IBD Sharing between Africans, Neandertals, and Denisovans

Gundula Povysil and Sepp Hochreiter*

Institute of Bioinformatics, Johannes Kepler University Linz, Austria

*Corresponding author: E-mail: hochreit@bioinf.jku.at.

Accepted: September 19, 2016

Data deposition: Data and code used for creating figures and tables has been deposited at http://www.bioinf.jku.at/research/IBDNeandertalDenisovan/analysisBox.

Abstract

Interbreeding between ancestors of humans and other hominins outside of Africa has been studied intensively, while their common history within Africa still lacks proper attention. However, shedding light on human evolution in this time period about which little is known, is essential for understanding subsequent events outside of Africa. We investigate the genetic relationships of humans, Neandertals, and Denisovans by identifying very short DNA segments in the 1000 Genomes Phase 3 data that these hominins share identical by descent (IBD). By focusing on low frequency and rare variants, we identify very short IBD segments with high confidence. These segments reveal events from a very distant past because shorter IBD segments are presumably older than longer ones. We extracted two types of very old IBD segments that are not only shared among humans, but also with Neandertals and/or Denisovans. The first type contains longer segments that are found primarily in Asians and Europeans where more segments are found in South Asians than in East Asians for both Neandertal and Denisovan. These longer segments indicate complex admixture events outside of Africa. The second type consists of shorter segments that are shared mainly by Africans and therefore may indicate events involving ancestors of humans and other ancient hominins within Africa. Our results from the autosomes are further supported by an analysis of chromosome X, on which segments that are shared by Africans and match the Neandertal and/or Denisovan genome were even more prominent. Our results indicate that interbreeding with other hominins was a common feature of human evolution starting already long before ancestors of modern humans left Africa.

Key words: identity by descent, interbreeding, gene flow, Neandertal, Denisova, human evolution.

Introduction

One of the most fundamental questions of humanity is: "Where do we come from?" Recent advances in biotechnology made whole genome sequencing feasible and therefore helped us to get closer to an answer. In the 1000 Genomes Project (1000 Genomes Project Consortium 2010), thousands of individuals were sequenced, which gave new insights into the population structure of humans. Application of these sequencing technologies was extended to ancient DNA and made it possible to assemble the genome of hominins that lived tens of thousands of years ago (Green et al. 2010; Reich et al. 2010; Meyer et al. 2012; Prüfer et al. 2014). The outcomes of whole genome sequencing of Neandertals (Green et al. 2010; Prüfer et al. 2014) and a Denisovan (Reich et al. 2010; Meyer et al. 2012) changed the view on the history of humans and other hominins. Previous evidence from mtDNA (Krings et al. 1997; Currat and Excoffier 2004; Serre et al. 2004) and the Y chromosome (Krause et al. 2007) had suggested that Neandertals lived isolated in Europe and Asia until they were replaced by anatomically modern humans. However, current findings hint at numerous admixture events between hominin groups (Green et al. 2010; Reich et al. 2010; Meyer et al. 2012; Sankararaman et al. 2012, 2014, 2016; Yang et al. 2012; Lohse and Frantz 2014; Prüfer et al. 2014). Several studies have found that non-Africans, especially Asians, share more alleles with the Neandertal genome than sub-Saharan Africans (Green et al. 2010; Wall et al. 2013; Khrameeva et al. 2014; Prüfer et al. 2014). For the Denisovan genome, enrichment only within Oceanian populations was initially reported (Reich et al. 2010, 2011; Mendez et al. 2012, 2013). Later studies also detected DNA regions of Denisovan origin in East and Southeast Asians as well as in populations of the Americas, but only very few such regions in Europeans (Skoglund and Jakobsson 2011; Meyer et al. 2012; Prüfer et al. 2014). The

controversy over the origin of these differences in allele sharing between particular modern populations and the genomes of Neandertals and Denisovans is ongoing. While the differences may be attributed to the existence of an ancient population substructure within Africa (Eriksson and Manica 2012, 2014; Lowery et al. 2013), ancient admixture events between Neandertals, Denisovans and anatomically modern humans outside of Africa are considered as a more plausible explanation (Green et al. 2010; Sankararaman et al. 2012, 2014, 2016; Yang et al. 2012; Lohse and Frantz 2014; Prüfer et al. 2014).

Since African populations show less allele sharing with Neandertals and Denisovans than non-Africans, they are often used as a baseline for the absence of Neandertal or Denisovan ancestry (Yang et al. 2012; Sankararaman et al. 2014, 2016; Vernot and Akey 2014). Only few studies have looked at Neandertal ancestry among African populations (Lachance et al. 2012; Sanchez-Quinto et al. 2012; Wang et al. 2013; Llorente et al. 2015) and mostly came to the conclusion that Neandertal ancestry in Africans is probably due to Eurasian backflow admixture. However, archaic hominins lived longer side by side within than outside of Africa and this led to plentiful opportunities for gene flow within Africa that should not be dismissed (Hammer et al. 2011). Knowing DNA regions of earlier admixture would both help to better identify later gene flow events and shed light on adaptation processes that were supported by interbreeding between hominin groups at different points in time.

We use HapFABIA (Hochreiter 2013), our recently developed method for detecting DNA segments that are identical by descent (IBD), on the 1000 Genomes Phase 3 data with the aim to gain insights into the genetic relationships between humans, Neandertals and Denisovans (see "HapFABIA for Extracting Short IBD Segments" section). IBD detection methods have already been successful in inferring population structure (Gusev et al. 2012; Palamara et al. 2012; Botigué et al. 2013; Carmi et al. 2013; Gravel et al. 2013; Ralph and Coop 2013). Previous studies were limited to detecting relatively recent genetic relationships because the methods used were not able to reliably identify short IBD segments, which are presumably older than longer ones (Chapman and Thompson 2003). HapFABIA accurately detects these very short IBD segments via low frequency and rare variants that tag them because rare minor alleles are highly unlikely to arise independently (Strachan and Read 2004, Ch. 15.3, p. 441).

In the Phase 3 data of the 1000 Genomes Project, we found IBD-based indications of interbreeding between ancestors of humans and other ancient hominins within Africa (see examples of an IBD segment that matches the Denisovan and Neandertal genome, respectively, in figs. 1 and 2). First, we present results of IBD segment sharing between Africans and the ancient genomes of Neandertal and Denisovan. Then, we show that these shared IBD segments are older than those originating from introgression events outside of Africa. Finally, in the "Discussion" section, we consider different theories concerning the origin of IBD segments shared between modern Africans, Neandertals and Denisovans.

Material and Methods

Data

We used HapFABIA to extract short IBD segments from the 1000 Genomes Project Phase 3 data (1000 Genomes Project Consortium 2015) (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/, last accessed 31 October 2014). This dataset consists of low-coverage whole genome sequences from 2,504 individuals (661 Africans, 347 Admixed Americans, 504 East Asians, 489 South Asians, and 503 Europeans). Compared with the 1000 Genomes Phase 1 data (1000 Genomes Project Consortium 2012) this dataset includes additional subpopulations from Africa, the Americas and East Asia, as well as South Asian populations (see supplementary table S2, Supplementary Material online, for details about the subpopulations). We removed 11 individuals because they showed cryptic first degree relatedness to others (see "Removal of Related Individuals" section). The final dataset consisted of 2,493 individuals (654 Africans—AFR, 347 Admixed Americans—AMR, 504 East Asians—EAS, 485 South Asians—SAS, and 503 Europeans—EUR). We also excluded Admixed American individuals from the group of individuals that share an IBD segment in order to avoid confounding influences from their admixed ancestry. For the same reason we removed individuals with African Ancestry in Southwest US and African Caribbeans in Barbados for most of the analyses.

Since previous analyses revealed many phasing errors (Hochreiter 2013), we applied HapFABIA with default parameters to the unphased genotype data. Since short indels and larger copy number variants can still not be called as reliably as SNVs we removed them from the provided VCF files. Furthermore, we split multiallelic variants into multiple lines. We also filtered out repeat regions and CpGs using bed files downloaded from the UCSC genome browser (Meyer et al. 2013). HapFABIA is based on low frequency and rare variants, therefore we removed common and private variants prior to the analysis. Afterwards, all chromosomes were divided into intervals of 10,000 variants with adjacent intervals overlapping by 5,000 variants.

About 31× coverage whole genome sequencing data of the Denisovan and the Altai Neandertal genome of 52× coverage were provided by the Max Planck Institute for Evolutionary Anthropology (Meyer et al. 2012; Prüfer et al. 2014) (http://cdna.eva.mpg.de/denisova/, last accessed 2 February 2012 and http://cdna.eva.mpg.de/neandertal/altai/, last accessed 23 May 2013). If we only consider variants that are present in at least one individual of the 1000 Genomes Project Phase 3, 0.9% of the Denisovan bases and 0.4% of the Neandertal bases were marked as not determined. Of the

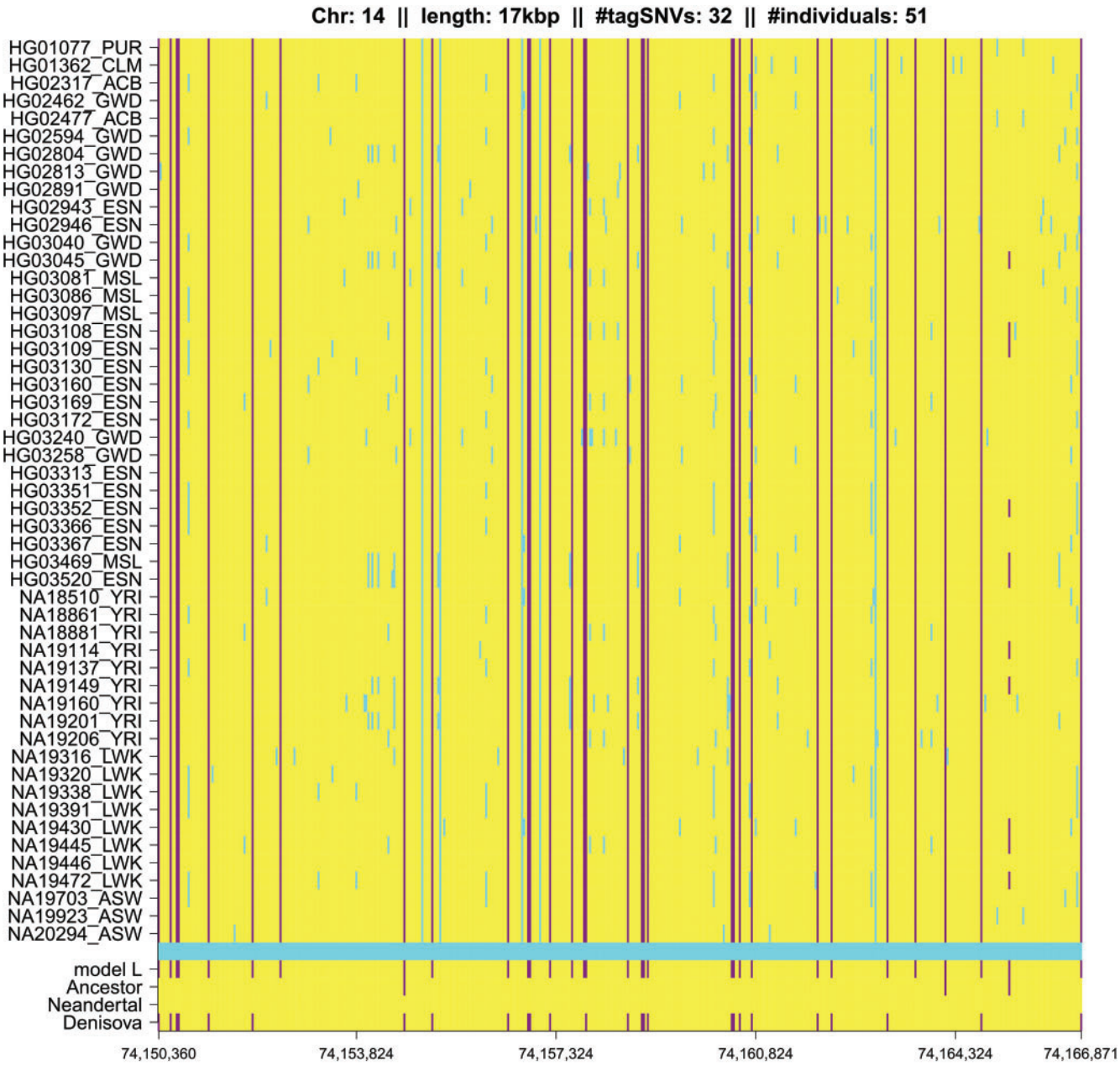

Fig. 1.—Example of a Denisovan-matching IBD segment shared among Africans. The rows represent all individuals that have the IBD segment, and columns represent consecutive variants. Major alleles are shown in yellow, minor alleles of variants that tag the IBD segment (tagSNVs) in violet, and minor alleles of other variants in cyan. The row labeled "model L" indicates tagSNVs identified by HapFABIA 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.

remaining Denisovan and Neandertal bases 96.0% and 95.6%, respectively, matched bases of the human reference, and 4.0% and 4.4%, 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 included. These are named ancestral genome of humans and other primates in the following.

HapFABIA for Extracting Short IBD Segments

In large sequencing data, HapFABIA (Hochreiter 2013) identifies very short IBD segments that are tagged by low frequency or rare variants (so-called tagSNVs) with a minor allele frequency (MAF) of $\leq$5%. 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

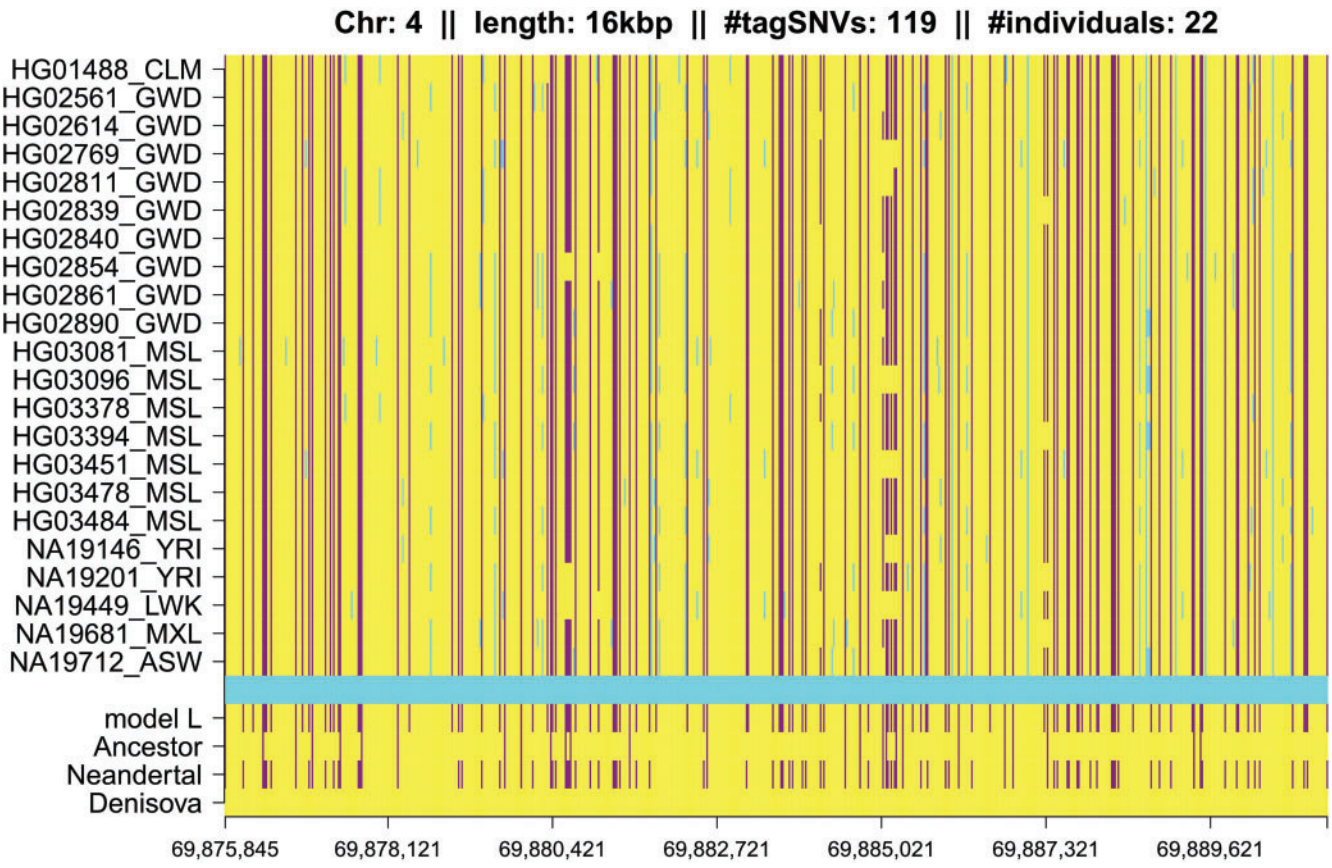

Fig. 2.—Example of a Neandertal-matching IBD segment shared among Africans. This segment is tagged by more than 100 minor alleles. See figure 1 for a detailed description.

the same ancestral origin in these individuals. Rare variants can be utilized to distinguish 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 (Strachan and Read 2004; Browning SR and Browning BL 2012). Furthermore, rare variants make juxtapositions of smaller IBD segments unlikely, which precludes their aggregation into one long IBD segment. Consequently, the length of IBD segments is estimated more accurately than with previous methods.

HapFABIA identifies 100 times smaller IBD segments than current state-of-the-art methods: 10 kb for HapFABIA versus 1 Mb for state-of-the-art methods. This was verified in experiments with artificial, simulated, and real genotyping data, where HapFABIA outperformed its competitors in detecting short IBD segments (Hochreiter 2013). It was also shown that HapFABIA is very robust to sequencing errors, therefore, the low coverage data of the 1000 Genomes Project do not pose a problem. HapFABIA is based on biclustering (Hochreiter et al. 2010), which in turn uses modern machine learning techniques derived from maximizing the posterior in a Bayes framework (Hochreiter et al. 2006; Clevert et al. 2011; Klambauer et al. 2012; Munro et al. 2014; Su et al. 2014). Preliminary results showing the ability of HapFABIA to detect IBD segments shared with Neandertals and Denisovans were already published in Hochreiter (2013).

Exponentially Distributed IBD Lengths

The length of an IBD segment is exponentially distributed with a mean of $100/(2g)$ cM (centiMorgans), where g is the number of generations which separate the two haplotypes that share a segment from their common ancestor (Thomas et al. 1994, 2008; Browning 2008; Gusev et al. 2012; Palamara et al. 2012). Therefore, on average, shorter IBD segments are older than longer ones. We exploit this relation in order to distinguish between older admixture within Africa and more recent interbreeding outside of Africa.

Ulgen and Li (2005) recommended using a recombination rate (cM-to-Mb ratio) of 1, but it varies from 0 to 9 along a chromosome (Yu et al. 2001). For chromosome X, the sex-averaged recombination rate is 2/3 of the female recombination rate on the X chromosome (Hedrick 2007). The female

recombination rate on the X chromosome has approximately the same value as the sex-averaged autosomal recombination rate (Kong et al. 2002). Thus, we assume that the sex-averaged recombination rate on the X chromosome is 2/3 times the autosomal recombination rate. The average length of IBD segments on chromosome X should therefore be 3/2 times the average length of the segments on autosomes (Povysil and Hochreiter 2014). Our length distributions show this expected outcome (see supplementary fig. S1, Supplementary Material online). This corroboration of our theoretical considerations justifies using the length for a relative estimation of the age of IBD segments. However, we are not confident in absolute age estimations of the IBD segments based on their length (see explanation in Povysil and Hochreiter 2014).

Removal of Related Individuals

The amount of IBD sharing between two individuals reflects their degree of relatedness. For a first degree relationship approximately 50% of the IBD segments are expected to be shared between two individuals. Therefore, we looked for the fraction of IBD segments each pair of individuals in the 1000 Genomes data shares. Pairs of individuals that have more than 50% of their IBD segments in common on several chromosomes were considered to be related. One individual of each pair was randomly selected to be removed and the IBD detection algorithm was rerun using the reduced data.

IBD Segments Matching Ancient Genomes

First we identified IBD segments in the 1000 Genomes data for human populations only. Then we checked whether the identified human IBD segments match ancient genomes. This matching is indicated by the presence of human minor alleles of tagSNVs in the corresponding ancient genome. In general, not the whole IBD segment is observed in ancient genomes but only a part thereof. The part of the IBD segment which carries the minor alleles that are observed in an ancient genome is called "ancient part" of the IBD segment. We define matching of an IBD segment and an ancient genome by three criteria: (1) at least 15% of the tagSNVs of the IBD segment matches the ancient genome, that is, also the ancient genome carries the human minor allele, (2) the "ancient part" of the IBD segment contains at least 8 tagSNVs, and (3) 30% of the tagSNVs in the "ancient part" of the IBD segment match the ancient genome. Finally, we corrected the IBD segment lengths to obtain the length of the "ancient part" that matches a particular ancient genome (see details in Povysil and Hochreiter 2014). We tried different thresholds for determining whether an IBD segment matches a particular ancient genome. Although the resulting number of shared IBD segments depends on these thresholds the distribution over different populations and the densities of the length of the IBD segments show no recognizable differences (see supplementary figs. S2 and S3 and table S1, Supplementary Material online).

Since the high coverage ancient genomes are only compared with previously detected human IBD segments, it is not a problem that the 1000 Genomes Project whole genome sequences are of low coverage. On the contrary, we have a high coverage verification that the IBD segments found are not random findings but indeed present in the genomes. Thus, the ancient genomes not only serve to look up ancient DNA segments (similar to a database) but additionally support IBD segments that are found in human genomes.

Removal of Ancestral IBD Segments

Human IBD segments that are found in ancient genomes may 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. Thus for most of the analyses, we removed IBD segments of which at least 30% of the tagSNVs that are shared by an ancient genome also match the reconstructed common ancestor sequence that was provided as part of the 1000 Genomes Project data. Again, adjusting this threshold did not affect the overall results. As can be seen in supplementary figure S4, Supplementary Material online, IBD segments shared with ancient genomes that are also present in the ancestral sequence are shorter than segments shared with ancient genomes that do not match the ancestral sequence. This confirms that, by removing IBD segments that are already present in the ancestral genome, we get rid of older IBD segments that originate from a common ancestor of all primates.

Results

We report results from our analysis of the 1000 Genomes Project Phase 3 (1000 Genomes Project Consortium 2015) whole genome sequencing data, where we focus on very short IBD segments that are tagged by low frequency and rare variants (tagSNVs). First, we extracted IBD segments shared by continental populations from Africa, East Asia, South Asia, and Europe. In the next step, we examined whether the IBD segments match the Neandertal and/or Denisovan sequence using the high-coverage whole genome sequencing data of the Denisovan and the Altai Neandertal (Meyer et al. 2012; Prüfer et al. 2014). For more details on the data, see "Data" section. We removed IBD segments that match the reconstructed common ancestor sequence of humans and other primates that was provided as part of the 1000 Genomes Project data (see "Removal of Ancestral IBD Segments" section), since they would confound the interbreeding analysis based on IBD sharing between modern human populations and Neandertals or Denisovans. We especially investigated IBD segments shared between Africans and Neandertals and/or Denisovans in contrast to

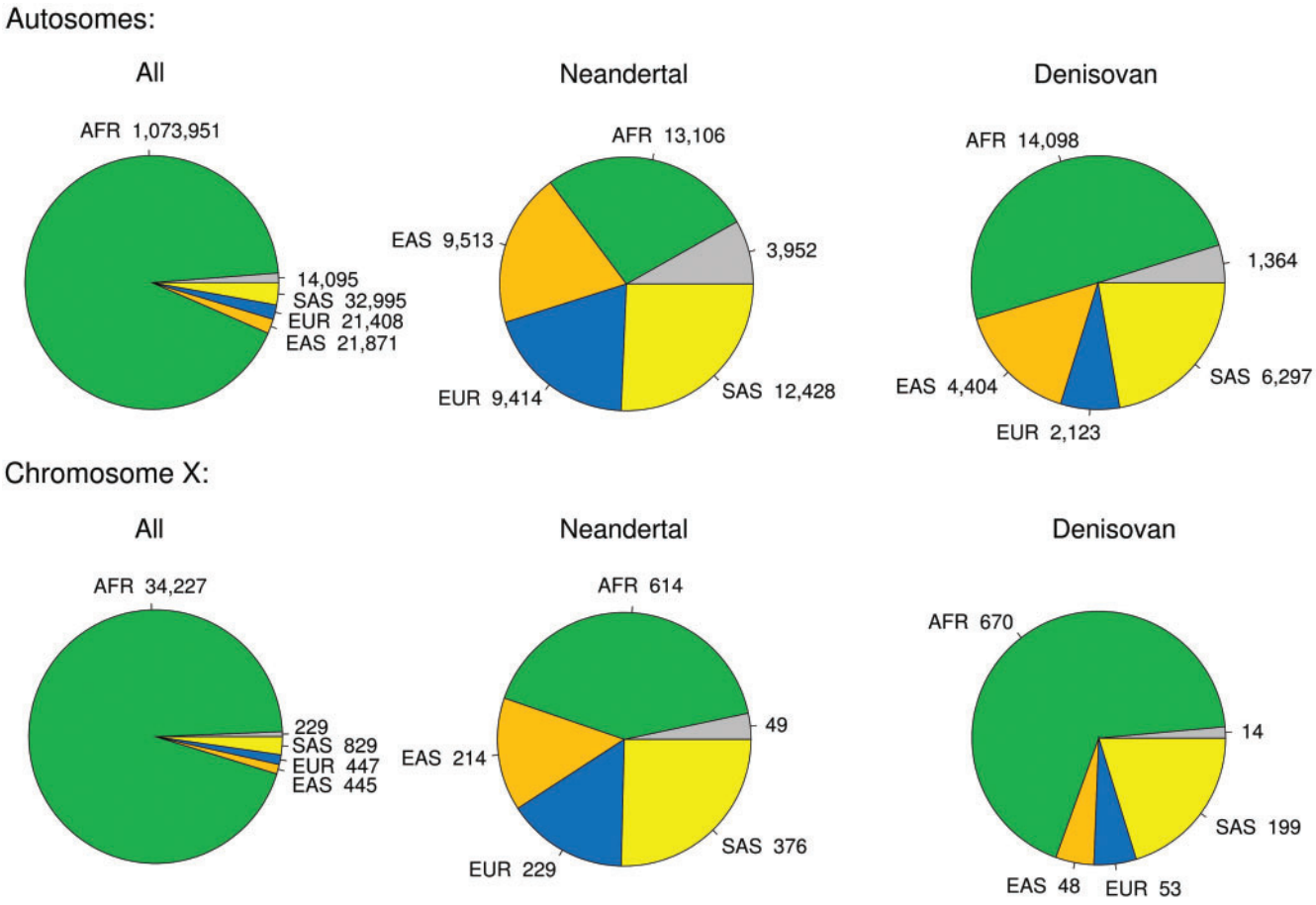

Fig. 3.—Predominant population for an IBD segment and genome. For each IBD segment, the population in which it is predominantly found is determined. "All" stands for all IBD segments found, while "Neandertal" and "Denisovan" denote all IBD segments matching the Neandertal and Denisovan sequence, respectively. The color indicates the population to which more than 50% of the individuals with the corresponding IBD segment are assigned to. Gray slices indicate IBD segments for which neither of the populations has a share of more than 50%. The pie charts show prominent IBD sharing between Africans and ancient genomes, visualized by the green slices of the pies.

segments shared between non-Africans and these ancient genomes. By comparing the average length of IBD segments that particular populations share with these ancient genomes, we were able to distinguish between very early events within Africa and later admixture outside of Africa as the origin of shared segments (see "Exponentially Distributed IBD Lengths" section). Further, differences between the autosomes and chromosome X were of particular interest because chromosome X is under especially high selective pressure and may show different levels of introgression if gene flow is primarily male-driven (Nielsen et al. 2005; Lambert et al. 2010; Veeramah et al. 2014).

In total, more than 25,000 IBD segments that match ancient genomes are shared predominantly by Africans. On chromosome X, there are more than 1,000 of these segments which amounts to more than half of the X-chromosomal segments that match ancient genomes. First, we investigate the frequency of IBD segments that are shared with ancient genomes across different continental populations. An IBD segment is assigned to an ancient genome if at least 30% of it matches the respective genome and to a population if more than 50% of the individuals with the corresponding IBD segment belong to it (see "IBD Segments Matching Ancient Genomes" section for more details). We visualize the results in figure 3, where different pie charts reflect the different genomes and the color reflects the population to which more than 50% of the individuals with the corresponding IBD segment belong. Gray slices represent IBD segments where neither of the populations has a share of more than 50%. Compared with all IBD segments found in the 1000 Genomes Project data (labeled "All"), those that match ancient genomes are enriched in Europeans and Asians. Interestingly, South Asians do not only share more segments with the Denisovan genome than East Asians and Europeans, but also the number of segments shared with the Neandertal genome is higher for South Asians than for other non-African

Table 1

Number of Neandertal- and Denisovan-Matching IBD Segments Shared Exclusively by a Particular Continental Population for Each Chromosome

Chr.	Neandertal					Denisovan				
	ALL	EAS	SAS	EUR	AFR	ALL	EAS	SAS	EUR	AFR
1	3,906	271 (6.9%)	332 (8.5%)	130 (3.3%)	751 (19.2%)	2,284	165 (7.2%)	219 (9.6%)	15 (0.7%)	865 (37.9%)
2	4,094	219 (5.3%)	398 (9.7%)	125 (3.1%)	943 (23.0%)	2,477	106 (4.3%)	354 (14.3%)	24 (1.0%)	997 (40.3%)
3	3,299	203 (6.2%)	453 (13.7%)	75 (2.3%)	679 (20.6%)	1,811	82 (4.5%)	273 (15.1%)	13 (0.7%)	709 (39.1%)
4	3,675	283 (7.7%)	397 (10.8%)	93 (2.5%)	830 (22.6%)	2,332	251 (10.8%)	382 (16.4%)	14 (0.6%)	892 (38.3%)
5	3,168	225 (7.1%)	366 (11.6%)	153 (4.8%)	573 (18.1%)	1,978	159 (8.0%)	260 (13.1%)	34 (1.7%)	679 (34.3%)
6	3,424	268 (7.8%)	270 (7.9%)	128 (3.7%)	533 (15.6%)	1,993	146 (7.3%)	219 (11.0%)	28 (1.4%)	692 (34.7%)
7	2,607	146 (5.6%)	202 (7.7%)	83 (3.2%)	627 (24.1%)	1,472	64 (4.3%)	129 (8.8%)	10 (0.7%)	707 (48.0%)
8	2,295	75 (3.3%)	228 (9.9%)	136 (5.9%)	675 (29.4%)	1,380	39 (2.8%)	180 (13.0%)	27 (2.0%)	649 (47.0%)
9	1,893	123 (6.5%)	211 (11.1%)	76 (4.0%)	335 (17.7%)	1,245	45 (3.6%)	269 (21.6%)	12 (1.0%)	472 (37.9%)
10	2,393	144 (6.0%)	242 (10.1%)	74 (3.1%)	400 (16.7%)	1,470	76 (5.2%)	259 (17.6%)	11 (0.7%)	536 (36.5%)
11	2,778	116 (4.2%)	275 (9.9%)	171 (6.2%)	445 (16.0%)	1,587	82 (5.2%)	227 (14.3%)	42 (2.6%)	555 (35.0%)
12	2,298	139 (6.0%)	275 (12.0%)	59 (2.6%)	298 (13.0%)	1,058	77 (7.3%)	133 (12.6%)	5 (0.5%)	406 (38.4%)
13	1,891	127 (6.7%)	242 (12.8%)	55 (2.9%)	338 (17.9%)	1,080	97 (9.0%)	152 (14.1%)	5 (0.5%)	352 (32.6%)
14	1,898	181 (9.5%)	190 (10.0%)	72 (3.8%)	256 (13.5%)	904	55 (6.1%)	94 (10.4%)	9 (1.0%)	321 (35.5%)
15	1,350	126 (9.3%)	142 (10.5%)	42 (3.1%)	230 (17.0%)	773	31 (4.0%)	130 (16.8%)	2 (0.3%)	289 (37.4%)
16	1,213	55 (4.5%)	131 (10.8%)	41 (3.4%)	326 (26.9%)	782	21 (2.7%)	112 (14.3%)	3 (0.4%)	382 (48.8%)
17	1,018	46 (4.5%)	137 (13.5%)	42 (4.1%)	266 (26.1%)	637	30 (4.7%)	71 (11.1%)	3 (0.5%)	298 (46.8%)
18	1,423	70 (4.9%)	135 (9.5%)	58 (4.1%)	308 (21.6%)	817	44 (5.4%)	135 (16.5%)	1 (0.1%)	309 (37.8%)
19	1,061	63 (5.9%)	98 (9.2%)	33 (3.1%)	247 (23.3%)	677	43 (6.4%)	63 (9.3%)	2 (0.3%)	277 (40.9%)
20	1,116	46 (4.1%)	115 (10.3%)	55 (4.9%)	287 (25.7%)	631	11 (1.7%)	66 (10.5%)	9 (1.4%)	318 (50.4%)
21	628	35 (5.6%)	72 (11.5%)	19 (3.0%)	166 (26.4%)	378	24 (6.3%)	62 (16.4%)	0 (0.0%)	153 (40.5%)
22	559	30 (5.4%)	43 (7.7%)	17 (3.0%)	94 (16.8%)	391	30 (7.7%)	85 (21.7%)	1 (0.3%)	127 (32.5%)
1-22	47,987	2,991 (6.2%)	4,954 (10.3%)	1,737 (3.6%)	9,607 (20.0%)	28,157	1,678 (6.0%)	3,874 (13.8%)	270 (1.0%)	10,985 (39.0%)
X	1,479	84 (5.7%)	204 (13.8%)	83 (5.6%)	512 (34.6%)	983	8 (0.8%)	123 (12.5%)	23 (2.3%)	563 (57.3%)

NOTE.—The column labeled "Chr." gives the chromosome, and "Neandertal" and "Denisovan" group IBD segments matching the Neandertal and Denisovan genomes, respectively. "ALL" gives the total number of IBD segments matching the respective ancient genome, "EAS", "SAS", "EUR", and "AFR" report the number of matching IBD segments shared exclusively by East Asians, South Asians, Europeans, and Africans, respectively, and the percentage compared to the total number of IBD segments matching the respective ancient genome.

populations. Furthermore, many segments matching ancient genomes are shared predominantly by Africans — especially on chromosome X. About 13,106 of 47,987 Neandertal-matching IBD segments on the autosomes and 614 of 1,479 on chromosome X are primarily found in Africans. Denisovan-matching IBD segments are predominantly shared by Africans in 14,098 out of 28,157 cases for the autosomes and 670 out of 983 for chromosome X.

The IBD segments we considered above are often observed in more than one continental population, which means that, although they are found primarily in Africans, some Asians or Europeans may also share the same segment. Later admixture events between human populations may thus be the reason for parts of ancient genomes being observed in Africans. Therefore, we additionally investigated IBD segments that are observed exclusively in one continental population. Table 1 shows the number of Neandertal- and Denisovan-matching IBD segments per chromosome that are shared exclusively by different continental populations. Neandertal-matching IBD segments are most often exclusive to Africans across all chromosomes. About 20% of all Neandertal-matching IBD segments on the autosomes are shared exclusively by Africans, compared with less than 7%, 11% and 4% for East Asians, South Asians and Europeans, respectively. On chromosome X, almost 35% of all Neandertal-matching IBD segments are in fact shared exclusively by Africans. The results are even more impressive for sharing with the Denisovan genome. Almost 40% of all Denisovan-matching IBD segments on the autosomes are found exclusively in Africans followed by South Asians with less than 14% of the shared segments. Europeans share less than 300 IBD segments exclusively with Denisovans indicating that IBD segments shared between Europeans and Denisovans probably were introduced into European genomes via admixture with other human populations. Interestingly, on chromosome X, East Asians and Europeans have very few exclusive IBD segments that are shared with the Denisovan individual. About 563 out of 983 IBD segments on chromosome X that match the Denisovan genome are exclusive to the African population, and a large proportion of the remaining segments is also observed in Africans. Explanations for these findings are: (1) that little Denisovan DNA entered the human X chromosome

Table 2

Number of Neandertal- and Denisovan-Matching IBD Segments Shared Exclusively by Africans that are Observed in a Particular Subpopulation

	AFR	ESN	GWD	LWK	MSL	YRI
Neandertal 1–22	9,607	2,987	3,135	2,994	3,191	3,133
Neandertal X	512	139	117	148	123	132
Denisovan 1–22	10,985	3,312	3,494	3,410	3,602	3,489
Denisovan X	563	167	154	187	163	166

Note.—The rows present results for Neandertal on the autosomes and chromosome X, as well as, results for the Denisovan on the autosomes and chromosome X, respectively. The column labeled "AFR" gives the total number of IBD segments that are exclusively shared by Africans and match the respective genome, "ESN" (Esan, Nigeria), "GWD" (Gambian, Western Division), "LWK" (Luhya, Kenya), "MSL" (Mende, Sierra Leone), and "YRI" (Yoruba, Nigeria) report the number of IBD segments exclusively shared by Africans that are also shared by at least one individual from the respective subpopulation. See Supplementary Material for details about the subpopulations.

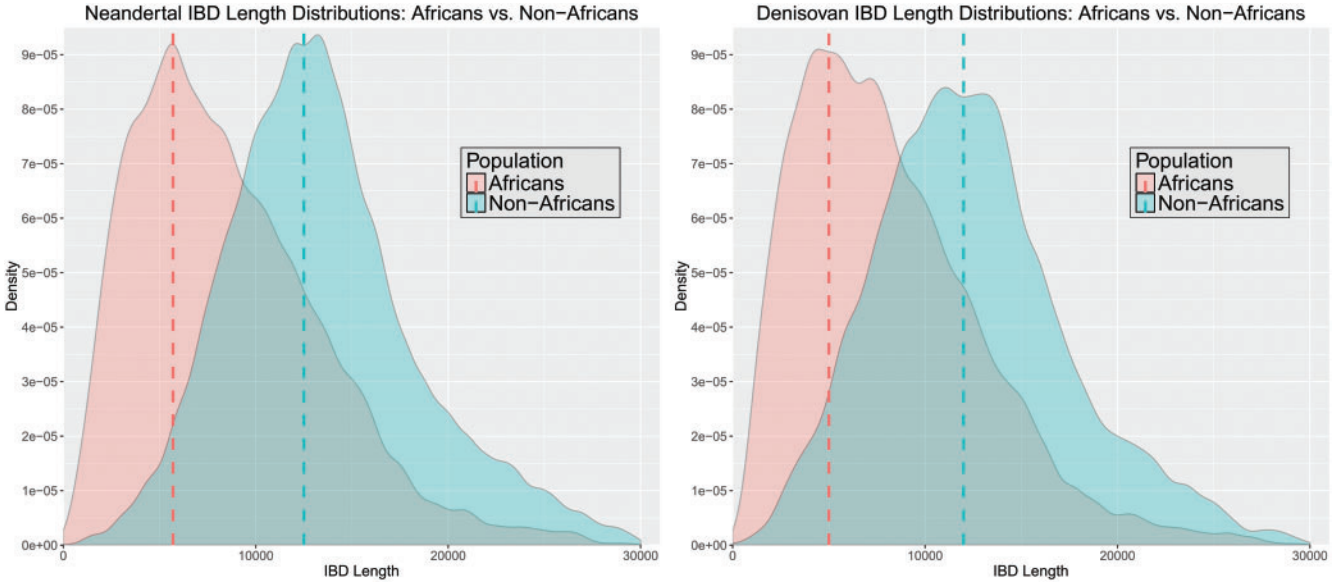

Fig. 4.—Population specific IBD segment lengths for Neandertal- and Denisovan-matching segments. Densities of lengths of IBD segments on the autosomes that match the Neandertal (left) or Denisovan (right) genome and are private to Africans (red) versus IBD segments matching ancient genomes that are not observed in Africans (blue). The dotted lines emphasize peaks of the densities. Compared with those of non-Africans, African IBD segments that match an ancient genome are enriched in regions of shorter segment lengths.

outside of Africa or (2) that due to selective pressure, modern Eurasians lost most of the Denisovan sequence on this chromosome. IBD segments that are exclusively shared between Africans and Neandertals or Denisovans occur in similar quantities in all African subpopulations (see table 2). The uniform distribution of ancient genome sharing across Africa contradicts a recent Eurasian backflow to Africa as sole origin of IBD segments shared between Africans and ancient genomes.

Next, we consider the lengths of IBD segments of different continental populations in order to compare their age. Since, as mentioned in "Exponentially Distributed IBD Lengths" section, shorter segments are assumed to be older than longer ones, IBD segments that originate from events involving ancestors of modern humans and ancestors of Neandertals and/or Denisovans within Africa should be shorter than those from later interbreeding outside of Africa. Figure 4 shows densities of lengths of IBD segments that match the Neandertal/Denisovan genome and are private to Africans versus IBD segments that are not observed in Africans. With the major density peak (the mode of the density) at about 5,700 and 5,000 bp, respectively, Neandertal- and Denisovan-matching IBD segments are shorter when they are shared by Africans than when they are shared by non-Africans, where the peaks are at about 12,500 and 12,000 bp, respectively. Neandertal- and Denisovan-matching IBD segments that are only shared by non-African populations are clearly longer and therefore younger than those shared between Africans and the ancient genomes. This supports previous reports of interbreeding after migration out of Africa (Green et al. 2010; Reich et al. 2010; Meyer et al. 2012; Sankararaman et al. 2012, 2014, 2016; Yang et al. 2012; Lohse and Frantz 2014; Prüfer et al. 2014). On the other

hand, the short Neandertal- and Denisovan-matching IBD segments found exclusively in Africans are considerably older and therefore probably originate from within Africa.

If we compare the densities of lengths of IBD segments that match the Neandertal/Denisovan genome and are private to Africans versus IBD segments that are observed in Africans but not exclusively, we also see a peak at 12,000 and 11,000 bp for Neanderthal- and Denisovan-matching IBD segments, respectively, that are observed in Africans but not exclusively (see supplementary fig. S5, Supplementary Material online). These peaks may represent segments that stem from later interbreedings outside of Africa but entered the African genome via a recent Eurasian backflow as reported previously (Wang et al. 2013; Prüfer et al. 2014; Llorente et al. 2015). Similarly, if we compare the densities of lengths of IBD segments that match the Neandertal/Denisovan genome and are private to Africans versus IBD segments that are observed outside of Africa but not exclusively, we see that the densities of IBD segments observed outside of Africa are shifted to the left if compared with densities of IBD segments not observed in Africans (see supplementary fig. S6, Supplementary Material online vs. fig. 4). The shorter segments that cause this shift may indicate remnants of the older segments that originate from within Africa.

Discussion

There are different explanations for our results that concern the origin of segments that are shared between Africans and the Neandertal and/or Denisovan genome.

One interpretation of our results is that ancestors of humans and ancient hominins interbred within Africa. The hypothesis of ancient substructures in Africa with limited gene flow between subpopulations of hominins (Plagnol and Wall 2006; Slatkin 2008; Yang et al. 2012) does not contradict this interbreeding. Neandertals and Denisovans could be more closely related to Africans than to out-of-Africa populations because of more interactions between their ancestors. In this case, since the ancestors of Africans and Neandertals/Denisovans were not clearly separated, this could be considered "admixture" rather than "interbreeding".

Another interpretation of the extensive IBD sharing between Africans and ancient genomes is that these shared IBD segments originate from a common ancestor of Neandertals/Denisovans and humans. They can only be found in modern Africans due to incomplete lineage sorting. According to this scenario, the detected IBD segments arose first in the population that existed prior to the ancient separation of Neandertals, Denisovans and modern humans, but were relatively rare. Consequently they survived in both archaic humans and in present-day Africans, while drifting to a very low frequency in non-Africans. On average, these IBD segments are tagged by more than 20 rare variants. Surprisingly the individuals analyzed possess either all of these variants or none of them while individuals with only some of these rare variants are missing. In our opinion, it is unlikely that, for thousands of IBD segments, only these extremes survived while all the intermediate cases died out completely. Consequently, we assume that the source population was separated from all the other populations for a long time and, therefore, acquired such a high number of mutations. We do not know, whether the separated population was already a Neandertal, a Denisovan, their ancestor, or a different hominin. We cannot rule out, that the IBD segments also existed in ancestors of modern Eurasians and were lost due to strong genetic drift. However, our results suggest an interbreeding within Africa that involved a population that was isolated for an extended period of time. This early interbreeding can still be detected via IBD segments that are shared between Africans and Neandertals and/or Denisovans.

Another explanation without introgression in Africa is back-to-Africa gene flow as suggested by Prüfer et al. (2014), Wang et al. (2013), and Llorente et al. (2015) for Neandertal ancestry in Africans. However, (1) the large number of IBD segments shared exclusively between the ancient genomes and Africans, (2) their even distribution across all African populations analyzed, and (3) their short length contradict this hypothesis.

Besides the short IBD segments shared with Africans, the other interesting finding is, that there are more IBD segments shared between South Asians and Neandertals and/or Denisovans than between other non-African populations and these ancient genomes. Recent investigations found that South Asians share a surprisingly high amount of DNA with the Denisovan genome (Sankararaman et al. 2016). In our analysis the amount is even higher. Possible explanations for this finding are as follows: (1) additional interbreeding events with ancestors of South Asians, (2) introduction of IBD segments from ancient genomes into other non-African populations via South Asians and not directly, and (3) combinations of bottlenecks, genetic drift, and different selective pressures.

As supplementary figure S7, Supplementary Material online, shows, the length distributions of IBD segments shared between Neandertals or Denisovans and different non-African populations are very similar. Thus, a separate interbreeding event at a different time point seems to be unlikely. However, based on the IBD segment length distributions, we cannot rule out multiple interbreeding events with different human populations in a relatively short time period. Because the IBD segment lengths are determined by the number of crossovers since the interbreeding, a single cannot be distinguishing from multiple events.

According to the second explanation, Denisovans and Neandertals interbred only with the ancestors of South Asians. In the following, South Asians introduced parts of the ancient genomes into the genomes of Europeans and East Asians via later admixture events. This explanation

might account for the low number of Denisovan-matching segments found in Europeans. However, we think it is unlikely that it is the reason for the considerably larger number of IBD segments shared exclusively between Neandertals and Europeans/East Asians and between Denisovans and East Asians.

The third explanation is based on the idea that the interbreeding occurred between the ancient hominins and the common ancestors of Eurasians. Consequently, different populations went through different bottlenecks, experienced various selective pressures and admixtures with other human populations. Because of the aforementioned factors, ancient DNA that had entered the human gene pool survived in one population while it has been lost in another population, or has been diluted in the third population. This would explain why some of the detected IBD segments are present in all three non-African populations, some only in two out of three, and others are exclusively found in a single continental population. Previous results (Vernot and Akey 2014, 2015; Kim and Lohmueller 2015) showed that above factors probably are not the only reason for varying levels of Neandertal- and Denisovan-ancestry in different non-African populations, however, we assume that at least they affected IBD sharing between modern humans and ancient genomes.

Conclusion

Using the 1000 Genomes Phase 3 data to represent the human genome and the sequenced genomes of a Neandertal and a Denisovan, we analyzed genetic relationships between these hominins. We characterized these relations by means of shared DNA segments that are identical by descent, focusing on low frequency and rare variants to identify very short IBD segments with high confidence.

Since shorter IBD segments are presumably older than longer ones, the segments we extracted reveal events from the very distant past. We found short IBD segments that match the Neandertal and/or Denisovan genome and are shared mainly by Africans. These segments may either stem from a common ancestor with subsequent incomplete lineage sorting or more likely from an interbreeding of ancestors of humans and other ancient hominins within Africa.

DNA segments which originate from a common history within Africa can also help to improve the reconstruction of human evolution outside of Africa. Furthermore, these segments have the potential to lead to new insights into adaptation processes at different time points in human history.

Supplementary Material

Supplementary figures S1–S7 and tables S1 and S2 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgements

We thank Günter Klambauer and Ulrich Bodenhofer for helpful discussions, proofreading, and comments on this manuscript. Furthermore, we thank the NVIDIA corporation for the GPU donation that facilitated our work. Funding for open access charge: Funds from the Institute of Bioinformatics, Johannes Kepler University Linz.

Literature Cited

1000 Genomes Project Consortium. 2010. A map of human genome variation from population-scale sequencing. Nature 467:1061–1073.
1000 Genomes Project Consortium. 2012. An integrated map of genetic variation from 1,092 human genomes. Nature 491:56–65.
1000 Genomes Project Consortium. 2015. A global reference for human genetic variation. Nature 526:68–74.
Botigué LR, et al. 2013. Gene flow from North Africa contributes to differential human genetic diversity in southern Europe. Proc Natl Acad Sci U S A. 110:11791–11796.
Browning SR. 2008. Estimation of pairwise identity by descent from dense genetic marker data in a population sample of haplotypes. Genetics 178:2123–2132.
Browning SR, Browning BL. 2012. Identity by descent between distant relatives: detection and applications. Annu Rev Genet. 46:617–633.
Carmi S, et al. 2013. The variance of identity-by-descent sharing in the Wright–Fisher model. Genetics 193:911–928.
Chapman NH, Thompson EA. 2003. A model for the length of tracts of identity by descent in finite random mating populations. Theor Popul Biol. 64:141–150.
Clevert DA, et al. 2011. cn.FARMS: a latent variable model to detect copy number variations in microarray data with a low false discovery rate. Nucleic Acids Res. 39:e79.
Currat M, Excoffier L. 2004. Modern humans did not admix with Neanderthals during their range expansion into Europe. PLoS Biol. 2:e421.
Eriksson A, Manica A. 2012. Effect of ancient population structure on the degree of polymorphism shared between modern human populations and ancient hominins. Proc Natl Acad Sci U S A. 109:13956–13960.
Eriksson A, Manica A. 2014. The doubly conditioned frequency spectrum does not distinguish between ancient population structure and hybridization. Mol Biol Evol. 31:1618–1621.
Gravel S, et al. 2013. Reconstructing Native American migrations from whole-genome and whole-exome data. PLoS Genet. 9:e1004023.
Green RE, et al. 2010. A draft sequence of the Neandertal genome. Science 328:710–722.
Gusev A, et al. 2012. The architecture of long-range haplotypes shared within and across populations. Mol Biol Evol. 29:473–486.
Hammer MF, Woerner AE, Mendez FL, Watkins JC, Wall JD. 2011. Genetic evidence for archaic admixture in Africa. Proc Natl Acad Sci U S A. 10:15123–15128.
Hedrick PW. 2007. Sex: differences in mutation, recombination, selection, gene flow, and genetic drift. Evolution 61:2750–2771.
Hochreiter S. 2013. HapFABIA: identification of very short segments of identity by descent characterized by rare variants in large sequencing data. Nucleic Acids Res. 41:e202.
Hochreiter S, Clevert DA, Obermayer K. 2006. A new summarization method for Affymetrix probe level data. Bioinformatics 22:943–949.
Hochreiter S, et al. 2010. FABIA: factor analysis for bicluster acquisition. Bioinformatics 26:1520–1527.
Khrameeva EE, et al. 2014. Neanderthal ancestry drives evolution of lipid catabolism in contemporary Europeans. Nat Commun. 5:3584.

Kim BY, Lohmueller KE. 2015. Selection and reduced population size cannot explain higher amounts of Neandertal ancestry in East Asian than in European human populations. Am J Hum Genet. 96:454–461.
Klambauer G, et al. 2012. 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:e69.
Kong A, et al. 2002. A high-resolution recombination map of the human genome. Nat Genet. 31:241–247.
Krause J, et al. 2007. The derived FOXP2 variant of modern humans was shared with Neandertals. Curr Biol. 17:1908–1912.
Krings M, et al. 1997. Neandertal DNA sequences and the origin of modern humans. Cell 90:19–30.
Lachance J, et al. 2012. Evolutionary history and adaptation from high-coverage whole-genome sequences of diverse African hunter-gatherers. Cell 150:457–469.
Lambert CA, et al. 2010. Highly punctuated patterns of population structure on the X chromosome and implications for African evolutionary history. Am J Hum Genet. 86:34–44.
Llorente MG, et al. 2015. Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent. Science 350:820–822.
Lohse K, Frantz LAF. 2014. Neandertal admixture in Eurasia confirmed by maximum likelihood analysis of three genomes. Genetics 196:1241–1251.
Lowery RK, et al. 2013. Neanderthal and Denisova genetic affinities with contemporary humans: introgression versus common ancestral polymorphisms. Gene 530:83–94.
Mendez FL, Watkins JC, Hammer MF. 2012. Global genetic variation at OAS1 provides evidence of archaic admixture in Melanesian populations. Mol Biol Evol. 29:1513–1520.
Mendez FL, Watkins JC, Hammer MF. 2013. Neandertal origin of genetic variation at the cluster of OAS immunity genes. Mol Biol Evol. 30:798–801.
Meyer LR, et al. 2013. The UCSC Genome Browser database: extensions and updates 2013. Nucleic Acids Res. 41:D64–D69.
Meyer M, et al. 2012. A high-coverage genome sequence from an archaic Denisovan individual. Science 338:222–226.
Munro SA, et al. 2014. Assessing technical performance in differential gene expression experiments with external spike-in RNA control ratio mixtures. Nat Commun. 5:5125.
Nielsen R, et al. 2005. A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol. 3:e170.
Palamara PF, Lencz T, Darvasi A, Pe'er I. 2012. Length distributions of identity by descent reveal fine-scale demographic history. Am J Hum Genet. 91:809–822.
Plagnol V, Wall JD. 2006. Possible ancestral structure in human populations. PLoS Genet. 2:e105.
Povysil G, Hochreiter S. 2014. Sharing of very short IBD segments between humans, Neandertals, and Denisovans. bioRxiv. Unpublished data [cited 2015 Apr 4]. Available from: http://biorxiv.org/content/early/2014/07/15/003988.
Prüfer K, et al. 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature 505:43–49.
Ralph P, Coop G. 2013. The geography of recent genetic ancestry across Europe. PLoS Biol. 11:e1001555.
Reich D, et al. 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature 468:1053–1060.
Reich D, et al. 2011. Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. Am J Hum Genet. 89:516–528.
Sanchez-Quinto F, et al. 2012. North African populations carry the signature of admixture with Neandertals. PLoS One 7:e47765.
Sankararaman S, et al. 2014. The genomic landscape of Neanderthal ancestry in present-day humans. Nature 507:354–357.
Sankararaman S, Mallick S, Patterson N, Reich D. 2016. The combined landscape of Denisovan and Neanderthal ancestry in present-day humans. Curr Biol. 26:1241–1247.
Sankararaman S, Patterson N, Li H, Pääbo S, Reich D. 2012. The date of interbreeding between Neandertals and modern humans. PLoS Genet. 8:e1002947.
Serre D, et al. 2004. No evidence of Neandertal mtDNA contribution to early modern humans. PLoS Biol. 2:e57.
Skoglund P, Jakobsson M. 2011. Archaic human ancestry in East Asia. Proc Natl Acad Sci U S A. 108:18301–18306.
Slatkin M. 2008. Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. Nat Rev Genet. 9:477–485.
Strachan T, Read AP. 2004. Human molecular genetics. 3rd ed. London and New York: Garland Science/Taylor & Francis Group.
Su Z, et al. 2014. A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. Nat Biotech. 32:903–914.
Thomas A, Camp NJ, Farnham JM, Allen-Brady K, Cannon-Albright LA. 2008. Shared genomic segment analysis. mapping disease predisposition genes in extended pedigrees using SNP genotype assays. Ann Hum Genet. 72:279–287.
Thomas A, Skolnick MH, Lewis CM. 1994. Genomic mismatch scanning in pedigrees. Math Med Biol. 11:1–16.
Ulgen A, Li W. 2005. Comparing single-nucleotide-polymorphism marker-based and microsatellite marker-based linkage analyses. BMC Genet. 6:S13.
Veeramah KR, Gutenkunst RN, Woerner AE, Watkins JC, Hammer MF. 2014. Evidence for increased levels of positive and negative selection on the X chromosome versus autosomes in humans. Mol Biol Evol. 31:2267–2282.
Vernot B, Akey JM. 2014. Resurrecting surviving Neandertal lineages from modern human genomes. Science 343:1017–1021.
Vernot B, Akey JM. 2015. Complex history of admixture between modern humans and Neandertals. Am J Hum Genet. 96:448–453.
Wall JD, et al. 2013. Higher levels of Neanderthal ancestry in East Asians than in Europeans. Genetics 194:199–209.
Wang S, Lachance J, Tishkoff SA, Hey J, Xing J. 2013. Apparent variation in Neanderthal admixture among African populations is consistent with gene flow from Non-African populations. Genome Biol Evol. 5:2075–2081.
Yang MA, Malaspinas AS, Durand EY, Slatkin M. 2012. Ancient structure in Africa unlikely to explain Neanderthal and Non-African genetic similarity. Mol Biol Evol. 29:2987–2995.
Yu A, et al. 2001. Comparison of human genetic and sequence-based physical maps. Nature 409:951–953.

Associate editor: Bill Martin