Research Report

Institute of Bioinformatics, Johannes Kepler University Linz

BIOINF

Sharing of Very Short IBD Segments between Humans, Neandertals, and Denisovans

Gundula Povysil and Sepp Hochreiter

Institute of Bioinformatics, Johannes Kepler University, Linz, Austria

Abstract

We analyze the sharing of very short identity by descent (IBD) segments between humans, Neandertals, and Denisovans to gain new insights into their demographic history. Short IBD segments convey information about events far back in time because the shorter IBD segments are, the older they are assumed to be. The identification of short IBD segments becomes possible through next generation sequencing (NGS), which offers high variant density and reports variants of all frequencies. However, only recently HapFABIA has been proposed as the first method for detecting very short IBD segments in NGS data. HapFABIA utilizes rare variants to identify IBD segments with a low false discovery rate.

We applied HapFABIA to the 1000 Genomes Project whole genome sequencing data to identify IBD segments which are shared within and between populations. Some IBD segments are shared with the reconstructed ancestral genome of humans and other primates. These segments are tagged by rare variants, consequently some rare variants have to be very old. Other IBD segments are also old since they are shared with Neandertals or Denisovans, which explains their shorter lengths compared to segments that are not shared with these ancient genomes. The Denisova genome most prominently matched IBD segments that are shared by Asians. Many of these segments were found exclusively in Asians and they are longer than segments shared between other continental populations and the Denisova genome. Therefore, we could confirm an introgression from Deniosvans into ancestors of Asians after their migration out of Africa. While Neandertal-matching IBD segments are most often shared by Asians, Europeans share a considerably higher percentage of IBD segments with Neandertals compared to other populations, too. Again, many of these Neandertal-matching IBD segments are found exclusively in Asians, whereas Neandertal-matching IBD segments that are shared by Europeans are often found in other populations, too. Neandertal-matching IBD segments that are shared by Asians or Europeans are longer than those observed in Africans. This hints at a gene flow from Neandertals into ancestors of Asians and Europeans after they left Africa. Interestingly, many Neandertal- or Denisova-matching IBD segments are predominantly observed in Africans — some of them even exclusively. IBD segments shared between Africans and Neandertals or Denisovans are strikingly short, therefore we assume that they are very old. This may indicate that these segments stem from ancestors of humans, Neandertals, and Denisovans and have survived in Africans.

Note: *We present preliminary results on chromosome 1 of the 1000 Genomes Project. This preprint will be soon completed with results from the other chromosomes.*

Contents

1   Introduction

The recent advent of next generation sequencing technologies made whole genome sequencing of thousands of individuals feasible (47). These sequencing techniques have been extended to allow sequencing of ancient DNA which enabled researchers to assemble the DNA of hominid individuals that lived ten thousands of years ago (15, 31, 35, 37). These advances in biotechnology could help to answer one of the most fundamental questions of humanity: "Where do we come from?"

In 2012, the 1000 Genomes Project Consortium (48) published first results on the genomes of more than thousand individuals from 14 populations stemming from Europe, East Asia, Africa, and the Americas. Reviewing the differences in genetic variation within and between populations, the authors found evidence of past events such as bottlenecks or admixtures, but also regions which seemed to be under a strong selective pressure. Gravel et al. (14) used the sequence data of the individuals from the Americas of the 1000 Genomes Project to quantify the contributions of European, African, and Native American ancestry to these populations, to estimate migration rates and timings, as well as to develop a demographic model.

Green et al. (15) were the first to analyze a draft sequence of the Neandertal genome derived from the bones of three individuals from Vindija Cave in Croatia that died about 40,000 years ago. They found that non-African individuals share more alleles with the Neandertal genome than sub-Saharan Africans. Until then evidence from mtDNA (9, 25, 42) and the Y chromosome (24) suggested that Neandertals lived isolated in Europe and Asia until they were replaced by anatomically modern humans. Although this theory still holds if differences in allele sharing are attributed to the existence of an ancient population substructure within Africa (10, 29), ancient admixture events between Neandertals and anatomically modern humans outside of Africa are considered as a more plausible explanation (15, 28, 40, 56). Further studies (31, 55) reported more Neandertal DNA preserved in modern East Asians than Europeans, hinting at some admixture after the separation of ancestors of Europeans and Asians. Prüfer et al. (35) confirmed higher rates of sharing for non-Africans using a high-quality genome sequence of a Neandertal from Denisova Cave in the Altai Mountains of Siberia. The authors further report that populations in Asia and America have more regions of Neandertal origin than populations in Europe.

A bone of a sister group of Neandertals, named Denisovans after the cave in the Altai Mountains of Siberia, was sequenced first at low (37) and later at high coverage (31). Studies on the low coverage draft sequence, as well as on the high coverage sequence, reported contributions from Denisovans to the gene pool of present-day individuals in Southeast Asia and Oceania (31, 37, 38, 43). Additionally to the clear signal of the Denisova genome found in Oceanians, Prüfer et al. (35) detected some regions of Denisovan origin in East and Southeast Asians and also in populations of the Americas, but only very few in Europeans. The authors further report gene flow from Neandertals and an unknown archaic population into Denisovans indicating that interbreeding among distinct hominid populations was more common than previously thought.

Using the high coverage Altai Neandertal genome Sankararaman et al. (39) screened the 1000 Genomes Project data for Neandertal ancestry with the genomes of the West African Yoruba individuals from Ibadan, Nigeria, as a reference panel that are assumed to harbor no Neandertal ancestry. They again found more Neandertal ancestry in East-Asians than in Europeans, but also very low levels in African Luhya individuals from Webuye, Kenya. The authors further looked for regions in the genome with very high or very low Neandertal ancestry that both might be due to

selective pressure. Vernot and Akey (53) searched for signatures of introgression in the sequences of Asian and European individuals of the 1000 Genomes Project and compared them with the Altai Neandertal genome. They confirmed more Neandertal DNA in East Asians than Europeans and looked for different levels of introgression along the genome.

In this study we propose to use short segments of identity by descent (IBD) to infer the population structure of humans and to gain insights into the genetic relationship of humans, Neandertals and Denisovans.

1.1   IBD for Inferring Population Structure

A DNA segment is *identical by state (IBS)* in two or more individuals if they all have identical nucleotide sequences in this segment. An IBS segment is *identical by descent (IBD)* in two or more individuals if they have inherited it from a common ancestor, that is, the segment has the same ancestral origin in these individuals. Rare variants can be utilized for distinguishing IBD from IBS without IBD because independent origins are highly unlikely for such variants. In other words, IBS generally implies IBD for rare variants, which is not true for common variants (44, Ch. 15.3, p. 441).

IBD detection methods have already been successfully used for inferring population structure. Gusev et al. (16) looked for long IBD segments shared within and between populations to estimate the demographic history of Ashkenazi Jewish individuals. Using similar models Palamara et al. (34) and Carmi et al. (7) reconstructed the demographic history of Ashkenazi Jewish and Kenyan Maasai individuals. Botigué et al. (3) confirmed gene flow from North Africans into Southern Europeans via patterns of long shared IBD segments. Ralph and Coop (36) tried to quantify the recent shared ancestry of different European populations by looking for long segments of shared DNA. Gravel et al. (14) similarly tried to draw conclusions of the genetic history of populations in the Americas using the respective data of the 1000 Genomes Project.

Except for Gravel et al. all of these studies were performed on SNP microarray data as IBD segments could not reliably be detected in large sequencing data. Sequencing data has higher marker density and also captures rare variants in contrast to SNP microarray data, therefore they would allow for a finer resolution of the length of IBD segments. Furthermore all previous studies based on microarrays were limited to long IBD segments that stem from a very recent common ancestor. However, shorter IBD segments would convey information about events farther back in time because the shorter IBD segments are, the older they are assumed to be. Therefore, existing studies were not able to resolve demographic histories at a fine scale and very far back into the past.

1.2   HapFABIA for Extracting Short IBD Segments

We recently developed HapFABIA (19) (see `http://dx.doi.org/10.1093/nar/gkt1013`) to identify very short segments of identity by descent (IBD) that are tagged by rare variants (the so called tagSNVs) in large sequencing data. HapFABIA identifies **100 times smaller IBD segments than current state-of-the-art methods: 10 kbp for HapFABIA vs. 1 Mbp for state-of-the-art methods**. HapFABIA utilizes rare variants ($\leq$5% MAF) to distinguish IBD from IBS without IBD because independent origins of rare minor alleles are highly unlikely (44, Ch. 15.3,p. 441). More

importantly, rare variants make juxtapositions of smaller IBD segments unlikely which prevents the summary of several small IBD segment into one long IBD segment. Consequently, the length of IBD segments is estimated more accurately than with previous methods.

In experiments with artificial, simulated, and real genotyping data HapFABIA outperformed its competitors in detecting short IBD segments (19). HapFABIA is based on biclustering (20) which in turn uses machine learning techniques derived from maximizing the posterior in a Bayes framework (8, 21, 22, 23, 30, 41, 45, 46).

HapFABIA is designed to detect short IBD segments in genotype data that were obtained from next generation sequencing (NGS), but can also be applied to DNA microarray data. Especially in NGS data, HapFABIA exploits rare variants for IBD detection. Rare variants convey more information on IBD than common variants, because random minor allele sharing is less likely for rare variants than for common variants (5). In order to detect short IBD segments, both the information supplied by rare variants and the information from IBD segments that are shared by more than two individuals should be utilized (5). HapFABIA uses both. The probability of randomly sharing a segment depends

- (a) on the allele frequencies within the segment, where lower frequency means lower probability of random sharing, and

- (b) on the number of individuals that share the allele, where more individuals result in lower probability of random segment sharing.

The shorter the IBD segments, the higher the likelihood that they are shared by more individuals. A segment that contains rare variants and is shared by more individuals has higher probability of representing IBD (18, 27). These two characteristics are our basis for detecting short IBD segments by HapFABIA.

2   IBD Segments on Chr. 1 of the 1000 Genomes Data

2.1   The 1000 Genomes Data

We used HapFABIA to extract short IBD segments from the 1000 Genomes Project genotyping data (48), more specifically, the phase 1 integrated variant call set (version 3) containing phased genotype calls for SNVs, short indels, and large deletions. This data set consists of 1,092 individuals (246 Africans, 181 Admixed Americans, 286 East Asians, and 379 Europeans), 38.2M SNVs, 2.4M short indels, and 14k large deletions. We removed 36 individuals because they showed cryptic relatedness to others. The final data set consisted of 1056 individuals (228 Africans, 175 Admixed Americans, 277 East Asians, and 376 Europeans). Chromosome 1 contains 3,007,196 SNVs that are on average 83 bp apart and have an average minor allele frequency (MAF) of 0.066. 2,280,175 (75.8%) SNVs are rare (MAF $\leq$ 0.05 including privates), 664,225 (22.1%) are private (minor allele is observed only once), 33,100 (1.1%) have an MAF of zero, and 693,921 (23.1%) are common (MAF $>$ 0.05). We kept only the rare SNVs for IBD detection and excluded private ones.

Chromosome 1 was divided into intervals of 10,000 SNVs with adjacent intervals overlapping by 5,000 SNVs. After removing common and private SNVs, we applied HapFABIA with default parameters to these intervals ignoring phase information since previous analysis revealed many phasing errors.

2.2   Summary Statistics of the Detected Short IBD Segments

HapFABIA found 105,652 different very short IBD segments on chromosome 1. These contained 635,280 rare tagSNVs, which amounts to 27.9% of the rare variants (39.3% if private SNVs are excluded) and 21% of all SNVs. The distance between centers of IBD segments had a median of 1.1 kbp and a mean of 2.4 kbp and ranged from 0 (overlapping IBD segments) to several Mbp. The number of individuals that shared the same IBD segment was between 2 and 113, with a median of 5 and a mean of 10.7. IBD segments were tagged by 8 to 223 tagSNVs, with a median of 13 and a mean of 18.8. The length of IBD segments ranged from 62 base pairs to 21 Mbp, with a median of 25 kbp and a mean of 27 kbp. IBD lengths are computed as described in Section 4.1, to match the assumptions for the distribution of IBD segment lengths as derived in other publications (4, 16, 50, 51).

3   Sharing of IBD Segments Between Populations and With Ancient Genomes

3.1   Sharing of IBD Segments Between Human Populations

3.1.1   Sharing Between All Human Populations

We were interested in the distribution of IBD segments among different populations. The main continental population groups are Africans (AFR), Asians (ASN), Europeans (EUR), and Admixed Americans (AMR), where AMR consist of Colombian, Puerto Rican, and Mexican individuals. Table 1 lists the number of IBD segments that are shared between particular continental populations. The vast majority (101,209) of the detected IBD segments (105,652) is shared by Africans (at least one African possesses the segment), of which 64,894 are exclusively found in Africans. Only 9,003 and 4,793 IBD segments are shared by Europeans and Asians, respectively. 558 IBD segments are exclusively found in Europeans and 1,452 exclusively in Asians. Admixed Americans share 207 IBD segments with Asians, but 1,008 with Europeans. Gravel et al. (14) reported recently that Admixed Americans, especially Colombians and Puerto Ricans, show a large proportion of European ancestry, which is also reflected in our results. If we additionally consider sharing with AFR, we obtain similar figures that are again consistent with the AMR admixture: 4,325 IBD segments have AFR/AMR/EUR sharing while only 621 IBD segments have AFR/AMR/ASN sharing. According to results of the 1000 Genomes Project Consortium (48), individuals with African ancestry carry much more rare variants than those of European or Asian ancestry supporting our finding that most IBD segments are shared by Africans. We found that few IBD segments are shared between two continental populations (Table 1 "Pairs of Populations") confirming recently published results (13, 48) (see also Section 3.1.2). The relatively large number of shared IBD segments between Africans and Europeans was due to many shared IBD segments between the AFR sub-group ASW (Americans with African ancestry from SW US) and Europeans. This tendency was also observed in the 1000 Genomes Project via the fixation index F_{ST} estimated by Hudson ratio of averages and via shared haplotype length around f_2 variants (48). The high content of European DNA segments in ASW is consistent with the finding that in African Americans a median proportion of 18.5% is European (6).

Overall, we conclude that IBD segments that are shared across continental populations, in particular by Africans, date back to a time before humans moved out of Africa. Consequently, the rare variants that tag these short IBD segments arose before this time. See Section A, for a discussion of the question whether rare variants are recent or old.

3.1.2   Sharing of IBD Segments Between Continental Populations

We found that most IBD segments are observed in only one continental population, which are, in most cases, Africans. However, for non-Africans, we saw a considerable sharing of IBD segments across continental groups. The sharing between continental populations seems to contradict statements in other publications. For example, Gravel et al. (13) report that sharing of SNVs between continental groups is low.

Nevertheless, sharing between continental groups has also been reported in the publication of the 1000 Genomes Project (48). The authors observed 17% of low-frequency variants in the range

Table 1: Number of IBD segments that are shared by particular continental populations.

Single Population					All Populations	
AFR	AMR	ASN	EUR		AFR/AMR/ASN/EUR	
64,894	378	1,452	558		1,127	
Pairs of Populations				Triplets of Populations		
AFR/AMR	AFR/ASN	AFR/EUR		AFR/AMR/ASN		AFR/AMR/EUR
28,710	387	986		621		4,325
AMR/ASN	AMR/EUR	ASN/EUR		AFR/ASN/EUR		AMR/ASN/EUR
207	1,008	351		159		489

AFR = Africans (228), AMR = Admixed Americans (175), ASN = East Asians (277), and EUR = Europeans (376).

0.5–5% and 53% of rare variants at 0.5% in a single continental group, which means that 83% of the low-frequency variants and 47% of the rare variants are shared between continental populations. They also describe sharing of f_2 variants between Finnish (FIN) and African individuals. f_2 variants are those variants for which their minor allele is observed in exactly 2 individuals. They are rarer than most SNVs that tag our extracted IBD segments because, on average, 6 individuals possess an IBD segment. According to Gravel et al. (13), sharing between continental groups decreases with rarer variants. Thus, in our extracted IBD segments, we would expect to find sharing between continental groups considerably more often than for f_2 variants. Furthermore, the 1000 Genomes Project Consortium (48) reports that f_2 variants that are shared between populations with no recent common ancestry are present on very short haplotypes with a median length of 15 kb. They conclude that these haplotypes are likely to reflect recurrent mutations and chance ancient coalescent events. This statement also supports our findings of IBD sharing between continental groups, but we can exclude recurrent mutations as a likely explanation because IBD segments are shared between multiple individuals.

3.2   Sharing of IBD Segments Between Human and Ancient Genomes

Since short IBD segments are thought to be very old, we wondered whether some IBD segments match bases of ancient genomes, such as Neandertal and Denisova. Ancient short IBD segments may reveal gene flow between ancient genomes and ancestors of modern humans and, thereby, may shed light on different out-of-Africa hypotheses (2). 31X coverage whole genome sequencing data for the Denisova, as well as, the Altai Neandertal genome of 52X coverage were provided by the Max Planck Institute for Evolutionary Anthropology (31, 35). Considering only the variants of the 1000 Genomes Project, 0.6% of the Denisova bases and 0.3% of the Neandertal bases were not determined, 92.0% and 91.3%, respectively, matched bases of the human reference, and 8% and 9%, respectively, matched either the human minor allele or were different from human alleles. As additional information in the 1000 Genomes Project data, bases of the reconstructed common ancestor of human, chimpanzee, gorilla, orang utan, macaque, and marmoset genomes were given.

3.2.1   IBD Sharing Between Human Populations and Ancient Genomes

We tested whether IBD segments that match particular ancient genomes to a large extent are found more often in certain populations than expected randomly. For each IBD segment, we computed two values: The first value was the proportion of tagSNVs that match a particular ancient genome, which we call "genome proportion" of an IBD segment (e.g. "Denisova proportion"). The second value was the proportion of individuals that possess an IBD segment and are from a certain population as opposed to the overall number of individuals that possess this IBD segment. We call this value the "population proportion" of an IBD segment (e.g. "Asian proportion"). Consider the following illustrative examples. If an IBD segment has 20 tagSNVs of which 10 match Denisova bases with their minor allele, then we obtain $10/20 = 0.5 = 50\%$ as the Denisova proportion. If an IBD segment is observed in 6 individuals of which 4 are Africans and 2 Europeans, then the African proportion is $4/6 = 0.67 = 67\%$ and the European proportion is $0.33 = 33\%$. A correlation between a genome proportion and a population proportion would indicate that this ancient genome is overrepresented in this specific population. Often human IBD segments that are found in ancient genomes have already been present in the ancestral genome of humans and other primates. These IBD segments would confound the interbreeding analysis based on IBD sharing between modern human populations and Neandertals or Denisovans. Therefore, we removed those IBD segments from the data before the following analysis.
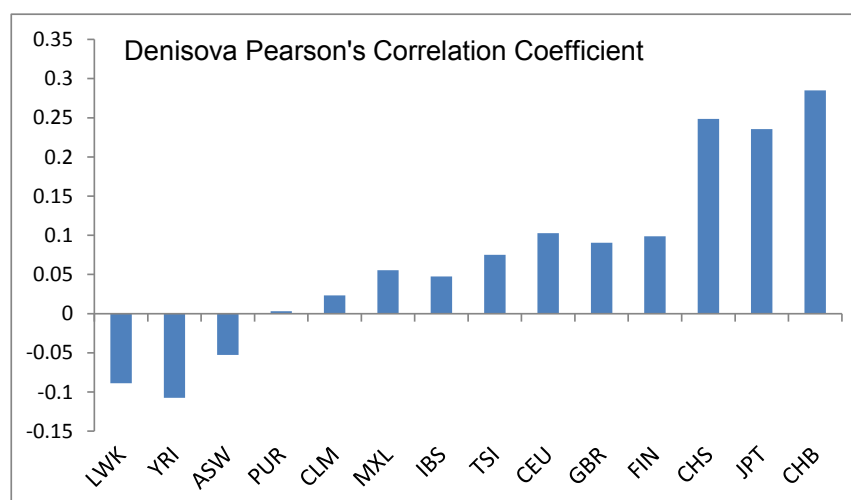

Figure 1: Pearson's correlation between population proportions and the Denisova genome proportion. Asians have a significantly larger correlation to the Denisova genome than other populations. Many IBD segments that match the Denisova genome are exclusively found in Asians, which has a large effect on the correlation coefficient. Europeans still have a significantly larger correlation to the Denisova genome than the average. Mexicans (MXL) have also a surprisingly high correlation to the Denisova genome while Iberians (IBS) have a low correlation compared to other Europeans.

Pearson's product moment correlation test and Spearman's rank correlation test both showed highly significant correlations between the Denisova genome and Asians, the Denisova genome and Europeans, the Neandertal genome and Asians, and the Neandertal genome and Europeans.
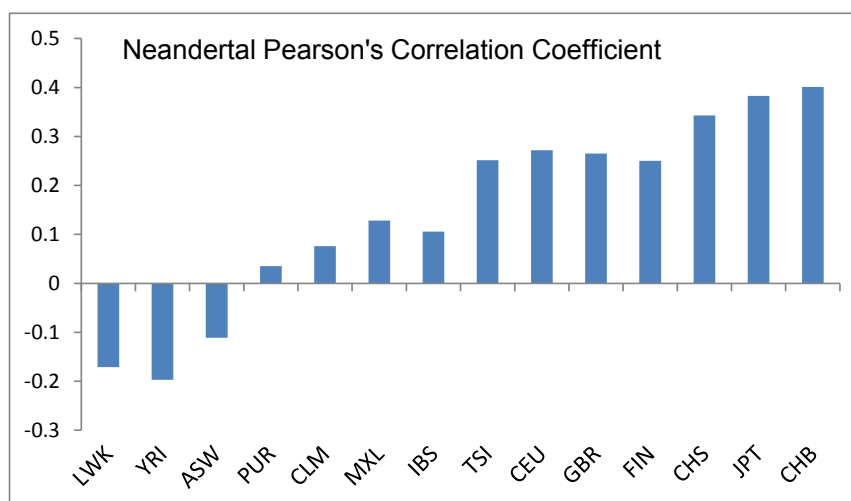
Figure 2: Pearson's correlation between population proportions and the Neandertal genome proportion. Europeans and Asians have significantly larger correlations to the Neandertal genome than other populations. Asians have slightly higher correlation coefficients than Europeans. Again, Mexicans (MXL) have a surprisingly high correlation to the Neandertal genome while Iberians (IBS) have a low correlation compared to other Europeans.

Figure 1 shows Pearson's correlation coefficients for the correlation between population proportion and the Denisova genome proportion. Asians have a significantly larger correlation to the Denisova genome than other populations (p-value $<$ 5e-324). Many IBD segments that match the Denisova genome are exclusively found in Asians, which has a large effect on the correlation coefficient. Europeans have still a significantly larger correlation to the Denisova genome than the average (p-value $<$ 5e-324). Surprisingly, Mexicans (MXL) have a high correlation to the Denisova genome too, while Iberians (IBS) have a low correlation compared to other Europeans. These unexpected correlations can be explained through the fact that of all European populations Iberians show the highest rates of African gene flow (1, 3, 32) whereas Mexicans show a high proportion of Native American ancestry which in turn might also reflect gene flow from Asia (14, 33). Figure 2 shows Pearson's correlation coefficients for the correlation between population proportion and the Neandertal genome proportion. Europeans and Asians have a significantly larger correlation to the Neandertal genome than other populations (p-value $<$ 5e-324). Asians have slightly higher correlation coefficients than Europeans. Again Mexicans (MXL) have surprisingly high correlation to the Neandertal genome while Iberians (IBS) have a low correlation compared to other Europeans.

However, correlation tests are sensitive to accumulations of minor effects. Therefore, we focused subsequently on strong effects, i.e. large values of genome proportions and large values of population proportions.

We define an IBD segment to match a particular ancient genome if the genome proportion is 30% or higher. Only 8% of the Denisova and 9% of the Neandertal bases (about 10% of the called bases) match the minor allele of the human genome on average. Therefore, we require an odds ratio of 3 to call an IBD segment to match an ancient genome. We found many more

IBD segments that match the Neandertal or the Denisova genome than expected randomly. This again supports the statement that the detected short IBD segments are old and some of them date back to times of the ancestors of humans, Neandertals, and Denisovans. IBD segments that match the Denisova genome often match the Neandertal genome, too, thus these segments cannot be attributed to either one of these genomes. Therefore, we introduce the "Archaic genome" (genome of archaic hominids ancestral to Denisova and Neandertal), to which IBD segments are attributed if they match both the Denisova and the Neandertal genome.

To test for an enrichment of Denisova-, Neandertal-, or Archaic-matching IBD segments in different subpopulations, we used Fisher's exact test for count data. Again, we first excluded those IBD segments that were also present in the reconstructed ancestral genome of humans and other primates. IBD segments were then classified as being shared by at least one individual of a certain population or not and either matching or not matching the ancient genome.
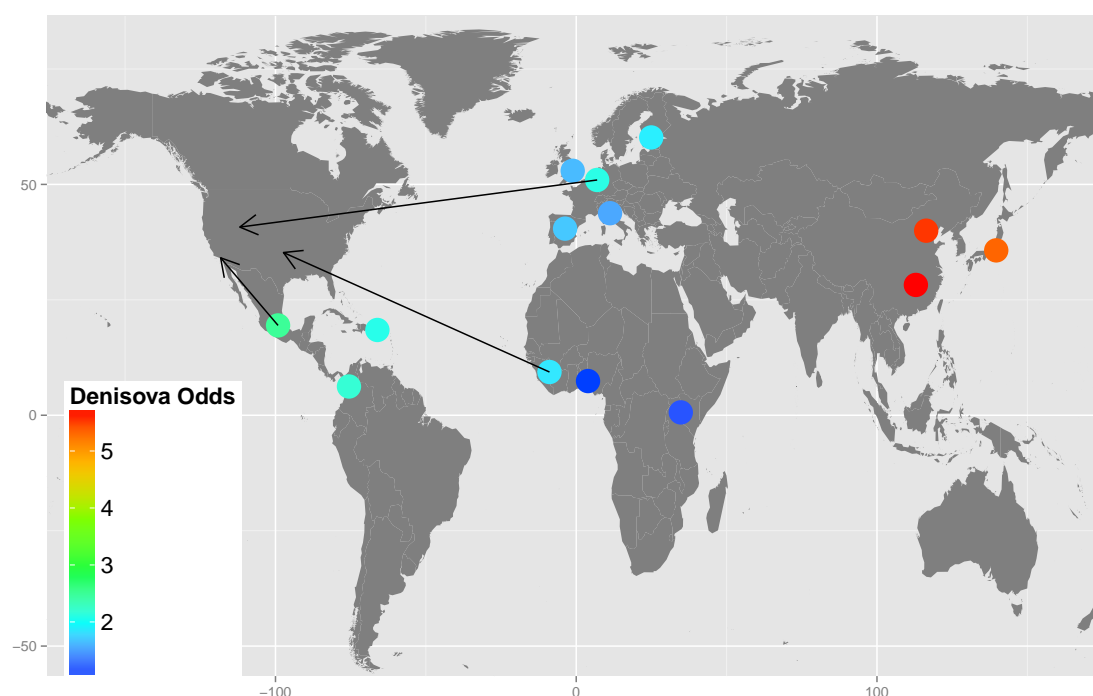

Figure 3: Odds scores of Fisher's exact test for an enrichment of Denisova-matching IBD segments in different populations are represented by colored dots. The arrows point from the region the populations stem from to the region of sample collection. IBD segments that are shared by Asians match the Denisova genome significantly more often than IBD segments that are shared by other populations (red dots). Africans show the lowest matching with the Denisova genome (dark blue dots). Surprisingly, Admixed Americans have higher odds for Denisova sharing than Europeans (green and turquoise vs. light blue dots).

Figure 3 shows the odds scores of the Fisher's exact tests for an enrichment of Denisova-matching IBD segments in different populations. As expected, Asians show the highest odds for IBD segments matching the Denisova genome (odds ratio of 5.44 and p-value of 1.2e-102), while Africans have the lowest odds (odds ratio of 0.22 and p-value of 9.4e-71). Surprisingly,

Admixed Americans show more sharing than Europeans (odds ratio of 2.63 vs. 2.11), although the difference is not very prominent. In general these results reflect previous findings. Meyer et al. (31) and Prüfer et al. (35) noted that Denisovans show more sharing with modern East Asians and South Americans (called Admixed Americans here) than with Europeans. Within the Asian populations Han Chinese from South (CHS) have slightly higher odds for matching the Denisova genome than Han Chinese from Beijing (CHB), while Japanese (JPN) have the lowest odds (Figure 4). Skoglund et al. (43) also report a particularly high affinity between Southeast Asians and the Denisova genome, in contrast to Meyer et al. (31) who find higher levels of sharing for individuals from North China compared to those of the south. Within the European populations we see the highest odds for Utah residents with ancestry from northern and western Europe (CEU) followed by Finnish (FIN), Iberians from Spain (IBS), and British from England and Scotland (GBR). Toscani in Italy (TSI) show the lowest levels of sharing, but all odds are between 1.56 and 2.11 (Figure 5).
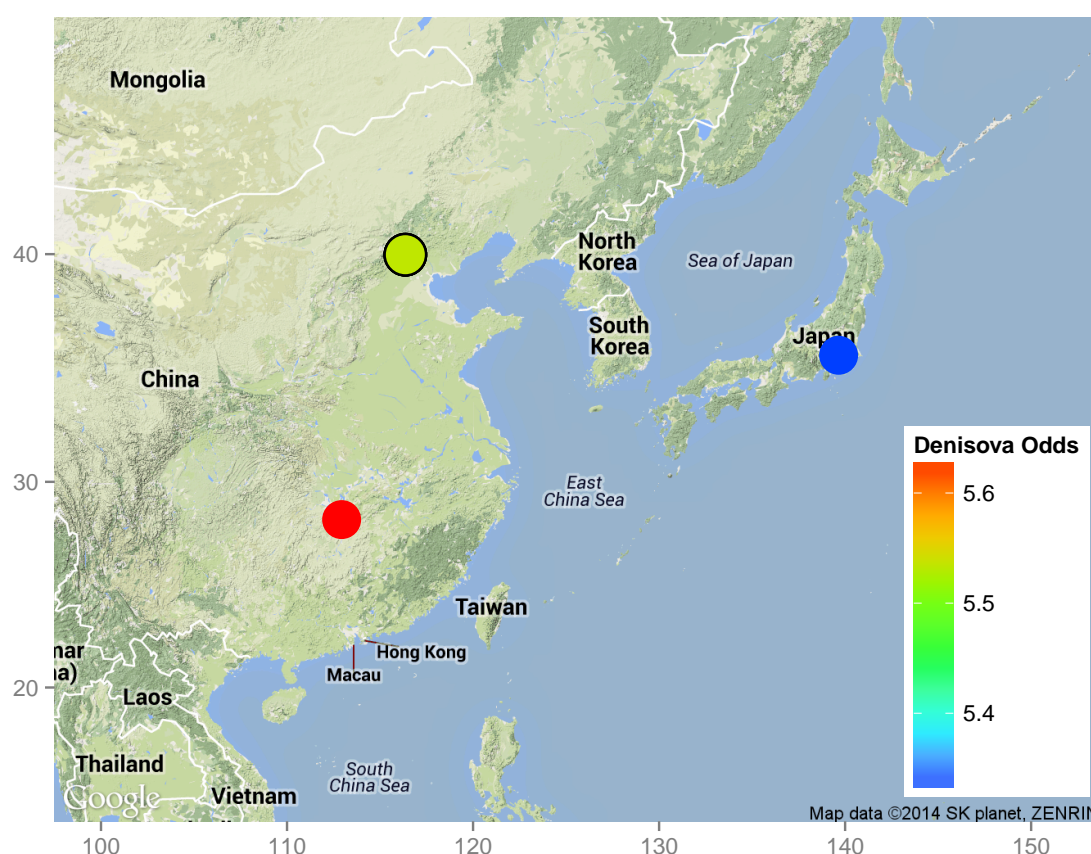

Figure 4: Odds scores of Fisher's exact test for an enrichment of Denisova-matching IBD segments in Asian populations are represented by colored dots. Within the Asian populations Han Chinese from South (CHS) have slightly higher odds for matching the Denisova genome than Han Chinese from Beijing (CHB) (red dot vs. green dot), while Japanese (JPN) have the lowest odds (blue dot).
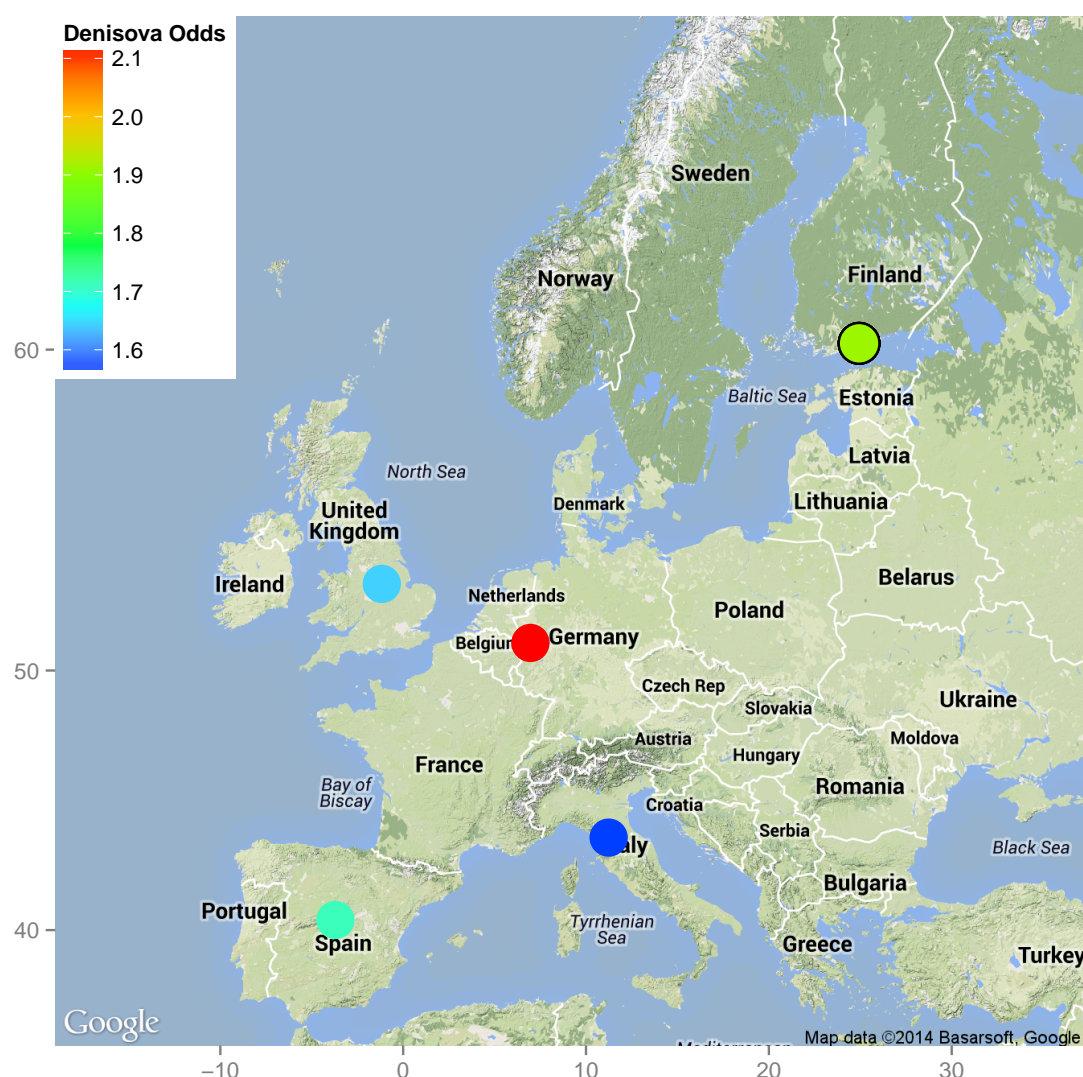
Figure 5: Odds scores of Fisher's exact test for an enrichment of Denisova-matching IBD segments in European populations are represented by colored dots. Utah residents with ancestry from northern and western Europe (CEU) are symbolized by a dot in Central Europe. Within the European populations we see the highest odds (red dot) for Utah residents with ancestry from northern and western Europe (CEU) followed by Finnish (FIN), Iberians from Spain (IBS), and British from England and Scotland (GBR). Toscani in Italy (TSI) show the lowest levels of sharing (blue dot), but all odds are between 1.56 and 2.11.

Figure 6 shows the odds scores of the Fisher's exact tests for an enrichment of Neandertal-matching IBD segments in different populations. As expected, Asians again show the highest odds for IBD segments matching the Neandertal genome (odds ratio of 27.49 and p-value $< 5e$-324), while Africans have the lowest odds (odds ratio of 0.03 and p-value $< 5e$-324). In contrast to the Denisova results, here Europeans show clearly more matching with the Neandertal genome than Admixed Americans (odds ratio of 12.66 vs. 2.90). In general these results reflect previous

findings. Other studies (15, 28) have reported significantly more sharing between Neandertals and non-Africans than with Africans. Wall et al. (55) noted that Neandertals contributed more DNA to modern East Asians than to Europeans. Within the Asian populations Han Chinese from Beijing (CHB) have slightly higher odds for matching the Neandertal genome than Han Chinese from South (CHS), while Japanese (JPN) lie in between (Figure 4). Within the European populations a north to south decline can be seen with Finnish (FIN) having the highest odds for matching the Neandertal genome followed by British from England and Scotland (GBR) and Utah residents with ancestry from northern and western Europe (CEU). Toscani in Italy (TSI) and Iberians from Spain (IBS) show the lowest levels of matching (Figure 8). An explanation for the latter two low odds for Neandertal matching is that according to Ralph et al. (36) Iberians and Italians show reduced rates of shared ancestry compared to the rest of Europe. Also Botigué et al. (3) found higher IBD sharing between North Africans and individuals from Southern Europe which would decrease the amount of DNA sharing with Neandertals. Sankararaman et al. (39) achieved a similar ranking for Neandertal ancestry using a different approach, but stated that because of the high standard deviation among individuals from the same population, small differences may be due to statistical noise. The same holds true for our results since a single individual with very high Neandertal ancestry can lead to inflated odds for a population. Nevertheless, since all Asian populations show especially high odds, we can exclude that all results are just due to noise.
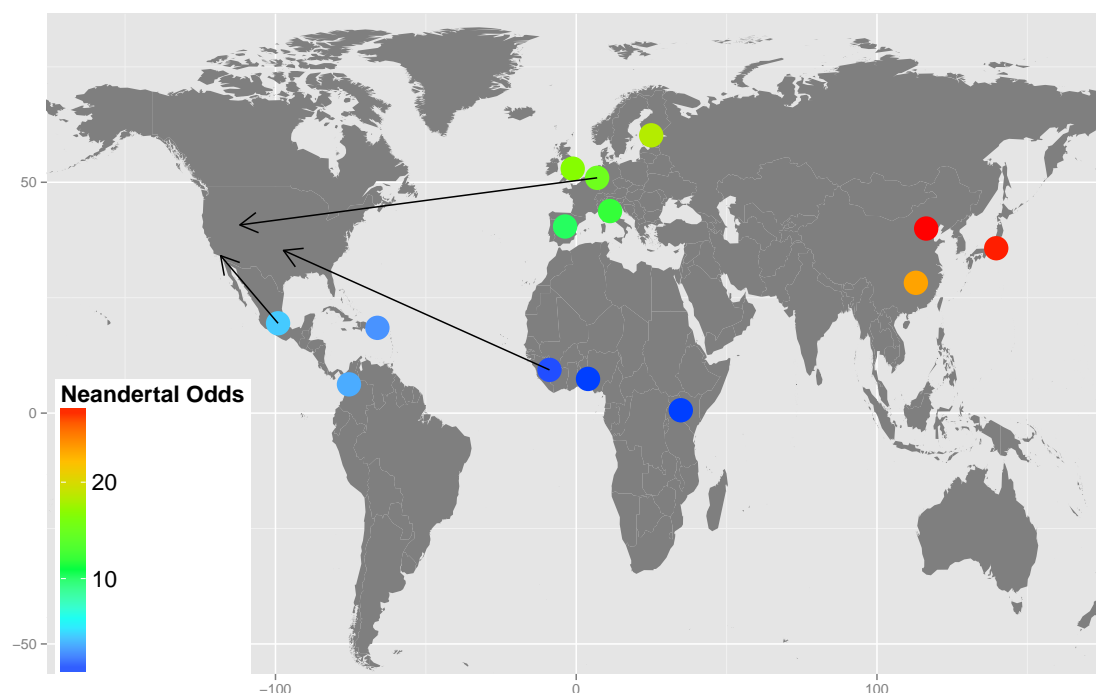
Figure 6: Odds scores of Fisher's exact test for an enrichment of Neandertal-matching IBD segments in different populations are represented by colored dots. The arrows point from the region the populations stem from to the region of sample collection. IBD segments that are shared by Asians match the Denisova genome significantly more often than IBD segments that are shared by other populations (red dots). Africans show the lowest matching with the Neandertal genome (dark blue dots). Europeans show more matching than Admixed Americans (green vs. light blue dots).
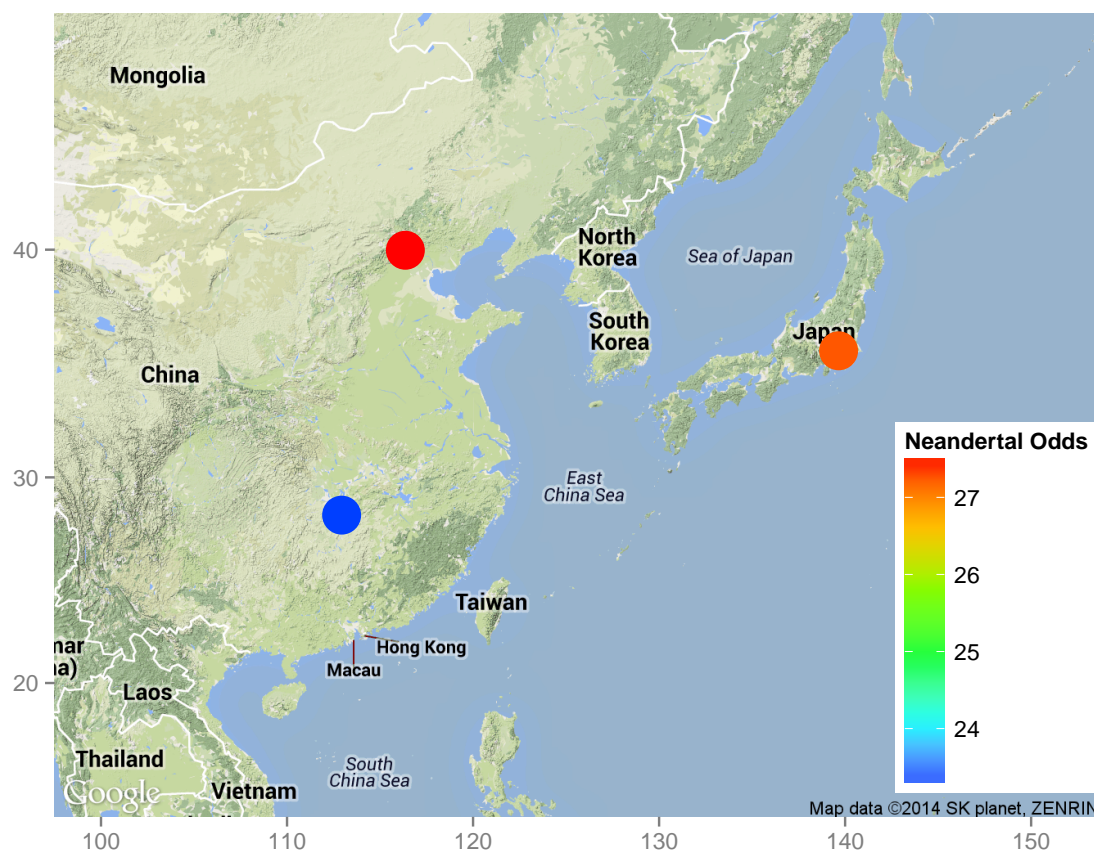
Figure 7: Odds scores of the Fisher's exact tests for an enrichment of Neandertal-matching IBD segments in Asian populations are represented by colored dots. Within the Asian populations Han Chinese from Beijing (CHB) have slightly higher odds for matching the Neandertal genome than Han Chinese from South (CHS) (red dot vs. blue dot), while Japanese (JPN) lie in between (orange dot).
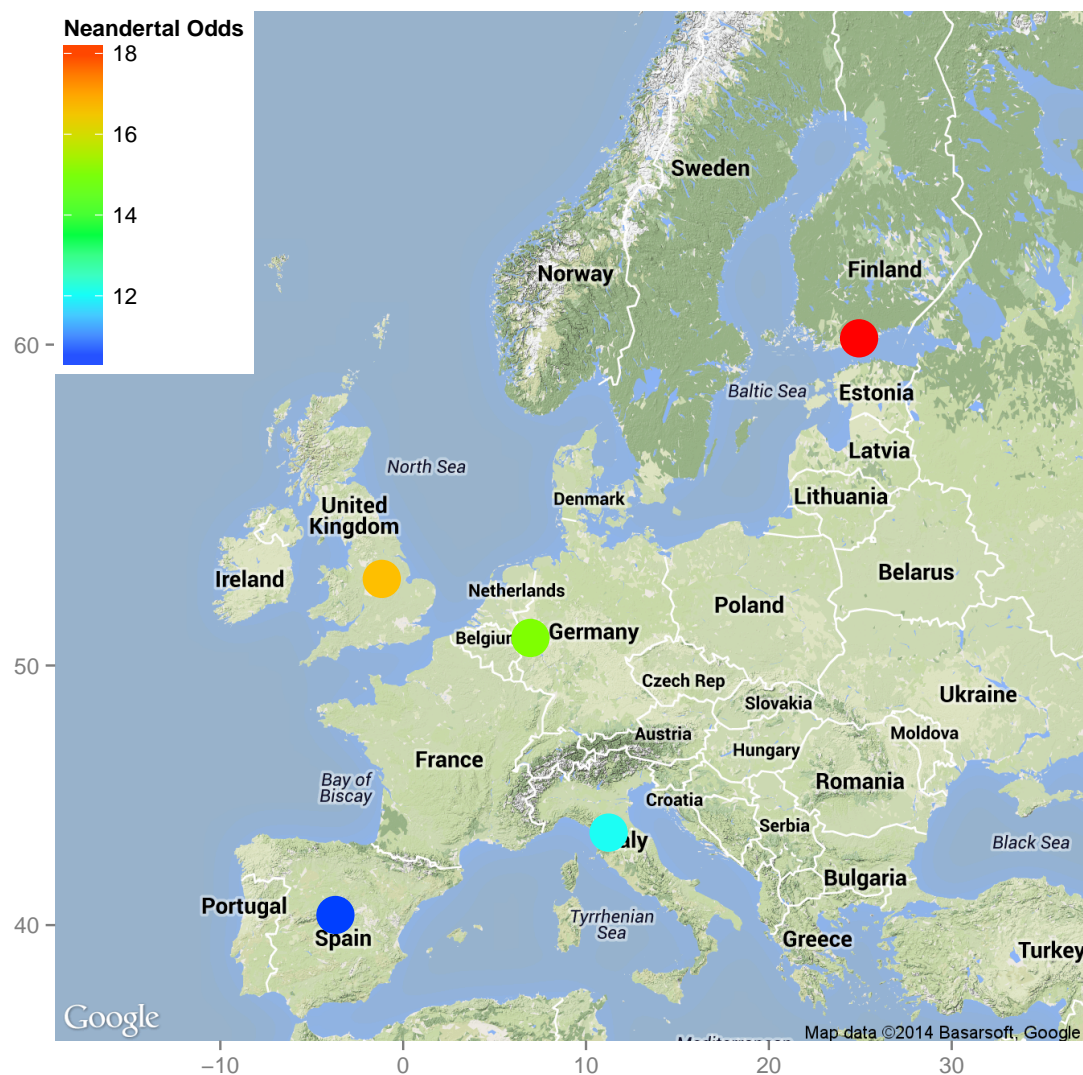
Figure 8: Odds scores of the Fisher's exact tests for an enrichment of Neandertal-matching IBD segments in European populations are represented by colored dots. Utah residents with ancestry from northern and western Europe (CEU) are symbolized by a dot in Central Europe. Within the European populations a north to south decline can be seen with Finnish (FIN) having the highest odds for matching the Neandertal genome (red dot) followed by British from England and Scotland (GBR) and Utah residents with ancestry from northern and western Europe (CEU) (orange and green dots). Toscani in Italy (TSI) and Iberians from Spain (IBS) show the lowest levels of matching (blue dots).

In order to concentrate on strong effects in terms of population proportions, we investigated which population has the majority proportion for an IBD segment that matches a particular genome — the population to which the majority of the individuals possessing this segment belong to. Figure 9 shows the population with the majority proportion for each IBD segment. The IBD segments are presented for each genome, where the colors show the populations with majority pro-

portion for the according IBD segment. More than half of the Neandertal-matching IBD segments have Asians or Europeans as majority population proportions. However, still many Neandertal-matching segments are mainly shared by Africans which to some extend contradicts the hypothesis of Prüfer et al. (35) that Neandertal ancestry in Africans is due to back-to-Africa gene flow. For the Archaic genome (intersection of Neandertal- and Denisova-matching IBD segments), IBD segments dominated by Asians or Europeans are also enriched if compared to all IBD segments found on chromosome 1 of the 1000 Genomes Project data (we will call the set of these segments "human genome"). The enrichment by Asian or European IBD segments is lower for the Denisova genome, but still significant, especially for Asian segments. Furthermore, as can be seen in Figure 9, more IBD segments are found that match the Neandertal genome than segments that match the Denisova genome. One explanation for this is that the unknown ancient gene flow into Denisovans mentioned by Prüfer et al. (35) replaced some segments from the common ancestor of Neandertals and Denisovans in the Denisova genome. Therefore, these segments are only shared by modern humans and Neandertals although they might have been introduced by the aforementioned common ancestor.

Next we wanted to know which populations share an IBD segment that matches a particular genome, that is, we asked whether this IBD segment is found in this population or not. Figure 10 shows for each genome (human and ancient) and each IBD segment, whether a population shares this IBD segment or not. IBD segments that match the Neandertal or the Archaic genome are found more often in Asians and Europeans than all IBD segments (human genome). This effect is not as prominent for IBD segments that match the Denisova genome, but still significant.
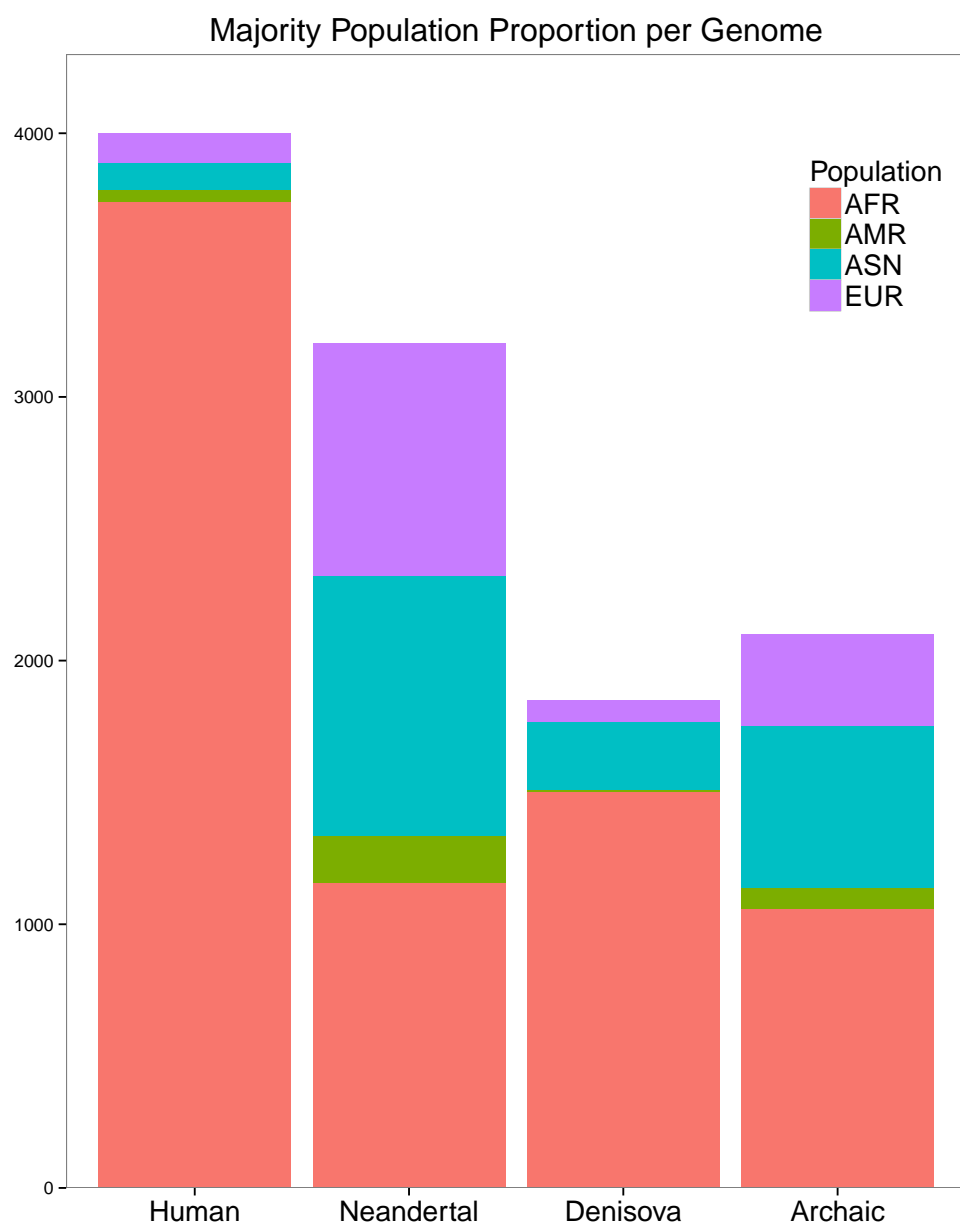
Figure 9: For each IBD segment, the population with the majority proportion is determined. IBD segments are given for each matching genome, where the color indicates the population that has majority proportion. For the human genome, 4,000 random IBD segments were chosen. More than half of the Neandertal-matching IBD segments have Asians or Europeans as majority population. The Archaic genome (Neandertal and Denisova) shows also an enrichment of IBD segments that are found mostly in Asians or Europeans. Denisova-matching IBD segments are often shared mainly by Asians.
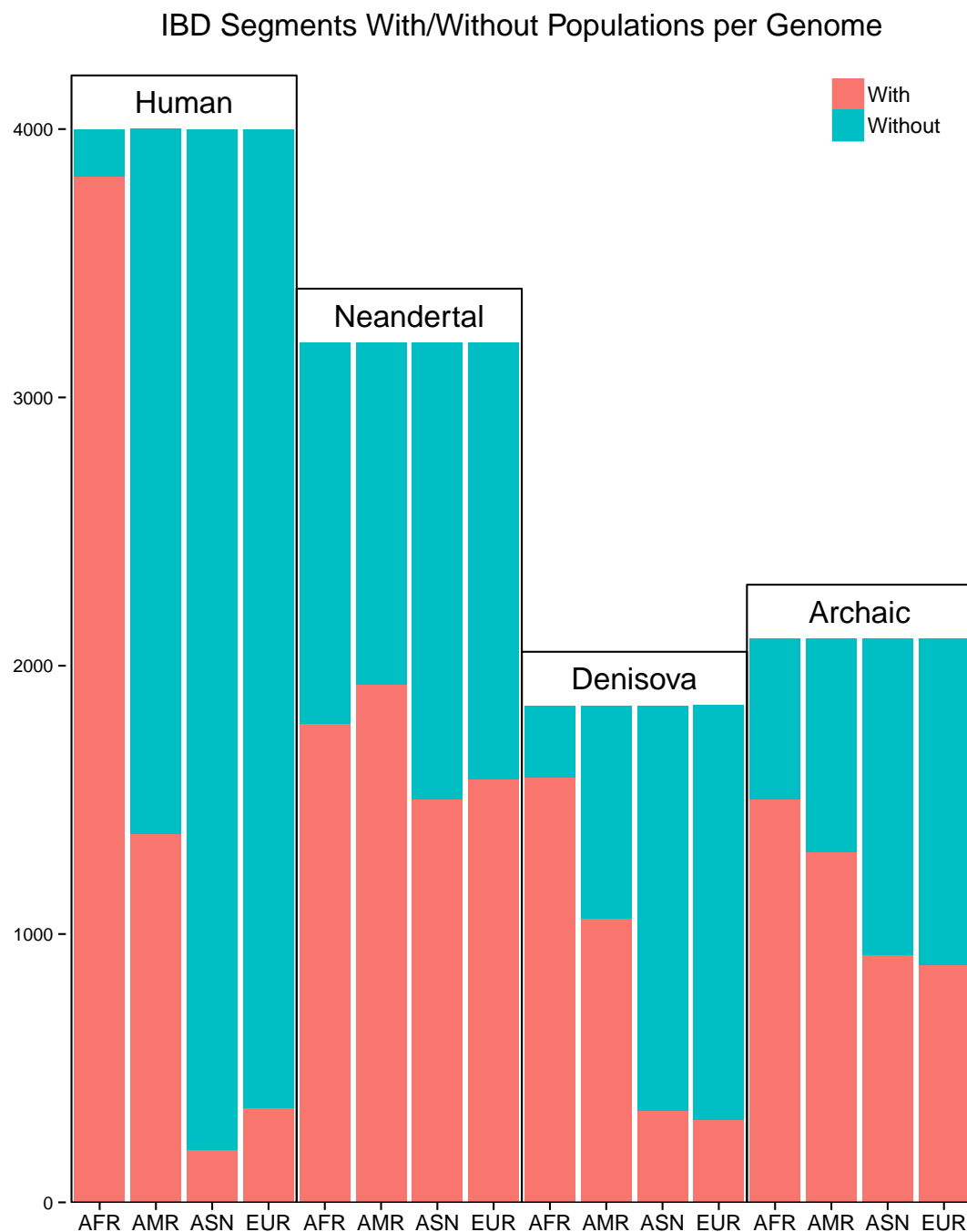
Figure 10: For each genome and each IBD segment, the color indicates whether a population contains this segment ("With") or not ("Without"). For the human genome, 4,000 random IBD segments were chosen. IBD segments that match the Neandertal or the Archaic genome are found more often in Asians and Europeans than all IBD segments (human). This effect is not as prominent for IBD segments that match the Denisova genome.

3.2.2   Densities of Population Proportions and Ancient Genomes

We plotted densities of population proportions for IBD segments that match a particular ancient genome (30% or more SNVs match) and for those that do not match that genome. Figure 11 shows the density of Asian proportions of Denisova-matching IBD segments (pink) vs. the density of Asian proportions of non-Denisova-matching IBD segments (cyan). Figure 12 shows analogous densities for Europeans. In comparison to all populations the density of Denisova-matching IBD segments that are observed in Asians and Europeans is higher than for non-matching IBD segments. This can be seen by the higher densities of matching IBD segments (pink) compared to densities of non-matching IBD segments (cyan) if the population proportions are not very close to zero — or, conversely, it can be seen at the lower peak at zero (less IBD segments that match the Denisova genome without Asian or European sharing). Many Denisova-matching IBD segments are shared exclusively among Asians which is indicated by the high density at a population proportion of 1 (pink density in Figure 11). Population proportion of 1 means that IBD segments are only shared among Asians.

Figures 14 and 15 show analogous densities as in Figures 11 and 12, but for the Neandertal genome. The differences we already observed for the Denisova genome are even more prominent for the Neandertal: Neandertal-matching IBD segments are observed even more often in Asians and Europeans than non-matching IBD segments if compared to all segments of the respective category. The higher densities (pink) at a population proportion not close to zero are now more prominent — or conversely, the lower peak at zero for Neandertal-matching IBD segments becomes clearer (less IBD segments that match the Neandertal genome without Asian or European sharing). For Asians, the peak at 1 in Figure 14 is even higher, representing IBD segments that are shared exclusively among Asians. IBD segment sharing exclusively within one continental population is very common, as the blue peaks at 1 in Figure 11 and Figure 14 show.

Figures 13 and 16 show the same densities for Africans. Both the density of African proportions of Denisova-matching and the same density for Neandertal-matching IBD segments have two peaks: one at a low and one at a high proportion of Africans. For Neandertal-matching IBD segments, the density at low proportions of Africans is even larger than for high proportions. Thus, IBD segments that match ancient genomes are either shared by a very low or a very high proportion of Africans. The low proportion of African density peak hints at admixture of ancestors of modern humans and Denisovans / Neandertals outside of Africa or the back to Africa gene flow mentioned by Prüfer et al. (35). The density peak at high proportions of Africans may be due to ancient DNA segments shared by hominid groups that were lost in other continental populations. This finding could support the hypothesis that ancient population substructure in Africa also allowed for the occurrence of different continental patterns of DNA sharing between modern humans and ancient genomes (10, 15).

IBD segments that match the "Archaic genome" are those IBD segments that match both the Denisova and Neandertal genome. Population proportion densities for the Archaic genome are presented in Figures 17, 18, and 19 for Asians, Europeans, and Africans, respectively. For the Archaic genome we see the same figure as for the Neandertal and the Denisova genome: the African density is bimodal that means either it is dominated by Africans or it contains no or only few Africans.
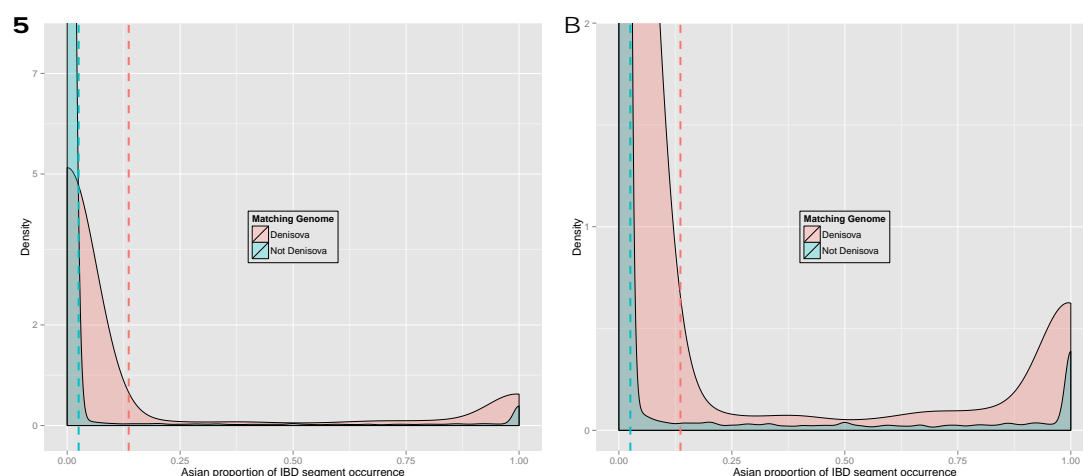
Figure 11: **Panel A**: Density of Asian proportions of Denisova-matching IBD segments (pink) vs. density of Asian proportions of non- Denisova-matching IBD segments (cyan). IBD segments were extracted from phased genotyping data of chromosome 1 of the 1000 Genomes Project. Dotted lines indicate the respective means. **Panel B**: The same densities as in Panel A but zoomed in.
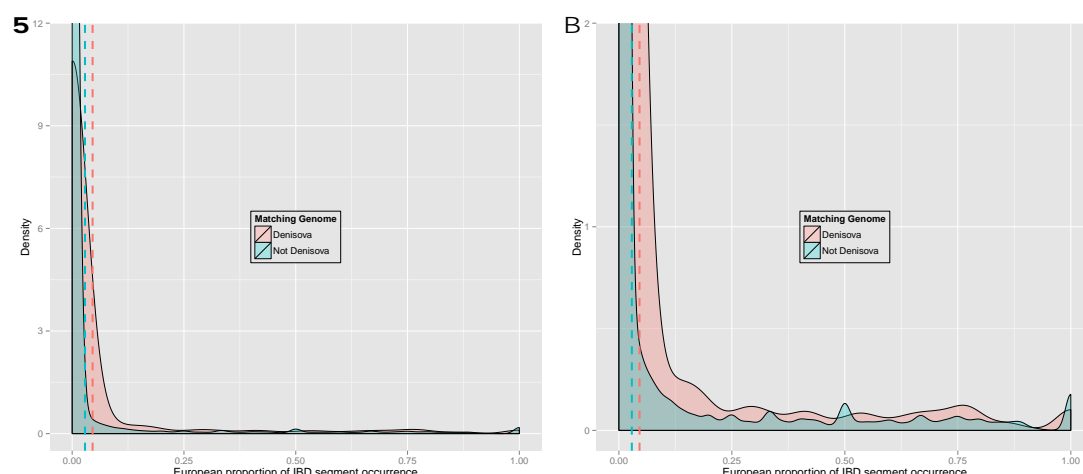

Figure 12: **Panel A**: Density of European proportions of Denisova-matching IBD segments (pink) vs. density of European proportions of non-Denisova-matching IBD segments (cyan). IBD segments were extracted from phased genotyping data of chromosome 1 of the 1000 Genomes Project. Dotted lines indicate the respective means. **Panel B**: The same densities as in Panel A but zoomed in.
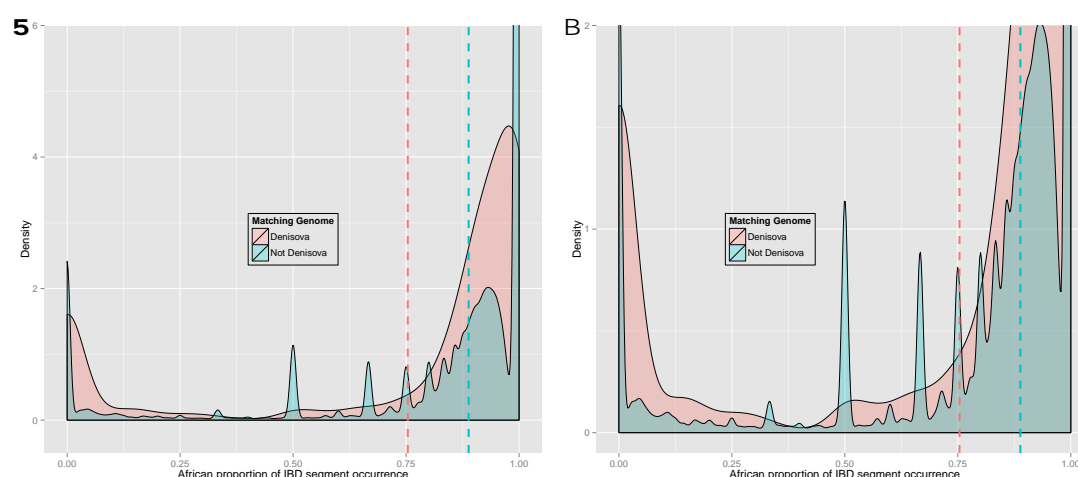
Figure 13: **Panel A**: Density of African proportions of Denisova-matching IBD segments (pink) vs. density of African proportions of non-Denisova-matching IBD segments (cyan). IBD segments were extracted from phased genotyping data of chromosome 1 of the 1000 Genomes Project. Dotted lines indicate the respective means. Peaks for non-Denisova-matching IBD segments are found at 0.5, 0.66, 0.33, 0.75, and 0.8, which corresponds to 1/2, 2/3, 1/3, 3/4, 4/5 (number of Africans / total number of individuals that have the IBD segment). The density of African proportions of Denisova-matching IBD segments has two peaks: one at a low and one at a high proportion of Africans. **Panel B**: The same densities as in Panel A but zoomed in.
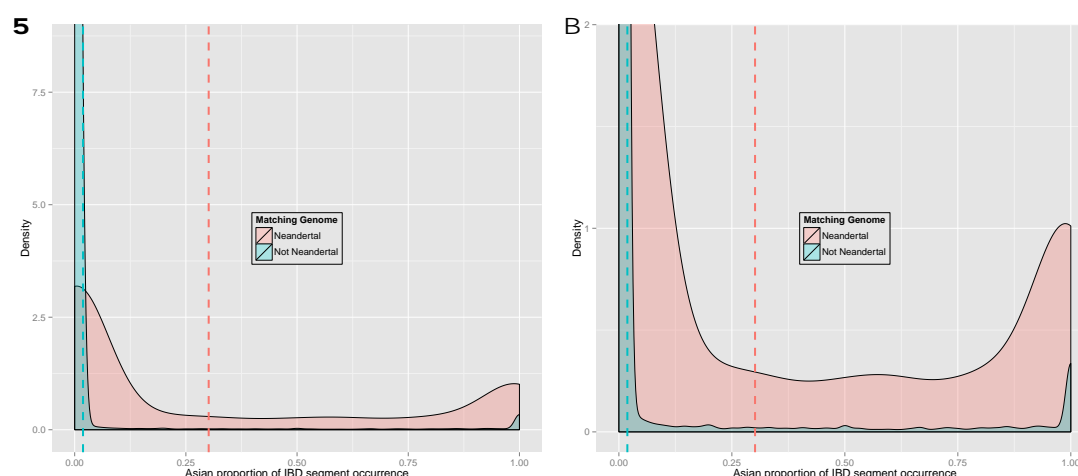

Figure 14: **Panel A**: Density of Asian proportions of Neandertal-matching IBD segments (pink) vs. density of Asian proportions of non-Neandertal-matching IBD segments (cyan). IBD segments were extracted from phased genotyping data of chromosome 1 of the 1000 Genomes Project. Dotted lines indicate the respective means. Many Neandertal-enriched IBD segments are shared mainly or exclusively by Asians as the peak at a proportion close to 1 shows. **Panel B**: The same densities as in Panel A but zoomed in.
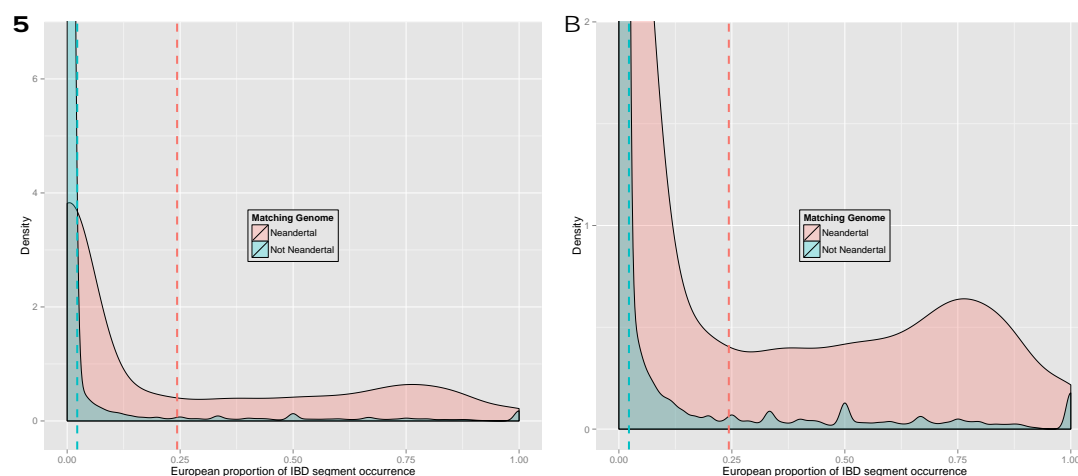
Figure 15: **Panel A**: Density of European proportions of Neandertal-matching IBD segments (pink) vs. density of European proportions of non-Neandertal-matching IBD segments (cyan). IBD segments were extracted from phased genotyping data of chromosome 1 of the 1000 Genomes Project. Dotted lines indicate the respective means. Many Neandertal-enriched IBD segments are shared mainly or exclusively by Asians as the peak at a proportion close to 1 shows. Neandertal-enriched IBD segments are shared by Europeans but fewer segments are shared exclusively among Europeans than for Asians (see Figure 14). **Panel B**: the same density zoomed in.
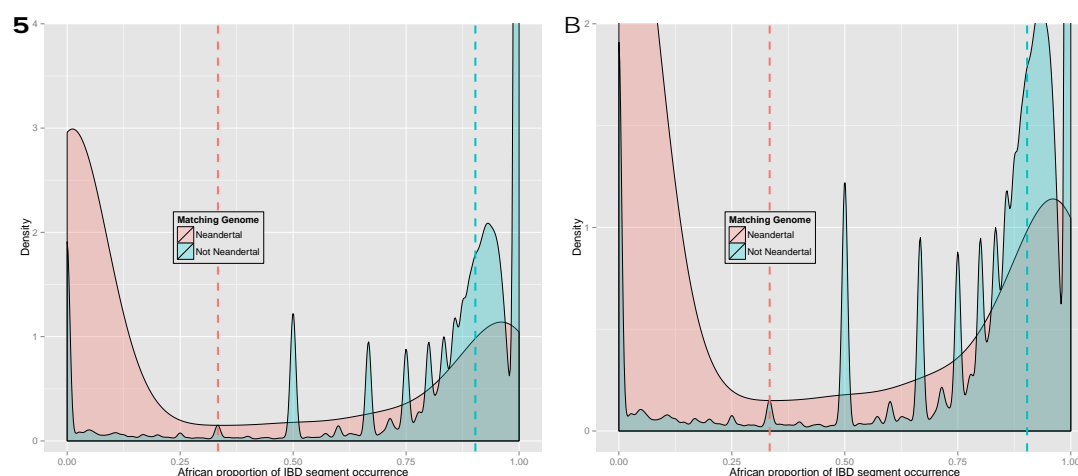

Figure 16: **Panel A**: Density of African proportions of Neandertal-matching IBD segments (pink) vs. density of African proportions of non-Neandertal-matching IBD segments (cyan). IBD segments were extracted from phased genotyping data of chromosome 1 of the 1000 Genomes Project. Dotted lines indicate the respective means. Peaks for non-Neandertal-matching IBD segments are found at 0.5, 0.66, 0.33, 0.75, and 0.8, which corresponds to 1/2, 2/3, 1/3, 3/4, 4/5 (number of Africans / total number of individuals that have the IBD segment). The density of African proportions of Neandertal-matching IBD segments has two peaks: one at a low and one at a high proportion of Africans. The density of low proportions of Africans is even larger than the density of high proportions. **Panel B**: the same density zoomed in.

Figure 17: **Panel A**: Density of Asian proportions of Archaic-genome-matching IBD segments (pink) vs. density of Asian proportions of non-Archaic-genome-matching IBD segments (cyan). IBD segments were extracted from phased genotyping data of chromosome 1 of the 1000 Genomes Project. Dotted lines indicate the respective means. **Panel B**: the same density zoomed in.

Figure 18: **Panel A**: Density of European proportions of Archaic-genome- matching IBD segments (pink) vs. density of European proportions of non- Archaic-genome-matching IBD segments (cyan). IBD segments were extracted from phased genotyping data of chromosome 1 of the 1000 Genomes Project. Dotted lines indicate the respective means. Many Archaic-genome-enriched IBD segments are shared mainly or exclusively by Asians as the peak at a proportion close to 1 shows. **Panel B**: the same densities zoomed in.
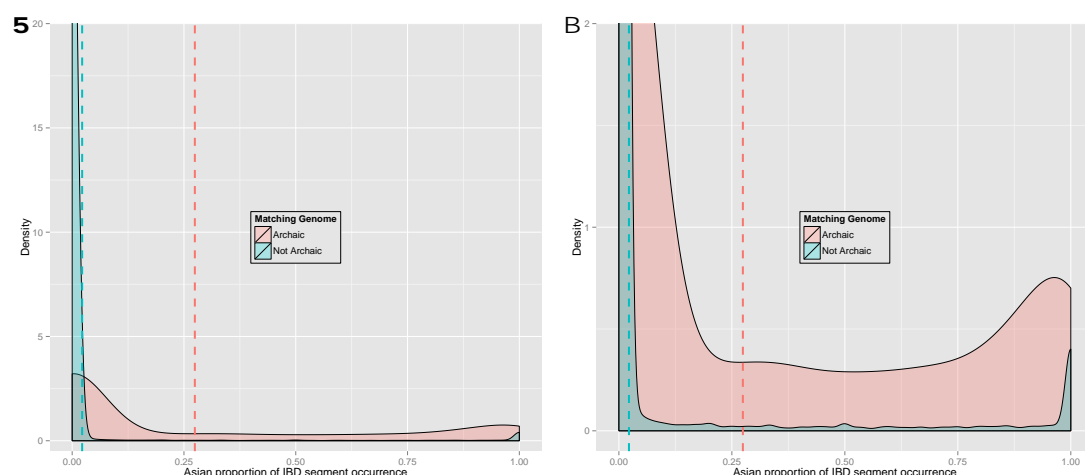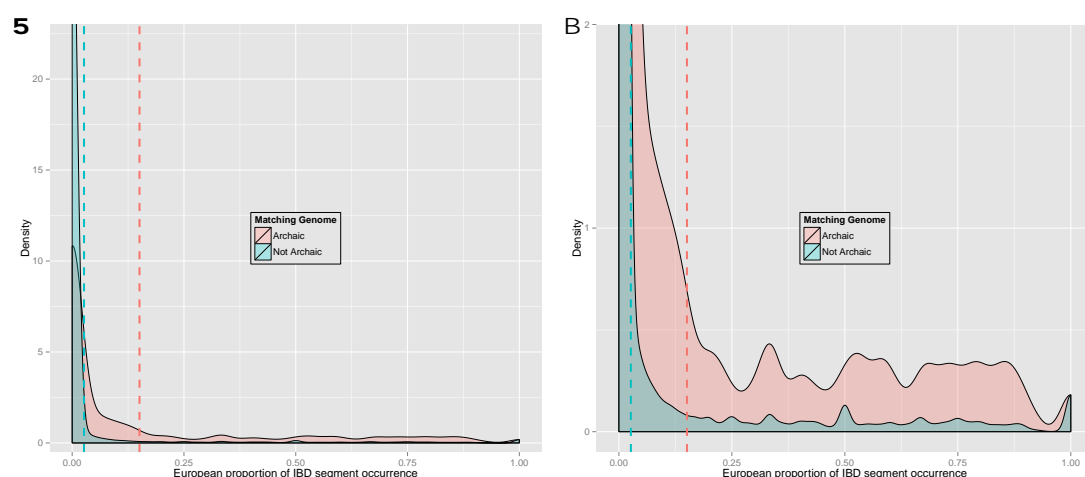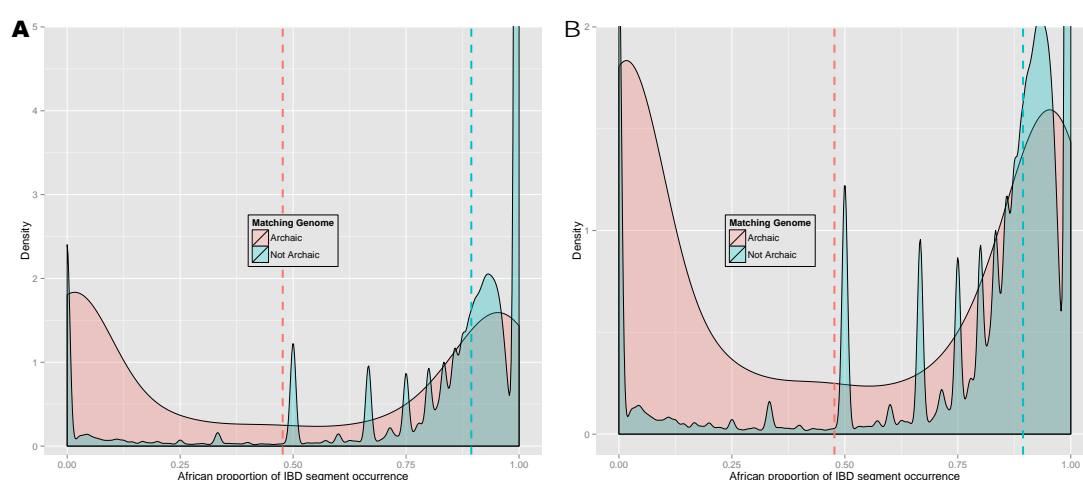
Figure 19: **Panel A**: Density of African proportions of Archaic-genome-matching IBD segments (pink) vs. density of African proportions of non-Archaic-genome- matching IBD segments (cyan). IBD segments were extracted from phased genotyping data of chromosome 1 of the 1000 Genomes Project. Dotted lines indicate the respective means. Peaks for non-Archaic-genome-matching IBD segments are found at 0.5, 0.66, 0.33, 0.75, and 0.8, which corresponds to 1/2, 2/3, 1/3, 3/4, 4/5 (number of Africans / total number of individuals that have the IBD segment). The density of African proportions of Archaic-genome-matching IBD segments has two peaks: one at a low and one at a high proportion of Africans. **Panel B**: the same density zoomed in.

4   Analyses of Lengths of IBD Segments

4.1   Relating the IBD Length to Years from Present

We aim to establish a relation between the length of an IBD segment and the time of the most recent common ancestor of the individuals that possess the IBD segment. The shorter the IBD segment is, the older it is assumed to be, the further in the past the most recent common ancestor should be found. For IBD length distributions, mathematical models have already been established. However, these models assume IBD segment sharing between only two individuals.

4.1.1   Exponential Distributed IBD Lengths

The length of an IBD segment is exponentially distributed with a mean of $100/(2g)$ cM (centi-Morgans), where g is the number of generations which separate the two haplotypes that share a segment from their common ancestor (4, 16, 34, 50, 51). Ulgen and Li (52) recommend to use a recombination rate, cM-to-Mbp ratio, of 1, however it varies from 0 to 9 along a chromosome (57).

We are not able to perform reliable age estimations of the IBD segments based on their length.

We encountered severe problems in estimating the age of IBD segments based on their length:

- The original ancestor DNA sequence is assumed to have a length of 1 Morgan before it is broken up by recombinations. However, founder genomes cannot be assumed to be distinguishable across the length of 1 Morgan.

- It is assumed that recombinations are random and all resulting segments have the same chance to survive. However, e.g. after population admixture or introgression of ancient genomes into ancestors of humans, recombined segments may have different fitness and some may vanish due to the high selective pressure. Thus, after such events the selective pressure leads to a bias of the IBD length distribution which makes the estimation of their age intractable.

- The age estimations are based on the mean, thus it is assumed that there are enough recombination events in each generation to average out random effects. Therefore, for few admixture/introgression events (few matings/offspring) these estimations are not reliable.

Due to these problems, we do not present age estimation at this point of our investigation.

4.1.2   Correction for the Assumptions of IBD Length Distributions

The IBD length distribution was derived from sharing between two individuals, but we consider IBD sharing among many individuals and compute the raw IBD segment length as the maximal IBD sharing of any two individuals that possess the IBD segment. This results in overestimation of the lengths, because it is the maximum of all pairwise sharings. We also observed a second cause for raw IBD segments being longer than expected by the exponential distribution. The more

individuals share an IBD segment, the more likely it is to find two individuals that share random minor alleles which would falsely extend the IBD segment.

Therefore, we corrected the raw lengths of IBD segments by locating the first tagSNV from the left (upstream) which is shared by at least 3/4 of the individuals that possess the IBD segment. This tagSNV is the left break point for the IBD segment. Analogously, we determined the right break point by the first tagSNV from the right (downstream) that is shared by at least 3/4 of the individuals. The distance between these break points is the (corrected) length of an IBD segment.

4.1.3   Length Correction for IBD with Ancient Genomes

We are interested in IBD between modern human and ancient genomes. However, the human IBD segment length is not an appropriate measure for the length of IBD with ancient genomes because only a part of the IBD segment may match an ancient genome (see Figure 33).

We corrected the IBD segment lengths to obtain the IBD lengths between human and ancient genomes. The corrected length of an IBD segment is the length of the "ancient part" that matches a particular ancient genome. This "ancient part" must contain at least 8 tagSNVs, which is the minimum number of tagSNVs per IBD segments. First, the left (upstream) break point of the "ancient part" of an IBD segment genome is detected. This left break point was defined as the first location in the IBD segment from the left (upstream), where at least 4 out of 8 tagSNVs match the ancient genome. From the right (downstream), the right break point of the "ancient part" of an IBD segment was detected analogously. Since not all bases of the ancient genomes were called, we modified the definition of the break points and required at least 6 bases of the 8 tagSNVs to be called of which 3, have to match the ancient genome. If either the left or right break point of an "ancient part" could not be found, then this IBD segment does not contain an "ancient part" and was excluded from all further analyses.

Matching of an IBD segment and an ancient genome for IBD segment lengths analyses was defined as:

1. at least 15% of the tagSNVs of the IBD segment must match the ancient genome,

2. the "ancient part" of the IBD segment must contain at least 8 tagSNVs, and

3. 30% of the tagSNVs in the "ancient part" of the IBD segment must match the ancient genome.

4.2   Histograms of Lengths of IBD Segments for the Different Genomes

Figure 20 shows the histograms of IBD segment lengths for all IBD segments (human genome) and for IBD segments that match the Neandertal genome. For the human genome a peak at 24,200 bp is visible, whereas, for the Neandertal genome, peaks are at 6,000 and 22,000 bp. It can be seen that IBD segments that match the Neandertal genome are shorter, thus also older.

Figure 21 shows the histograms of IBD segment lengths for all IBD segments that match the Denisova and the "Archaic" genome ("Archaic genome" contains IBD segments that match both the Denisova and Neandertal genome). For the Denisova genome, we have a peak at 4,000 bp, whereas, for the Archaic genome, peaks are at 11,000 bp and 24,000 bp. The peaks for the
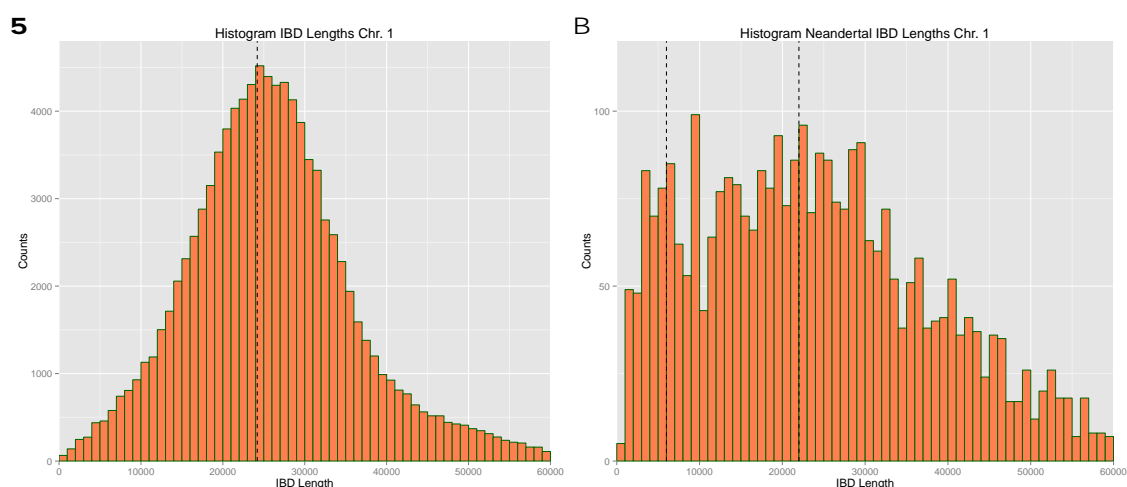
Figure 20: **Panel A**: Histogram of the IBD segment lengths for all IBD segments found in the 1000 Genomes Project data (human genome). The global peak is at 24,200 bp. **Panel B**: Histogram of the IBD segment lengths for IBD segments that match the Neandertal genome. Peaks at 6,000 bp and 22,000 bp are indicated.
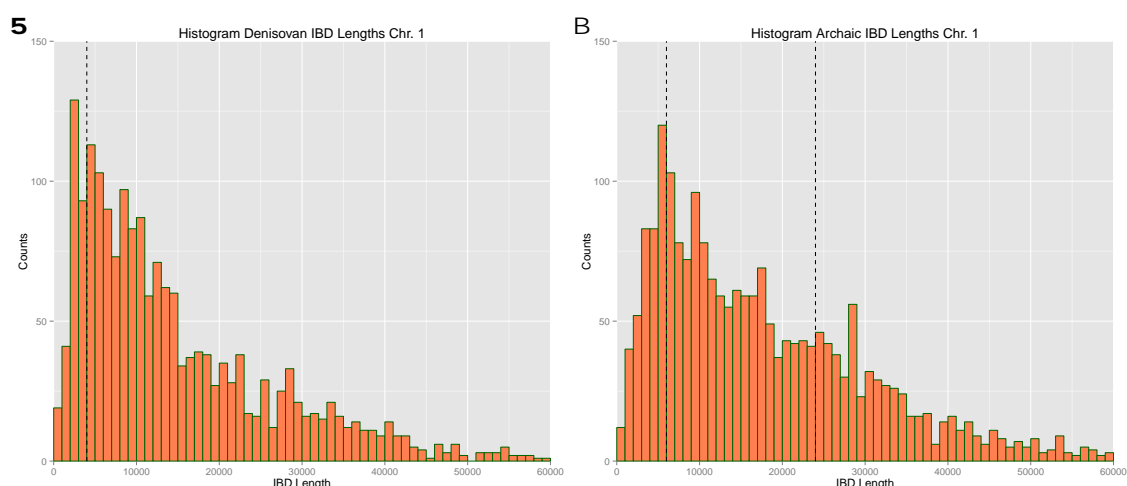

Figure 21: **Panel A**: Histogram of the IBD segment lengths for IBD segments that match the Denisova genome. A peak at 4,000 bp is indicated. **Panel B**: Histogram of the IBD segment lengths for IBD segments that match both the Neandertal and the Denisova genome ("Archaic genome"). Peaks at 6,000 bp and 24,000 bp are indicated.

Archaic genome are almost at the same positions as the corresponding peaks for the Denisova and Neandertal genome. IBD segment sharing between humans, Neandertals, and Denisovans may have two different reasons. First, the IBD segments may stem from a common ancestor and are passed on in each of these hominid groups. Secondly, IBD segment sharing may be caused by an introgression of one hominid group into another. For the former the IBD segments are supposed to be on average shorter than for the latter scenario. Therefore, the peaks at shortest, and hence oldest, IBD segments must be assumed to stem from a common ancestor of Neandertals and Denisovans. Some of these IBD segments may have been lost either in Neandertals or Denisovans (at least in the specimens analyzed) and are therefore not attributed to the Archaic genome but to one of the other two groups. Introgression of Neandertals into the Denisova genome or vice versa and a subsequent gene flow into humans can explain long, therefore more recent, IBD segments that are attributed to the Archaic genome. This hypothesis is supported by the results of Prüfer et al. (35) which show evidence of Neandertal gene flow into Denisovans.

4.3   IBD Segment Lengths of Human Populations

Figure 22 shows the density of the lengths of IBD segments that are private to Asians vs. the density of the lengths of IBD segments that are private to Europeans. Since IBD segments are private to each continental population, the densities are based on disjoint IBD segment sets. Both show three peaks at similar lengths. Asians show a global peak at 26,500 bp, while Europeans show a global peak at 25,300 bp. The other peaks are around 21,000 bp and 49,000 bp. For Asians the peak at 49,000 bp is larger than for Europeans resulting in a lower global peak.

Figure 23A shows the density of lengths of IBD segments that are private to Asians vs. lengths of IBD segments that are only shared between Asians and Africans. Figure 23B shows the same plot for Europeans instead of Asians. A small difference is visible in the global peaks of length distributions of IBD segments that are private to a continental population and those shared with Africans (blue dashed lines): 26,500 vs. 19,000 bp for Asians and 25,300 vs. 23,500 bp for Europeans. Segments that are also shared with Africans show an enrichment at shorter lengths compared to segments that are private to either Asians or Europeans.

Figure 24A shows the density of lengths of IBD segments that are private to Asians vs. lengths of IBD segments that are shared between Asians, Europeans, and Africans. Figure 24B shows the same plot for Europeans instead of Asians. Of course, the peaks for Asians and Europeans are the same as in Figure 23. Again segments that are shared by all three continental populations show an enrichment at shorter lengths compared to segments that are private to either Asians or Europeans. We already assumed that IBD segments that are shared by all population groups predate the Out-of-Africa split and therefore have to be very old. These results confirm our assumptions since we show that these segments are very short.

Next we investigated the effect on the length distribution if IBD segments are removed that are shared by Africans. Figure 25A shows the density of lengths of IBD segments that are private to Asians vs. the density of IBD segment lengths shared by Asians and Europeans, but not by Africans. In Figure 25B, the same plot as in Figure 25A is shown, but now compared to IBD segments that are private to Europeans. In Figure 24A and B, the density for IBD segments that are shared with Africans, has high values for shorter IBD segments (blue). In Figure 25A and B, this range of high values vanishes, because IBD segments that are shared with Africans are removed. A higher density region at longer segment length becomes visible for IBD segments that

are shared by Europeans and Asians, but not by Africans compared to segments that are private to either Asians or Europeans.
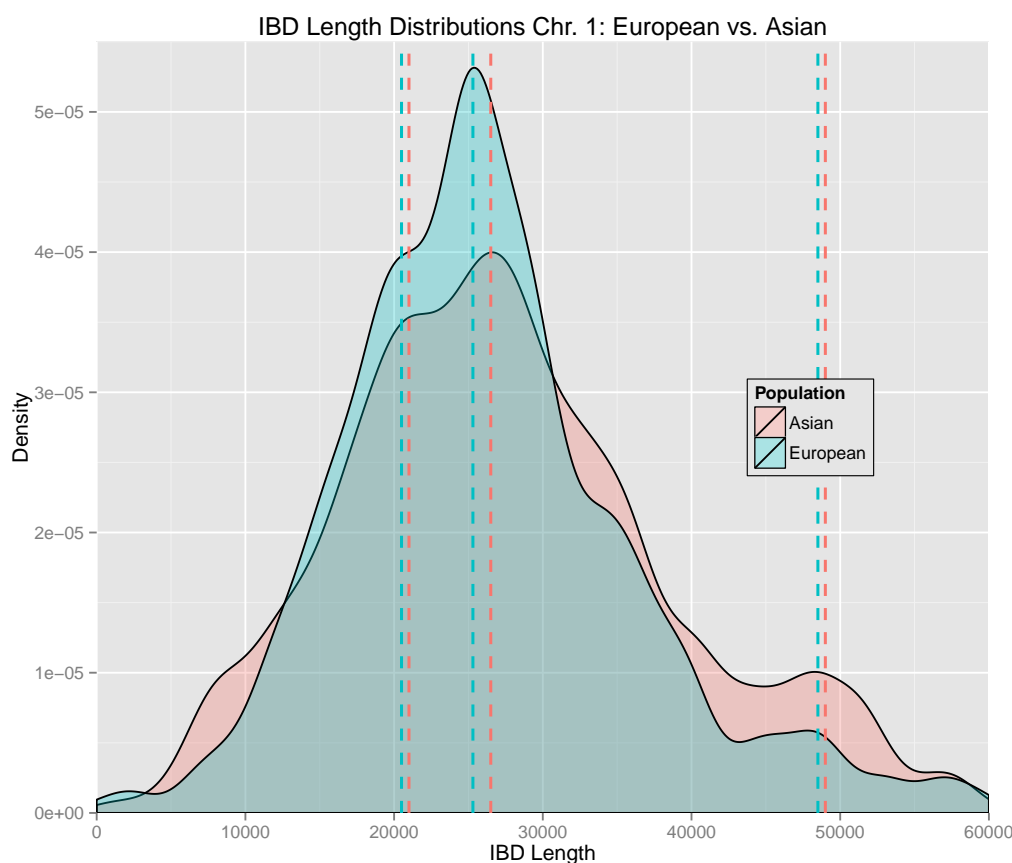

Figure 22: Density of lengths of IBD segments that are private to Asians vs. the analogous density for Europeans. Interesting peaks are marked by dashed lines. Asians have a global peak at 26,500 bp (red), while Europeans have the global peaks at 25,300 bp (blue). Both have smaller peaks at 21,000 bp and 49,000 bp.
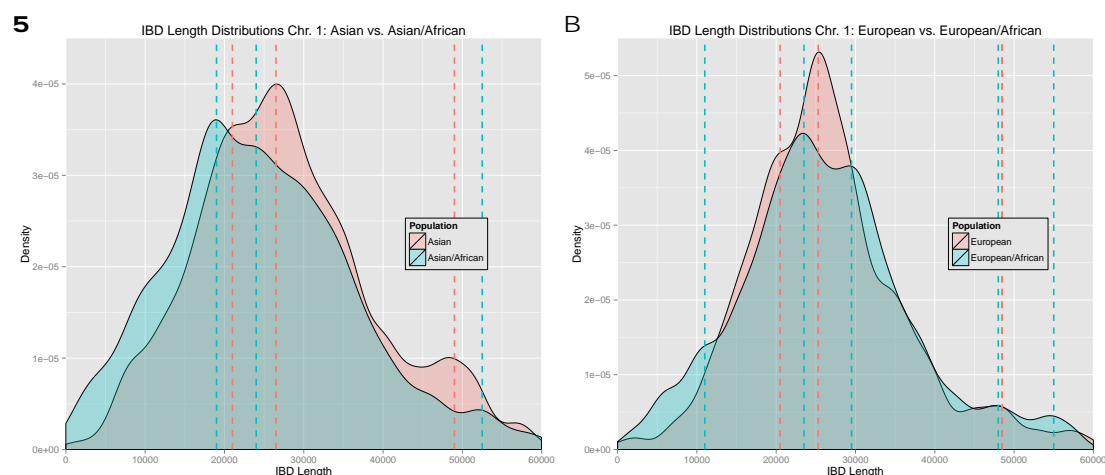
Figure 23: **Panel A**: Density of lengths of IBD segments that are private to Asians vs. density of IBD segment lengths shared only by Asians and Africans. **Panel B**: Density of lengths of IBD segments that are private to Europeans vs. density of IBD segment lengths shared only by Europeans and Africans. African-Asian IBD segments have peaks at 19,000 bp, 24,000 bp, and 52,500 bp (blue dashed lines in panel B). African-European IBD segments have a global peak at 23,500 bp and smaller peaks at 11,000 bp, 29,500 bp, 48,000 bp, and 52,000 bp (blue dashed lines in panel B). The global Asian peak is at 26,500 bp (red dashed line in panel A), while the global peak for Europeans is at 25,300 bp (red dashed line in panel B). Both have smaller peaks at 21,000 bp and 49,000 bp.

Figure 24: **Panel A**: Density of lengths of IBD segments that are private to Asians vs. density of IBD segment lengths shared by Asians, Europeans, and Africans. **Panel B**: Density of lengths of IBD segments that are private to Europeans vs. density of IBD segment lengths shared by Asians, Europeans, and Africans. Again segments that are shared by all three continental populations show an enrichment at shorter lengths compared to segments that are private to either Asians or Europeans (blue vs. red).
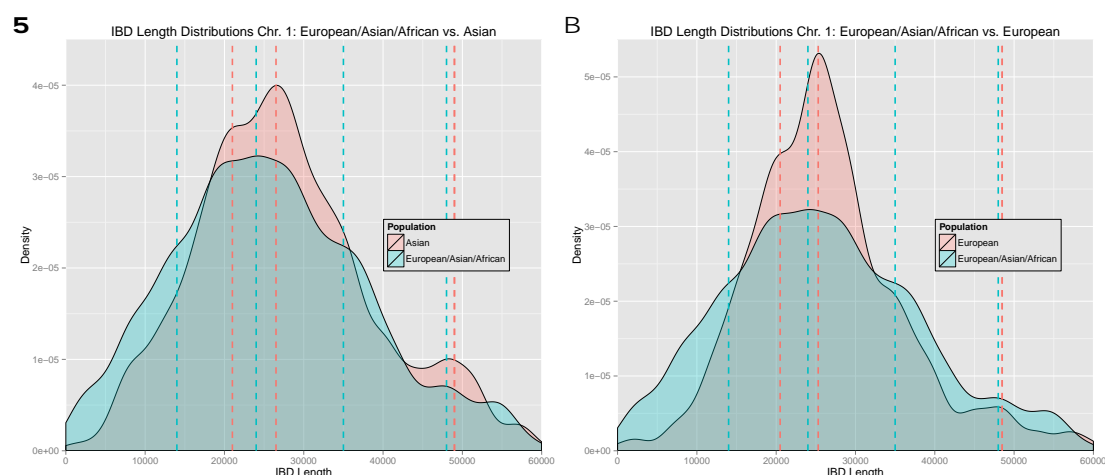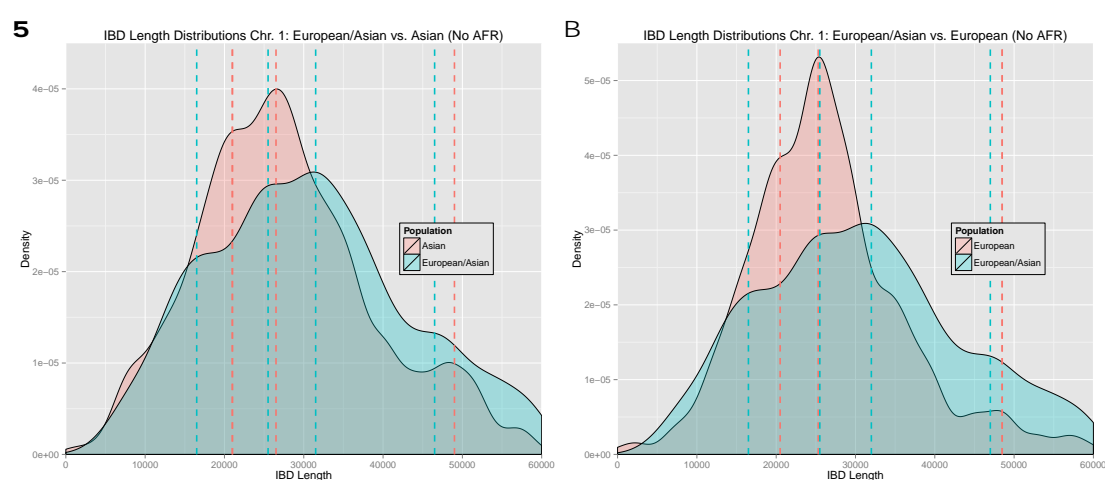
Figure 25: **Panel A**: Density of lengths of IBD segments that are private to Asians vs. density of IBD segment lengths shared by Asians and Europeans, but not by Africans meaning African-sharing IBD segments are removed as opposed to Figure 24A. **Panel B**: The same plot as in panel A, but IBD segments that are shared by Europeans and Asians are compared to IBD segments that are private to Europeans. African-sharing IBD segments are removed as opposed to Figure 24B. Both panels show that removing IBD segments shared with Africans in turn removes the higher densities at lower segment lengths that were present in Figure 24. Instead the densities for longer segment lengths are higher for IBD segments shared between Europeans and Asians (no Africans) compared to segments that are private to either Asians or Europeans.

4.4   Lengths of IBD Segments that Match the Denisova or Neandertal Genome

4.4.1   Lengths of IBD Segments that Match the Denisova Genome

Figure 26A shows the density of the lengths of all IBD segments (human genome) vs. the density of the lengths of IBD segments that match the Denisova genome. The length density of all IBD segments has a peak at 25,000 bp, while the length distribution of IBD with the Denisova genome has a peak at 5,000 bp. Clearly, segments of IBD with the Denisova genome are shorter and therefore older than those solely shared among present day humans (red density above blue at the left hand side). We were interested in how different populations share the Denisova genome. Figure 26B shows densities of lengths of IBD segments that match the Denisova genome and are enriched by a particular continental population. For all populations combined as well as for Africans and Europeans, a density peak is visible between 5,000 and 6,000 bp. Europeans have their global peak at 13,500 bp and additional peaks at 19,500 bp, 28,000 bp, and 37,500 bp. Denisova-matching segments that are shared by Asians are generally longer with a global peak at 31,000 bp and additional smaller peaks at 9,500 bp, 20,000 bp, and 39,500 bp. These densities seem to reveal that there was a gene flow from Denisovans into the Asian genome outside Africa. However, also Europeans show some hints of introgression from the Denisovans after migration out of Africa.
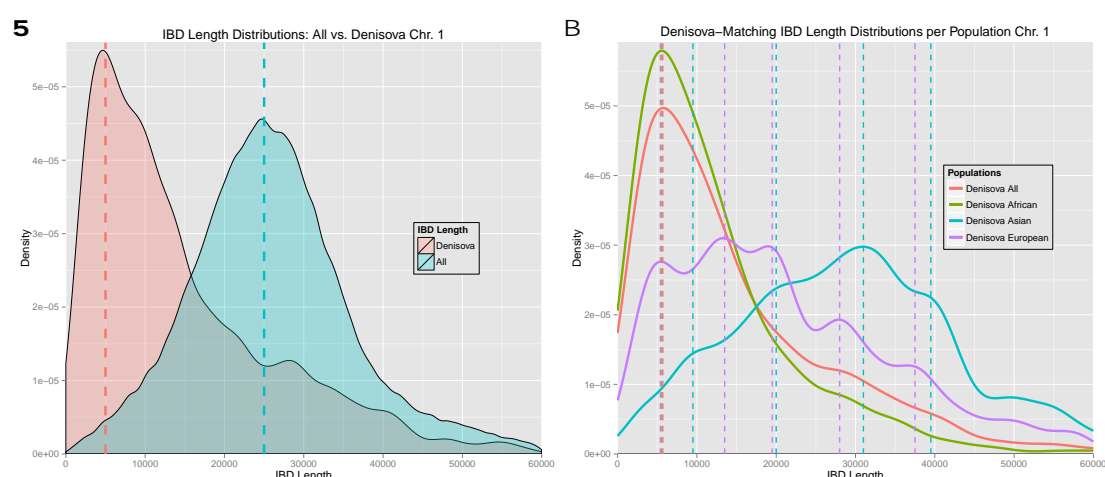

Figure 26: **Panel A**: Density of the lengths of all IBD segments vs. the lengths of IBD segments that match the Denisova genome. The dashed lines indicate the peaks at 9,000 bp for the Denisova and 24,200 bp for the human genome. **Panel B**: Densities of lengths of IBD segments that match the Denisova genome and are enriched in a particular population. The dashed lines indicate density peaks. For all populations combined as well as for Africans and Europeans, a density peak is visible between 5,000 and 6,000 bp. Europeans have their global peak at 13,500 bp and additional peaks at 19,500 bp, 28,000 bp, and 37,500 bp. Denisova-matching segments that are shared by Asians are generally longer with a global peak at 31,000 bp and additional smaller peaks at 9,500 bp, 20,000 bp, and 39,500 bp.

These peaks can be seen more clearly if only IBD segments are considered that are private to a population. Figure 27A shows densities of lengths of IBD segments that match the Denisova genome and are private to a population. For all populations together, the peak of the length density

is at 6,000 bp; for IBD segments private to Africans, it is at 7,000 bp; for IBD segments private to Europeans, it is at 12,000bp; and for IBD segments private to Asians, it is at 27,000 bp. The density of IBD lengths that match the Denisova genome and are private to Asians is also high around 40,000 bp and 54,000 bp. Europeans have an additional peak at 46,000 bp. Figure 27B shows densities of lengths of IBD segments that match the Denisova genome and are private to Africans vs. IBD segments that are not observed in Africans. The global peak for Africans is at 5,000 bp, while the density of lengths of IBD segments that are not observed in Africans has a global peak at 27,000 bp. Africans have older segments probably stemming from common ancestors of Denisovans and humans. For non-African populations, the high densities for longer IBD segments hint at an introgression from Denisovans after migration out of Africa.
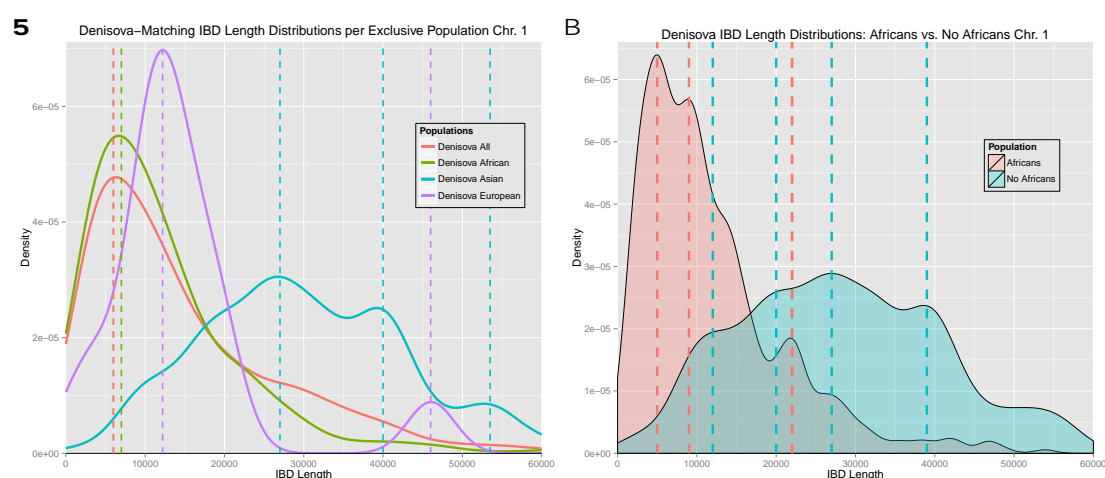

Figure 27: **Panel A**: Densities of lengths of IBD segments that match the Denisova genome and are private to a population. The peaks are at 6,000 for all IBD segments matching the Denisova genome, 7,000 for Africans, 12,000 for Europeans, and 27,000 for Asians. **Panel B**: Densities of lengths of IBD segments that match the Denisova genome and are private to Africans vs. IBD segments that are not observed in Africans. The global peak for Africans is at 5,000 bp, while the density of lengths of IBD segments that are not observed in Africans has a global peak at 27,000 bp.

4.4.2   Lengths of IBD Segments that Match the Neandertal Genome

Figure 28A shows the density of the lengths of all IBD segments (human genome) vs. the density of the lengths of IBD segments that match the Neandertal genome. The length density of all IBD segments has a peak at 25,000 bp, while the length distribution of IBD segments with the Neandertal genome (the part of the IBD segment that matches the Neandertal genome) has peaks at 5,500 bp, 23,000 bp, and 41,000 bp. Figure 28B shows densities of lengths of IBD segments that match the Neandertal genome and are enriched in a particular continental population. The peaks of the length distribution for IBD segments that are shared by all populations combined are at 6,000 bp and 23,000 bp. Africans have their global peak also around 6,000 bp. Europeans have their global peak at 24,000 bp and additional peaks at 39,000 bp and 52,000 bp. The density of lengths of IBD segments that are private to Asians and match the Neandertal genome has a global peak at 27,000 bp and a second peak around 42,000 bp. The density peak for Africans is clearly separated

from the density peaks for Europeans and Asians which almost match each other. This hints at introgression from the Neandertals into anatomically modern humans that were the ancestors of Europeans and Asians after these humans left Africa. The higher density of short IBD segments which are prominent in Africans in the range of 5,000–15,000 bp hints at old DNA segments that humans share with the Neandertal genome. Again our results contradict the hypothesis of Prüfer et al. (35) that Neandertal ancestry in Africans is due to back-to-Africa admixture, but instead hint at an earlier gene-flow within Africa.
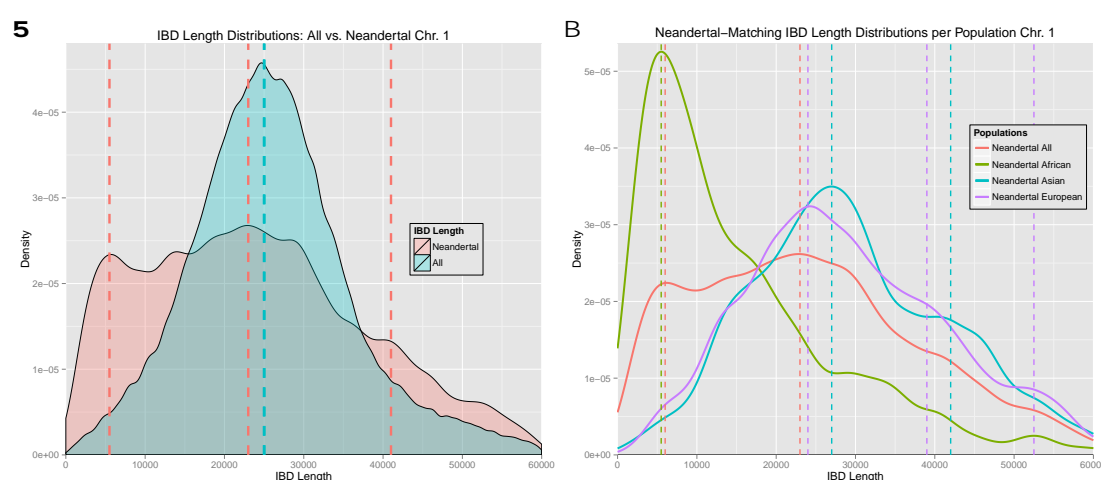

Figure 28: **Panel A**: Density of the lengths of all IBD segments vs. the lengths of IBD segments that match the Neandertal genome. The dashed lines indicate the peaks of densities at 25,000 bp for all IBD segments (blue) and at 5,500 bp, 23,000 bp, and 41,000 bp for Neandertal-matching segments (red). **Panel B**: Densities of lengths of IBD segments that match the Neandertal genome and are enriched in a particular population. The dashed lines indicate the density peaks around 6,000 bp for Africans, 6,000 bp and 23,000 bp for all populations combined, 27,000 bp and 42,000 bp for Asians, and 24,000 bp, 39,000 bp, and 52,000 bp for Europeans.

Next we considered IBD segments that are private to a continental population. Figure 29A shows densities of lengths of IBD segments that match the Neandertal genome and are private to a population. The peaks are at 6,000 bp for Africans and at 7,000 bp and 24,000 bp for all humans combined. The density for Europeans has peaks at 15,500 bp, 26,500 bp, 42,000 bp and 55,500 bp. Asians have density peaks at 26,000 bp and 41,000 bp. The densities of Asians and Europeans agree well with each other. They have peaks at longer segment lengths compared to Africans. Therefore, we were interested in IBD segments that match the Neandertal genome and that are private to Africans and those which are not observed in Africans hence shared by either Asians or Europeans or both. Figure 29B shows densities of lengths of IBD segments that match the Neandertal genome and are private to Africans vs. IBD segments that are not observed in Africans. The peak for Africans is around 5,000 bp, while IBD segments that are not observed in Africans have a global length density peak at 25,000 bp and smaller peaks at 14,000 bp and 41,000 bp. Most prominently, non-African IBD segments that match the Neandertal genome are enriched in regions of longer segment lengths.

Figure 29: **Panel A**: Densities of lengths of IBD segments that match the Neandertal genome and are private to a population. The major peaks are at 6,000 bp for Africans, 7,000 bp and 24,000 bp for all humans combined, 15,500 bp, 26,500 bp for Europeans, and 26,000 bp for Asians. **Panel B**: Densities of lengths of IBD segments that match the Neandertal genome and are private to Africans vs. IBD segments that are not observed in Africans. The peak for Africans is around 5,000 bp, while IBD segments that are not observed in Africans have a global length density peak at 25,000 and smaller peaks at 14,000 bp and 41,000 bp. Most prominently, non-African IBD segments that match the Neandertal genome are enriched in regions of longer segment lengths (blue density).
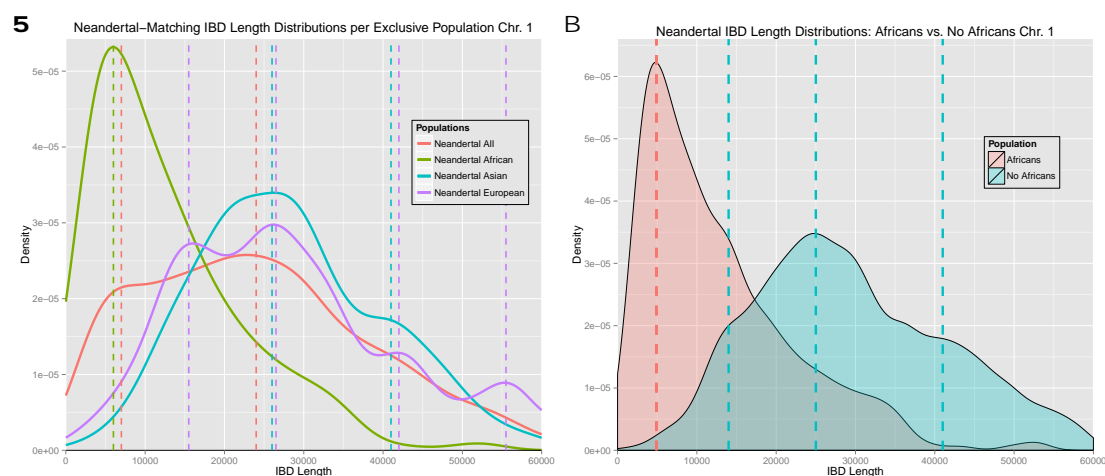
4.4.3   Lengths of IBD Segments that Match Neandertal & Denisova

IBD segments that match the "Archaic genome" are IBD segments that match both the Neandertal and the Denisova genome. Segments matching the Archaic genome stem either from a genome of archaic hominids which were ancestors of Neandertals and Denisovans or they stem from introgression of one hominid group into another. Garrigan et al. (12) were the first to present evidence for a prolonged period of ancestral population subdivision followed by relatively recent interbreeding in the history of human populations. Later, Wall et al. (54) found evidence that suggests that admixture between diverged hominid groups may have been a common feature of human evolution. Recently Prüfer et al. (35) published findings that suggest gene flow from Neandertals into Denisovans as well as an additional ancestral component in Denisovans from an unknown ancient population.

Figure 30A shows the density of the lengths of all IBD segments (human genome) vs. the density of the lengths of IBD segments that match the Archaic genome. For the human genome, the global density peak is again at 25,000 bp. For IBD with the Archaic genome, the global density peak is at 6,000 bp, but smaller peak at 15,500 bp and 24,500 bp can be observed. Figure 30B shows densities of lengths of IBD segments that match the Archaic genome and are enriched in a particular continental population. The global density peak for all populations as well as for Africans can be seen around 6,000 bp. Europeans and Asians have peaks around 9,000 bp, 16,000 bp, 26,000 bp, and 42,000 bp. Similar to IBD segments matching the Denisova or Neandertal genome separately, segments shared by Europeans or Asians are longer than segments shared by
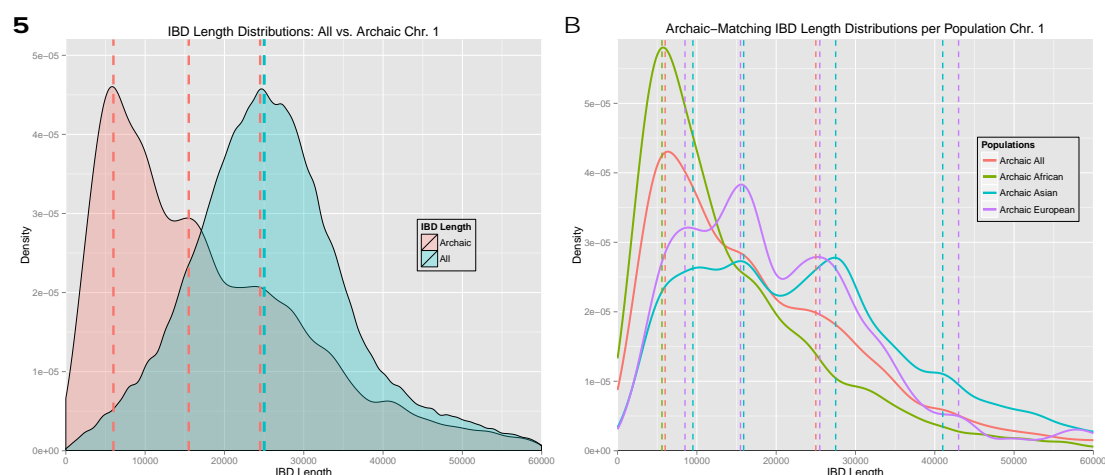
Africans.

Figure 30: **Panel A**: Density of the lengths of all IBD segments vs. the lengths of IBD segments that match the Archaic genome. The dashed lines indicate the peaks of densities at 6,000 bp and 25,000 bp as well as the smaller peaks at 15,500 bp and 24,500 bp. **Panel B**: Densities of lengths of IBD segments that match the Archaic genome and are enriched in a particular population. The dashed lines indicate the density peaks with the global peak for all populations combined as well as for Africans around 6,000 bp. Europeans and Asians have peaks at longer segment lengths ranging from 9,000 bp to 42,000 bp.

Next we considered IBD segments that are private to a continental population. Figure 31A shows densities of lengths of IBD segments that match the Archaic genome and are private to a population. All populations combined, as well as each population separately have a peak between 7,000 bp and 10,000 bp. Asians have their global peak between 17,500 bp and 26,000 bp. IBD segments that are private to Europeans have an additional peak at 34,000 bp. Figure 31B shows densities of lengths of IBD segments that match the Archaic genome and are private to Africans vs. IBD segments that are not observed in Africans. The global peak for Africans is at 5,000 bp, while lengths of IBD segments that are not observed in Africans have peaks at 15,000 bp and 26,500 bp. However, Africans have also a peak at 33,000 bp. Most prominently, non-African IBD segments that match the Archaic genome are enriched at larger segment lengths. This enrichment seems to be caused by events after humans migrated out of Africa. The introgression from the Neandertal into ancestors of modern humans may also have introduced a part of the Denisova genome that has been contained in the Neandertal genome, or vice versa. We would consider this part of the human genome as matching the Archaic genome.
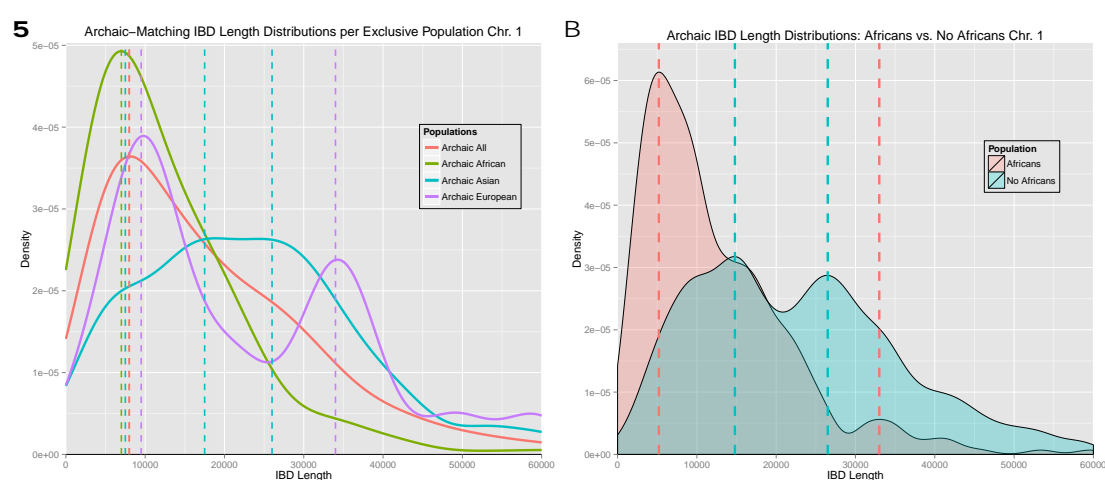
Figure 31: **Panel A**: Densities of lengths of IBD segments that match the Archaic genome and are private to a population. The peaks are between 7,000 bp and 10,000 bp for all populations combined, as well as each population separately. Asians have their global peak between 17,500 bp and 26,000 bp while Europeans have an additional peak at 34,000 bp. **Panel B**: Densities of lengths of IBD segments that match the Archaic genome and are private to Africans vs. IBD segments that are not observed in Africans. The global peak for Africans is at 5,000 bp, while lengths of IBD segments that are not observed in Africans have peaks at 15,000 bp and 26,500 bp. However, Africans have also a peak at 33,000 bp. Non-African IBD segments that match the Archaic genome are enriched at larger segment lengths (blue density).

5   Examples of IBD Segments that Match Ancient Genomes

Figure 32 shows a typical example of an IBD segment that matches the Denisova genome and is shared exclusively among Asians. Figure 33 shows an IBD segment that matches the Denisova genome and is shared by Africans and one Admixed American.

Figure 34 shows an example of an IBD segment that matches the Neandertal genome and is shared by Europeans, Asians, Admixed Americans, and Americans with African ancestry from SW US. Figure 35 shows an example of an IBD segment that matches the Neandertal genome and is shared by Europeans and Asians. Figure 36 shows an IBD segment that matches the Neandertal genome and is found exclusively in Asians.

Figure 37 shows an example of an IBD segment matching the Neandertal and the Denisova genome shared by Asians, Admixed Americans, and Americans with African ancestry from SW US. Figure 38 shows an example of an IBD segment matching the Neandertal and the Denisova genome shared by Africans and Admixed Americans.
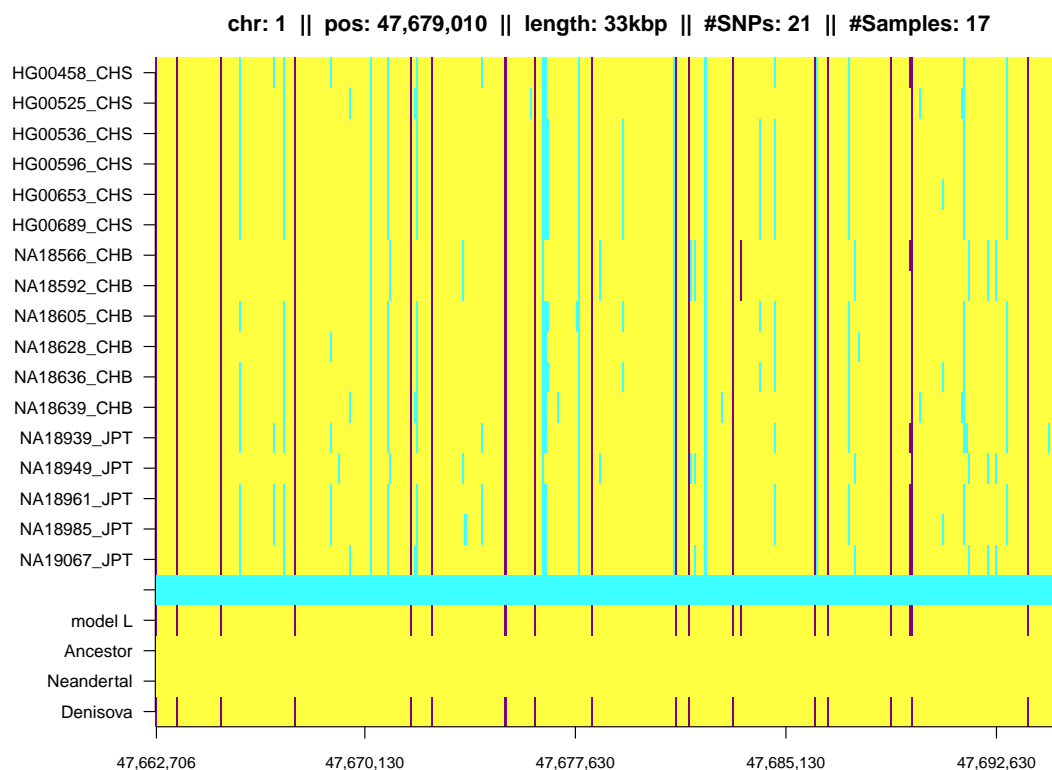
Figure 32: Example of an IBD segment matching the Denisova genome shared exclusively among Asians. The data analyzed by HapFABIA were genotypes from chromosome 1 of the 1000 Genomes Project. The rows give all individuals that contain the IBD segment and columns consecutive SNVs. Major alleles are shown in yellow, minor alleles of tagSNVs in violet, and minor alleles of other SNVs in cyan. The row labeled "model L" indicates tagSNVs identified by Hap-FABIA in violet. The rows "Ancestor", "Neandertal", and "Denisova" show bases of the respective genomes in violet if they match the minor allele of the tagSNVs (in yellow otherwise). For the "Ancestor genome" we used the reconstructed common ancestor sequence that was provided as part of the 1000 Genomes Project data.
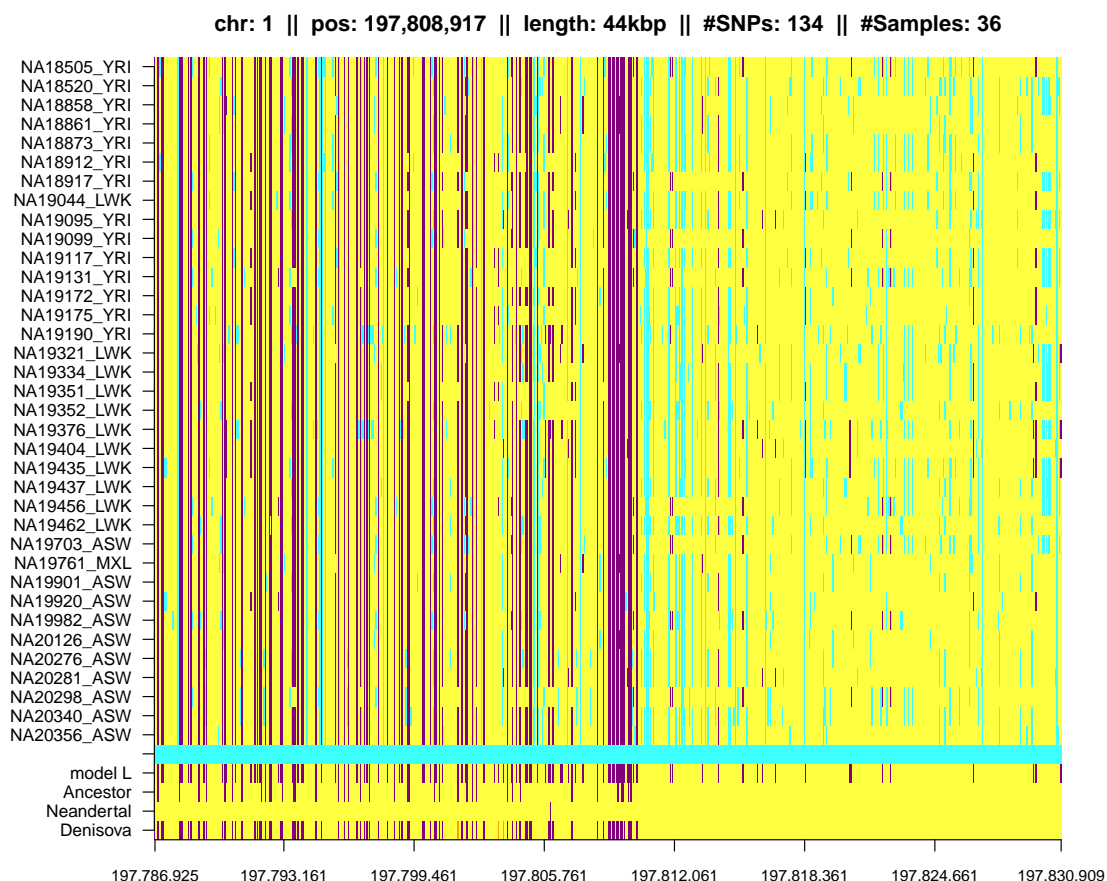
Figure 33: Example of an IBD segment matching the Denisova genome shared by Africans and one Admixed American. Only the first part of the IBD segment matches the Denisova genome and part of the IBD segment is shared by the majority of individuals. Many tagSNVs are also present in the reconstructed ancestor sequence, but they are not present in the Neandertal genome. See Figure 32 for a description.

Figure 34: Example of an IBD segment matching the Neandertal genome shared by Europeans, Asians, Admixed Americans, and Americans with African ancestry from SW US. See Figure 32 for a description.

Figure 35: Example of an IBD segment matching the Neandertal genome shared by Europeans and Asians. See Figure 32 for a description.
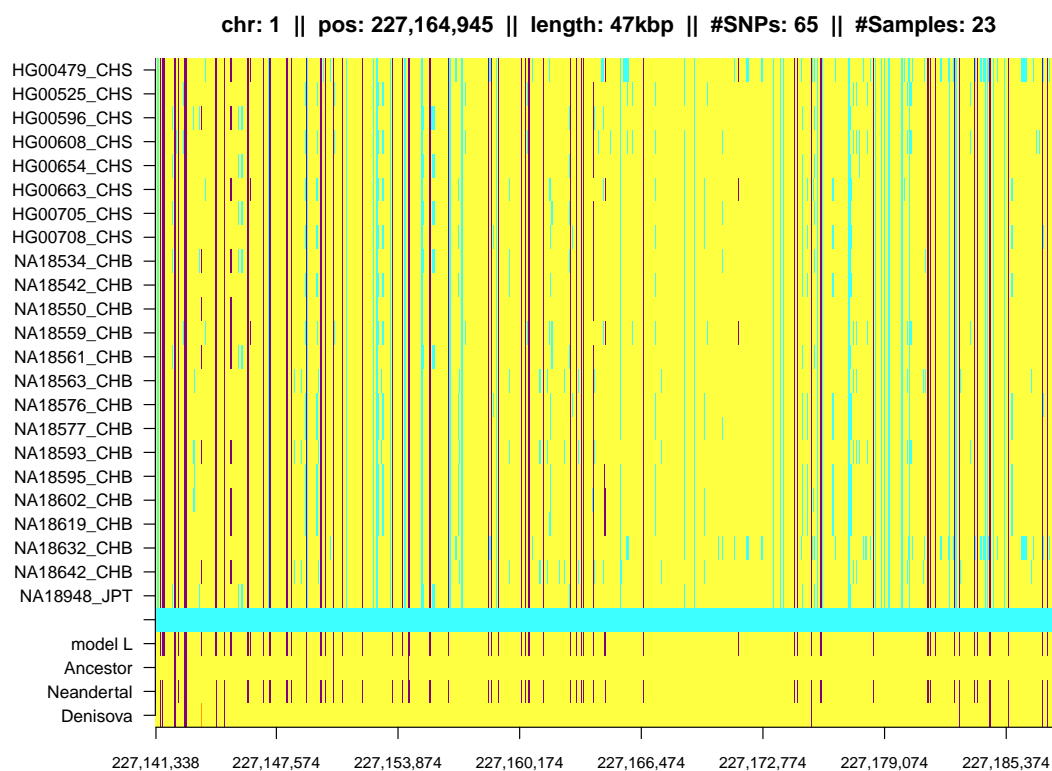
Figure 36: Example of an IBD segment matching the Neandertal genome shared exclusively among Asians. See Figure 32 for a description.

Figure 37: Example of an IBD segment matching the Neandertal and the Denisova genome shared by Asians, Admixed Americans, and Americans with African ancestry from SW US. See Figure 32 for a description.
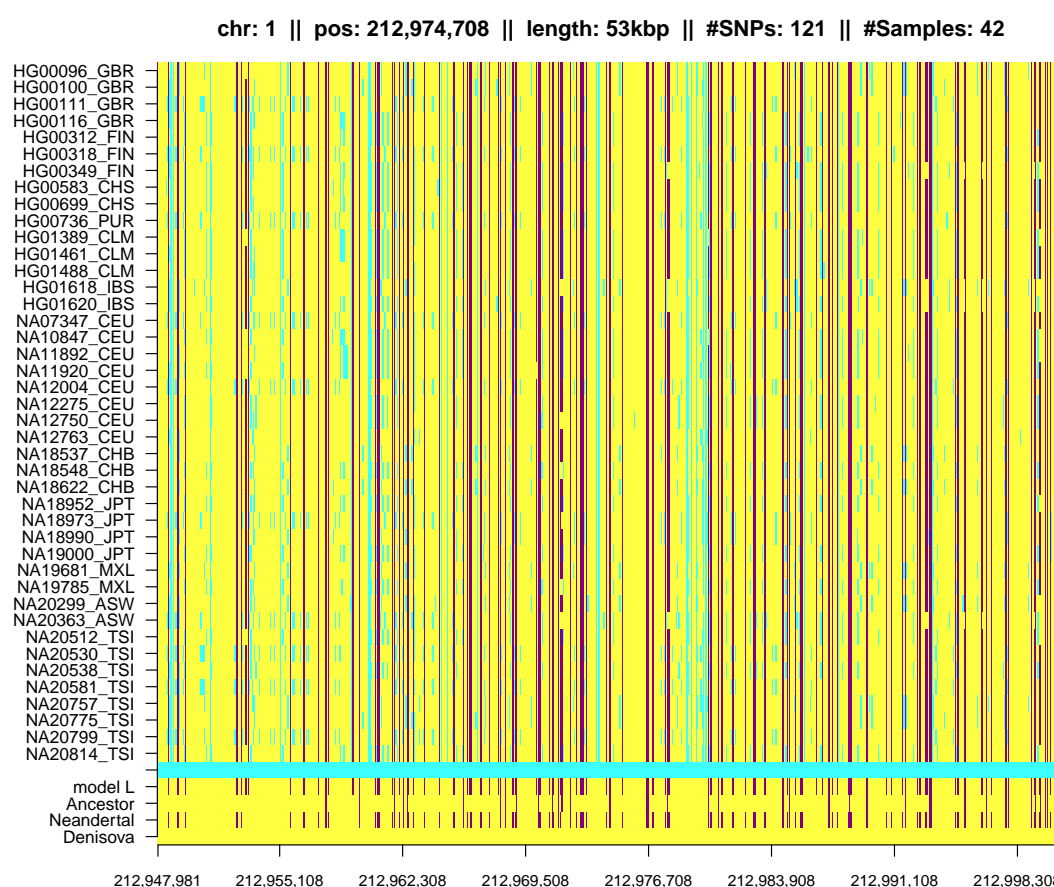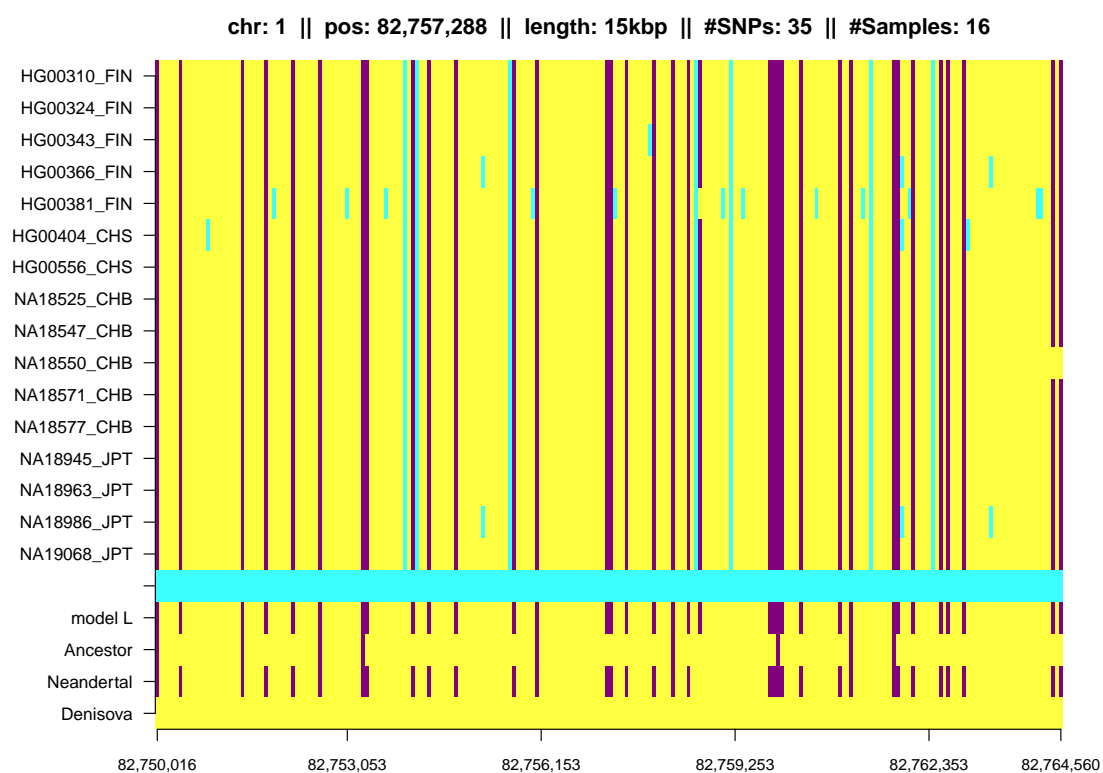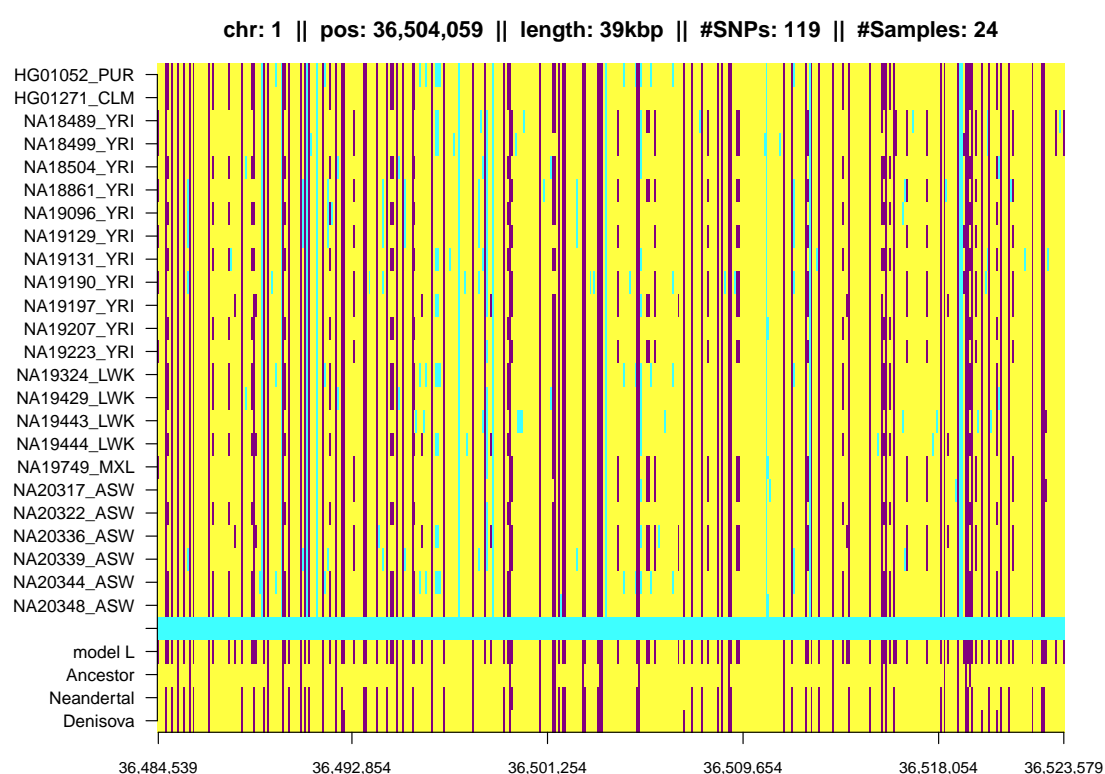
Figure 38: Example of an IBD segment matching the Neandertal and the Denisova genome shared by Africans and Admixed Americans. See Figure 32 for a description.
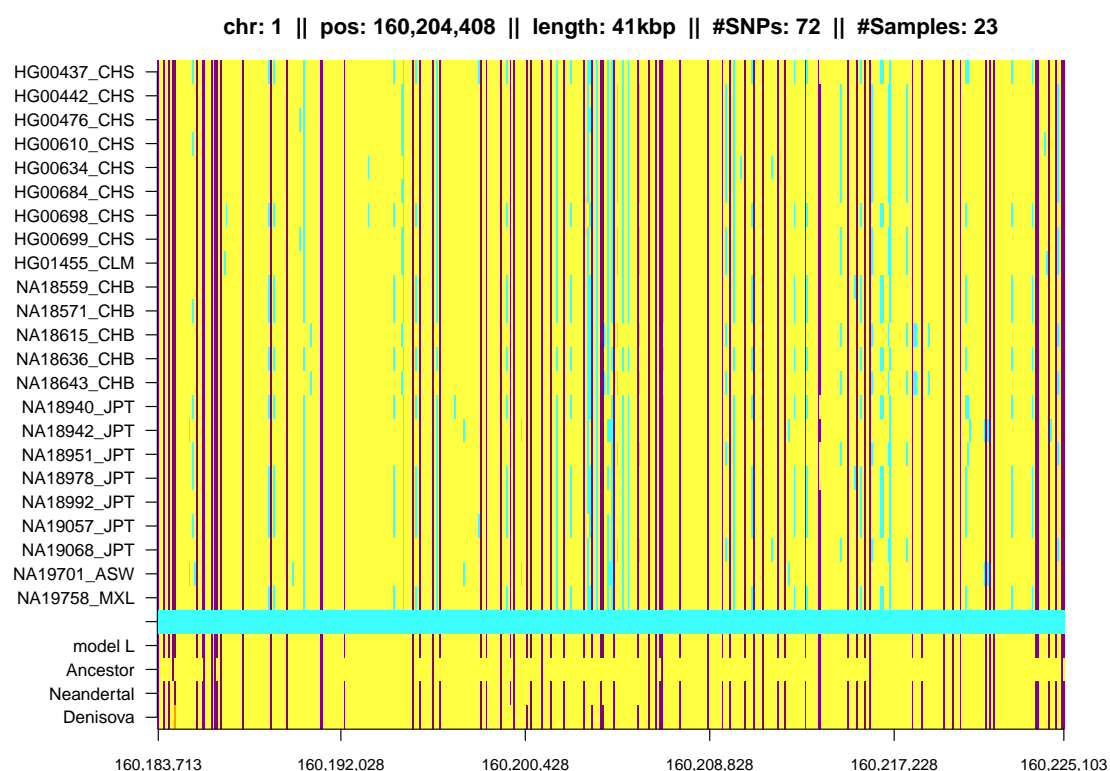
6   Conclusion

We applied HapFABIA to the 1000 Genomes Project data to extract very short identity by descent (IBD) segments and analyzed the IBD sharing patterns of human populations, Neandertals, and Denisovans.

Some IBD segments are shared with (1) the reconstructed ancestral genome of humans and other primates, (2) the Neandertal genome, and/or (3) the Denisova genome. We could confirm an introgression from Denisovans into ancestors of Asians after their migration out of Africa. Furthermore, IBD sharing hints at a gene flow from Neandertals into ancestors of Asians and Europeans after they left Africa (to a larger extend into ancestors of Asians). Interestingly, many Neandertal- or Denisova-matching IBD segments are predominantly observed in Africans — some of them even exclusively. IBD segments shared between Africans and Neandertals or Denisovans are strikingly short, therefore we assume that they are very old. This may indicate that these segments stem from ancestors of humans, Neandertals, and Denisovans and have survived in Africans.

In the near future the 1000 Genomes Project will be completed by genome sequences from additional populations, the UK10K project extents the subset of European genomes, and, most importantly, more ancient genomes will be successfully sequenced (Ust-Ishim, El Sidrón, etc.). Therefore, we expect that the analysis of the population structure by sharing patterns of very short IBD segments will become an increasingly important method in population genetics and its results will become more and more fine-grained.

Appendices

A   Are Rare Variants Recent or Old?

Rare variants are found to be old which might seem to contradict results of other researchers.

However, the conclusion that rare variants are, in general, recent does not contradict our results. On the contrary, that the majority of IBD segments extracted by HapFABIA is found in Africans can explain why three times as many variants with 0.5–5% minor allele frequency are found in Africans as in Europeans or Asians in the 1000 Genomes Project (48). In particular, if rare variants are caused by a recent population growth, then the numbers found by the 1000 Genomes Project (48) are difficult to interpret. We want to discuss two reasons why rare variants may be old.

(I) Many rare variants are old, however, most rare variants may be recent, e.g. because of a recent growth in population. We did not consider private variants and have a bias toward less rare variants because, the more individuals share an IBD segment, the higher its significance, the more likely it is detected by HapFABIA. In our analysis, only 39% of the rare variants (not counting private variants) are in IBD segments. Therefore, the remaining majority of rare and private variants might be recent.

The vast majority of IBD segments is found in Africans. This would explain why Africans have more rare and low-frequency variants than Europeans or Asians. The publication of the 1000 Genomes Project (48) reports:

> "individuals from populations with substantial African ancestry (YRI, LWK and ASW) carry up to three times as many low-frequency variants (0.5–5% frequency) as those of European or East Asian origin,"

Table S14 in the supplementary information of the publication of the 1000 Genomes Project (48) provides the number of derived variants per individual in each population in its last rows (DAF denotes "derived allele frequency", i.e. the frequency of the mutation):

	African			European				
	ASW	LWK	YRI	CEU	FIN	GBR	IBS	TSI
DAF <0.5%	116,216	152,267	122,978	30,034	32,918	31,944	36,129	33,425
DAF 0.5-5%	558,996	649,303	683,239	143,987	146,737	146,802	155,270	146,066

	Asian			Admixed American		
	CHB	CHS	JPT	CLM	MXL	PUR
DAF <0.5%	36,081	40,974	45,107	45,184	38,654	48,892
DAF 0.5-5%	119,668	118,137	120,115	189,257	163,737	214,430

Recent variants are supposed to be derived variants. The table shows about four times more variants with derived allele frequency of 0.5-5% in Africans than in other populations (if we ignore the Admixed Americans that have African admixture). For derived allele frequency < 0.5%, we observe 2.5 times more variants in Africans than in Europeans or Asians.

(II) Many rare variants are old, however, compared to common SNPs, they are recent, i.e. the temporal relations remain, but variants are dated further back. We state in the manuscript that many rare variants are old and from times before humans migrated out of Africa. However, many common SNPs are even older and stem from common ancestors of human and chimpanzee. This means that many rare variants are old, but compared to common SNPs they are recent.

The fact that common SNPs are old is supported by findings of the Chimpanzee Consortium. In Hacia et al. (17) it was found that, of 397 human SNP sites, 214 were ancestral (shared with common ancestors of chimpanzee and human). Of the ancestral SNPs, 1/4 had the minor allele as ancestral allele. For the chimpanzee genome (49), it was found that

> "Of ∼7.2 million SNPs mapped to the human genome in the current public database, we could assign the alleles as ancestral or derived in 80% of the cases according to which allele agrees with the chimpanzee genome sequence"

and that

> "a significant proportion of derived alleles have high frequencies: 9.1% of derived alleles have frequency $\geq 80\%$."

According to (49, Suppl. Fig. S9) about 25% of the derived alleles have frequency $\geq 50\%$ in which case the minor allele is the ancestral allele.

That some SNPs are very old and that some haplotypes are shared between humans and chimpanzee was also found in Leffler et al. (26):

> "We conducted a genome-wide scan for long-lived balancing selection by looking for combinations of SNPs shared between humans and chimpanzees. In addition to the major histocompatibility complex, we identified 125 regions in which the same haplotypes are segregating in the two species,"

In a recent publication (11), it was found that

> "The average age across all SNVs was $34,200 \pm 900$ years ($\pm$ s.d.) in European Americans and $47,600 \pm 1,500$ years in African Americans, and these estimates were robust to sequencing errors [...]".

The authors further write

> "SNVs shared between European Americans and African Americans were significantly older (104,400 years and 115,800 years for European Americans and African Americans, respectively)".

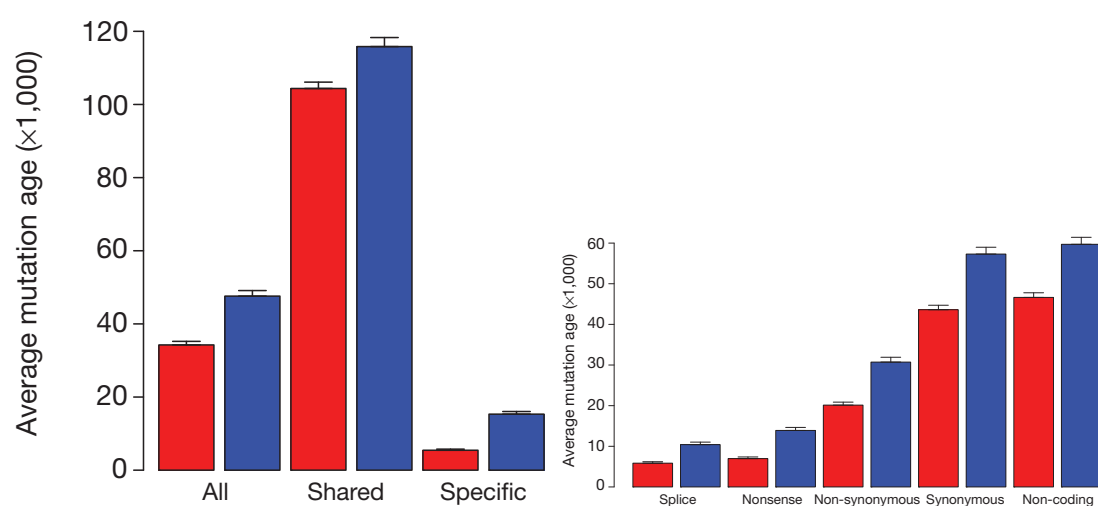These figures are visualized as bar plots in Figure 39.

Figure 39: Left: Average age of all SNVs, "Shared" SNVs found both in European Americans (red) and in African Americans (blue), and SNVs found in only one population ("Specific"). Right: Average age for different functional types of variants. Both plots are taken from Fu et al. (11).

B   Population Groups of the 1000 Genomes Project

The subpopulations of the 1000 Genomes Project are given in Table 2 and the locations, where the individuals reside, are shown on the map in Figure 40.

Table 2: Overview of population groups of the 1000 Genomes Project.

Pop	#Ind	Grp	Location
ASW	122	AFR	Americans with African ancestry from SW US
YRI	176	AFR	Yoruba in Ibadan, Nigeria
LWK	194	AFR	Luhya from Webuye, Kenya
CLM	120	AMR	Colombians in Medellin, Colombia
MXL	132	AMR	Mexicans from Los Angeles, California
PUR	110	AMR	Puerto Ricans
CEU	170	EUR	Utah residents with ancestry from northern and western Europe
FIN	186	EUR	Finnish
GBR	178	EUR	British from England and Scotland
IBS	28	EUR	Iberians from Spain
TSI	196	EUR	Toscani in Italy
CHB	194	ASN	Han Chinese from Beijing
CHS	200	ASN	Han Chinese from South
JPT	178	ASN	Japanese from Tokyo, Japan

The column labeled "Pop" gives the population, "#Ind" reports the number of individuals, "Grp" gives the continental group, and "Location" the location from where the individuals stem.
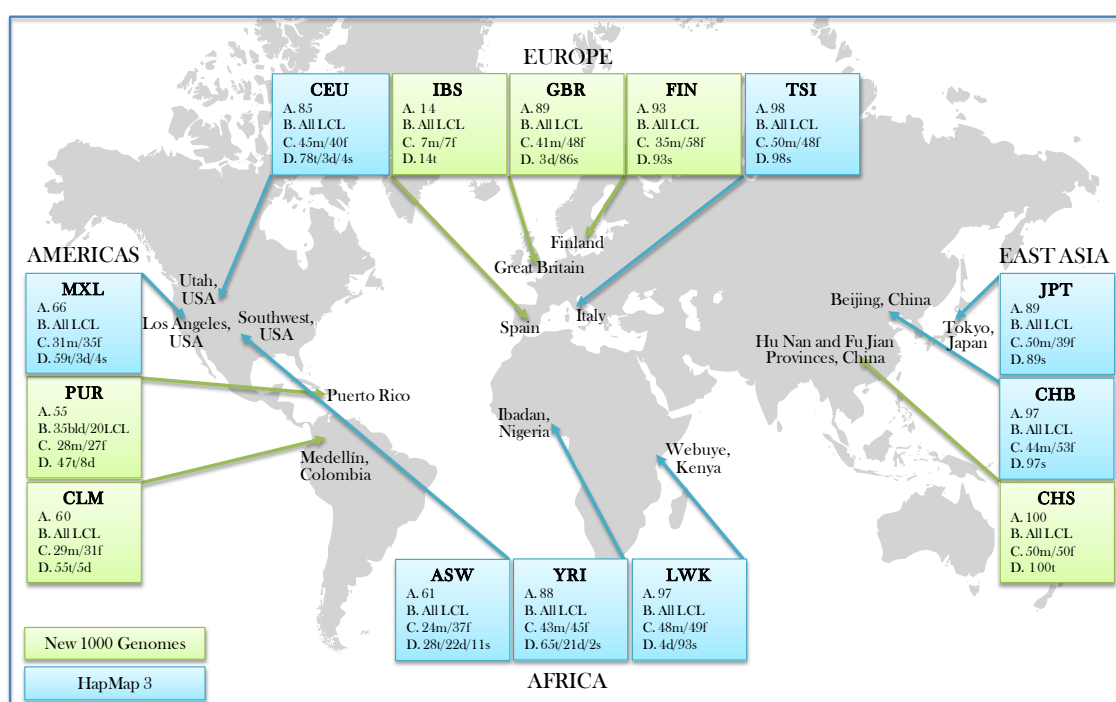
Figure 40: Map indicating the locations of the populations the 1000 Genomes Project phase I. Figure taken from (48).

References

[1] A. Auton, K. Bryc, A. R. Boyko, K. E. Lohmueller, J. Novembre, A. Reynolds, A. Indap, M. H. Wright, J. D. Degenhardt, R. N. Gutenkunst, K. S. King, M. R. Nelson, and C. D. Bustamante. Global distribution of genomic diversity underscores rich complex history of continental human populations. *Genome Research*, 19(5):795–803, 2009.

[2] M. G. B. Blum and M. Jakobsson. Deep divergences of human gene trees and models of human origins. *Mol. Biol. Evol.*, 28(2):889–898, 2011.

[3] L. R. Botigué, B. M. Henn, S. Gravel, B. K. Maples, C. R. Gignoux, E. Corona, G. Atzmon, E. Burns, H. Ostrer, C. Flores, J. Bertranpetit, D. Comas, and C. D. Bustamante. Gene flow from north africa contributes to differential human genetic diversity in southern europe. *Proceedings of the National Academy of Sciences*, 110(29):11791–11796, 2013.

[4] S. R. Browning. Estimation of pairwise identity by descent from dense genetic marker data in a population sample of haplotypes. *Genetics*, 178(4):2123–2132, 2008.

[5] S. R. Browning and B. L. Browning. Identity by descent between distant relatives: Detection and applications. *Annu. Rev. Genet.*, 46:617–633, 2012.

[6] K. Bryc, A. Auton, M. R. Nelson, J. R. Oksenberg, S. L. Hauser, S. Williams, A. Froment, J.-M. Bodo, C. Wambebe, S. A. Tishkoff, et al. Genome-wide patterns of population structure and admixture in west africans and african americans. *P. Natl. Acad. Sci. USA*, 107(2):786–791, 2009.

[7] S. Carmi, P. F. Palamara, V. Vacic, T. Lencz, A. Darvasi, and I. Pe'er. The variance of identity-by-descent sharing in the wright–fisher model. *Genetics*, 193(3):911–928, 2013.

[8] D.-A. Clevert, A. Mitterecker, A. Mayr, G. Klambauer, M. Tuefferd, A. DeBondt, W. Talloen, H. W. H. Göhlmann, and S. Hochreiter. cn.FARMS: a latent variable model to detect copy number variations in microarray data with a low false discovery rate. *Nucleic Acids Res.*, 39(12):e79, 2011.

[9] M. Currat and L. Excoffier. Modern humans did not admix with neanderthals during their range expansion into europe. *PLoS Biol*, 2(12):e421, 11 2004.

[10] A. Eriksson and A. Manica. Effect of ancient population structure on the degree of polymorphism shared between modern human populations and ancient hominins. *Proceedings of the National Academy of Sciences*, 109(35):13956–13960, 2012.

[11] W. Fu, T. D. O'Connor, G. Jun, H. M. Kang, G. Abecasis, S. M. Leal, S. Gabriel, D. Altshuler, J. Shendure, D. A. Nickerson, et al. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature*, 493(7431):216–220, 2013.

[12] D. Garrigan, Z. Mobasher, S. B. Kingan, J. A. Wilder, and M. F. Hammer. Deep haplotype divergence and long-range linkage disequilibrium at xp21.1 provide evidence that humans descend from a structured ancestral population. *Genetics*, 170(4):1849–1856, 2005.

[13] S. Gravel, B. M. Henn, R. N. Gutenkunst, A. R. Indap, G. T. Marth, A. G. Clark, F. Yu, R. A. Gibbs, The 1000 Genomes Project, and C. D. Bustamante. Demographic history and rare allele sharing among human populations. *P. Natl. Acad. Sci. USA*, 108(29):11983–11988, 2011.

[14] S. Gravel, F. Zakharia, A. Moreno-Estrada, J. K. Byrnes, M. Muzzio, J. L. Rodriguez-Flores, E. E. Kenny, C. R. Gignoux, B. K. Maples, W. Guiblet, J. Dutil, M. Via, K. Sandoval, G. Bedoya, T. K. Oleksyk, A. Ruiz-Linares, E. G. Burchard, J. C. Martinez-Cruzado, C. D. Bustamante, and The 1000 Genomes Project. Reconstructing native american migrations from whole-genome and whole-exome data. *PLoS Genet*, 9(12):e1004023, 2013.

[15] R. E. Green, J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, H. Li, W. Zhai, M. H. Fritz, et al. A draft sequence of the Neandertal genome. *Science*, 328(5979):710–722, 2010.

[16] A. Gusev, P. F. Palamara, G. Aponte, Z. Zhuang, A. Darvasi, P. Gregersen, and I. Pe'er. The architecture of long-range haplotypes shared within and across populations. *Mol. Biol. Evol.*, 29(2):473–486, 2012.

[17] J. G. Hacia, J.-B. Fan, O. Ryder, L. Jin, K. Edgemon, G. Ghandour, R. A. Mayer, B. Sun, L. Hsie, C. M. Robbins, et al. Determination of ancestral alleles for human single-nucleotide polymorphisms using high-density oligonucleotide arrays. *Nat. Genet.*, 22(2):164–167, 1999.

[18] B. V. Halldorsson, D. Aguiar, R. Tarpine, and S. Istrail. The Clark phaseable sample size problem: long-range phasing and loss of heterozygosity in GWAS. *J. Comput. Biol.*, 18(3):323–333, 2011.

[19] S. Hochreiter. HapFABIA: Identification of very short segments of identity by descent characterized by rare variants in large sequencing data. *Nucleic Acids Research*, 41(22):e202, 2013.

[20] S. Hochreiter, U. Bodenhofer, M. Heusel, A. Mayr, A. Mitterecker, A. Kasim, S. VanSanden, D. Lin, W. Talloen, L. Bijnens, H. W. H. Göhlmann, Z. Shkedy, and D.-A. Clevert. FABIA: factor analysis for bicluster acquisition. *Bioinformatics*, 26(12):1520–1527, 2010.

[21] S. Hochreiter, D.-A. Clevert, and K. Obermayer. A new summarization method for Affymetrix probe level data. *Bioinformatics*, 22(8):943–949, 2006.

[22] G. Klambauer, K. Schwarzbauer, A. Mayr, D.-A. Clevert, A. Mitterecker, U. Bodenhofer, and S. Hochreiter. cn.MOPS: mixture of poissons for discovering copy number variations in next generation sequencing data with a low false discovery rate. *Nucleic Acids Res.*, 40(9):e69, 2012.

[23] G. Klambauer, T. Unterthiner, and S. Hochreiter. Dexus: identifying differential expression in rna-seq studies with unknown conditions. *Nucleic Acids Research*, 41(21):e198, 2013.

[24] J. Krause, C. Lalueza-Fox, L. Orlando, W. Enard, R. E. Green, H. A. Burbano, J.-J. Hublin, C. Hänni, J. Fortea, Marco de la Rasilla, J. Bertranpetit, A. Rosas, and S. Pääbo. The derived foxp2 variant of modern humans was shared with neandertals. *Current Biology*, 17(21):1908–1912, 2007.

[25] M. Krings, A. Stone, R. W. Schmitz, H. Krainitzki, M. Stoneking, and S. Pääbo. Neandertal dna sequences and the origin of modern humans. *Cell*, 90(1):19–30, 1997.

[26] E. M. Leffler, Z. Gao, S. Pfeifer, L. Ségurel, A. Auton, O. Venn, R. Bowden, R. Bontrop, J. D. Wall, G. Sella, et al. Multiple instances of ancient balancing selection shared between humans and chimpanzees. *Science*, 339(6127):1578–1582, 2013.

[27] G. Leibon, D. N. Rockmore, and M. R. Polla. A SNP streak model for the identification of genetic regions identical-by-descent. *Stat. Appl. Genet. Mol.*, 7(1):Article v16, 2008.

[28] K. Lohse and L. A. F. Frantz. Neandertal admixture in eurasia confirmed by maximum likelihood analysis of three genomes. *Genetics*, 2014.

[29] R. K. Lowery, G. Uribe, E. B. Jimenez, M. A. Weiss, K. J. Herrera, M. Regueiro, and R. J. Herrera. Neanderthal and denisova genetic affinities with contemporary humans: Introgression versus common ancestral polymorphisms. *Gene*, 530(1):83–94, 2013.

[30] C. C. Mahrenholz, I. Abfalter, U. Bodenhofer, R. Volkmer, and S. Hochreiter. Complex networks govern coiled coil oligomerization — predicting and profiling by means of a machine learning approach. *Mol. Cell. Proteomics*, 10:M110.004994, 2011.

[31] M. Meyer, M. Kircher, M.-T. Gansauge, H. Li, F. Racimo, S. Mallick, J. G. Schraiber, F. Jay, K. Prüfer, C. deFilippo, et al. A high-coverage genome sequence from an archaic denisovan individual. *Science*, 338(6104):222–226, 2012.

[32] P. Moorjani, N. Patterson, J. N. Hirschhorn, A. Keinan, L. Hao, G. Atzmon, E. Burns, H. Ostrer, A. L. Price, and D. Reich. The history of african gene flow into southern europeans, levantines, and jews. *PLoS Genet*, 7(4):e1001373, 2011.

[33] A. Moreno-Estrada, S. Gravel, F. Zakharia, J. L. McCauley, J. K. Byrnes, C. R. Gignoux, P. A. Ortiz-Tello, R. J. Martínez, D. J. Hedges, R. W. Morris, C. Eng, K. Sandoval, S. Acevedo-Acevedo, P. J. Norman, Z. Layrisse, P. Parham, J. C. Martínez-Cruzado, E. G. Burchard, M. L. Cuccaro, E. R. Martin, and C. D. Bustamante. Reconstructing the population genetic history of the caribbean. *PLoS Genet*, 9(11):e1003925, 2013.

[34] P. F. Palamara, T. Lencz, A. Darvasi, and I. Pe'er. Length distributions of identity by descent reveal fine-scale demographic history. *Am. J. Hum. Genet.*, 91(5):809–822, 2012.

[35] K. Prüfer, F. Racimo, N. Patterson, F. Jay, S. Sankararaman, S. Sawyer, A. Heinze, G. Renaud, P. H. Sudmant, C. d. Filippo, H. Li, S. Mallick, M. Dannemann, Q. Fu, M. Kircher, M. Kuhlwilm, M. Lachmann, M. Meyer, M. Ongyerth, M. Siebauer, C. Theunert, A. Tandon, P. Moorjani, J. Pickrell, J. C. Mullikin, S. H. Vohr, R. E. Green, I. Hellmann, P. L. F. Johnson, H. Blanche, H. Cann, J. O. Kitzman, J. Shendure, E. E. Eichler, E. S. Lein, T. E. Bakken, L. V. Golovanova, V. B. Doronichev, M. V. Shunkov, A. P. Derevianko, B. Viola, M. Slatkin, D. Reich, J. Kelso, and S. Paabo. The complete genome sequence of a neanderthal from the altai mountains. *Nature*, 505(7481):43–49, 2014.

[36] P. Ralph and G. Coop. The geography of recent genetic ancestry across europe. *PLoS Biol*, 11(5):e1001555, 2013.

[37] D. Reich, R. E. Green, M. Kircher, J. Krause, N. Patterson, E. Y. Durand, B. Viola, A. W. Briggs, U. Stenzel, P. L. Johnson, et al. Genetic history of an archaic hominin group from denisova cave in siberia. *Nature*, 468:1053–1060, 2010.

[38] D. Reich, N. Patterson, M. Kircher, F. Delfin, M. R. Nandineni, I. Pugach, A. M. Ko, Y. C. Ko, T. A. Jinam, M. E. Phipps, et al. Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. *Am. J. Hum. Genet.*, 89(4):516–528, 2011.

[39] S. Sankararaman, S. Mallick, M. Dannemann, K. Prufer, J. Kelso, S. Paabo, N. Patterson, and D. Reich. The genomic landscape of neanderthal ancestry in present-day humans. *Nature*, 507(7492):354–357, 2014.

[40] S. Sankararaman, N. Patterson, H. Li, S. Pääbo, and D. Reich. The date of interbreeding between neandertals and modern humans. *PLoS Genet*, 8(10):e1002947, 2012.

[41] K. Schwarzbauer, U. Bodenhofer, and S. Hochreiter. Genome-wide chromatin remodeling identified at GC-rich long nucleosome-free regions. *PLoS ONE*, 7(11):e47924, 2012.

[42] D. Serre, A. Langaney, M. Chech, M. Teschler-Nicola, M. Paunovic, P. Mennecier, M. Hofreiter, G. Possnert, and S. Pääbo. No evidence of neandertal mtdna contribution to early modern humans. *PLoS Biol*, 2(3):e57, 03 2004.

[43] P. Skoglund and M. Jakobsson. Archaic human ancestry in east asia. *Proc. Natl. Acad. Sci. USA*, 108(45):18301–18306, 2011.

[44] T. Strachan and A. P. Read. *Human Molecular Genetics*. Garland Science/Taylor & Francis Group, London and New York, 3rd edition, 2004.

[45] W. Talloen, D.-A. Clevert, S. Hochreiter, D. Amaratunga, L. Bijnens, S. Kass, and H. Göhlmann. I/NI-calls for the exclusion of non-informative genes: a highly effective filtering tool for microarray data. *Bioinformatics*, 23(21):2897–2902, 2007.

[46] W. Talloen, S. Hochreiter, L. Bijnens, A. Kasim, Z. Shkedy, and D. Amaratunga. Filtering data from high-throughput experiments based on measurement reliability. *Proc. Natl. Acad. Sci. USA*, 107(46):173–174, 2010.

[47] The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature*, 467(7319):1061–1073, 2010.

[48] The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491(7422):56–65, 2012.

[49] The Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*, 437:69–87, 2005.

[50] A. Thomas, N. J. Camp, J. M. Farnham, K. Allen-Brady, and L. A. Cannon-Albright. Shared genomic segment analysis. mapping disease predisposition genes in extended pedigrees using SNP genotype assays. *Ann. Hum. Genet.*, 72(Pt 2):279–287, 2008.

[51] A. Thomas, M. H. Skolnick, and C. M. Lewis. Genomic mismatch scanning in pedigrees. *Math. Med. Biol.*, 11(1):1–16, 1994.

[52] A. Ulgen and W. Li. Comparing single-nucleotide-polymorphism marker-based and microsatellite marker-based linkage analyses. *BMC Genet.*, 6:S13, 2005.

[53] B. Vernot and J. M. Akey. Resurrecting surviving neandertal lineages from modern human genomes. *Science*, 343(6174):1017–1021, 2014.

[54] J. D. Wall, K. E. Lohmueller, and V. Plagnol. Detecting ancient admixture and estimating demographic parameters in multiple human populations. *Mol. Biol. Evol.*, 26(8):1823–1827, 2009.

[55] J. D. Wall, M. A. Yang, F. Jay, S. K. Kim, E. Y. Durand, L. S. Stevison, C. Gignoux, A. Woerner, M. F. Hammer, and M. Slatkin. Higher levels of Neanderthal ancestry in East Asians than in Europeans. *Genetics*, 194(1):199–209, 2013.

[56] M. A. Yang, A.-S. Malaspinas, E. Y. Durand, and M. Slatkin. Ancient structure in africa unlikely to explain neanderthal and non-african genetic similarity. *Molecular Biology and Evolution*, 29(10):2987–2995, 2012.

[57] A. Yu, C. Zhao, Y. Fan, W. Jang, A. J. Mungall, P. Deloukas, A. Olsen, N. A. Doggett, N. Ghebranious, K. W. Broman, et al. Comparison of human genetic and sequence-based physical maps. *Nature*, 409:951–953, 2001.