Identical by Descent (IBD): Investigation of the Genetic Ties between Africans, Denisovans, and Neandertals

Takudzwa Fadziso[1], Mani Manavalan[2]

[1]Institute of Lifelong Learning and Development Studies, Chinhoyi University of Technology, **ZIMBABWE**
[2]Technical Project Manager, Larsen & Toubro Infotech (LTI), Mumbai, **INDIA**

ABSTRACT

Interbreeding between human ancestors and other hominins has been extensively studied outside of Africa, but their shared history within Africa has received less study. However, comprehending subsequent events outside of Africa requires shining light on human evolution during this period, about which little is known. We investigate the genetic relationships of humans. By finding relatively short DNA sequences that these hominins share in the 1000 Genomes Phase 3 data, researchers were able to distinguish between African, Neandertals, and Denisovans descent by identical (IBD). It was confidently detected very short IBD segments by focusing on low frequency and uncommon variations. These segments reflect occurrences from the distant past because small IBD segments are likely older than larger ones. There have been two types of very old IBD segments found that are shared by humans, Neandertals, and/or Denisovans. Longer segments are more common in Asians and Europeans, with more segments in the South. Asians exceed East Asians in both Neandertal and Denisovan cultures. These longer portions indicate complex admixture occurring outside of the admixture events. Africa, the second category comprises shorter pieces that are largely shared among Africans and hence may depict African-related events.

Key Words: Human evolution, identity by descent, gene flow, Neandertal, Interbreeding, Denisova

INTRODUCTION

The question of "Where did we originate from?" is one of humanity's most fundamental ones. Whole-genome sequencing has been possible thanks to recent breakthroughs in biotechnology, which has aided us in getting closer to a solution. Thousands of individuals were sequenced as part of the 1000 Genomes Project (1000 Genomes Project Consortium 2010), revealing new information on human population structure. Reich et al. (2010); Green et al. (2010); Meyer et al. (2012); Prufer et al. (2014) used these sequencing technologies to reconstruct the genome of tens of thousands of years ago hominins. The findings of whole-genome sequencing of Neandertals (Green et al. 2010; Prufer et al. 2014) and a Denisovan (Reich et al. 2010; Meyer et al. 2012) revolutionized the way people and other hominins were thought to have evolved. Previous data from the Y chromosome

(Krause et al. 2007) and mtDNA (Krings et al. 1997; Currat and Excoffier 2004; Serre et al. 2004) indicates that Neandertals lived in isolation in Europe and Asia until they were displaced by anatomically modern humans.

Current data, however, point to several mixing episodes between hominin lineages (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; Prufer et al. 2014; Lohse and Frantz 2014; Prufer et al. 2014). Non-Africans, particularly Asians, share more alleles with the Neandertal genome than sub-Saharan Africans, according to several studies (Green et al. 2010; Wall et al. 2013; Khrameeva et al. 2014; Prufer et al. 2014). Initially, enrichment of the Denisovan genome was exclusively seen in Oceanian populations (Reich et al. 2010, 2011; Mendez et al. 2012, 2013). Later research discovered Denisovan DNA areas in East and Southeast Asians, as well as populations from the Americas. However, there are just a handful of such areas in Europe (Skoglund and Meyer et al. 2012; Prufer et al. 2014). Jakobsson 2011; Meyer et al. 2012; Prufer et al. 2014). The genesis of these discrepancies in allele sharing is a source of debate between the genomes of modern people and the genomes of ancient populations. The study of Neandertals and Denisovans is still underway.

While the differences could 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 thought to be a more plausible explanation (Green et al. 2010; Sankararaman et al. 2012, 2014, 2016; Yang et al. 2012). Because African populations share fewer alleles with Neandertals and Denisovans than non-Africans, they are frequently used to rule out Neandertal or Denisovan ancestry (Yang et al. 2012; Sankararaman et al. 2014, 2016; Vernot and Akey 2014). Only a few research have looked at Neandertal ancestry in African populations (Lachance et al. 2012; Sanchez-Quinto et al. 2012; Wang et al. 2013; Llorente et al. 2015), with the majority concluding that Neandertal ancestry in Africans is most likely owing to Eurasian backflow admixture. Furthermore, because archaic hominins lived in Africa for longer than they did outside of it, there were numerous chances for gene flow within Africa that should not be overlooked (Hammer et al. 2011). Knowing the DNA regions of earlier admixture would aid in identifying subsequent gene flow events and shed light on adaptive mechanisms aided by interbreeding across hominin tribes at different times.

On the 1000 Genomes Phase 3 data, we use HapFABIA (Hochreiter 2013), a newly developed method for detecting DNA segments that are identical by descent (IBD), to gain insights into the genetic relationships between humans, Neandertals, and Denisovans (see the section "HapFABIA for Extracting Short IBD Segments"). Gusev et al. 2012; Palamara et al. 2012; Botigue et al. 2013; Carmi et al. 2013; Gravel et al. 2013; Ralph and Coop 2013) have already shown that IBD detection approaches may be used to infer population structure. Because the methods employed to identify short IBD segments, which are likely older than longer ones, previous investigations were limited to finding relatively recent genetic ties (Chapman and Thompson 2003). Because rare minor alleles are highly unlikely to arise independently, HapFABIA accurately detects these very short IBD segments via low frequency and rare variants that tag them (Strachan and Read 2004).

Objectives of the Study

This study will tend to discover IBD-based evidence of interbreeding between ancestors of humans and other ancient hominins within Africa in the 1000 Genomes Project Phase 3

data. The evidence of IBD segment sharing between Africans and Neandertal and Denisovan ancient genomes are reported in this paper. Also, show that these shared IBD segments are older than those originating from outside of Africa introgression events. Finally, in the "Discussion" section, we look at various possibilities for the origin of IBD segments that modern Africans, Neandertals, and Denisovans share.

LITERATURE REVIEW

If two or more individuals have identical nucleotide sequences in the same DNA region, it is called identical by state (IBS). If two or more individuals have inherited an IBS segment from a common ancestor, the segment is identical by descent (IBD), meaning the segment has the same ancestral origin (Manavalan, 2014). Because independent origins are exceedingly implausible for such variants, rare variants can be used to distinguish IBD from IBS without IBD. In other words, for rare variants, IBS generally predicts IBD, whereas, for common variants, this is not the case (Ralph and Coop, 2013). For inferring population structure, IBD detection approaches have already proved successful (Bynagari, 2014).

To assess the demographic history of Ashkenazi Jewish individuals, Gusev et al. (2012) sought lengthy IBD segments shared within and between communities. Palamara et al. (2012) and Carmi et al. (2013) both used comparable methods to recreate the demographic histories of Ashkenazi Jewish and Kenyan Maasai people. Botigué et al. (2013) used patterns of lengthy shared IBD segments to prove that genes flowed from North Africa to Southern Europe. By looking for long regions of shared DNA, Ralph and Coop (2013) attempted to quantify the recent shared ancestry of several European ethnicities.

IBD detection approaches have already been used to infer population structure with success. All of these investigations, except for Grave and coworkers used SNP microarray data because IBD segments could not be reliably found in big sequencing data (Manavalan & Bynagari, 2015). In comparison to SNP microarray data, sequencing data has a higher marker density and also captures uncommon mutations, allowing for a finer precision of IBD segment length. Furthermore, all previous microarray analyses were restricted to large IBD segments descended from a recent shared ancestor (Bynagari, 2015). Shorter IBD segments, on the other hand, would transmit information about events that occurred further back in time because shorter IBD segments are thought to be older. As a result, previous investigations were unable to clarify demographic histories at a fine scale and over a long period (Donepudi, 2016). We recently developed HapFABIA (Hochreiter, 2013) to identify extremely short segments of identity by descent (IBD) that are tagged by uncommon variants (so-called tagSNVs) in huge sequencing data (see http://dx.doi.org/10.1093/nar/gkt1013). HapFABIA finds IBD segments 100 times smaller than existing state-of-the-art approaches: 10 kbp vs. 1 Mbp for state-of-the-art methods. Because separate origins of rare minor alleles are exceedingly implausible, HapFABIA uses rare variants (5 percent MAF) to distinguish IBD from IBS without IBD (49, Ch. 15.3,p. 441). Rare mutations also make juxtapositions of smaller IBD segments improbable, preventing the summarization of numerous short IBD segments into one long IBD segment. As a result, the length of IBD segments is more correctly approximated than with earlier techniques.

HapFABIA beat its competitors in recognizing short IBD segments in tests using artificial, simulated, and real genotyping data (Hochreiter, 2013). HapFABIA is built on biclustering (Donepudi, 2014a) and machine learning approaches developed from maximizing the posterior in a Bayes framework (Hochreiter, 2006; Klambauer et al., 2012; 2013).

HapFABIA is a tool for detecting small IBD segments in genotype data from next-generation sequencing (NGS), although it may also be used with DNA microarray data (Manavalan, 2016). HapFABIA uses uncommon variations to detect IBD, especially in NGS data. Because random minor allele sharing is less likely for rare variants than for common variants, rare variants provide more information on IBD than common variants (Browning and Browning, 2012). Both the information provided by rare variations and the information from IBD segments shared by more than two individuals should be used to find short IBD segments (Browning and Browning, 2012). The likelihood of sharing a segment at random is determined by several factors. HapFABIA depends (a) on the frequency of alleles inside the segment, where a lower frequency indicates a lower chance of random segment sharing, and (b) on the number of individuals who share the allele, where a higher number indicates a lower chance of random segment sharing (Bynagari, 2016).

The shorter the IBD segments, the more likely they are to be shared by a larger number of people. A segment that contains rare variations and is shared by a larger number of people is more likely to exhibit IBD (Halldorsson et al., 2011; Leibon et al., 2008). These two traits serve as the foundation for HapFABIA's detection of short IBD segments (Donepudi, 2015).

IBD segments varied in length from 15 to 21 Mbp, with a median of 24 kbp and a mean of 26 kbp. To match the assumptions for the distribution of IBD segment lengths derived in other works, IBD lengths were computed (Thomas et al., 2008; Thomas et al., 1994; Browning, 2008).

METHODS

Data

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), we utilized HapFABIA to extract small IBD segments. This collection contains low-coverage whole-genome sequences from 2,504 people (661 Africans, 347 Admixed Americans, 504 East Asians, 489 South Asians, and 503 Europeans). In comparison to 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. We deleted 11 people because they had cryptic first-degree relationships with others (see the section "Removal of Related Individuals"). A total of 2,493 people (654 Africans—AFR, 347 Admixed Americans—AMR, 504 East Asians—EAS, 485 South Asians—SAS, and 503 Europeans—EUR) were included in the final dataset. To prevent complicating influences from their admixed ancestry, we also eliminated Admixed American individuals from the group of individuals who share an IBD section. For the same reason, we excluded people of African ancestry from the Southwest United States and African Caribbeans from Barbados from the majority of the analyses.

We used HapFABIA with default parameters to analyze the unphased genotype data because prior analyses revealed several phasing problems (Hochreiter 2013). We eliminated short indels and larger copy number variants from the VCF files since they cannot be called as reliably as SNVs. We also divided multiallelic variations into numerous lines. Using bed files retrieved from the UCSC genome browser (Meyer et al. 2013), we filtered out repetition areas and CpGs. We eliminated common and private variations from the analysis because HapFABIA is based on low frequency and unusual

variants. All chromosomes were then separated into 10,000-variant intervals, with adjacent intervals overlapping by 5,000 variants.

The Max Planck Institute for Evolutionary Anthropology (Meyer et al. 2012; Prufer et al. 2014) provided whole-genome sequencing data for the Denisovan (http://cdna.eva.mpg.de/denisova/, last accessed 2 February 2012 and http://cdna.eva.mpg.de/neandertal/altai/, last accessed 23 May 2021) and the Altai Neandertal genome (http://cd Only variations found in at least one member of the 1000 Genomes Project Phase 3 were designated as not determined, accounting for 0.9 percent of Denisovan bases and 0.4 percent of Neandertal bases. The remaining Denisovan and Neandertal bases matched bases of the human reference 96.0 percent and 95.6 percent, respectively, while 4.0 percent and 4.4 percent, respectively, matched either the human minor allele or were different from human alleles. Bases of the reconstructed common ancestor of human, chimp, gorilla, orang-utan, macaque, and marmoset genomes were added as extra information in the 1000 Genomes Project data.

Extracting Short IBD Segments using HapFABIA

HapFABIA (Hochreiter 2013) finds very short IBD segments tagged by low frequency or rare variants (so-called tagSNVs) with a minor allele frequency (MAF) of $\leq 5\%$ in big sequencing data. If the nucleotide sequences in a DNA segment are identical in two or more persons, it is called identical by state (IBS). If two or more individuals have inherited an IBS segment from a common ancestor, the segment is identical by descent (IBD), meaning the segment has the same ancestral origin in these individuals. Because independent origins are exceedingly implausible for such polymorphisms, they can be used to distinguish IBD from IBS without IBD.

IBD Lengths That Are Distributed Exponentially

The length of an IBD segment is exponentially dispersed, with a mean of $100/(2g)$ cM (centiMorgans) (Thomas et al. 1994, 2008; Browning 2008; Gusev et al. 2012; Palamara et al. 2012). As a result, shorter IBD segments have a higher average age than longer ones. We use this relationship to discriminate between ancient intra-African mixing and more modern interbreeding outside of Africa. The recombination rate (cM-to-Mb ratio) advocated by Ulgen and Li (2005) is 1, yet it fluctuates from 0 to 9 along a chromosome (Yu et al. 2001). On the X chromosome, the sex averaged recombination rate is 2/3 that of the female recombination rate (Hedrick 2007).

Removing Related People

The degree of relatedness is shown in the quantity of IBD sharing between two people. About half of the IBD pieces should be shared between two people in a first-degree connection. As a result, we searched the 1000 Genomes data for the fraction of IBD segments that each pair of people shared. Individuals that shared more than half of their IBD regions on multiple chromosomes were thought to be related (Manavalan & Donepudi, 2016). The IBD identification method was performed using the reduced data after one individual of each pair was randomly selected to be deleted.

Ancient Genomes and IBD Segments

We began by searching the 1000 Genomes data for IBD segments in human populations solely. Then we compared the human IBD segments to ancient genomes to see if they were the same. The existence of human minor alleles of tagSNVs in the ancient genome

indicates that the two genomes are related. In most cases, ancient genomes only include a portion of the IBD region. The "old component" of the IBD segment is the portion of the IBD segment that contains minor alleles seen in ancient genomes. Three criteria are used to define the match between an IBD region and an ancient genome:

(1) At least 15% of the tagSNVs in the IBD region match the ancient genome, indicating that the ancient genome also contains the human minor allele.

(2) At least 8 tagSNVs are present in the "ancient part" of the IBD segment, and

(3) 30% of the tagSNVs in the "ancient part" of the IBD segment match the ancient genome. Finally, we used the IBD segment lengths to calculate the length of the "ancient portion" that corresponds to a specific ancient genome (see Manavalan & Ganapathy (2014) for more information).

Ancestral IBD Segments Are Removed

Human IBD regions discovered in ancient genomes could have existed in human and other primates' ancestral genomes. Interbreeding studies based on IBD sharing between current human populations and Neandertals or Denisovans might be thrown off by these IBD segments. As a result, we excluded IBD segments when at least 30% of the tagSNVs shared by an ancient genome also match the 1000 Genomes Project's reconstructed common ancestor sequence.

RESULTS AND DISCUSSION

We present findings from a study of whole-genome sequencing data from the 1000 Genomes Project Phase 3 (1000 Genomes Project Consortium 2015), focusing on very short IBD segments with low frequency and uncommon variations (tagSNVs). First, we identified IBD segments that were shared across continental populations in Africa, East Asia, South Asia, and Europe. Using high-coverage whole-genome sequencing data from the Denisovan and the Altai Neandertal (Meyer et al. 2012; Prufer et al. 2014), we investigated whether the IBD segments matched the Neandertal and/or Denisovan sequence. The "Data" section has more information about the data.

We eliminated IBD segments that matched the 1000 Genomes Project's reconstructed common ancestor sequence of humans and other primates (see "Removal of Ancestral IBD Segments") because they would skew the interbreeding analysis based on IBD sharing between modern humans and Neandertals or Denisovans. In comparison to segments shared by non-Africans and these ancient genomes, we focused on IBD segments shared by Africans and Neandertals and/or Denisovans.

Africans share the majority of the over 25,000 IBD segments that match ancient genomes. There are about 1,000 of these segments on chromosome X, accounting for more than half of the ancient genome-matching X-chromosomal segments. First, we look at how common IBD segments are in ancient genomes across different continents. 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 (for more information, see the section "IBD Segments Matching Ancient Genomes"). Figure 1 shows the results, where different pie charts represent different genomes and the color represents the population that has more than 50% of the individuals with the appropriate IBD segment. Gray slices depict IBD segments in which neither of the populations has more than 50% of the population. Those that match ancient

genomes are enriched in Europeans and Asians when compared to all IBD segments reported in the 1000 Genomes Project data (labeled "All").
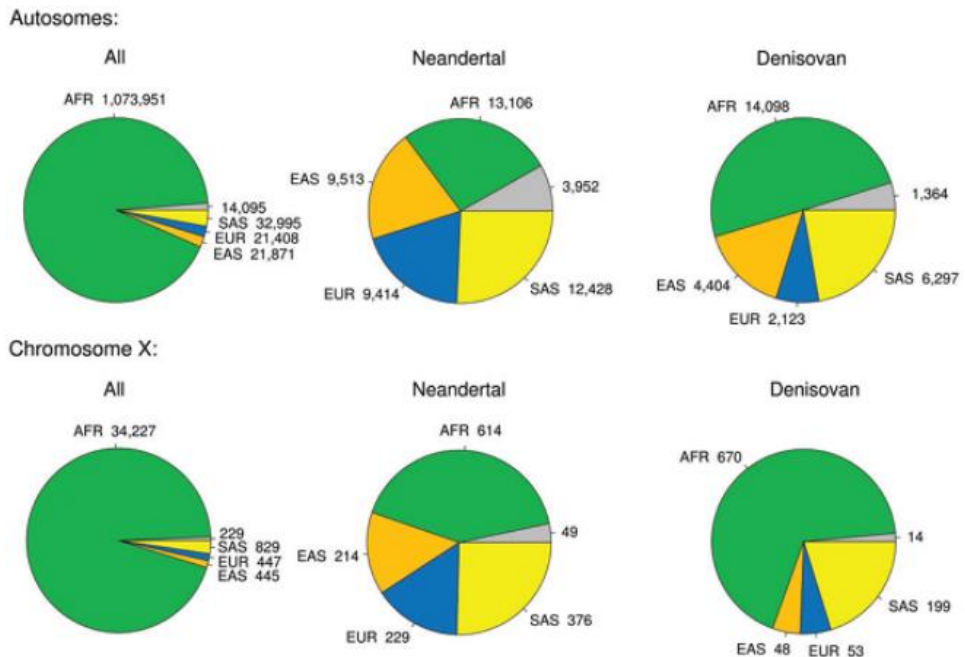

Figure 1: Predominant population for an IBD segment and genome.

South Asians have more segments in common with the Denisovan genome than East Asians and Europeans, and they also have more segments in common with the Neandertal genome than other non-African races. Furthermore, several ancient genome regions, particularly on chromosome X, are shared primarily by Africans. Africans are responsible for 13,106 of the 47,987 Neandertal-matching IBD segments on autosomes and 614 of the 1,479 on chromosomal X. In 14,098 out of 28,157 cases for autosomes and 670 out of 983 cases for chromosomal X, Denisovan-matching IBD regions are primarily shared by Africans.

The IBD segments we discussed above are frequently found in more than one continental group, which means that while they are mostly found in Africans, they may also be found in Asians or Europeans. Parts of ancient genomes have therefore been discovered in Africans, possibly due to later admixture events between human populations. As a result, we looked into IBD segments found only in one continental population. The number of Neandertal- and Denisovan-matching IBD segments per chromosome that are exclusively shared by various geographic groups are shown in Table 1. Across all chromosomes, Neanderthal-like IBD fragments are almost often found only in Africans. Only Africans share approximately 20% of all Neandertal-matching IBD segments on autosomes, compared to less than 7%, 11%, and 4% for East Asians, South Asians, and Europeans, respectively. Africans share around 35% of all Neandertal-matching IBD regions on chromosome X. The findings are even more striking when compared to the Denisovan genome (Donepudi, 2014b). Africans account for over 40% of all Denisovan-matching IBD segments on autosomes, followed by South Asians, who account for fewer than 14% of the

shared segments. Only about 300 IBD segments are shared uniquely by Europeans and Denisovans, implying that IBD segments shared by Europeans and Denisovans were likely introduced into European genomes through mixing with other human populations. Surprisingly, East Asians and Europeans share very few exclusive IBD sequences with the Denisovan person on chromosome X.

About 563 of the 983 IBD segments on chromosome X that match the Denisovan genome are only found in Africans, and a high percentage of the remaining segments are also found in Africans. The findings could be explained by one of two things: (1) little Denisovan DNA penetrated the human X chromosome outside of Africa, or (2) modern Eurasians lost most of the Denisovan sequence on this chromosome due to selective pressure. About 563 of the 983 IBD segments on chromosome X that match the Denisovan genome are only found in Africans, and a high percentage of the remaining segments are also found in Africans. The findings could be explained by one of two things: (1) little Denisovan DNA penetrated the human X chromosome outside of Africa, or (2) modern Eurasians lost most of the Denisovan sequence on this chromosome due to selective pressure. The lengths of IBD segments in different continental populations are then compared to determine their age.

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%)

Because shorter segments are assumed to be older than longer ones, IBD segments originating from events involving ancestors of modern humans and ancestors of Neandertals and/or Denisovans within Africa should be shorter than those originating from later interbreeding outside of Africa, as mentioned in the section "Exponentially Distributed IBD Lengths." Figure 2 compares the density of lengths of IBD segments that match the Neandertal/ Denisovan genome and are unique to Africans vs IBD segments not found in Africans.
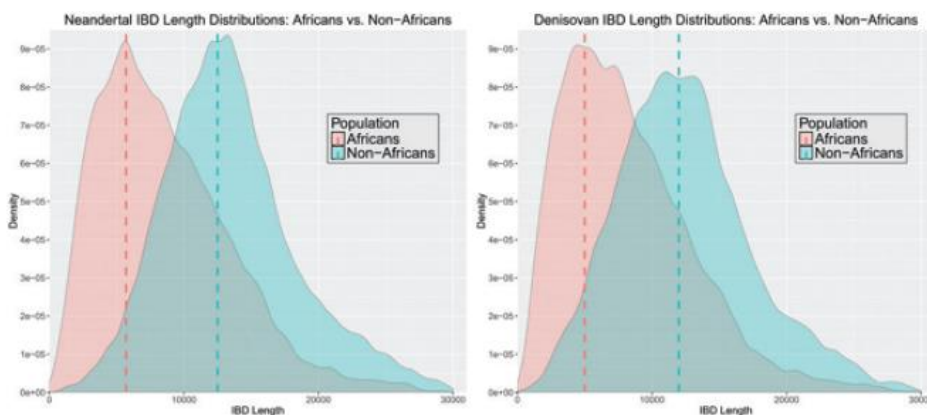
Figure 2: Population-specific IBD segment lengths for Neandertal- and Denisovan-matching segments

When Neandertal- and Denisovan-matching IBD segments are shared by Africans, the primary density peak (the mode of the density) is around 5,700 and 5,000 bp, respectively, whereas when they are shared by non-Africans, the peaks are about 12,500 and 12,000 bp, respectively. Only non-African populations share Neandertal and Denisovan-matching IBD segments, which are significantly longer and younger than those shared by Africans and ancient genomes. 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; Prufer et al. 2014) complement prior accounts of interbreeding after migration out of Africa.

The brief Neandertal- and Denisovan-matching IBD segments discovered solely in Africans are far older, and thus most likely come from inside Africa. When the densities of lengths of IBD segments that match the Neandertal/Denisovan genome and are private to Africans are compared to IBD segments that are observed in Africans but not exclusively, we see a peak at 12,000 and 11,000 bp for Neanderthal- and Denisovan-matching IBD segments that are observed in Africans but not exclusively, respectively (see supplementary fig. S5, Supplementary Material online). These peaks could be segments from later interbreedings outside of Africa that made their way into the African genome via a recent Eurasian backflow, as previously documented (Wang et al. 2013; Prufer et al. 2014; Llorente et al. 2015). Similarly, when we compare the densities of lengths of IBD segments that match the Neandertal/Denisovan genome and are exclusive to Africans versus IBD segments that are observed outside of Africa but not exclusively in Africans, we see that the densities of IBD segments observed outside of Africa are shifted to the left when compared to densities of IBD segments not observed in Africans (see supplementary figure 2). The shorter segments that are causing this shift could be the remains of older segments that originated in Africa.

CONCLUSION

We studied genetic links between Neandertals and Denisovans using 1000 Genomes Phase 3 data to represent the human genome and sequenced genomes of a Neandertal and a Denisovan. We identified very short IBD segments with high confidence utilizing shared DNA segments that are identical by descent, focusing on low frequency and uncommon variations.

REFERENCES

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, L. R., Henn, B. M., Gravel, S., Maples, B. K., Gignoux, C. R., Corona, E., Atzmon, G., Burns, E., Ostrer, H., Flores, ., C., Bertranpetit, J., Comas, D. and Bustamante, C. D. 2013. 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.

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

Browning SR. 2008. Estimation of pairwise identity by descent from dense genetic marker data in a population sample of haplotypes. Genetics 178:2123–2132.

Bynagari, N. B. (2014). Integrated Reasoning Engine for Code Clone Detection. *ABC Journal of Advanced Research*, *3*(2), 143-152. https://doi.org/10.18034/abcjar.v3i2.575

Bynagari, N. B. (2015). Machine Learning and Artificial Intelligence in Online Fake Transaction Alerting. *Engineering International*, *3*(2), 115-126. https://doi.org/10.18034/ei.v3i2.566

Bynagari, N. B. (2016). Industrial Application of Internet of Things. *Asia Pacific Journal of Energy and Environment*, *3*(2), 75-82. https://doi.org/10.18034/apjee.v3i2.576

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

Donepudi, P. K. (2014a). Technology Growth in Shipping Industry: An Overview. *American Journal of Trade and Policy*, *1*(3), 137-142. https://doi.org/10.18034/ajtp.v1i3.503

Donepudi, P. K. (2014b). Voice Search Technology: An Overview. *Engineering International*, *2*(2), 91-102. https://doi.org/10.18034/ei.v2i2.502

Donepudi, P. K. (2015). Crossing Point of Artificial Intelligence in Cybersecurity. *American Journal of Trade and Policy*, *2*(3), 121-128. https://doi.org/10.18034/ajtp.v2i3.493

Donepudi, P. K. (2016). Influence of Cloud Computing in Business: Are They Robust?. *Asian Journal of Applied Science and Engineering*, *5*(3), 193-196. Retrieved from https://journals.abc.us.org/index.php/ajase/article/view/1181

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

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

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., Clevert, D.A. and Obermayer, K. 2006. A new summarization method for Affymetrix probe level data. Bioinformatics, 22(8):943–949.

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

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 sizecannot explain higher amounts of Neandertal ancestry in East Asian than in European human populations. AmJ Hum Genet. 96:454–461.

Klambauer, G., Schwarzbauer, K., Mayr, A., Clevert, D.A., Mitterecker, A., Bodenhofer, U. and Hochreiter, S. 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(9):e69.

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

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 highcoverage 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.

Manavalan, M. (2014). Fast Model-based Protein Homology Discovery without Alignment. *Asia Pacific Journal of Energy and Environment*, *1*(2), 169-184. https://doi.org/10.18034/apjee.v1i2.580

Manavalan, M. (2016). Biclustering of Omics Data using Rectified Factor Networks. *International Journal of Reciprocal Symmetry and Physical Sciences*, *3*, 1–10. Retrieved from https://upright.pub/index.php/ijrsps/article/view/40

Manavalan, M., & Bynagari, N. B. (2015). A Single Long Short-Term Memory Network can Predict Rainfall-Runoff at Multiple Timescales. *International Journal of Reciprocal Symmetry and Physical Sciences*, *2*, 1–7. Retrieved from https://upright.pub/index.php/ijrsps/article/view/39

Manavalan, M., & Donepudi, P. K. (2016). A Sample-based Criterion for Unsupervised Learning of Complex Models beyond Maximum Likelihood and Density Estimation. *ABC Journal of Advanced Research*, *5*(2), 123-130. https://doi.org/10.18034/abcjar.v5i2.581

Manavalan, M., & Ganapathy, A. (2014). Reinforcement Learning in Robotics. *Engineering International*, *2*(2), 113-124. https://doi.org/10.18034/ei.v2i2.572

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 ofOAS 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.

Pru¨ fer 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, Pa¨a¨bo S, Reich D. 2012. The date of interbreeding between Neandertals and modern humans. PLoS Genet. 8:e1002947.

Schaffner, S. F. 2004. The x chromosome in population genetics. Nat Rev Genet, 5(1):43–51, 2004.

Serre D, et al. 2004. No evidence of Neandertal mtDNA contribution to early modern humans. PLoS Biol. 2:e57.

Skoglund P, JakobssonM. 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, LiW. 2005. Comparing single-nucleotide-polymorphism marker based and microsatellite marker-based linkage analyses. BMC Genet. 6:S13.

Unpublished data [cited 2015 Apr 4]. Available from: http://biorxiv. org/content/early/2014/07/15/003988.

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.

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.

--0--