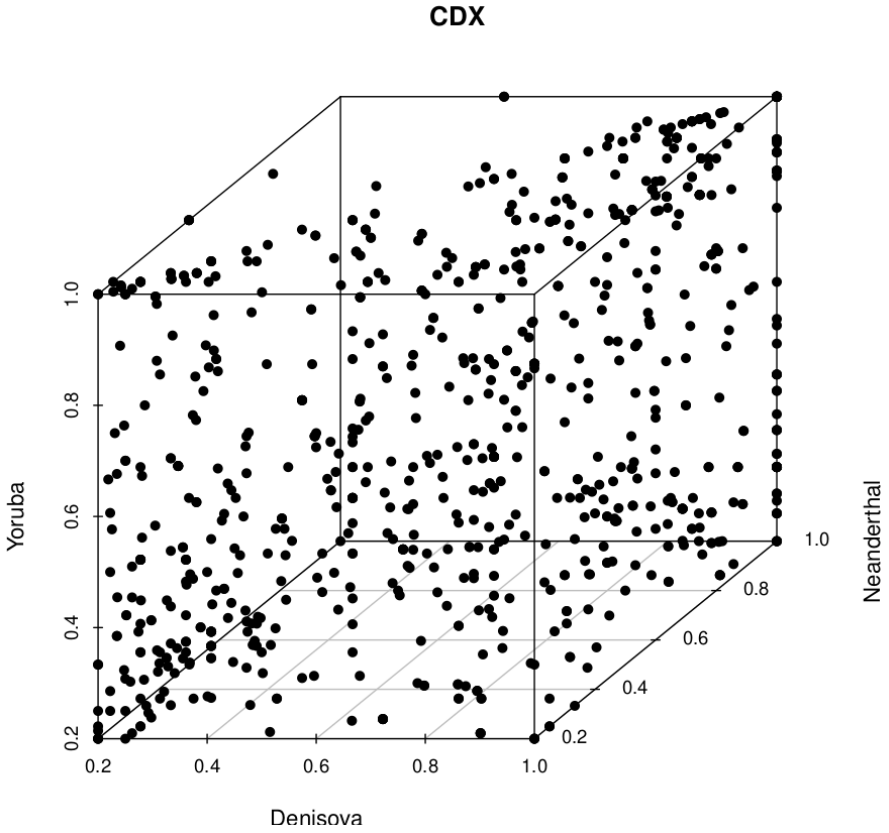
Three dimensional scatter plots and correlograms (two dimensional scatter plots) of the match rates to Denisova, Neanderthal and Yoruba for all populations analyzed in the Master Thesis.

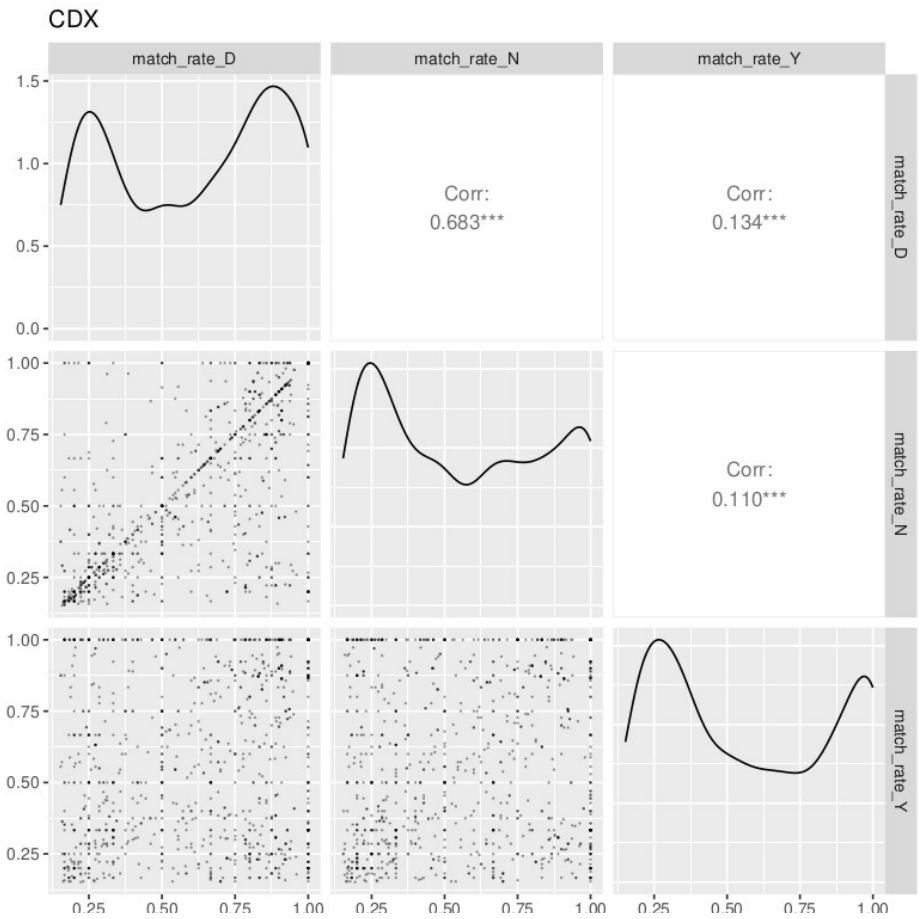
Appendix 1.1: 3d and 2d scatter plots of the match rates to Denisova, Neanderthal and Yoruba for Bengali in Bangladesh (BEB).



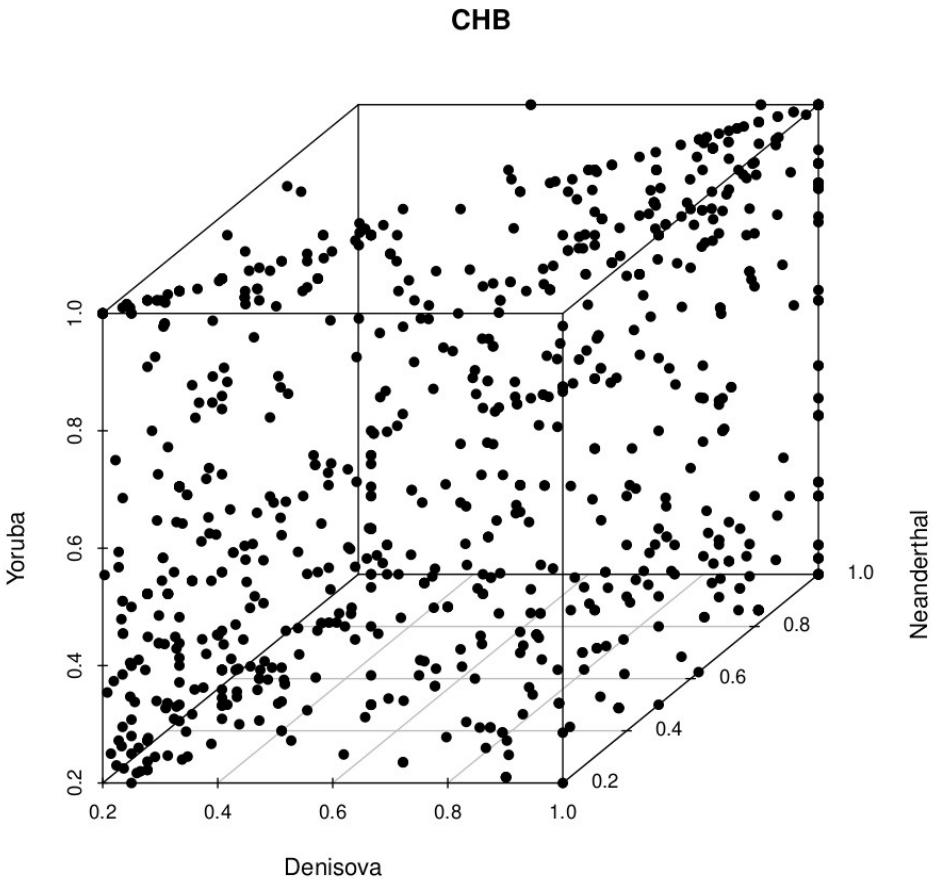

BEB
match_rate_D
match_rate_N
match_rate_Y
Corr:
0.690***
Corr:
0.277***
Corr:
0.235***
1.5
1.0
0.5
0.0
1.00
0.75
0.50
0.25
0.25
0.50
0.75
1.00

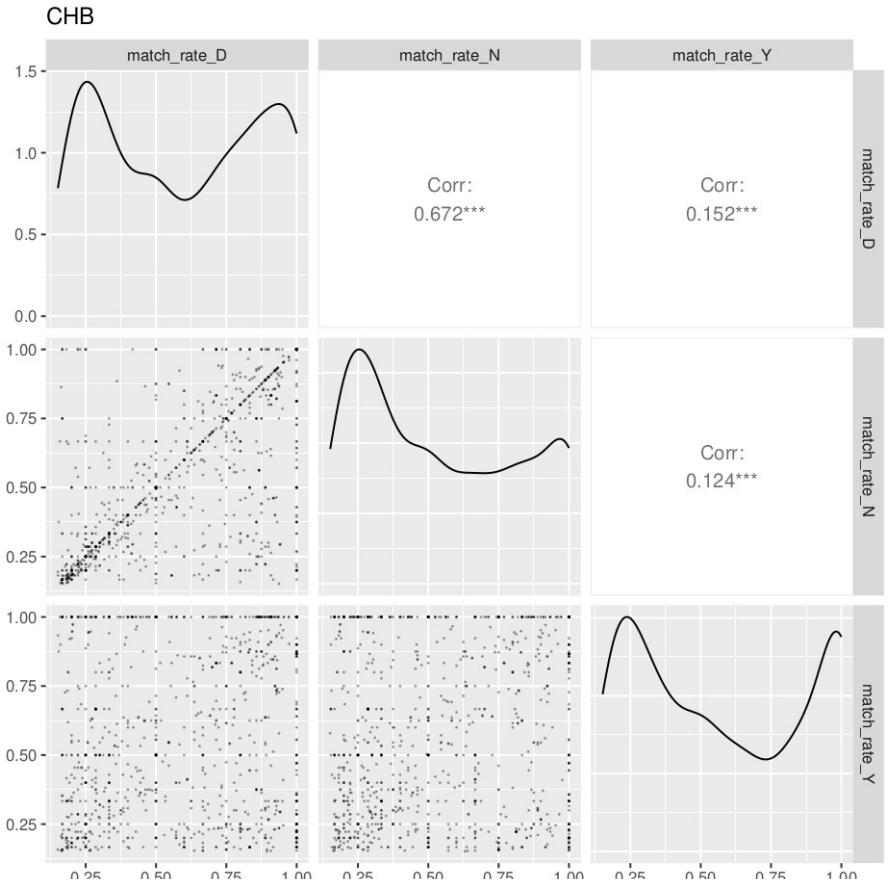
Appendix 1.2: 3d and 2d scatter plots of the match rates to Denisova, Neanderthal and Yoruba for Dai Chinese in Xishuangbanna, China (CDX).



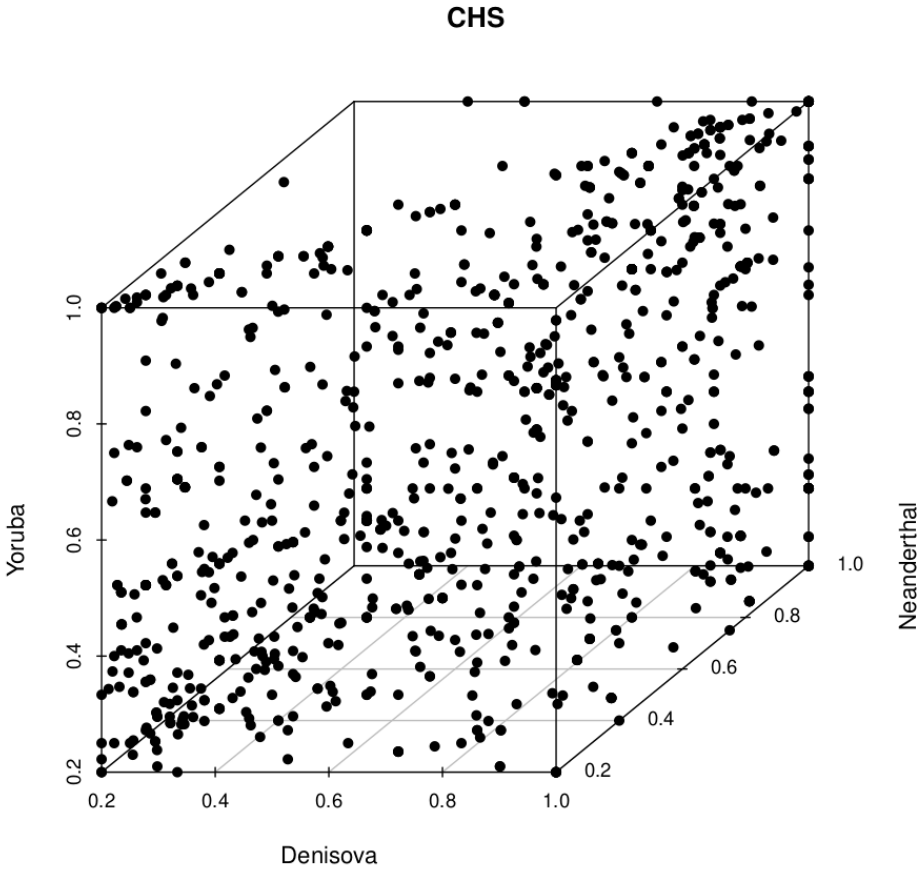

CDX
match_rate_D
match_rate_N
match_rate_Y
Corr:
0.683***
Corr:
0.134***
Corr:
0.110***
1.5
1.0
0.5
0.0
1.00
0.75
0.50
0.25
0.25
0.50
0.75
1.00

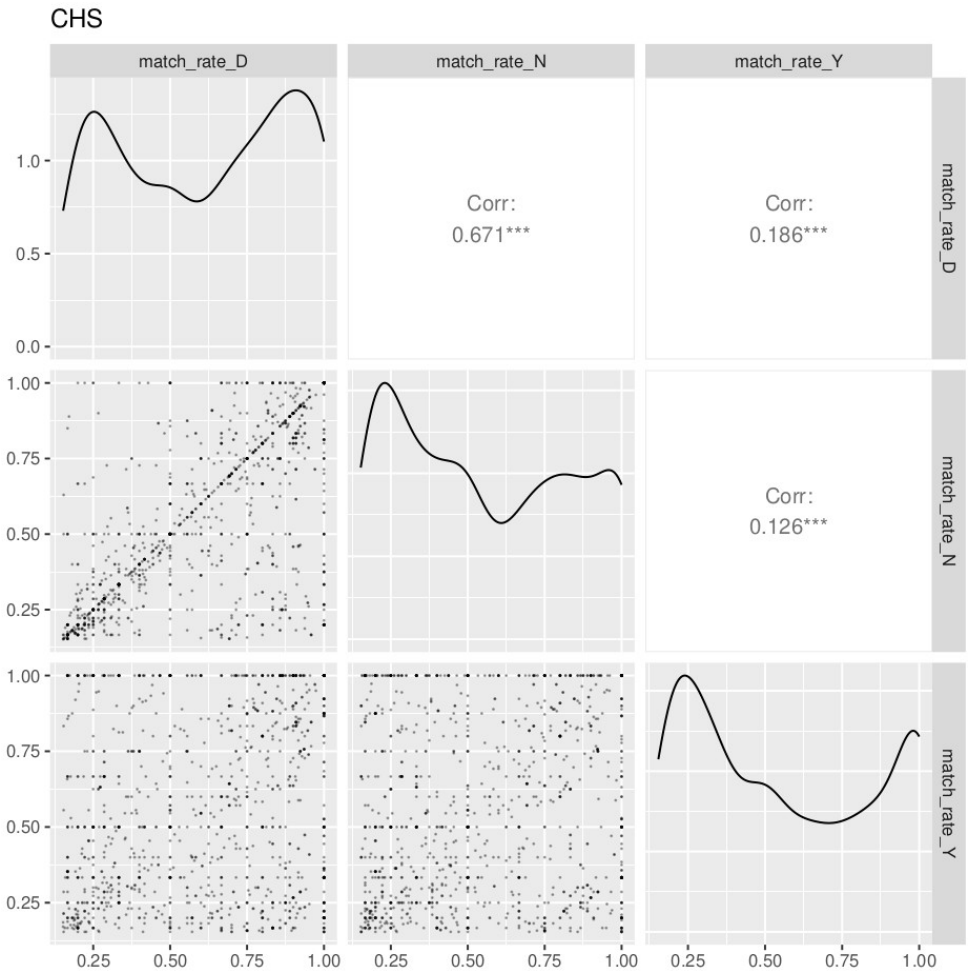
Appendix 1.3: 3d and 2d scatter plots of the match rates to Denisova, Neanderthal and Yoruba for Han Chinese in Beijing, China (CHB).



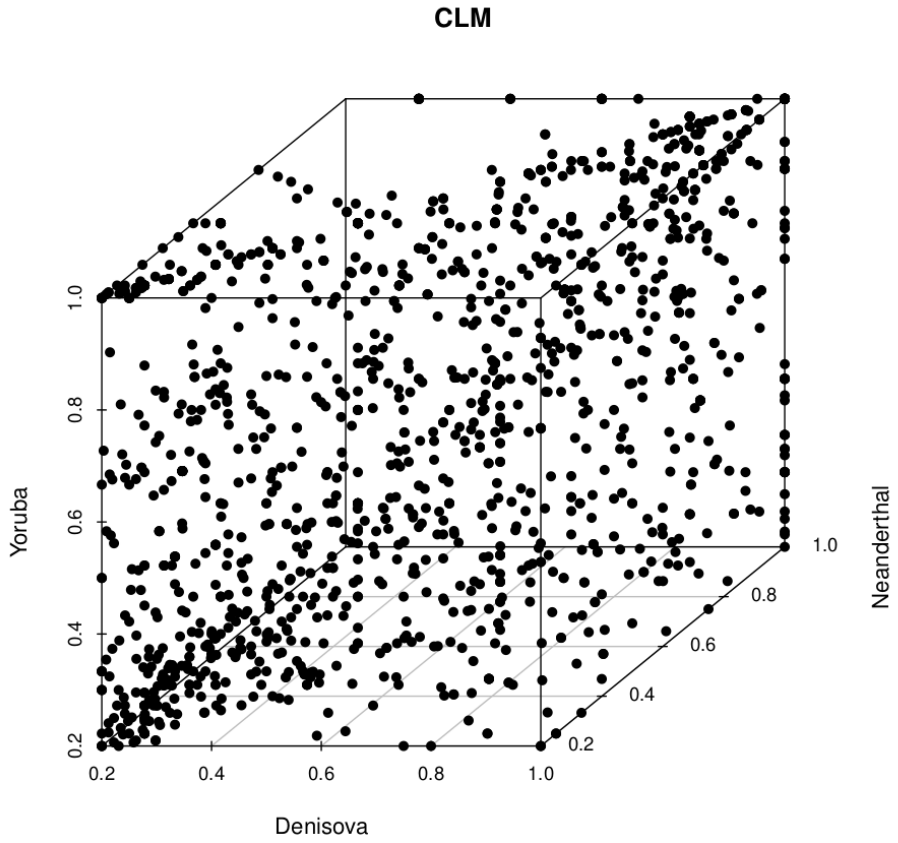

CHB
match_rate_D
match_rate_N
match_rate_Y
Corr:
0.672***
Corr:
0.152***
Corr:
0.124***
match_rate_D
match_rate_N
match_rate_Y

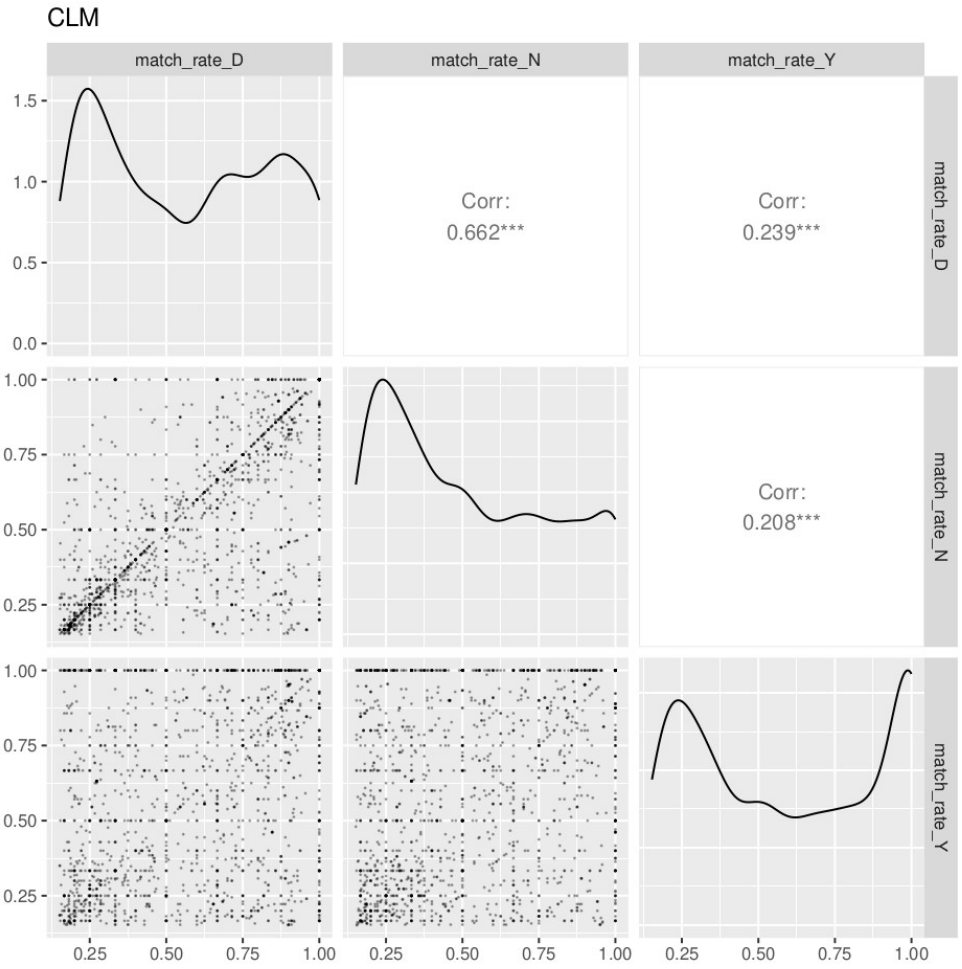
Appendix 1.4: 3d and 2d scatter plots of the match rates to Denisova, Neanderthal and Yoruba for Han Chinese South (CHS).



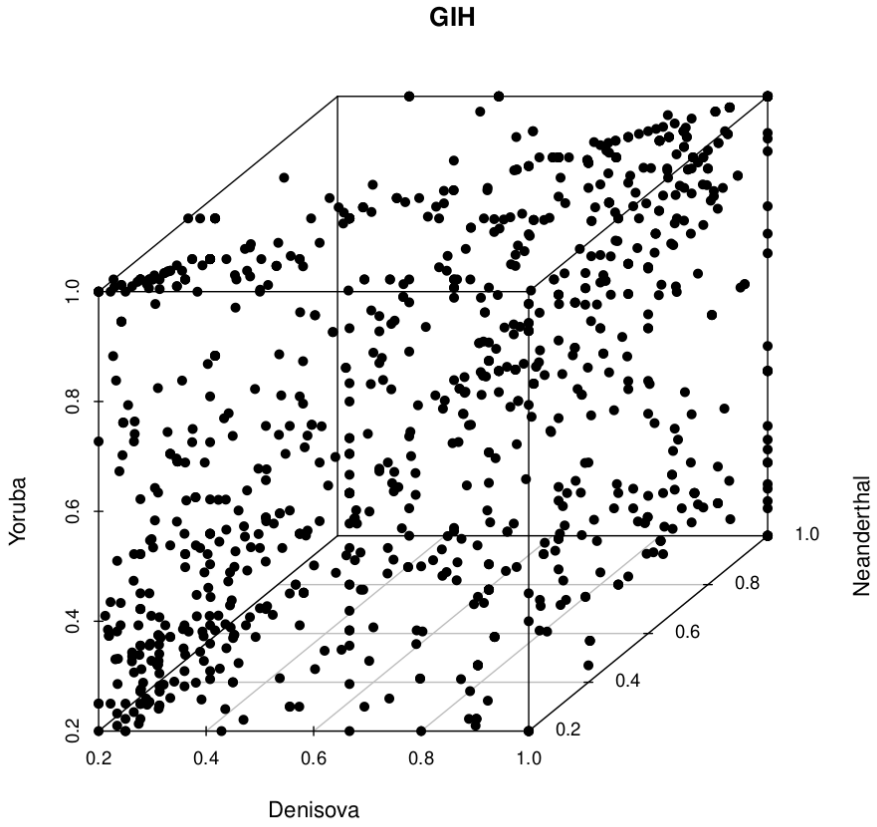

CHS
match_rate_D
match_rate_N
match_rate_Y
Corr:
0.671***
Corr:
0.186***
Corr:
0.126***
1.0
0.5
0.0
1.00
0.75
0.50
0.25

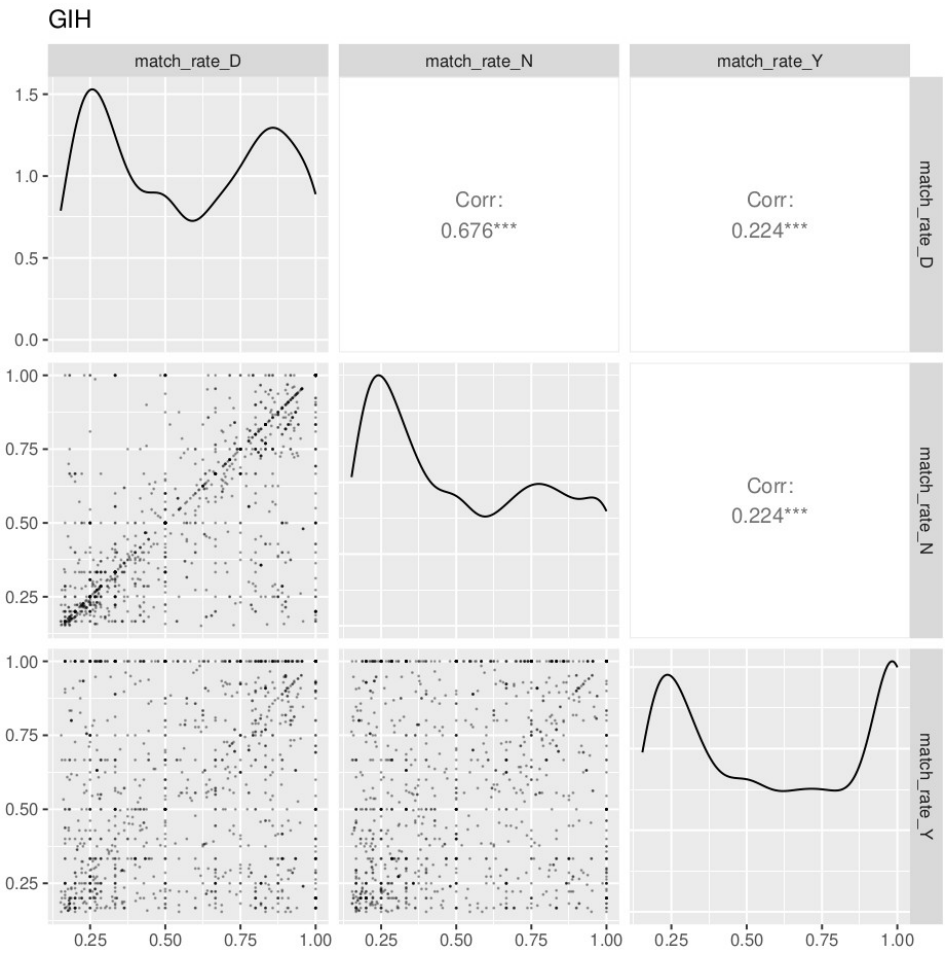
Appendix 1.5: 3d and 2d scatter plots of the match rates to Denisova, Neanderthal and Yoruba for Colombian in Medellín, Colombia (CLM).



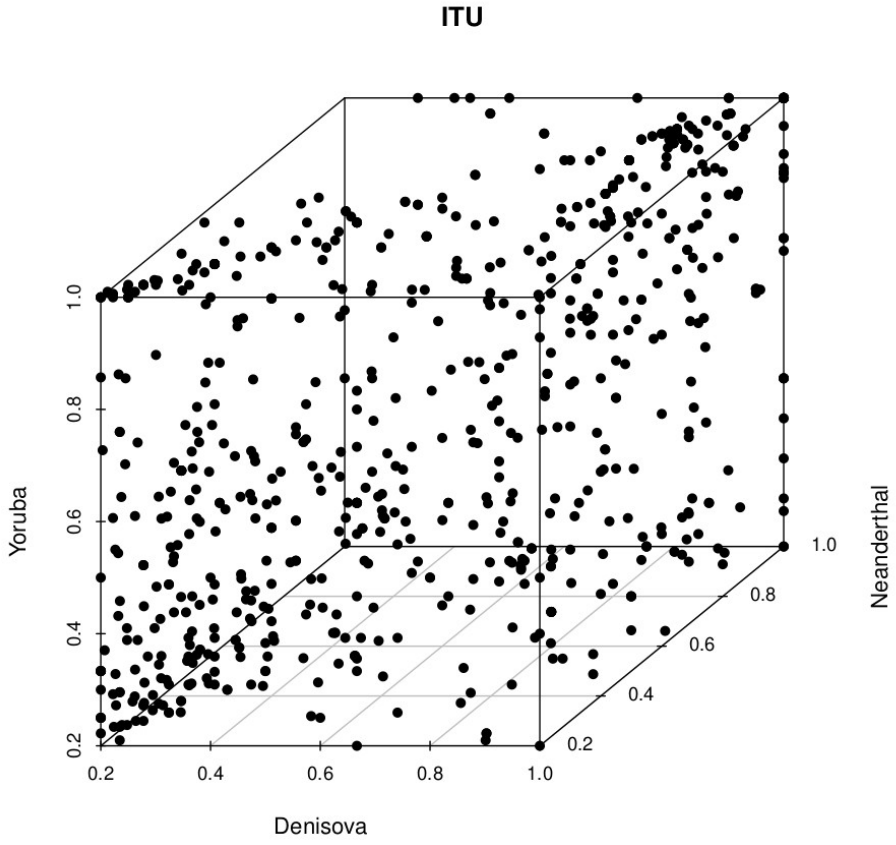

CLM
match_rate_D
match_rate_N
match_rate_Y
Corr:
0.662***
Corr:
0.239***
Corr:
0.208***
1.5
1.0
0.5
0.0
1.00
0.75
0.50
0.25
0.25
0.50
0.75
1.00

Appendix 1.6: 3d and 2d scatter plots of the match rates to Denisova, Neanderthal and Yoruba for Gujarati Indians in Houston, Texas, USA (GIH).

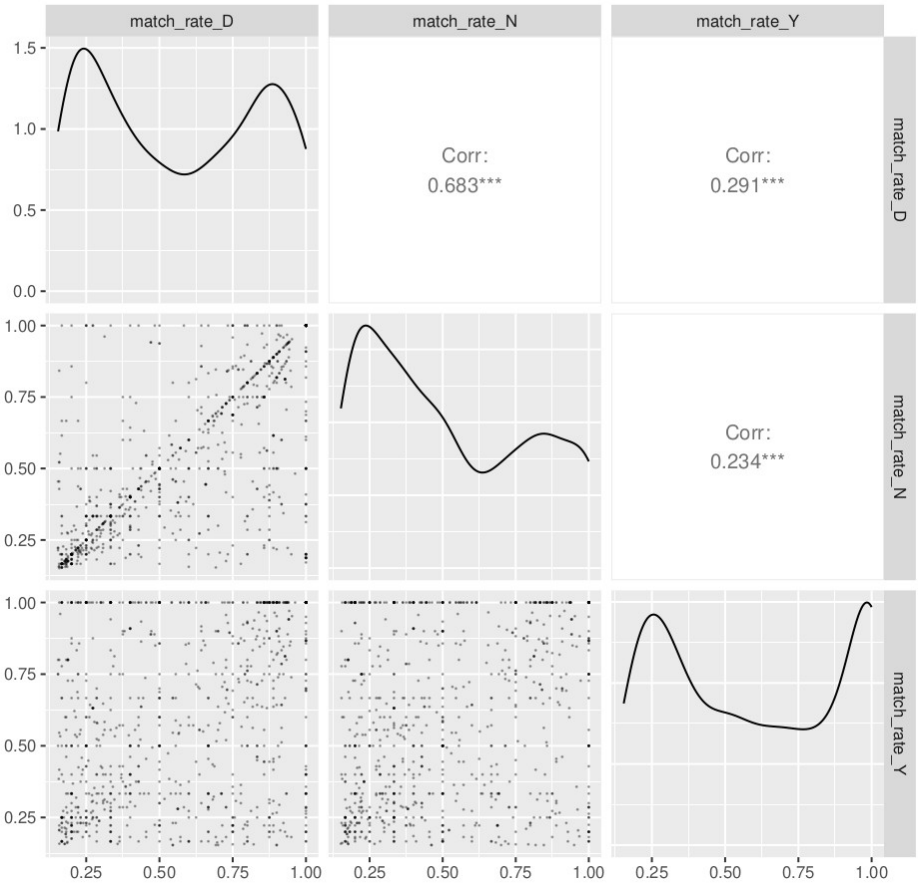



GIH
match_rate_D
match_rate_N
match_rate_Y
Corr:
0.676***
Corr:
0.224***
Corr:
0.224***
1.5
1.0
0.5
0.0
1.00
0.75
0.50
0.25
0.25
0.50
0.75
1.00

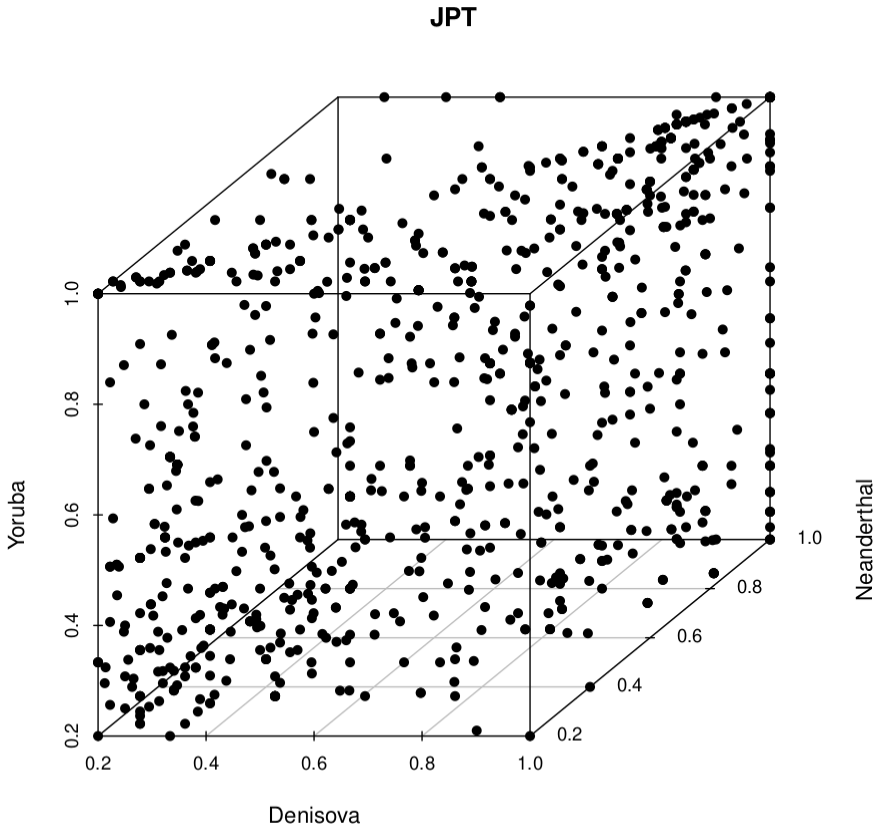
Appendix 1.7: 3d and 2d scatter plots of the match rates to Denisova, Neanderthal and Yoruba for Indian Telugu in the United Kingdom (ITU).

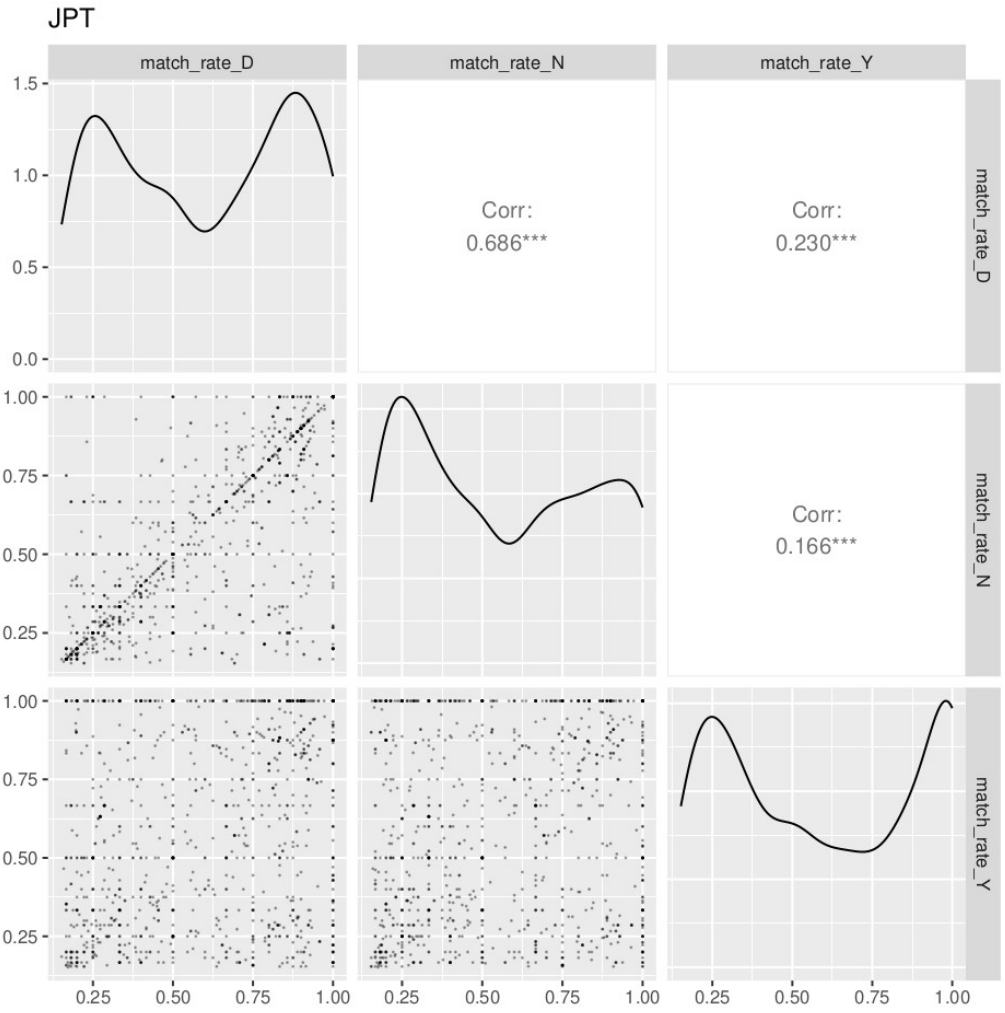


ITU

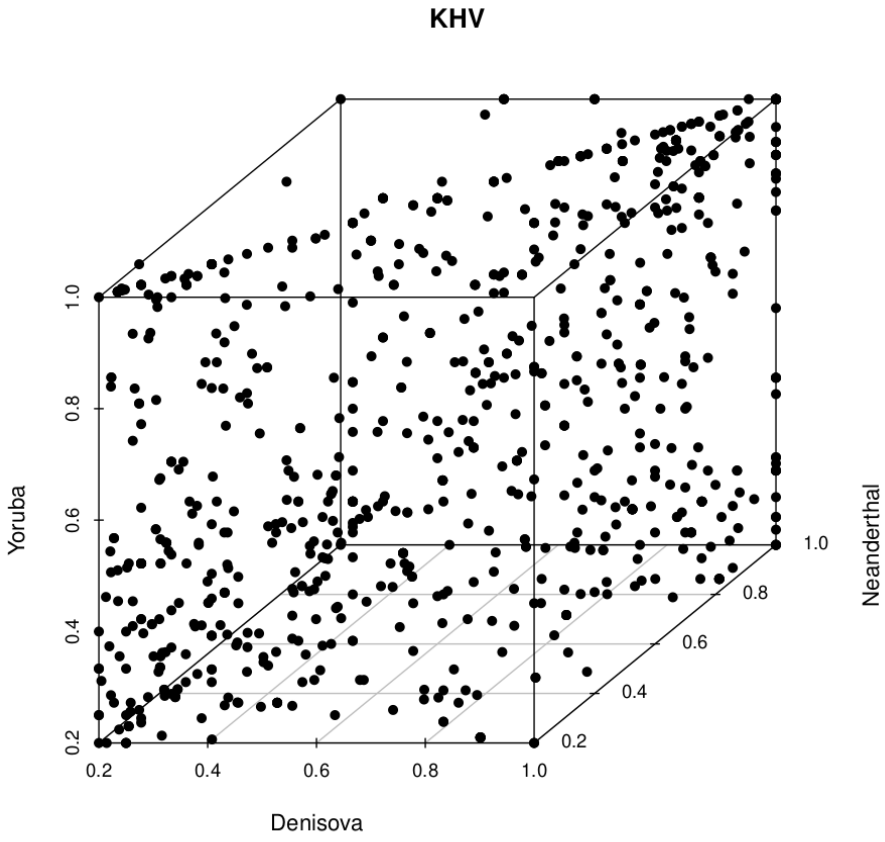

match_rate_D
match_rate_N
match_rate_Y
Corr:
0.683***
Corr:
0.291***
Corr:
0.234***
1.5
1.0
0.5
0.0
1.00
0.75
0.50
0.25
0.25
0.50
0.75
1.00

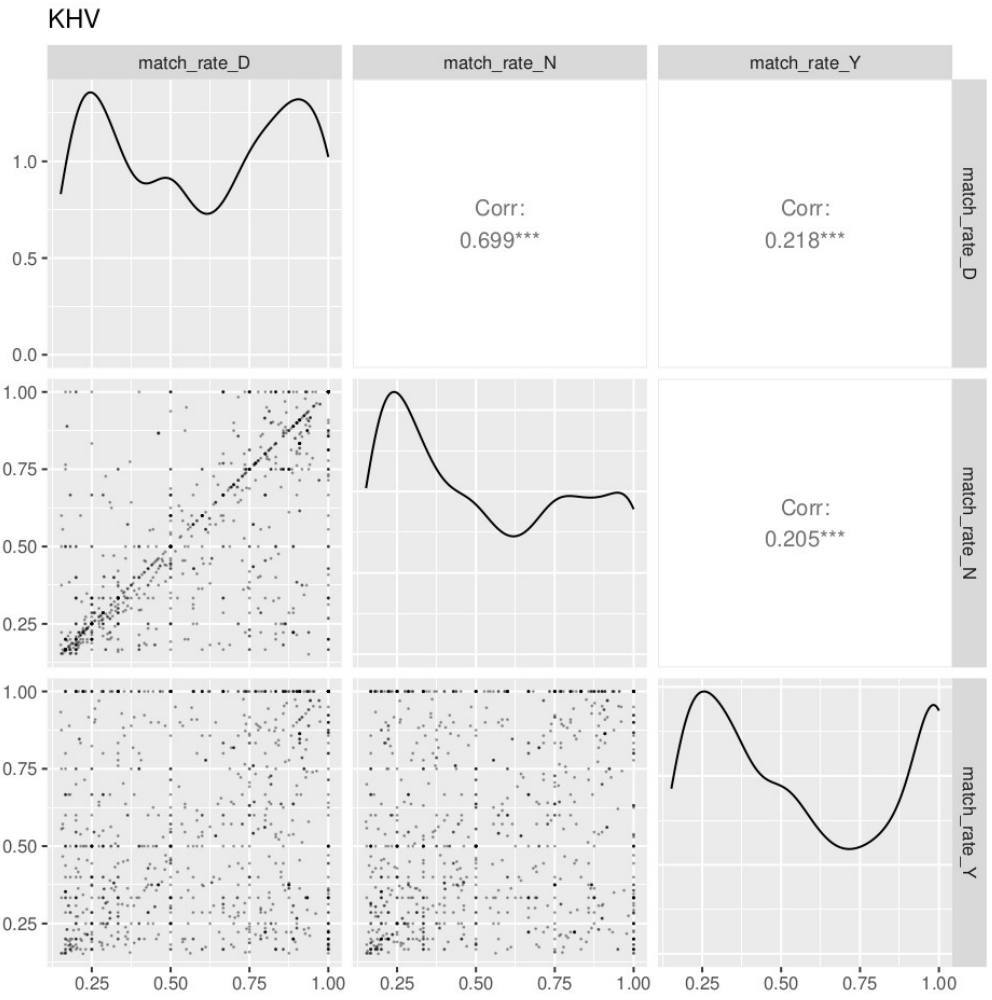
Appendix 1.8: 3d and 2d scatter plots of the match rates to Denisova, Neanderthal and Yoruba for Japanese in Tokyo, Japan (JPT).



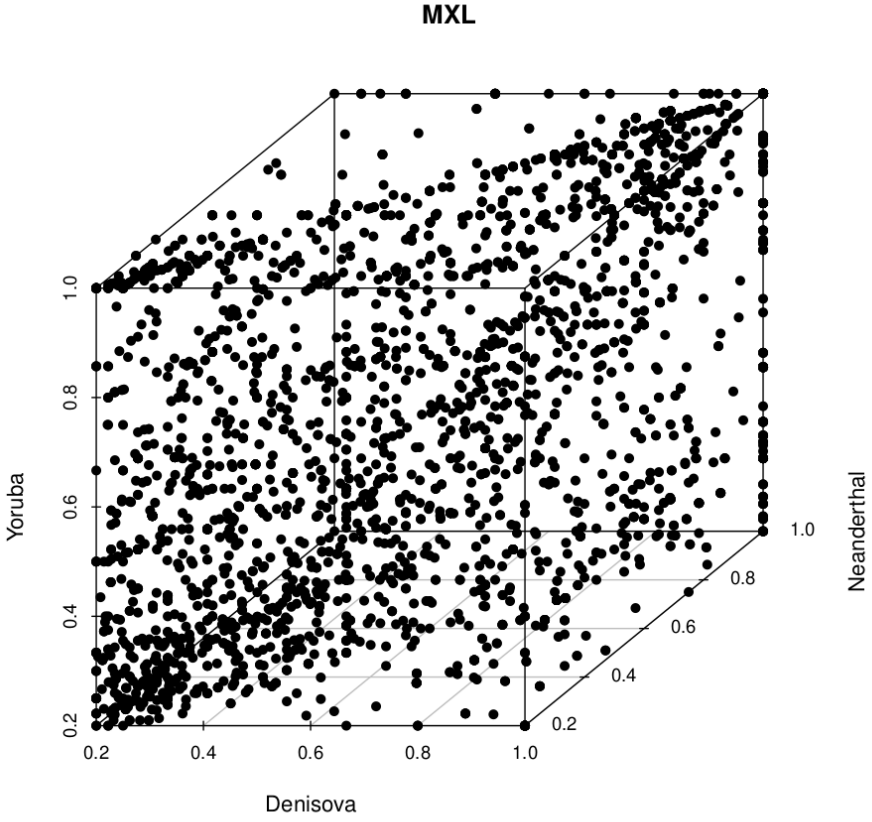

JPT
match_rate_D
match_rate_N
match_rate_Y
Corr:
0.686***
Corr:
0.230***
Corr:
0.166***
1.5
1.0
0.5
0.0
1.00
0.75
0.50
0.25
0.25
0.50
0.75
1.00

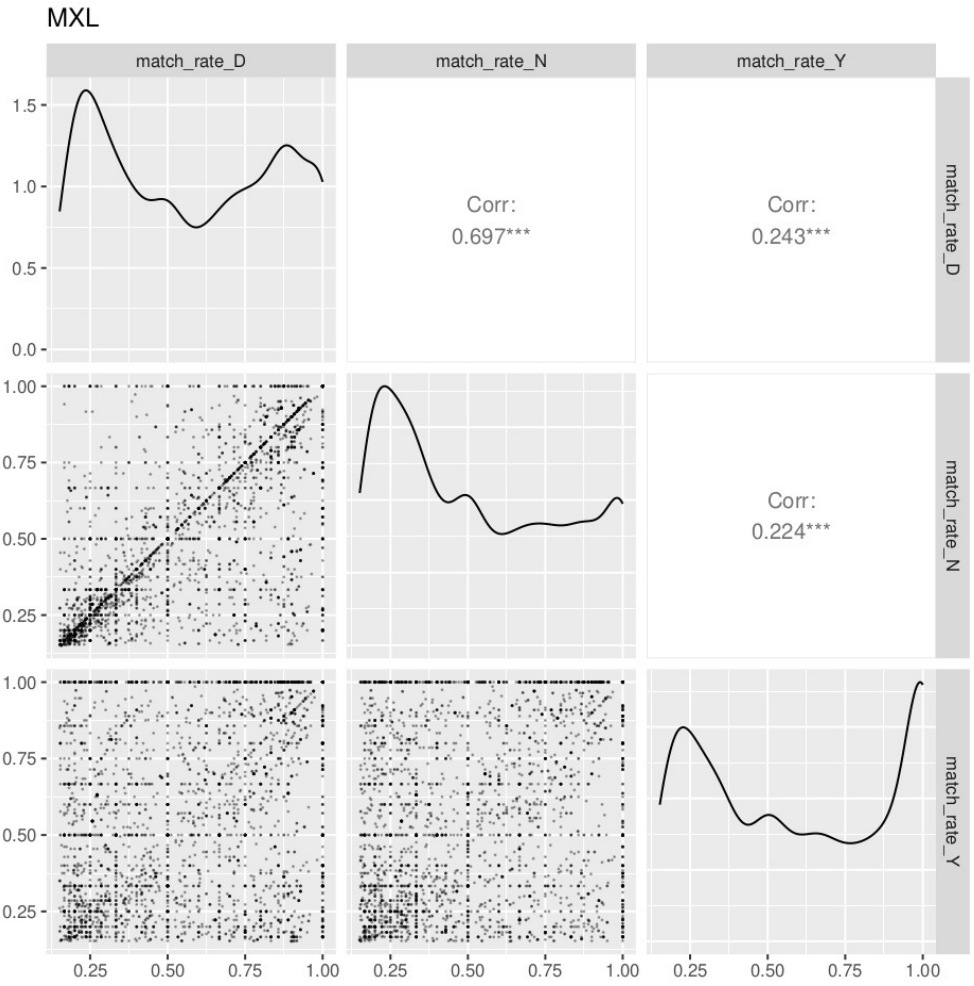
Appendix 1.9: 3d and 2d scatter plots of the match rates to Denisova, Neanderthal and Yoruba for Kinh in Ho Chi Minh City, Vietnam (KHV).



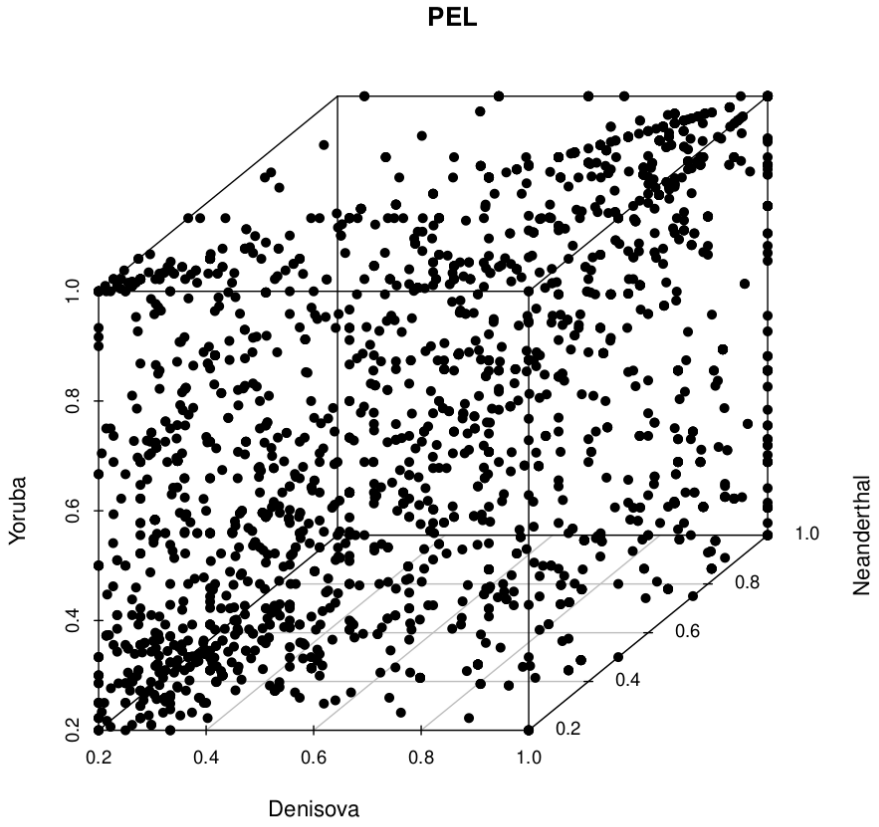

KHV
match_rate_D
match_rate_N
match_rate_Y
Corr:
0.699***
Corr:
0.218***
Corr:
0.205***
1.0
0.5
0.0
1.00
0.75
0.50
0.25
0.25
0.50
0.75
1.00

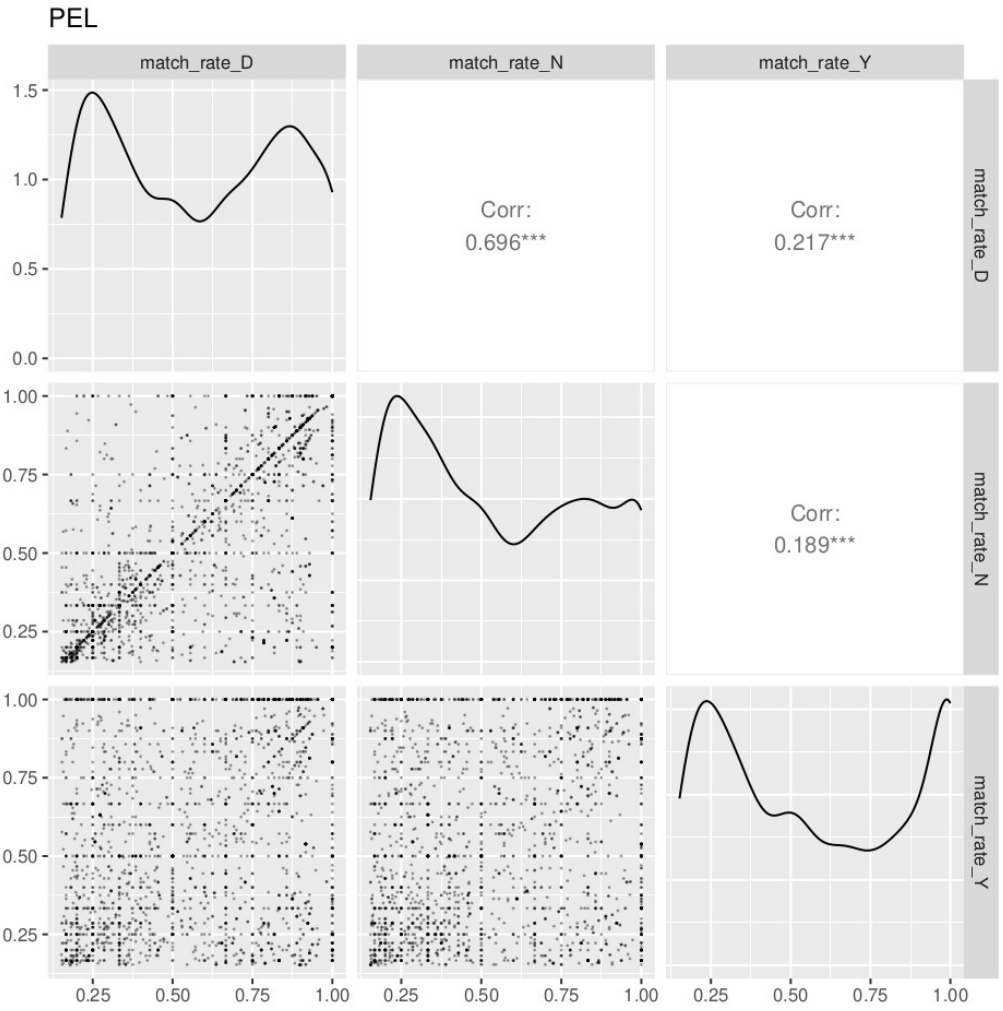
Appendix 1.10: 3d and 2d scatter plots of the match rates to Denisova, Neanderthal and Yoruba for Mexican Ancestry in Los Angeles, California, USA (MXL).



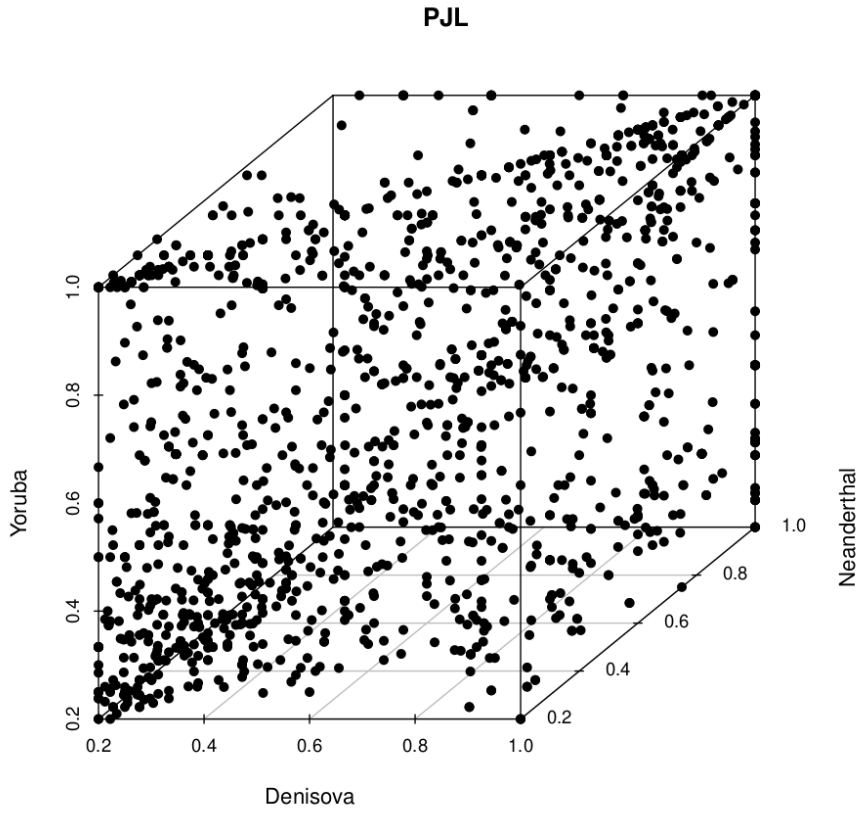

MXL
match_rate_D
match_rate_N
match_rate_Y
Corr:
0.697***
Corr:
0.243***
Corr:
0.224***
1.5
1.0
0.5
0.0
1.00
0.75
0.50
0.25
0.25
0.50
0.75
1.00

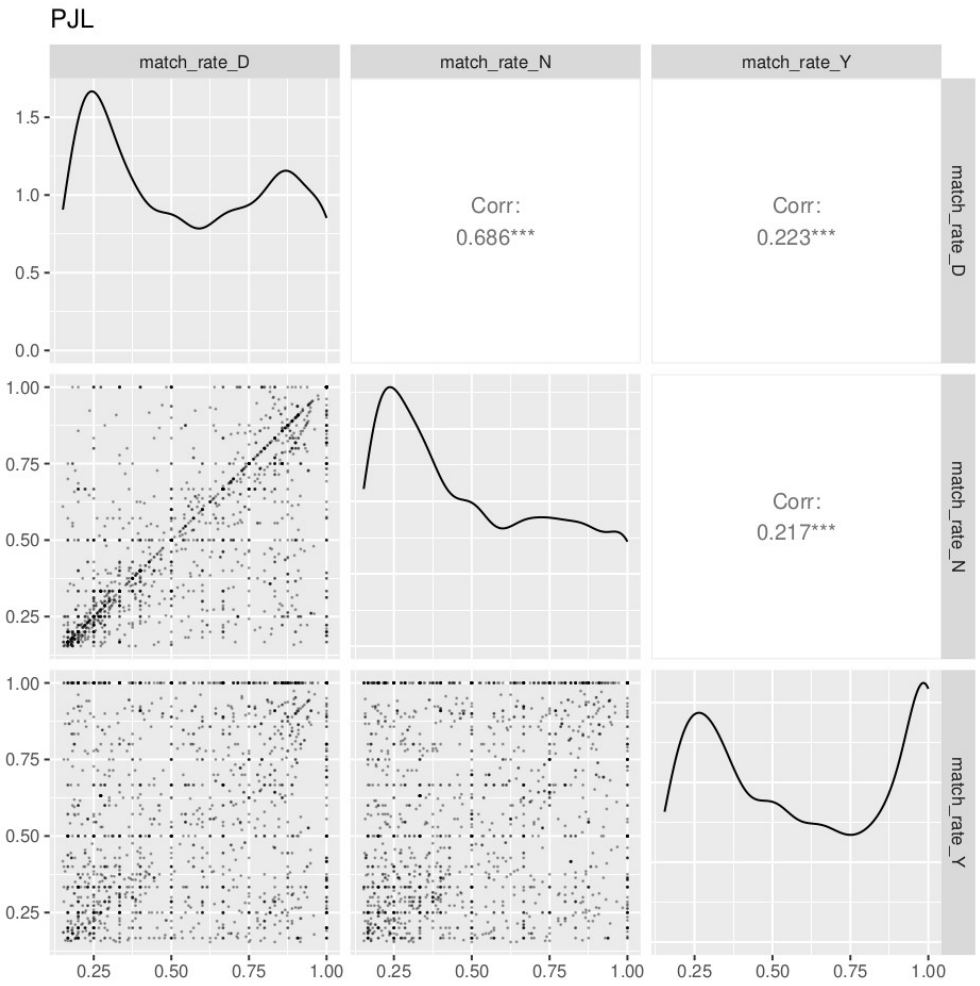
Appendix 1.11: 3d and 2d scatter plots of the match rates to Denisova, Neanderthal and Yoruba for Peruvian in Lima, Peru (PEL).



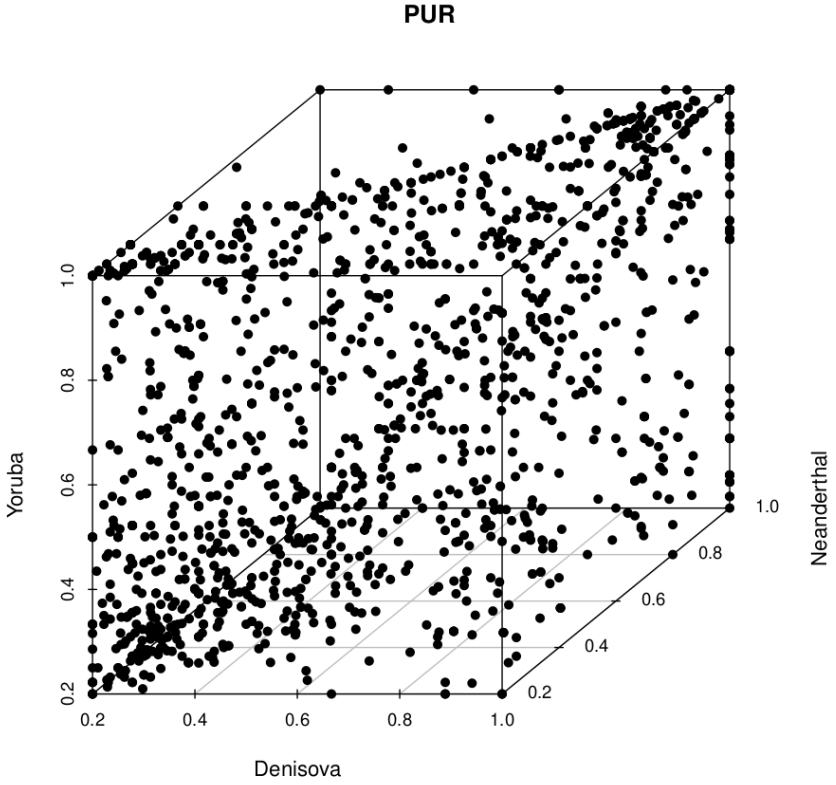

PEL
match_rate_D
match_rate_N
match_rate_Y
Corr:
0.696***
Corr:
0.217***
Corr:
0.189***
match_rate_D
match_rate_N
match_rate_Y
1.5
1.0
0.5
0.0
1.00
0.75
0.50
0.25
0.25
0.50
0.75
1.00

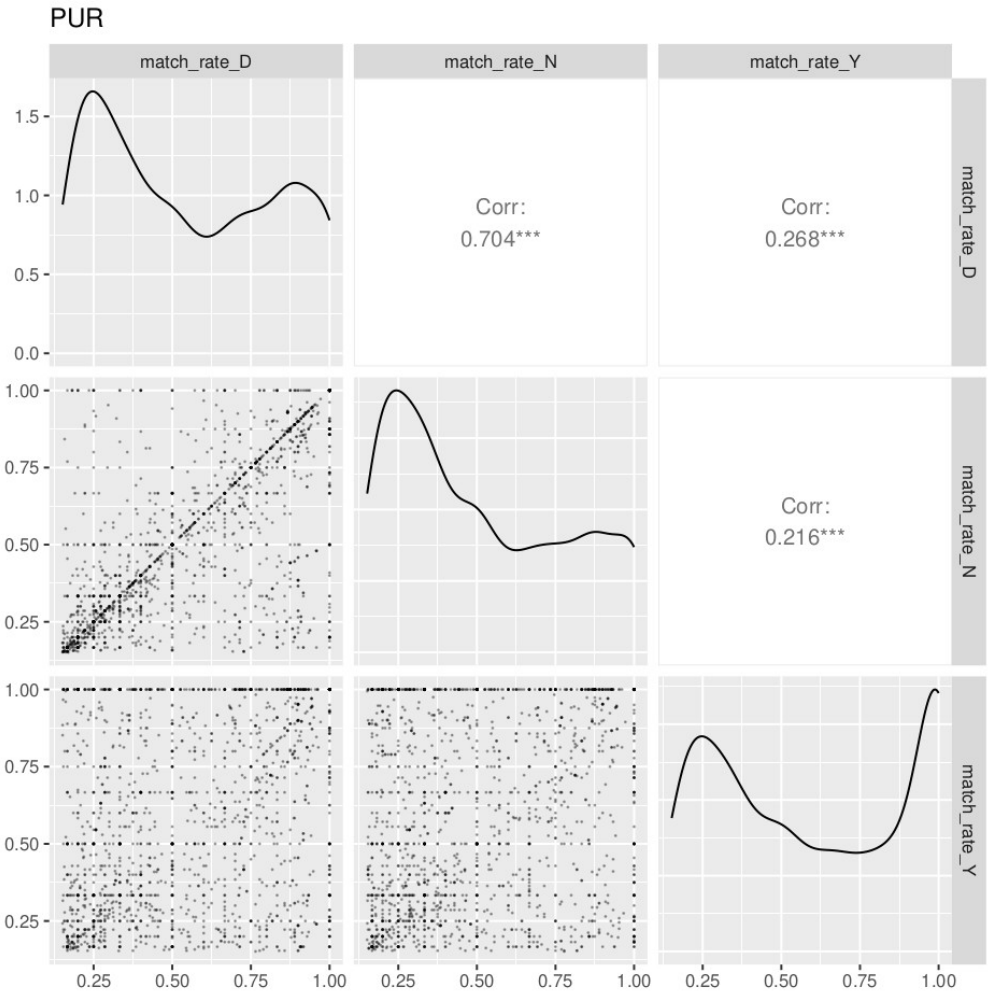
Appendix 1.12: 3d and 2d scatter plots of the match rates to Denisova, Neanderthal and Yoruba for Punjabi in Lahore, Pakistan (PJL).



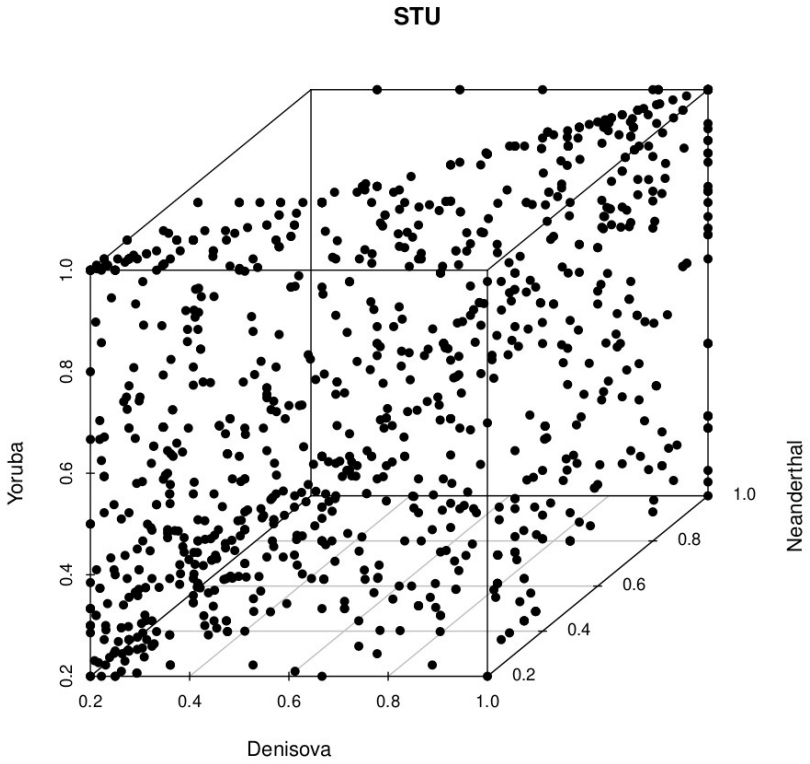

PJL
match_rate_D
match_rate_N
match_rate_Y
Corr:
0.686***
Corr:
0.223***
Corr:
0.217***
1.5
1.0
0.5
0.0
1.00
0.75
0.50
0.25
0.25
0.50
0.75
1.00

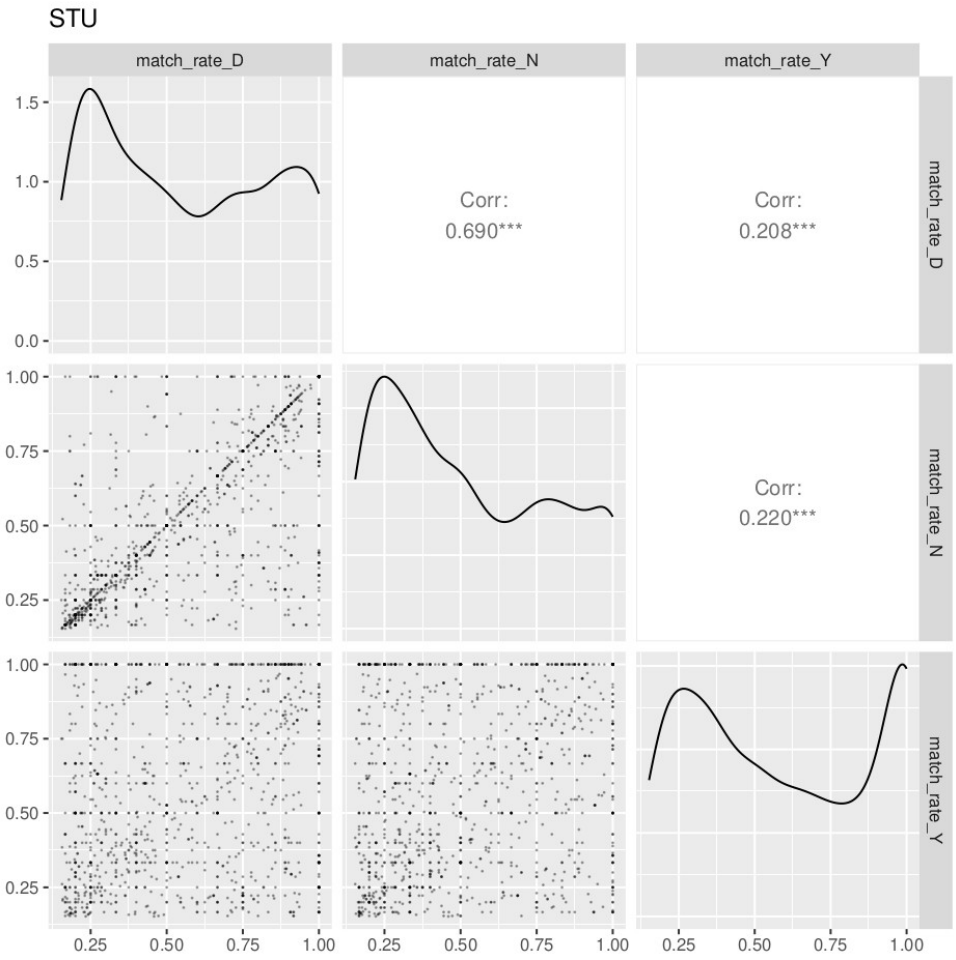
Appendix 1.13: 3d and 2d scatter plots of the match rates to Denisova, Neanderthal and Yoruba for Puerto Rican in Puerto Rico (PUR).



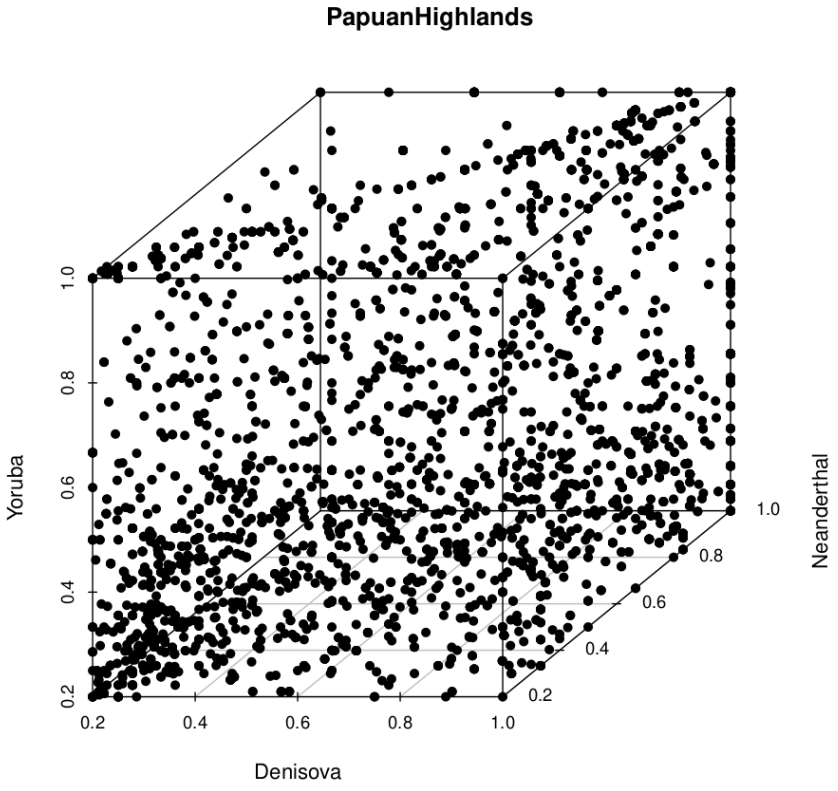

PUR
match_rate_D
match_rate_N
match_rate_Y
Corr:
0.704***
Corr:
0.268***
Corr:
0.216***
1.5
1.0
0.5
0.0
1.00
0.75
0.50
0.25
0.25
0.50
0.75
1.00

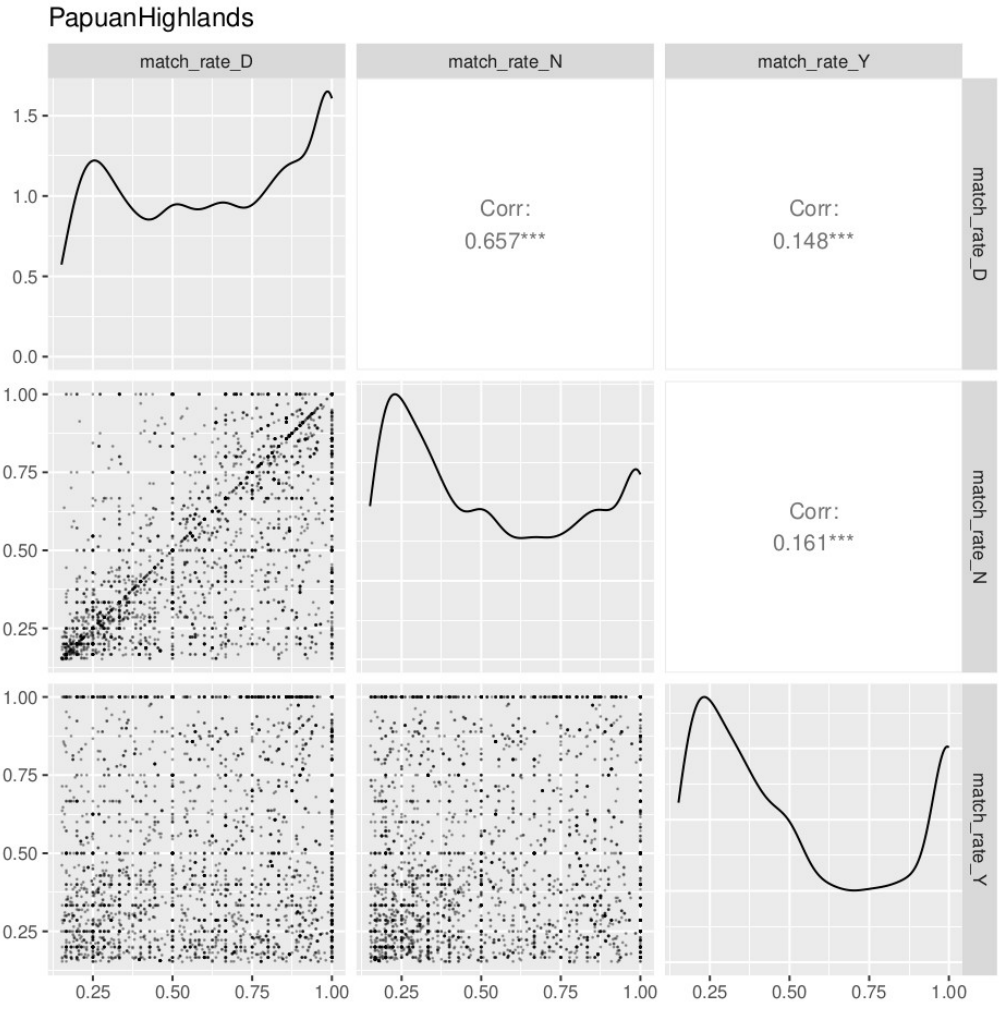
Appendix 1.14: 3d and 2d scatter plots of the match rates to Denisova, Neanderthal and Yoruba for Sri Lankan Tamil in the United Kingdom (STU).



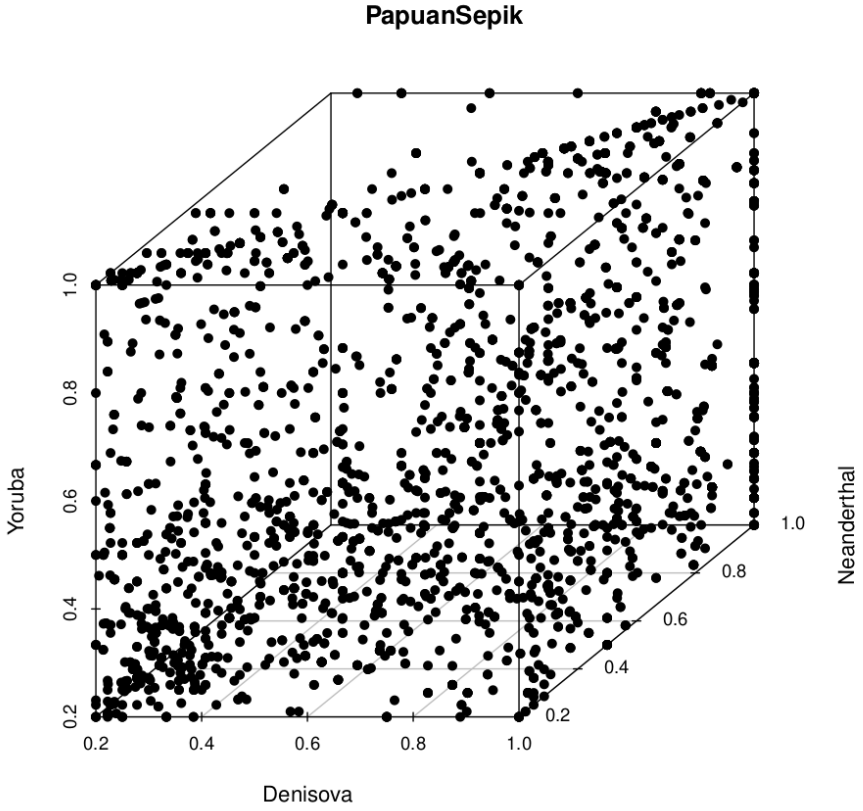

STU
match_rate_D
match_rate_N
match_rate_Y
Corr:
0.690***
Corr:
0.208***
Corr:
0.220***
1.5
1.0
0.5
0.0
1.00
0.75
0.50
0.25
0.25
0.50
0.75
1.00

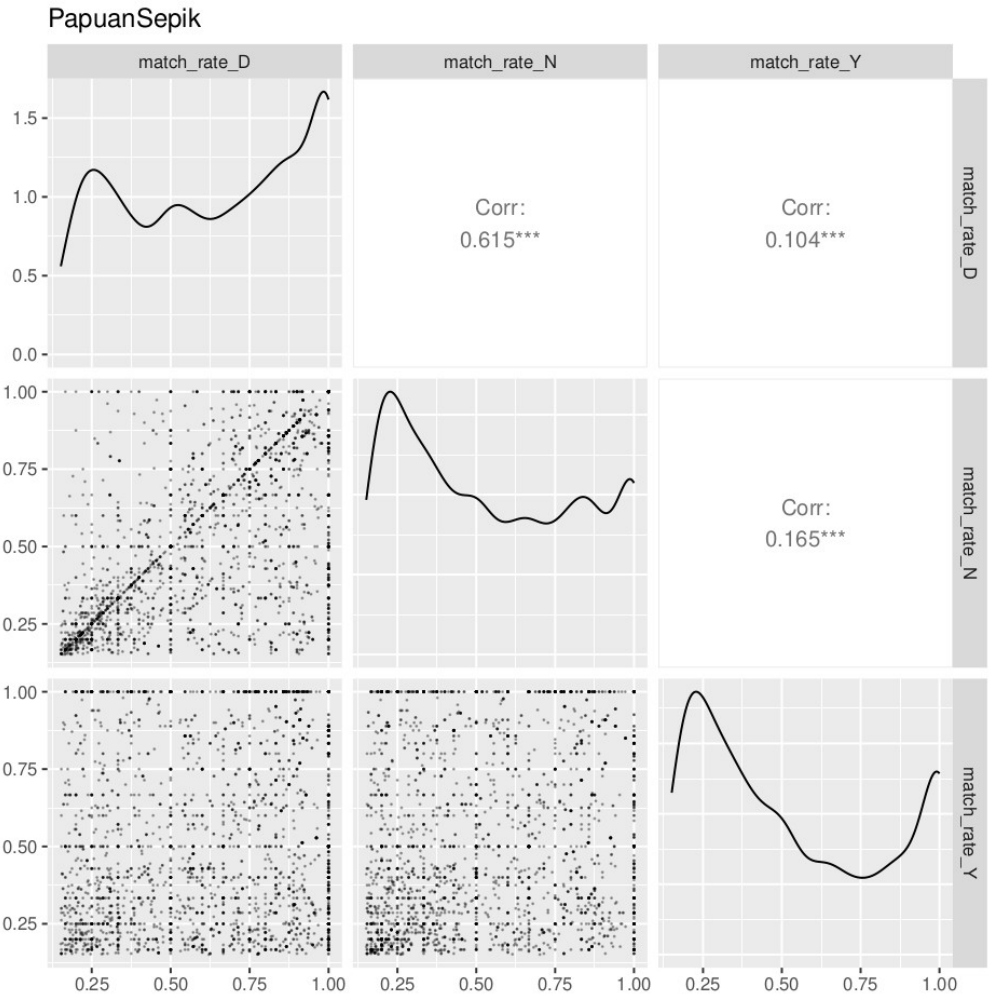
Appendix 1.15: 3d and 2d scatter plots of the match rates to Denisova, Neanderthal and Yoruba for Papuan Highlands in New Guinea (PapuanHighlands).



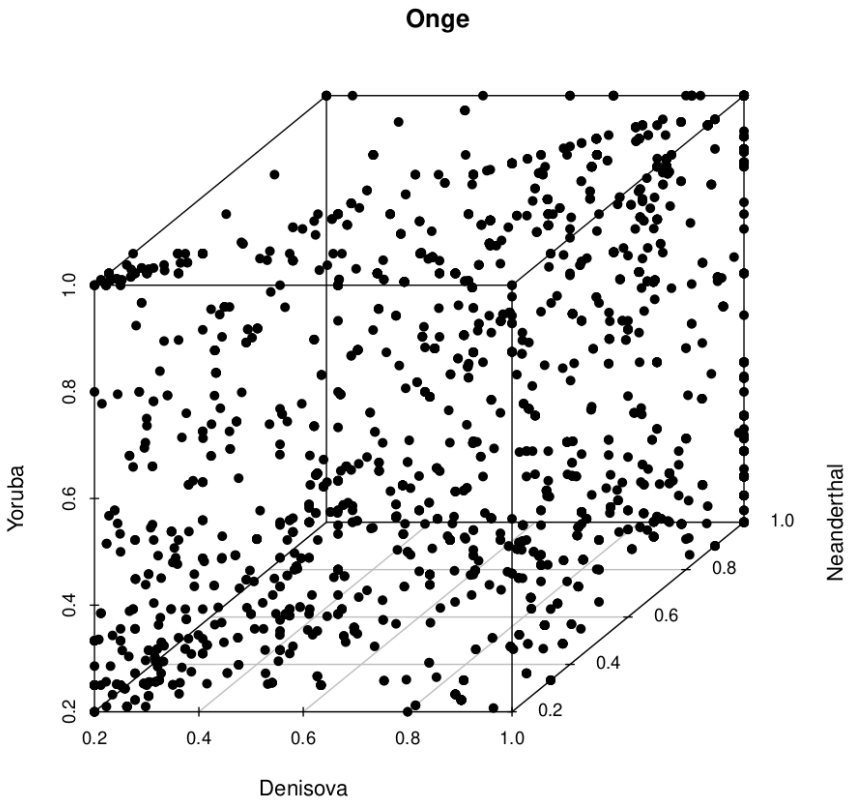

PapuanHighlands
match_rate_D
match_rate_N
match_rate_Y
Corr:
0.657***
Corr:
0.148***
Corr:
0.161***
1.5
1.0
0.5
0.0
1.00
0.75
0.50
0.25
0.25
0.50
0.75
1.00

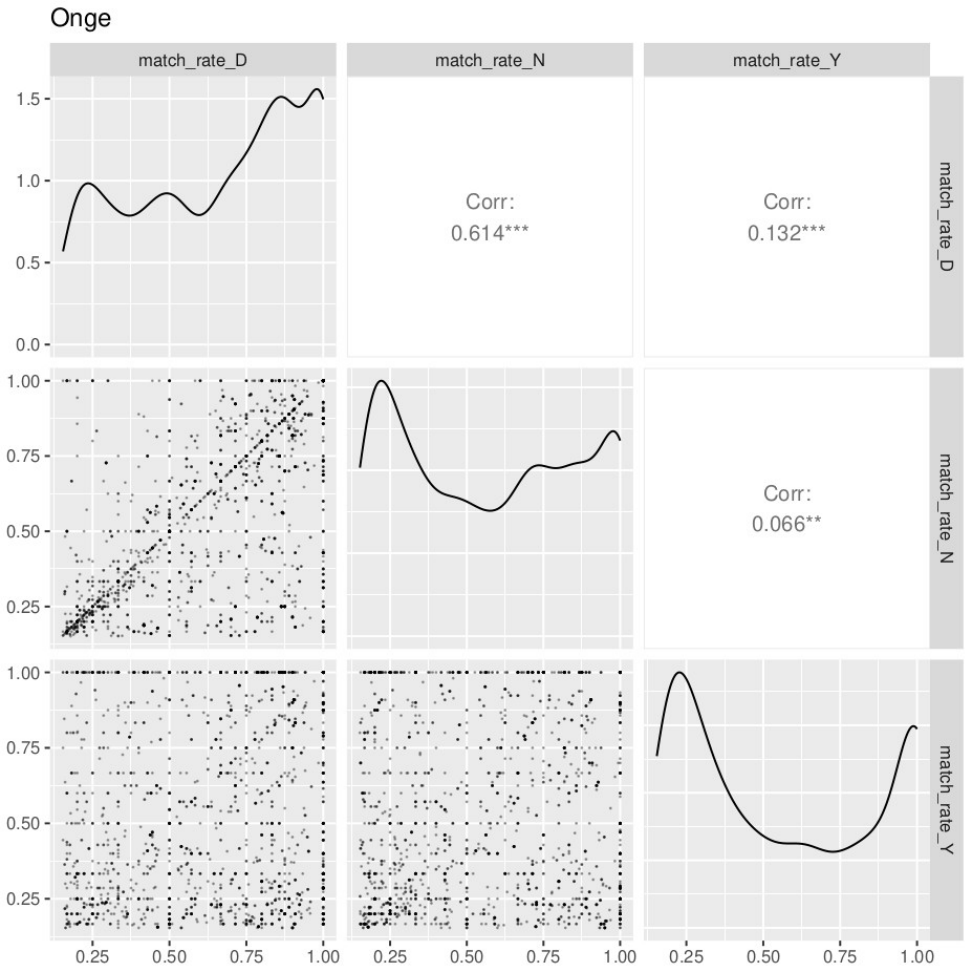
Appendix 1.16: 3d and 2d scatter plots of the match rates to Denisova, Neanderthal and Yoruba for Papuan Sepik in New Guinea (PapuanSepik).



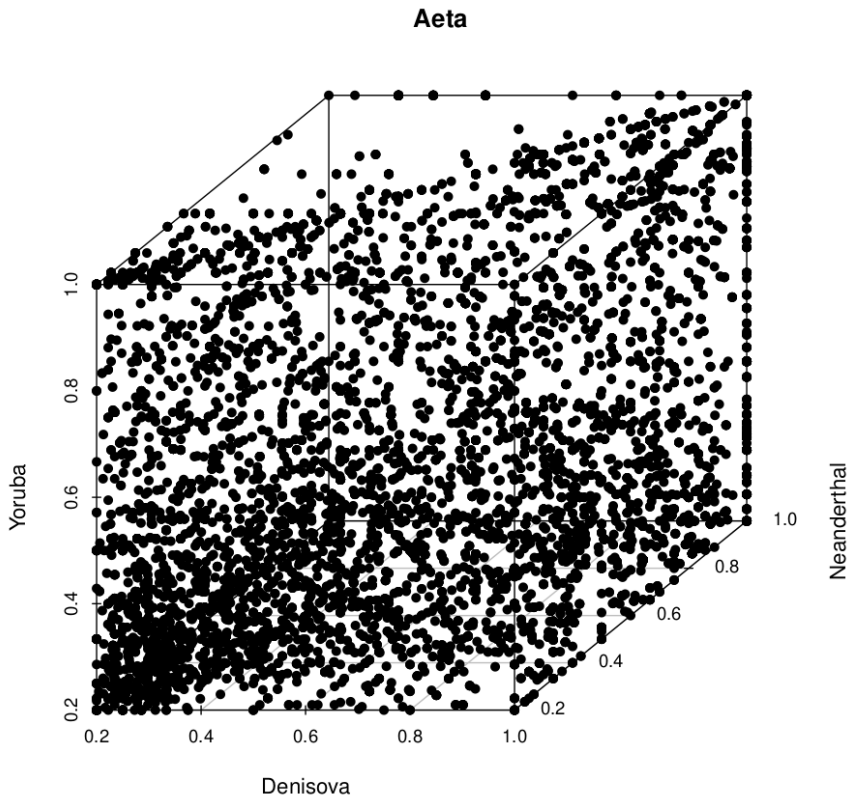

PapuanSepik
match_rate_D
match_rate_N
match_rate_Y
Corr:
0.615***
Corr:
0.104***
Corr:
0.165***
1.5
1.0
0.5
0.0
1.00
0.75
0.50
0.25
0.25
0.50
0.75
1.00

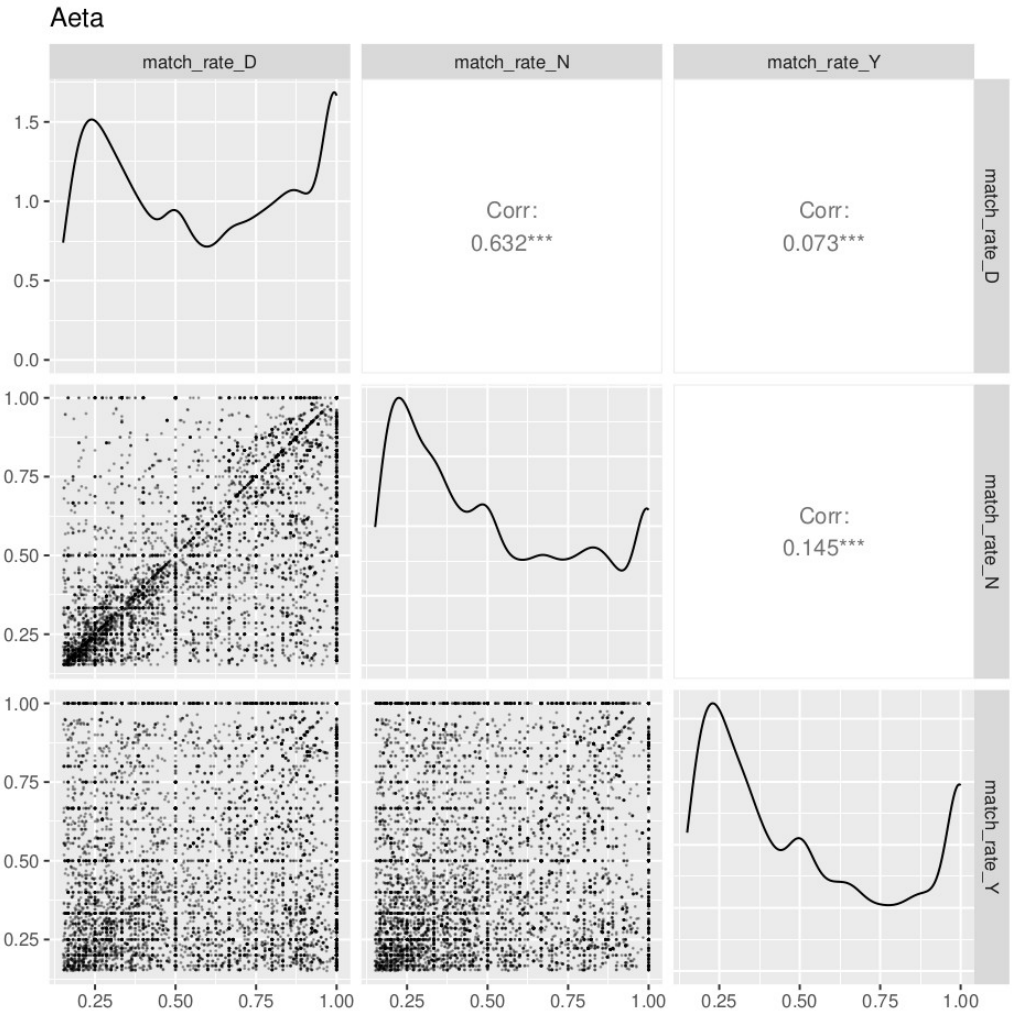
Appendix 1.17: 3d and 2d scatter plots of the match rates to Denisova, Neanderthal and Yoruba for Onge in Andaman Islands.



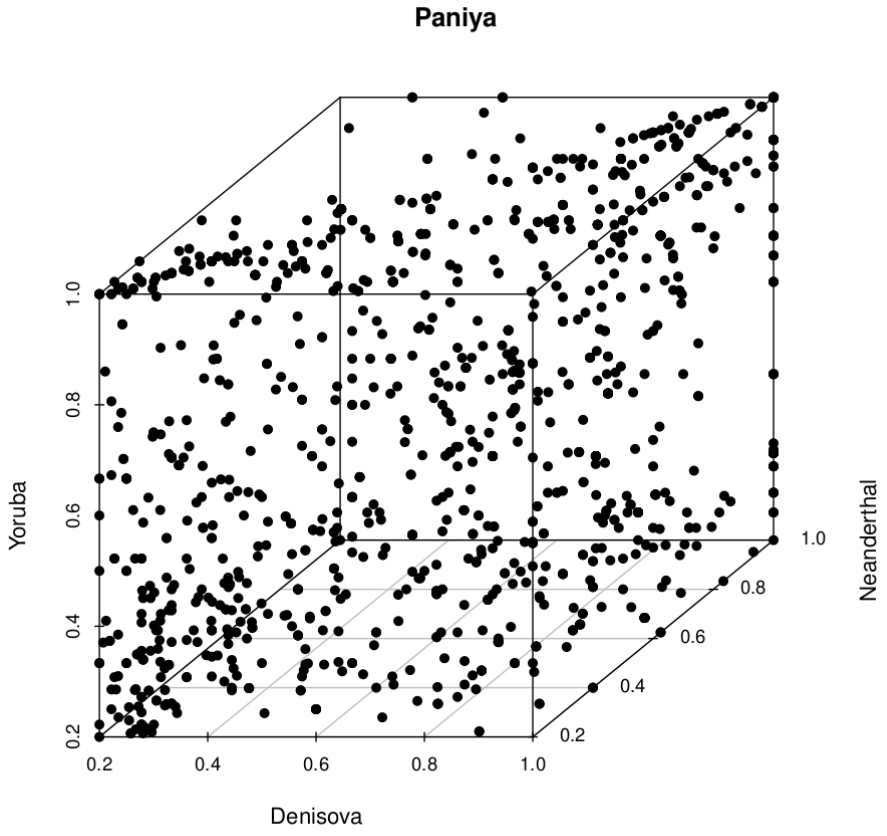

Onge
match_rate_D
match_rate_N
match_rate_Y
Corr:
0.614***
Corr:
0.132***
Corr:
0.066**
1.5
1.0
0.5
0.0
1.00
0.75
0.50
0.25
0.25
0.50
0.75
1.00

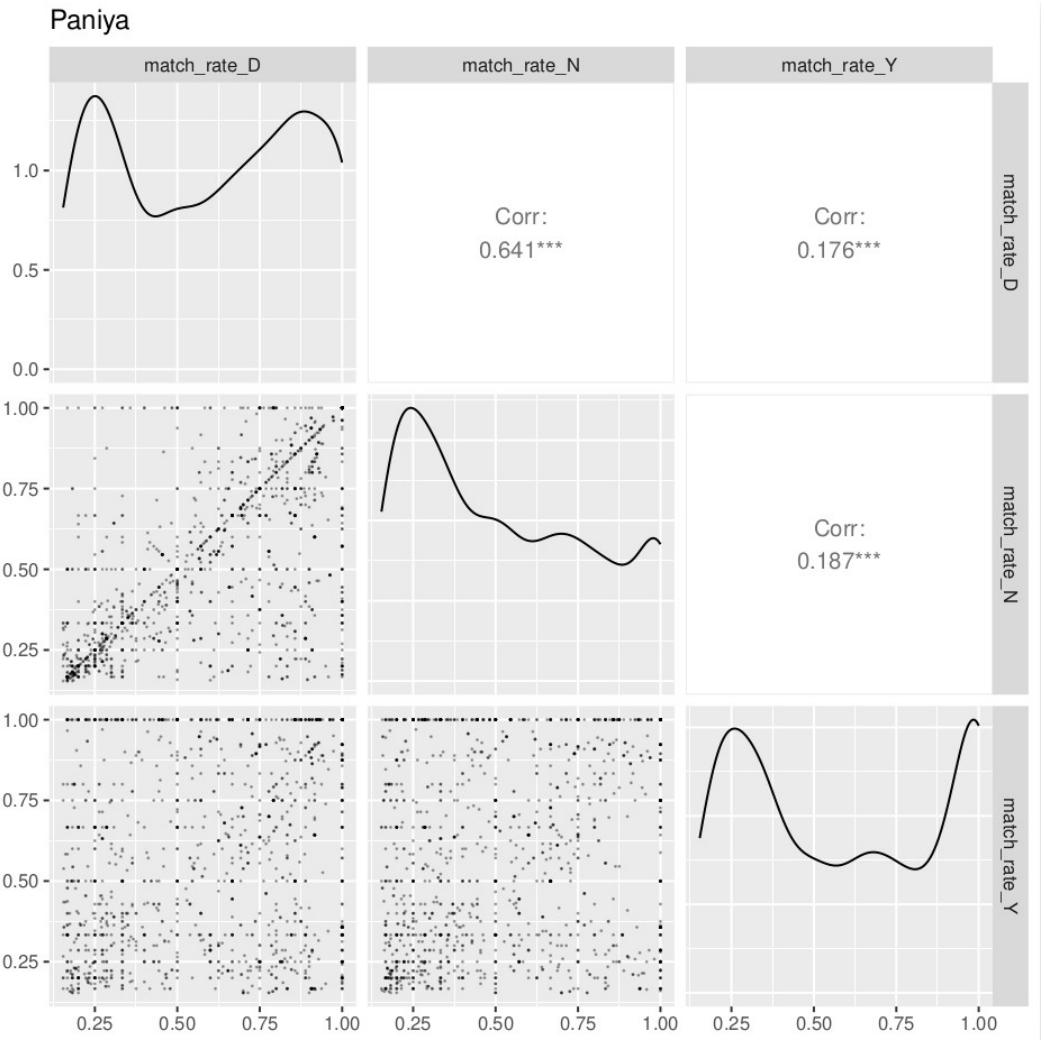
Appendix 1.18: 3d and 2d scatter plots of the match rates to Denisova, Neanderthal and Yoruba for Aeta in the Philippines.




Aeta
match_rate_D
match_rate_N
match_rate_Y
Corr:
0.632***
Corr:
0.073***
Corr:
0.145***
1.5
1.0
0.5
0.0
1.00
0.75
0.50
0.25
0.25
0.50
0.75
1.00

Appendix 1.19: 3d and 2d scatter plots of the match rates to Denisova, Neanderthal and Yoruba for Paniya in India.

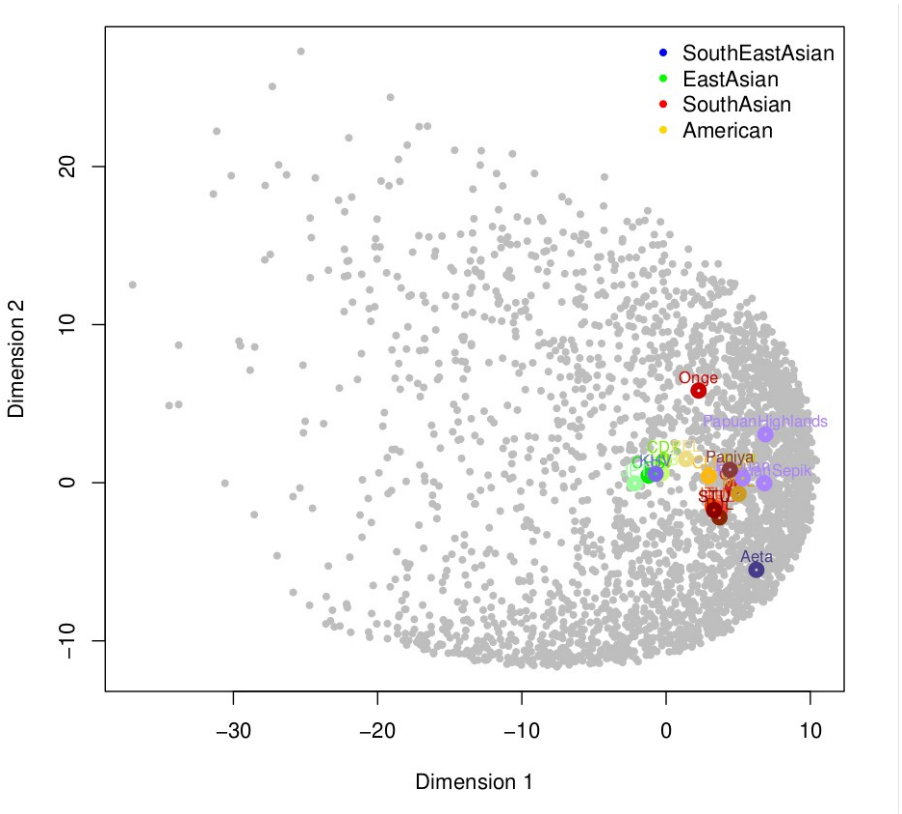



Paniya
match_rate_D
match_rate_N
match_rate_Y
Corr:
0.641***
Corr:
0.176***
Corr:
0.187***
1.0
0.5
0.0
1.00
0.75
0.50
0.25
0.25
0.50
0.75
1.00

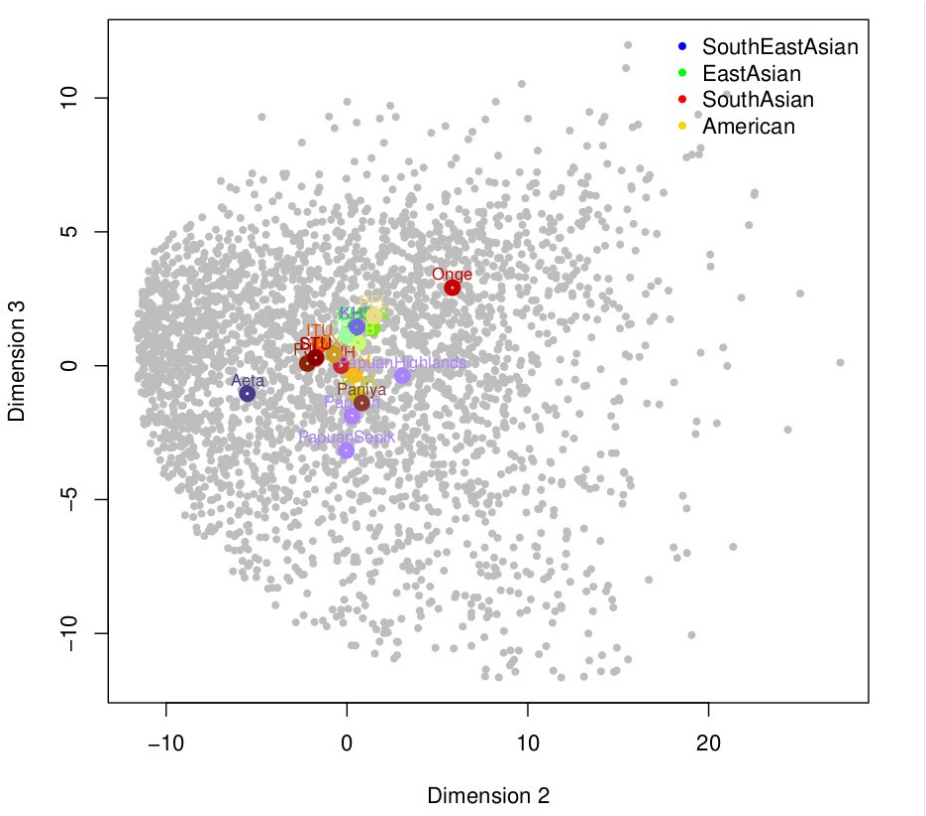
6.2 Appendix 2

Multidimensional Scatter Plots for all populations for the first and second dimensions on the left and for the second and third dimensions on the right. Populations' centroids are color coded according to the broad geographic region of provenance: blue for South East Asia, green for East Asia, red for South Asia, yellow for America.

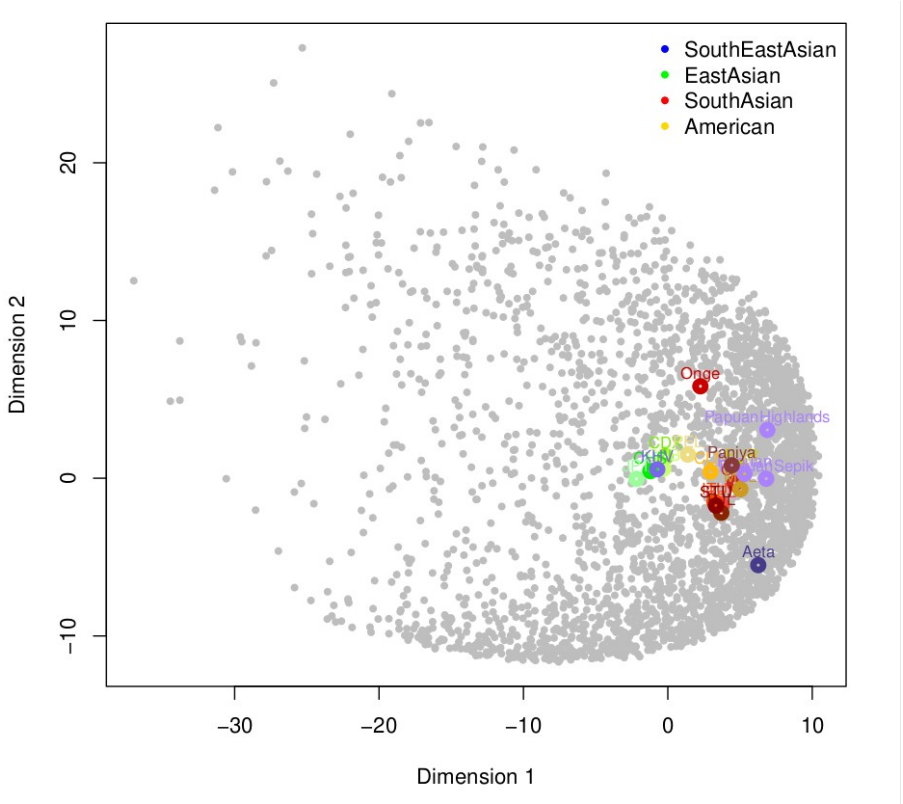
Appendix 2.1: Multidimensional Scaling plot for data divided in 10 bins, with match rate to Neanderthal <0.4 and for regions matching to the Denisovan genome at least 100 thousand bases long; first and second dimensions.



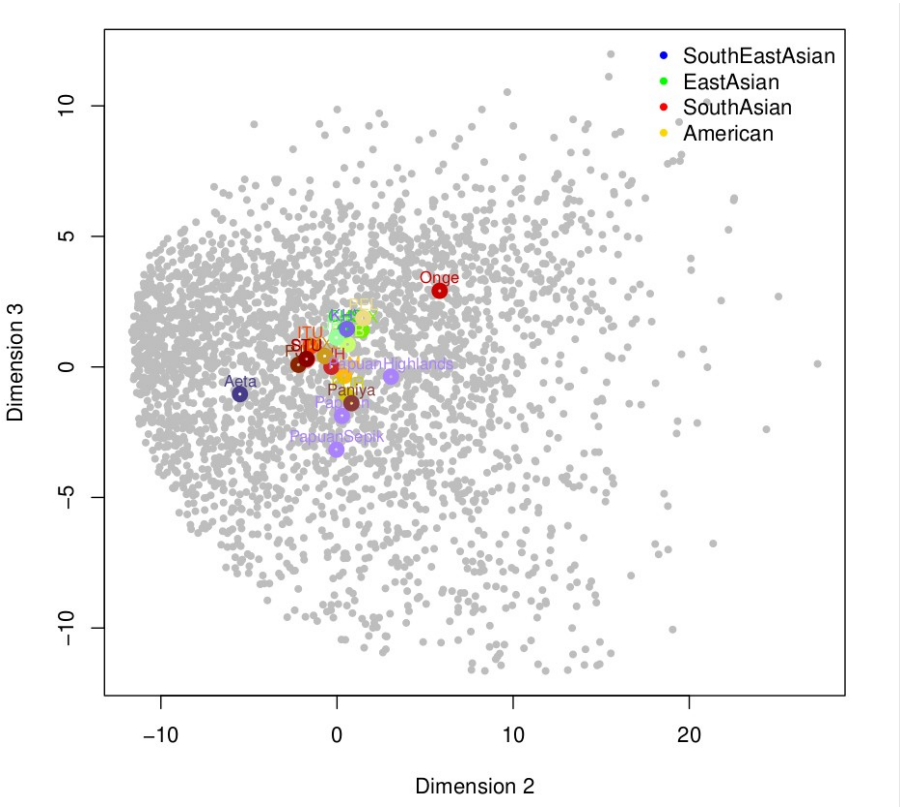
Appendix 2.2: Multidimensional Scaling plot for data divided in 10 bins, with match rate to Neanderthal <0.4 and for regions matching to the Denisovan genome at least 100 thousand bases long; second and third dimensions.



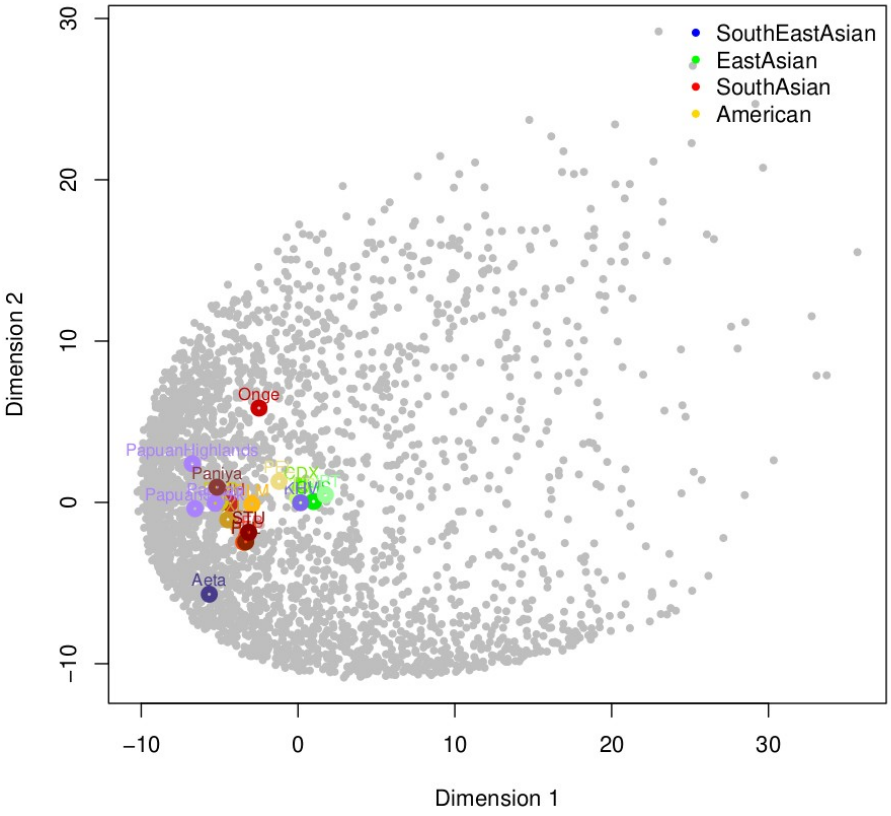
Appendix 2.3: Multidimensional scaling plot for data divided in 10 bins, with match rate to Neanderthal <0.5 and for regions matching to the Denisovan genome at least 100 thousand bases long; first and second dimensions.



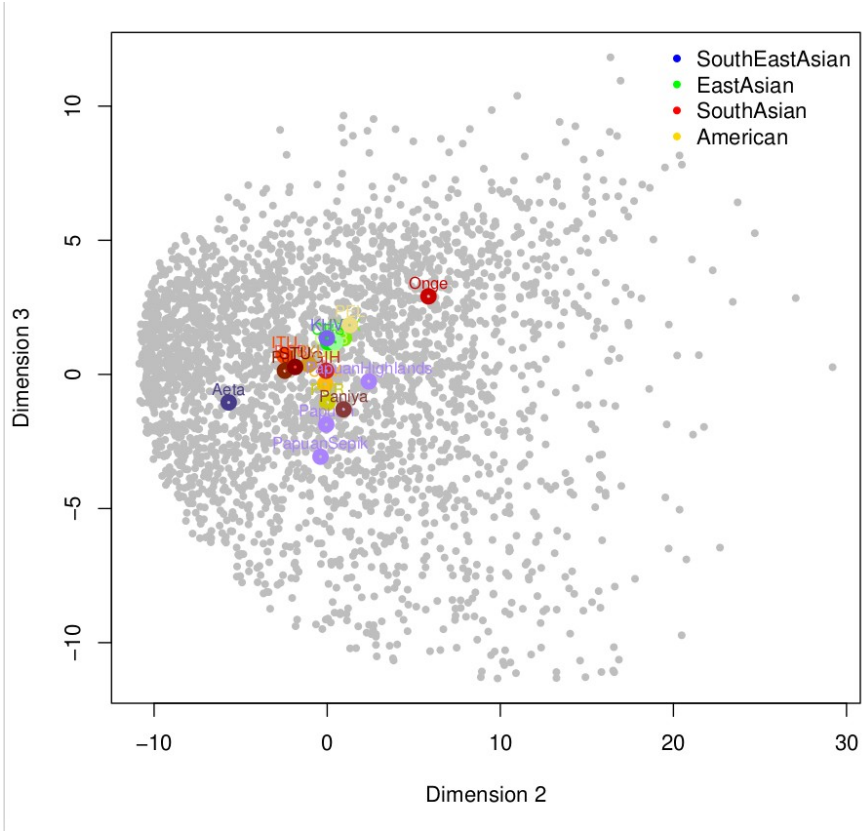
Appendix 2.4: Multidimensional scaling plot for data divided in 10 bins, with match rate to Neanderthal <0.5 and for regions matching to the Denisovan genome at least 100 thousand bases long; second and third dimensions.



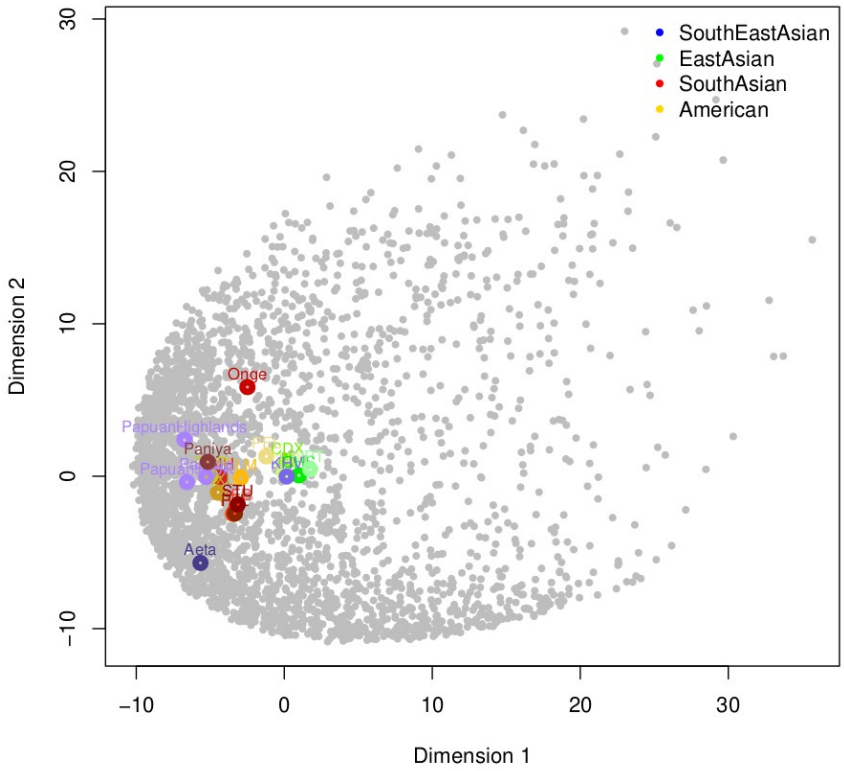
Appendix 2.5: Multidimensional scaling plot for data divided in 10 bins, with match rate to Neanderthal <0.3 and for regions matching to the Denisovan genome at least 500 thousand bases long; first and second dimensions.



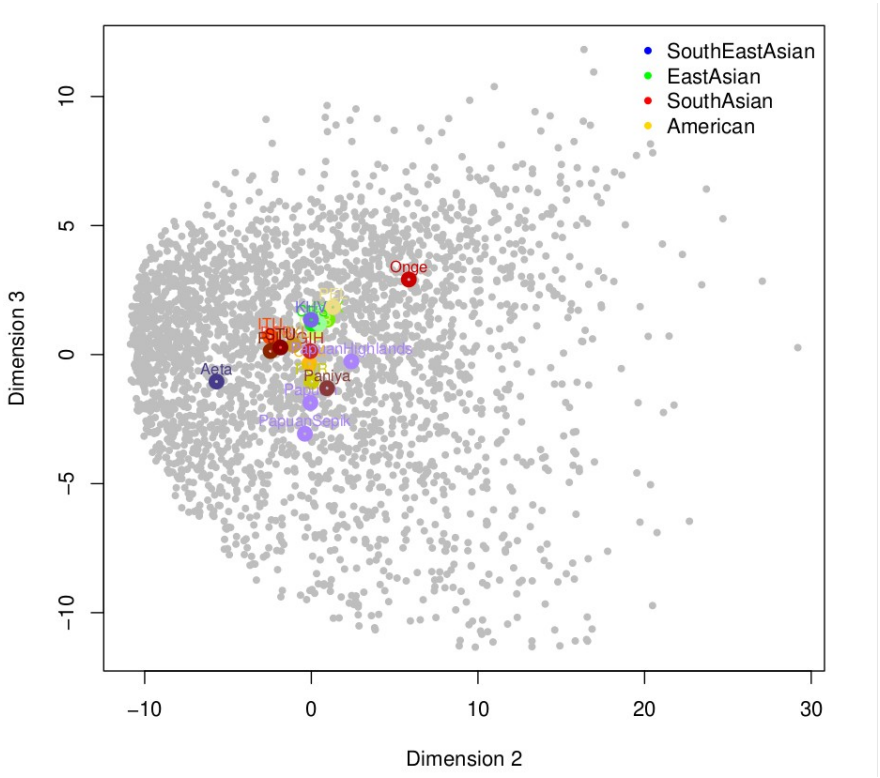
Appendix 2.6: Multidimensional scaling plot for data divided in 10 bins, with match rate to Neanderthal <0.3 and for regions matching to the Denisovan genome at least 500 thousand bases long; second and third dimensions.



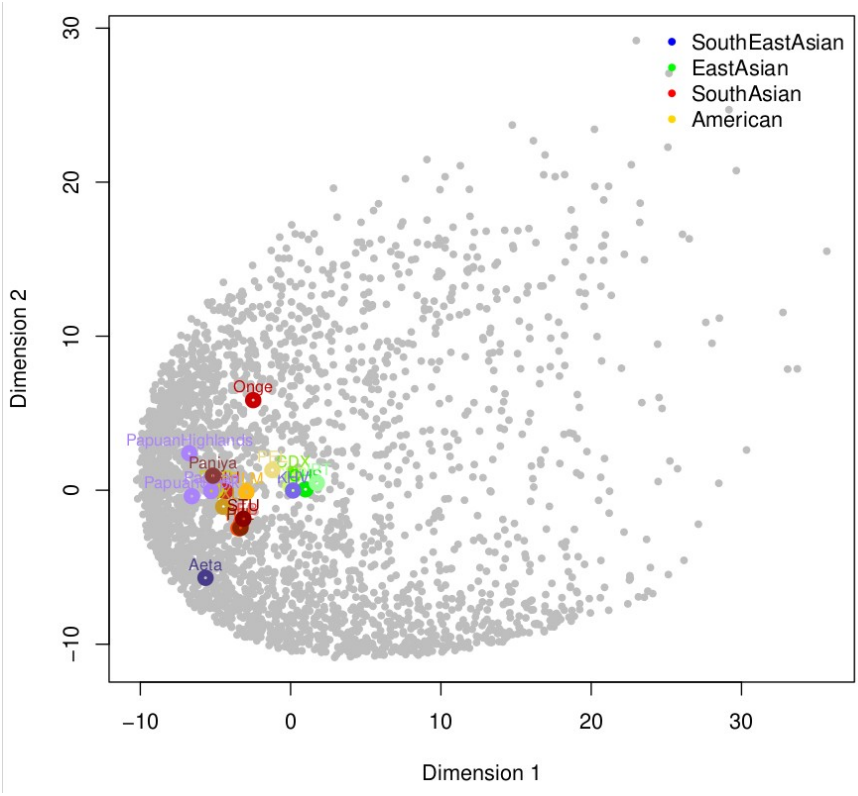
Appendix 2.7: Multidimensional scaling plot for data divided in 10 bins, with match rate to Neanderthal <0.4 and for regions matching to the Denisovan genome at least 500 thousand bases long; first and second dimensions.



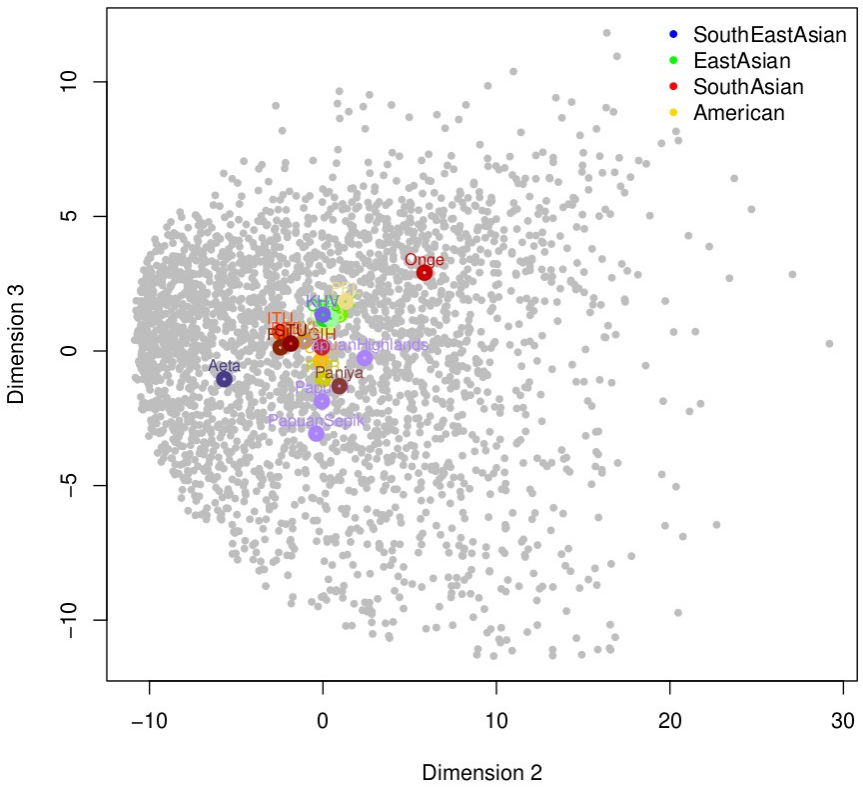
Appendix 2.8: Multidimensional scaling plot for data divided in 10 bins, with match rate to Neanderthal <0.4 and for regions matching to the Denisovan genome at least 500 thousand bases long; second and third dimensions.



Appendix 2.9: Multidimensional scaling plot for data divided in 10 bins, with match rate to Neanderthal <0.5 and for regions matching to the Denisovan genome at least 500 thousand bases long; first and second dimensions.



Appendix 2.10: Multidimensional scaling plot for data divided in 10 bins, with match rate to Neanderthal <0.5 and for regions matching to the Denisovan genome at least 500 thousand bases long; second and third dimensions.



7. Bibliography

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*(7571), 68–74.

Baba, H., Aziz, F., Kaifu, Y., Suwa, G., Kono, R. T., & Jacob, T. (2003). Homo erectus calvarium from the Pleistocene of Java. *Science*, *299*(5611), 1384–1388.

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., … Zheng, Y. T. (2010). Natural selection on EPAS1 (HIF2alpha) associated with low hemoglobin concentration in Tibetan highlanders. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(25), 11459–11464.

Bennett, E. A., Crevecoeur, I., Viola, B., Derevianko, A. P., Shunkov, M. V., Grange, T., Maureille, B., & Geigl, E.-M. (2019). Morphology of the Denisovan phalanx closer to

modern humans than to Neanderthals. *Science Advances*, *5*(9), eaaw3950.

Bergström, A., McCarthy, S. A., Hui, R., Almarri, M. A., Ayub, Q., Danecek, P., Chen, Y., Felkel, S., Hallast, P., Kamm, J., Blanché, H., Deleuze, J.-F., Cann, H., Mallick, S., Reich, D., Sandhu, M. S., Skoglund, P., Scally, A., Xue, Y., … Tyler-Smith, C. (2020). Insights into human genetic variation and population history from 929 diverse genomes. *Science*, *367*(6484). https://doi.org/10.1126/science.aay5012

Bischoff, J. L., Williams, R. W., Rosenbauer, R. J., Aramburu, A., Arsuaga, J. L., García, N., & Cuenca-Bescós, G. (2007). High-resolution U-series dates from the Sima de los Huesos hominids yields : implications for the evolution of the early Neanderthal lineage. *Journal of Archaeological Science*, *34*(5), 763–770.

Brantingham, P. J., & Xing, G. (2006). Peopling of the northern Tibetan plateau. *World Archaeology*, *38*(3), 387–414.

Browning, S. R., Browning, B. L., Zhou, Y., Tucci, S., & Akey, J. M. (2018). Analysis of Human Sequence Data Reveals Two Pulses of Archaic Denisovan Admixture. *Cell*, *173*(1), 53–61.e9.

Brown, P., Sutikna, T., Morwood, M. J., Soejono, R. P., Jatmiko,

Saptomo, E. W., & Due, R. A. (2004). A new small-bodied hominin from the Late Pleistocene of Flores, Indonesia. *Nature*, *431*(7012), 1055–1061.

Brown, S., Higham, T., Slon, V., Pääbo, S., Meyer, M., Douka, K., Brock, F., Comeskey, D., Procopio, N., Shunkov, M., Derevianko, A., & Buckley, M. (2016). Identification of a new hominin bone from Denisova Cave, Siberia using collagen fingerprinting and mitochondrial DNA analysis. *Scientific Reports*, *6*, 23559.

Brown, S., Massilani, D., Kozlikin, M. B., Shunkov, M. V., Derevianko, A. P., Stoessel, A., Jope-Street, B., Meyer, M., Kelso, J., Pääbo, S., Higham, T., & Douka, K. (2022). The earliest Denisovans and their cultural adaptation. *Nature Ecology & Evolution*, *6*(1), 28–35.

Burbano, H. A., Hodges, E., Green, R. E., Briggs, A. W., Krause, J., Meyer, M., Good, J. M., Maricic, T., Johnson, P. L. F., Xuan, Z., Rooks, M., Bhattacharjee, A., Brizuela, L., Albert, F. W., de la Rasilla, M., Fortea, J., Rosas, A., Lachmann, M., Hannon, G. J., & Pääbo, S. (2010). Targeted investigation of the Neandertal genome by array-based sequence capture. *Science*, *328*(5979), 723–725.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S.

M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, *4*, 7.

Chen, F., Welker, F., Shen, C.-C., Bailey, S. E., Bergmann, I., Davis, S., Xia, H., Wang, H., Fischer, R., Freidline, S. E., Yu, T.-L., Skinner, M. M., Stelzer, S., Dong, G., Fu, Q., Dong, G., Wang, J., Zhang, D., & Hublin, J.-J. (2019). A late Middle Pleistocene Denisovan mandible from the Tibetan Plateau. *Nature*, *569*(7756), 409–412.

Chen, L., Wolf, A. B., Fu, W., Li, L., & Akey, J. M. (2020). Identifying and Interpreting Apparent Neanderthal Ancestry in African Individuals. *Cell*, *180*(4), 677–687.e16.

Delaneau, O., Zagury, J.-F., Robinson, M. R., Marchini, J. L., & Dermitzakis, E. T. (2019). Accurate, scalable and integrative haplotype estimation. *Nature Communications*, *10*(1), 1–10.

Demeter, F., Zanolli, C., Westaway, K. E., Joannes-Boyau, R., Duringer, P., Morley, M. W., Welker, F., Rüther, P. L., Skinner, M. M., McColl, H., Gaunitz, C., Vinner, L., Dunn, T. E., Olsen, J. V., Sikora, M., Ponche, J.-L., Suzzoni, E., Frangeul, S., Boesch, Q., … Shackelford, L. (2022). A Middle Pleistocene Denisovan molar from the Annamite Chain of northern Laos. *Nature Communications*, *13*(1), 2557.

Di Vincenzo, F., & Manzi, G. (2023). Homo heidelbergensis as the Middle Pleistocene common ancestor of Denisovans, Neanderthals and modern humans. *Journal of Mediterranean Earth Sciences*, *15*. https://doi.org/10.13133/2280-6148/18074

Fisher, S. E., & Scharff, C. (2009). FOXP2 as a molecular window into speech and language. *Trends in Genetics: TIG*, *25*(4), 166–177.

Freidline, S. E., Westaway, K. E., Joannes-Boyau, R., Duringer, P., Ponche, J.-L., Morley, M. W., Hernandez, V. C., McAllister-Hayward, M. S., McColl, H., Zanolli, C., Gunz, P., Bergmann, I., Sichanthongtip, P., Sihanam, D., Boualaphane, S., Luangkhoth, T., Souksavatdy, V., Dosseto, A., Boesch, Q., … Demeter, F. (2023). Early presence of Homo sapiens in Southeast Asia by 86-68 kyr at Tam Pà Ling, Northern Laos. *Nature Communications*, *14*(1), 3193.

Fu, Q., Meyer, M., Gao, X., Stenzel, U., Burbano, H. A., Kelso, J., & Pääbo, S. (2013). DNA analysis of an early modern human from Tianyuan Cave, China. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(6), 2223–2227.

Fu, Q., Mittnik, A., Johnson, P. L. F., Bos, K., Lari, M.,

Bollongino, R., Sun, C., Giemsch, L., Schmitz, R., Burger, J., Ronchitelli, A. M., Martini, F., Cremonesi, R. G., Svoboda, J., Bauer, P., Caramelli, D., Castellano, S., Reich, D., Pääbo, S., & Krause, J. (2013). A revised timescale for human evolution based on ancient mitochondrial genomes. *Current Biology: CB*, *23*(7), 553–559.

Gabunia, L., Vekua, A., & Lordkipanidze, D. (2000). The environmental contexts of early human occupation of Georgia (Transcaucasia). *Journal of Human Evolution*, *38*(6), 785–802.

GenomeAsia100K Consortium. (2019). The GenomeAsia 100K Project enables genetic discoveries across Asia. *Nature*, *576*(7785), 106–111.

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., … Pääbo, S. (2010). A draft sequence of the Neandertal genome. *Science*, *328*(5979), 710–722.

Groucutt, H. S., Grün, R., Zalmout, I. A. S., Drake, N. A., Armitage, S. J., Candy, I., Clark-Wilson, R., Louys, J., Breeze, P. S., Duval, M., Buck, L. T., Kivell, T. L., Pomeroy, E.,

Stephens, N. B., Stock, J. T., Stewart, M., Price, G. J., Kinsley, L., Sung, W. W., … Petraglia, M. D. (2018). Homo sapiens in Arabia by 85,000 years ago. *Nature Ecology & Evolution*, *2*(5), 800–809.

Groucutt, H. S., White, T. S., Scerri, E. M. L., Andrieux, E., Clark-Wilson, R., Breeze, P. S., Armitage, S. J., Stewart, M., Drake, N., Louys, J., Price, G. J., Duval, M., Parton, A., Candy, I., Carleton, W. C., Shipton, C., Jennings, R. P., Zahir, M., Blinkhorn, J., … Petraglia, M. D. (2021). Multiple hominin dispersals into Southwest Asia over the past 400,000 years. *Nature*, *597*(7876), 376–380.

Grün, R., Stringer, C., McDermott, F., Nathan, R., Porat, N., Robertson, S., Taylor, L., Mortimer, G., Eggins, S., & McCulloch, M. (2005). U-series and ESR analyses of bones and teeth relating to the human burials from Skhul. *Journal of Human Evolution*, *49*(3), 316–334.

Hajdinjak, M., Fu, Q., Hübner, A., Petr, M., Mafessoni, F., Grote, S., Skoglund, P., Narasimham, V., Rougier, H., Crevecoeur, I., Semal, P., Soressi, M., Talamo, S., Hublin, J.-J., Gušić, I., Kućan, Ž., Rudan, P., Golovanova, L. V., Doronichev, V. B., … Kelso, J. (2018). Reconstructing the genetic history of late Neanderthals. *Nature*, *555*(7698), 652–656.

Harvati, K., Röding, C., Bosman, A. M., Karakostis, F. A., Grün, R., Stringer, C., Karkanas, P., Thompson, N. C., Koutoulidis, V., Moulopoulos, L. A., Gorgoulis, V. G., & Kouloukoussa, M. (2019). Apidima Cave fossils provide earliest evidence of Homo sapiens in Eurasia. *Nature*, *571*(7766), 500–504.

Hershkovitz, I., Duval, M., Grün, R., Mercier, N., Valladas, H., Ayalon, A., Bar-Matthews, M., Weber, G. W., Quam, R., Zaidner, Y., & Weinstein-Evron, M. (2018). Response to Comment on “The earliest modern humans outside Africa.” *Science*, *362*(6413). https://doi.org/10.1126/science.aat8964

Higham, T., Douka, K., Wood, R., Ramsey, C. B., Brock, F., Basell, L., Camps, M., Arrizabalaga, A., Baena, J., Barroso-Ruíz, C., Bergman, C., Boitard, C., Boscato, P., Caparrós, M., Conard, N. J., Draily, C., Froment, A., Galván, B., Gambassini, P., … Jacobi, R. (2014). The timing and spatiotemporal patterning of Neanderthal disappearance. *Nature*, *512*(7514), 306–309.

Huang, X., Kruisz, P., & Kuhlwilm, M. (2022). sstar: A Python Package for Detecting Archaic Introgression from Population Genetic Data with S. *Molecular Biology and Evolution*, *39*(11). https://doi.org/10.1093/molbev/msac212

Hubisz, M. J., Williams, A. L., & Siepel, A. (2020). Mapping gene

flow between ancient hominins through demography-aware inference of the ancestral recombination graph. *PLoS Genetics*, *16*(8), e1008895.

Hublin, J. J. (2009). Out of Africa: modern human origins special feature: the origin of Neandertals. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(38), 16022–16027.

Huerta-Sánchez, E., Jin, X., Asan, Bianba, Z., Peter, B. M., Vinckenbosch, N., Liang, Y., Yi, X., He, M., Somel, M., Ni, P., Wang, B., Ou, X., Huasang, Luosang, J., Cuo, Z. X. P., Li, K., Gao, G., Yin, Y., … Nielsen, R. (2014). Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature*, *512*(7513), 194–197.

Jacobs, G. S., Hudjashov, G., Saag, L., Kusuma, P., Darusallam, C. C., Lawson, D. J., Mondal, M., Pagani, L., Ricaut, F.-X., Stoneking, M., Metspalu, M., Sudoyo, H., Lansing, J. S., & Cox, M. P. (2019). Multiple Deeply Divergent Denisovan Ancestries in Papuans. *Cell*, *177*(4), 1010–1021.e32.

Jacobs, Z., Li, B., Shunkov, M. V., Kozlikin, M. B., Bolikhovskaya, N. S., Agadjanian, A. K., Uliyanov, V. A., Vasiliev, S. K., O'Gorman, K., Derevianko, A. P., & Roberts, R. G. (2019). Timing of archaic hominin occupation of Denisova Cave in

southern Siberia. *Nature*, *565*(7741), 594–599.

King, W. (n.d.). *The Reputed Fossil Man of the Neanderthal*.

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*(7290), 894–897.

Krause, J., Lalueza-Fox, C., Orlando, L., Enard, W., Green, R. E., Burbano, H. A., Hublin, J.-J., Hänni, C., Fortea, J., de la Rasilla, M., Bertranpetit, J., Rosas, A., & Pääbo, S. (2007). The derived FOXP2 variant of modern humans was shared with Neandertals. *Current Biology: CB*, *17*(21), 1908–1912.

Lai, C. S., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F., & Monaco, A. P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature*, *413*(6855), 519–523.

Larena, M., McKenna, J., Sanchez-Quinto, F., Bernhardsson, C., Ebeo, C., Reyes, R., Casel, O., Huang, J.-Y., Hagada, K. P., Guilay, D., Reyes, J., Allian, F. P., Mori, V., Azarcon, L. S., Manera, A., Terando, C., Jamero, L., Jr, Sireg, G., Manginsay-Tremedal, R., … Jakobsson, M. (2021). Philippine Ayta possess the highest level of Denisovan ancestry in the world. *Current Biology: CB*, *31*(19), 4219–4230.e10.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* , *27*(21), 2987–2993.

Malaspinas, A.-S., Westaway, M. C., Muller, C., Sousa, V. C., Lao, O., Alves, I., Bergström, A., Athanasiadis, G., Cheng, J. Y., Crawford, J. E., Heupink, T. H., Macholdt, E., Peischl, S., Rasmussen, S., Schiffels, S., Subramanian, S., Wright, J. L., Albrechtsen, A., Barbieri, C., … Willerslev, E. (2016). A genomic history of Aboriginal Australia. *Nature*, *538*(7624), 207–214.

Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., Tandon, A., Skoglund, P., Lazaridis, I., Sankararaman, S., Fu, Q., Rohland, N., Renaud, G., Erlich, Y., Willems, T., Gallo, C., … Reich, D. (2016). The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature*, *538*(7624), 201–206.

Maricic, T., Günther, V., Georgiev, O., Gehre, S., Curlin, M., Schreiweis, C., Naumann, R., Burbano, H. A., Meyer, M., Lalueza-Fox, C., de la Rasilla, M., Rosas, A., Gajovic, S., Kelso, J., Enard, W., Schaffner, W., & Pääbo, S. (2013). A

recent evolutionary change affects a regulatory element in the human FOXP2 gene. *Molecular Biology and Evolution*, *30*(4), 844–852.

Massilani, D., Skov, L., Hajdinjak, M., Gunchinsuren, B., Tseveendorj, D., Yi, S., Lee, J., Nagel, S., Nickel, B., Devièse, T., Higham, T., Meyer, M., Kelso, J., Peter, B. M., & Pääbo, S. (2020). Denisovan ancestry and population history of early East Asians. *Science*, *370*(6516), 579–583.

Meyer, M., Arsuaga, J.-L., de Filippo, C., Nagel, S., Aximu-Petri, A., Nickel, B., Martínez, I., Gracia, A., de Castro, J. M. B., Carbonell, E., Viola, B., Kelso, J., Prüfer, K., & Pääbo, S. (2016). Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature*, *531*(7595), 504–507.

Meyer, M., Kircher, M., Gansauge, M.-T., Li, H., Racimo, F., Mallick, S., Schraiber, J. G., Jay, F., Prüfer, K., de Filippo, C., Sudmant, P. H., Alkan, C., Fu, Q., Do, R., Rohland, N., Tandon, A., Siebauer, M., Green, R. E., Bryc, K., … Pääbo, S. (2012). A high-coverage genome sequence from an archaic Denisovan individual. *Science*, *338*(6104), 222–226.

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. *American Journal of Human Genetics*, *96*(6), 986–991.

Petr, M., Hajdinjak, M., Fu, Q., Essel, E., Rougier, H., Crevecoeur, I., Semal, P., Golovanova, L. V., Doronichev, V. B., Lalueza-Fox, C., de la Rasilla, M., Rosas, A., Shunkov, M. V., Kozlikin, M. B., Derevianko, A. P., Vernot, B., Meyer, M., & Kelso, J. (2020). The evolutionary history of Neanderthal and Denisovan Y chromosomes. *Science*, *369*(6511), 1653–1656.

Petr, M., Pääbo, S., Kelso, J., & Vernot, B. (2019). Limits of long-term selection against Neandertal introgression. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(5), 1639–1644.

Plagnol, V., & Wall, J. D. (2006). Possible ancestral structure in human populations. *PLoS Genetics*, *2*(7), e105.

Pope, G.G. (1992), Craniofacial evidence for the origin of modern humans in China. *American Journal of Physical Anthropol*gy 35: 243-298. https://doi.org/10.1002/ajpa.1330350610

Posth, C., Wißing, C., Kitagawa, K., Pagani, L., van Holstein, L.,

Racimo, F., Wehrberger, K., Conard, N. J., Kind, C. J., Bocherens, H., & Krause, J. (2017). Deeply divergent archaic mitochondrial genome provides lower time boundary for African gene flow into Neanderthals. *Nature Communications*, *8*, 16046.

Prüfer, K., de Filippo, C., Grote, S., Mafessoni, F., Korlević, P., Hajdinjak, M., Vernot, B., Skov, L., Hsieh, P., Peyrégne, S., Reher, D., Hopfe, C., Nagel, S., Maricic, T., Fu, Q., Theunert, C., Rogers, R., Skoglund, P., Chintalapati, M., … Pääbo, S. (2017). A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science*, *358*(6363), 655–658.

Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., Heinze, A., Renaud, G., Sudmant, P. H., de Filippo, C., Li, H., Mallick, S., Dannemann, M., Fu, Q., Kircher, M., Kuhlwilm, M., Lachmann, M., Meyer, M., Ongyerth, M., … Pääbo, S. (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*, *505*(7481), 43–49.

Qin, P., & Stoneking, M. (2015). Denisovan Ancestry in East Eurasian and Native American Populations. *Molecular Biology and Evolution*, *32*(10), 2665–2674.

Raghavan, M., Skoglund, P., Graf, K. E., Metspalu, M.,

Albrechtsen, A., Moltke, I., Rasmussen, S., 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., … Willerslev, E. (2013). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature*, *505*(7481), 87–91.

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., … Pääbo, S. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*, *468*(7327), 1053–1060.

Reich, D., Patterson, N., Kircher, M., Delfin, F., Nandineni, M. R., Pugach, I., Ko, A. M.-S., Ko, Y.-C., Jinam, T. A., Phipps, M. E., Saitou, N., Wollstein, A., Kayser, M., Pääbo, S., & Stoneking, M. (2011). Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. *American Journal of Human Genetics*, *89*(4), 516–528.

Rieux, A., Eriksson, A., Li, M., Sobkowiak, B., Weinert, L. A., Warmuth, V., Ruiz-Linares, A., Manica, A., & Balloux, F. (2014). Improved calibration of the human mitochondrial

clock using ancient genomes. *Molecular Biology and Evolution*, *31*(10), 2780–2792.

Rightmire, G. P. (2004). Brain size and encephalization in early to Mid-Pleistocene Homo. *American Journal of Physical Anthropology*, *124*(2), 109–123.

Sankararaman, S., Mallick, S., Patterson, N., & Reich, D. (2016). The Combined Landscape of Denisovan and Neanderthal Ancestry in Present-Day Humans. *Current Biology: CB*, *26*(9), 1241–1247.

Sawyer, S., Renaud, G., Viola, B., Hublin, J.-J., Gansauge, M.-T., Shunkov, M. V., Derevianko, A. P., Prüfer, K., Kelso, J., & Pääbo, S. (2015). Nuclear and mitochondrial DNA sequences from two Denisovan individuals. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(51), 15696–15700.

Schiffels, S., & Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. *Nature Genetics*, *46*(8), 919–925.

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*(5987), 72–75.

Skoglund, P., & Jakobsson, M. (2011). Archaic human ancestry in East Asia. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(45), 18301–18306.

Skoglund, P., Mallick, S., Bortolini, M. C., Chennagiri, N., Hünemeier, T., Petzl-Erler, M. L., Salzano, F. M., Patterson, N., & Reich, D. (2015). Genetic evidence for two founding populations of the Americas. *Nature*, *525*(7567), 104–108.

Slon, V., Mafessoni, F., Vernot, B., de Filippo, C., Grote, S., Viola, B., Hajdinjak, M., Peyrégne, S., Nagel, S., Brown, S., Douka, K., Higham, T., Kozlikin, M. B., Shunkov, M. V., Derevianko, A. P., Kelso, J., Meyer, M., Prüfer, K., & Pääbo, S. (2018). The genome of the offspring of a Neanderthal mother and a Denisovan father. *Nature*, *561*(7721), 113–116.

Slon, V., Viola, B., Renaud, G., Gansauge, M.-T., Benazzi, S., Sawyer, S., Hublin, J.-J., Shunkov, M. V., Derevianko, A. P., Kelso, J., Prüfer, K., Meyer, M., & Pääbo, S. (2017). A fourth Denisovan individual. *Science Advances*, *3*(7), e1700186.

Soares, P., Alshamali, F., Pereira, J. B., Fernandes, V., Silva, N. M., Afonso, C., Costa, M. D., Musilová, E., Macaulay, V., Richards, M. B., Cerny, V., & Pereira, L. (2012). The

Expansion of mtDNA Haplogroup L3 within and out of Africa. *Molecular Biology and Evolution*, *29*(3), 915–927.

Stringer, C. B., & Hublin, J. (1999). New age estimates for the Swanscombe hominid, and their significance for human evolution. *Journal of Human Evolution*, *37*(6), 873–877.

Swisher, C. C., 3rd, Rink, W. J., Antón, S. C., Schwarcz, H. P., Curtis, G. H., Suprijo, A., & Widiasmoro. (1996). Latest Homo erectus of Java: potential contemporaneity with Homo sapiens in southeast Asia. *Science*, *274*(5294), 1870–1874.

Telis, N., Aguilar, R., & Harris, K. (2020). Selection against archaic hominin genetic variation in regulatory regions. *Nature Ecology & Evolution*, *4*(11), 1558–1566.

Vargha-Khadem, F., Watkins, K., Alcock, K., Fletcher, P., & Passingham, R. (1995). Praxic and nonverbal cognitive deficits in a large family with a genetically transmitted speech and language disorder. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(3), 930–933.

Viola, B. T., Gunz, P., Neubauer, S., Slon, V., Kozlikin, M. B., Shunkov, M. V., Meyer, M., Paabo, S., & Derevianko, A. P. (2019). A parietal fragment from Denisova cave. *American Journal of Physical Anthropology*, *168*, 258-258.

Witt, K. E., Villanea, F., Loughran, E., Zhang, X., & Huerta-Sanchez, E. (2022). Apportioning archaic variants among modern populations. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *377*(1852), 20200411.

Wu, X. and Poirier, F.E. (1995) Human evolution in China: A metric description of the fossils and a review of the sites. Oxford University Press, New York.

Xia, H., Zhang, D., Wang, J., Fagernäs, Z., Li, T., Li, Y., Yao, J., Lin, D., Troché, G., Smith, G. M., Chen, X., Cheng, T., Shen, X., Han, Y., Olsen, J. V., Shen, Z., Pei, Z., Hublin, J.-J., Chen, F., & Welker, F. (2024). Middle and Late Pleistocene Denisovan subsistence at Baishiya Karst Cave. *Nature*. https://doi.org/10.1038/s41586-024-07612-9

Yang, M. A., Gao, X., Theunert, C., Tong, H., Aximu-Petri, A., Nickel, B., Slatkin, M., Meyer, M., Pääbo, S., Kelso, J., & Fu, Q. (2017). 40,000-Year-Old Individual from Asia Provides Insight into Early Population Structure in Eurasia. *Current Biology: CB*, *27*(20), 3202–3208.e9.

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., … Wang, J. (2010). Sequencing of 50 human exomes reveals adaptation to high altitude. *Science*, *329*(5987), 75–78.

Zhang, X., Witt, K. E., Bañuelos, M. M., Ko, A., Yuan, K., Xu, S., Nielsen, R., & Huerta-Sanchez, E. (2021). The history and evolution of the Denisovan- haplotype in Tibetans. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(22). https://doi.org/10.1073/pnas.2020803118

Zhu, Z.-Y., Dennell, R., Huang, W.-W., Wu, Y., Rao, Z.-G., Qiu, S.-F., Xie, J.-B., Liu, W., Fu, S.-Q., Han, J.-W., Zhou, H.-Y., Ou Yang, T.-P., & Li, H.-M. (2015). New dating of the Homo erectus cranium from Lantian (Gongwangling), China. *Journal of Human Evolution*, *78*, 144–157.