Supplement: ABO genetic variation in Neanderthals and Denisovans

Fernando A. Villanea, Emilia Huerta-Sanchez, and Keolu Fox

Compiled on February 23, 2021.

1 Sharing of Neanderthal ABO haplotypes through Incomplete Lineage Sorting

We calculate the probability of a genome fragment carrying the ABO gene, of length of 31kb, shared by modern humans and Neanderthals due to incomplete ancestral lineage sorting, as described in Huerta-Sánchez et al. [2014]. For this calculation, r is the recombination rate per generation per bp, of 3.39e-8, for a genome fragment at coordinates HG19 9:136125329-136157138, as described in the HapMap recombination map [Consortium et al., 2007] and found at: ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/technical/working/20110106_recombination_hotspots/.

The divergence time of the human and Altai Neanderthal branches t, is 980,000 years, which is calculated as twice the split between Neanderthals and modern humans (550,000 years) minus the time at sampling of the Altai Neanderthal (120,000 years) [Prüfer et al., 2014, 2017, Douka et al., 2019]. We used an estimated time for interbreeding between the two groups of 50,000 years ago, and a generation time of 29 years. Under these assumptions, we exclude that it derives from the common ancestor (p = 5.5e-15) and conclude that this region entered the human gene pool through admixture with Neanderthals. Furthermore, we recalculated this probability using an extremely conservative divergence time of 300,000 years, as used for Denisovans in Huerta-Sánchez et al. [2014]. We should mention, using a highly conservative Neanderthalhuman split time considerably increases the probability of incomplete lineage sorting. Using this extreme estimate, the probability of incomplete lineage sorting is still only p = 0.00017, further supporting that this region entered the human gene pool through admixture with Neanderthals.

Superpopulation	Population	Fragment ID	Chromosome 9 Position
Europe	GBR	15	136071592 - 136374869
Europe	IBS	37	136109740 - 136412483
Europe	TSI	24	136071592 - 136383617
Southeast Asia	BEB	18	135862746 - 136374869
Southeast Asia	GIH	23	136131316 - 136374869
Southeast Asia	ITU	15	136077559 - 136389868
East Asia	JPT	29	136131539 - 136368328

Table 1: Superpopulation, Population, ID, and hg19 coordinates for archaic genome fragments in the 1000 Genomes Project detected in Browning et al. [2018] which contain the ABO gene.

2 Supplementary Tables

Supplementary Table 1. Superpopulation, Population, ID, and hg19 coordinates for archaic genome fragments in the 1000 Genomes Project detected in Browning et al. [2018] which contain the ABO gene.

3 Admixed populations in the 1000 Genomes Panel

Supplementary Table 2. Chromosome, start and stop position, ancestry block (NAT=Native American, EUR=European, AFR=African),for all neanderthal introgressed ABO haplotypes in American individuals, identified by a sample ID, and 1000 Genomes Panel population. In all individuals the neanderthal introgressed haplotype was found in an EUR ancestry block, confirming that these haplotypes were inherited through European admixture post-American colonization. Find .txt file at: https://drive.google. com/file/d/1ckuFOMZJ-Ovmlofh_CqB1_Z1yNgsuU5P/view?usp=sharing

4 Haplostrips genetic distances (unsorted)

4.1 Denisovan

Find .txt file at: https://drive.google.com/file/d/1gKW9jQo7Sw0ycfnHG1BDgEql5r_ X1AVr/view?usp=sharing

4.2 Altai neanderthal

Find .txt file at: https://drive.google.com/file/d/10gpVr99xNBC0LbqJ4ZQYxzVLy14e5oX-/
view?usp=sharing

4.3 Vindija neanderthal

Find .txt file at: https://drive.google.com/file/d/1-gmqPSS35wiBYpBkGS6GCvEhIvCAc9z5/ view?usp=sharing

4.4 Chagyrskaya neanderthal

Find.txt file at: https://drive.google.com/file/d/1XFej1HItNtrN13m1H3dJeWPbxrTSn1D-/
view?usp=sharing

Supplementary Figures $\mathbf{5}$

Supplementary Figure 1

			Ŀ	igure	allele	varia	ınt ta	uble.po	it	
BSNP	position	reference	sample	functionGVS	rsID	amino acid	protein	polyPhen	GERP	EXAC
SNP 129	136131056	CG	с	frameshift	56392308	none	NA	unknown	-5.32	delG=5059/ref=40137
ne -	136131059	G	G/C	missense	0	ASN,LYS	353/355	0.923	-0.105	C=1/G=49241
ne -	136131060	т	T/G	missense	0	ASN,THR	353/355	0.597	-1.81	G=8/T=49500
SNP_129	136131064	G	G/A	missense	56390333	ARG,TRP	352/355	1	1.03	C=2/A=41/G=54483
SNP_135	136131065	G	G/A	synonymous	183748371	none	351/355	unknown	-0.912	A=3/G=56545
NP 116	136131069	G	G/A	missense	7466899	ALA,VAL	350/355	0.095	-2.65	A=78/G=62032
ж. —	136131086	с	C/T	synonymous	0	none	344/355	unknown	3.47	T=17/C=85651
SNP 129	136131109	т	T/C	missense	56231718	ARG,GLY	337/355	0.47	-0.002	C=29/T=109011
SNP 138	136131118	с	C/T	missense	369695939	ALA,THR	334/355	0.013	-8.76	T=2/C=113272
ie .	136131154	с	C/T	missense	0	GLU,LYS	322/355	1	4.38	T=5/C=119169
SNP 117	136131188	c	CT	synonymous	8176749	none	310/355	unknown	1.36	T=14613/C=106423
SNP 129	136131192	т	T/C	missense	56346931	TYR.CYS	309/355	1	3.19	C=115/T=120953
20	136131240	G	G/C	missense	0	ALA GLY	293/355	0.022	0.545	C=1/G=121047
SNP 117	136131289	c	CT	missense	8176748	VAL MET	277/355	1	4.38	A=1/T=31193/C=87126
SNP 117	136131315	c	C/G	missense	8176747	GLY.ALA	268/355	0.006	-8.39	G=14032/C=93378
SNP 127	136131316	c	CT	missense	41302905	GLY.ARG	268/355	0.99	3.28	T=1726/C=105746
)c	136131319	c	CT	missense	0	GLY.ARG	267/355	1	4.38	G=1/T=1/C=106254
NP 117	136131322	G	G/T	missense	8176746	LEU MET	266/355	0.045	-8.76	A=1/T=13903/G=91164
NP 117	136131347	G	A/G	synonymous	8176745	none	257/355	unknown	-1.59	A=25722/G=61360
NP 117	136131350	G	G/T	synonymous	8176744	none	256/355	unknown	-3.34	T=2927/G=70809
	136131375	c	C/G	missense	0	ARG PRO	248/355	1	1.42	T=2/G=7/C=57099
END 126	136121299	G	GA	CORDORATE CONTRACTOR	25404115	2000	242/255	unknoum	2.49	A-26/0-53259
NP 127	136131389	G	G/A	synonymous	201420225	none	245/555	unknown	3.46	A=49/G=50292
AND 117	13013140/	G	CT	synonymous	201437323	CLV SED	2511555	0.879	-0.929	A-4810-30272 T-0170 (C-11170)
2010_120	130131415	c	CT CT	mosense	81/0/43	CLV ACD	233/355	0.638	4.30	1-03/00-04020
50VF_129	130131429	0	07	intooense	30110432	OL LASP	250/533		4.39	1=225/C=30383
SNF_117	130131437	C	10	synonymous	8176742	none	21//535	unknown	-8.65	1-13400-C-43302
2ND 120	136131461	G	CIA	synonymous	61/0/41	ADC CVE	219/333	0.027	-4.2	A=9491/G=30041
SNF_127	130131409		UA	intooense	30408700	ARGUETS	217/333	0.037	-9.58	A-6030-36342
SNP_117	1361314/2	A	AVI	missense	81/6/40	PHE,ILE	216/355	1	3.33	1=16226/A=47/66
5NP_135	136131490	0	CI	missense	181536132	VAL,MET	210/355	0.934	-9.58	1=1/C=62309
SNP_117	136131523	6	G/A	missense	81/6/39	ARG,CYS	199/355	0.868	2.8	A=1405/G=66839
SNP_129	136131539	A	A/G	synonymous	55764262	none	193/355	unknown	3.79	G=74/A=77166
5NP_135	136131556	0	6/1	missense	184446112	AKG,SEK	188/355		2.75	unknown
SNP_129	136131576	C	C/T	stop-gained	55727303	TRP,stop	181/355	unknown	4.56	T=3278/C=98528
SNP_129	136131589	С	C/T	missense	55687199	ALA,THR	177/355	0.048	-6.69	T=96/C=107234
SNP_138	136131590	G	G/A	synonymous	371569951	none	176/355	unknown	-9.38	A=9/G=108075
SNP_129	136131591	с	C/T	missense	56039827	ARG,HIS	176/355	0.024	-1.05	T=143/C=108359
SNP_116	136131592	G	C/G	missense	7853989	ARG,GLY	176/355	0.011	-1.09	A=1/C=15935/G=94822
ne	136131595	с	C/T	missense	0	VAL,MET	175/355	1	3.77	T=51/C=112005
SNP_129	136131616	G	G/C	missense	56043861	ARG,GLY	168/355	1	-4.42	A=1/C=2/G=115095
SNP_129	136131621	GT	G	frameshift	56284703	none	NA	unknown	-8.32	delT=410/ref=115308
SNP_129	136131630	G	G/A	missense	55756402	THR,MET	163/355	0.896	0.193	A=82/G=116240
SNP_137	136131635	G	G/A	synonymous	200932155	none	161/355	unknown	-3.15	A=139/G=116133
SNP_117	136131636	С	C/T	missense	8176738	ARG,HIS	161/355	0	-9.38	A=1/T=84/C=116231
SNP_86	136131651	G	G/A	missense	1053878	PRO,LEU	156/355	0.95	3.77	A=10390/G=106232
SNP_129	136131664	A	A/G	missense	55687353	PHE,LEU	152/355	0.983	4.69	unknown
ne	136131704	С	C/T	synonymous	0	none	138/355	unknown	4.95	T=1/C=116333
ne	136131718	G	G/A	synonymous	0	none	134/355	unknown	4.16	A=2/G=116272
20	136132845	A	A/G	missense	0	PHE,LEU	109/355	1	4.33	G=4/A=122326
SNP_117	136132852	G	G/A	synonymous	8176721	none	106/355	unknown	-3.12	A=1601/G=120723
SNP_135	136132853	T	T/C	missense	181412963	ASN,SER	106/355	1	4.33	C=2/T=122322
ie .	136132864	G	G/A	synonymous	0	none	102/355	unknown	0.112	A=1/G=122311
SNP_117	136132873	т	C/T	synonymous	8176720	none	99/355	unknown	-8.39	C=48708/T=73496
SNP_117	136132908	т	TC	frameshift	8176719	none	NA	unknown	4.2	insC=45730/ref=75728
20	136133466	AG	Α	frameshift	0	none	NA	unknown	2.04	unknown
NP_83	136133506	A	G/A	missense	512770	SER,PRO	74/355	unknown	2.04	G=92871/A=29225
e	136135226	G	G/C	missense	0	PRO,ARG	67/355	0.02	1.3	C=1/G=122329
NP_138	136135232	G	G/A	missense	375733196	SER,LEU	65/355	0	-1.07	A=10/G=122320
NP 129	136135236	С	C/T	missense	56335272	VALILE	64/355	0.137	1.3	unknown
SNP 83	136135237	A	A/G	coding-unknown	549443	none	NA	unknown	-2.61	G=90060/A=31798
SNP 83	136135238	Т	C/T	missense	549446	HIS,ARG	63/355	unknown	-2.61	unknown
c	136136728	с	C/G	missense	0	GLY ARG	50/355	0.004	0.822	G=1/C=66129
SNP 83	136136770	A	A/C	missense	688976	PHE VAL	36/355	unknown	-3.95	C=66731/A=21321
NP 117	136136773	C	CT	missense	8176696	GLY ARG	35/355	0.806	2.04	A=2/T=1262/C=58966
NP 129	136137547	C	C/A	missense	55876802	ARGLEU	18/355	0.001	-2.13	T=1/A=1785/C=109688
	126127551	G	GA	missense	0	I EU PHE	17/255	0.007	1.27	A=210=100007
NP 179	136137331	C	CT	missance	55017062	ALA TUP	16/355	0.067	4.6	G=2/T=215/C=109557
NB 120	130137334	C	CIA	moscuse	(172(20)	ALA, IBK	10/333	0.002		0-2/1-213/C=108337
or_129	13013/555	0	UA	synonymous	01/30301	none	13/355	unanown	1.30	A-11003-10/202
SNP 138	136150600	G	110	synonymous	sty/x/4410	none	// 135	TEXTOWN	-1.64	ATT1/1/17841/

Figure 1: SNVs and indels which define ABO allele variation in the coding portion of the ABO gene. These 68 variants were identified and annotated in Yip [2002], Patnaik et al. [2012].

6 Neanderthal haplotypes resolved

Supplementary Figure 2



Figure 2: Steps taken to validate heterozyous sites, and link blocks of phased heterozygous sites

Supplementary Figure 3

]	F,	iĮ	<u>b</u> .	u	r	e	(ł	1	a	g	y	r	sł	52	ą	ya	а	ł	p	h	a	S	e	d	•]	0	d	t															
13010100	100121009	1315100	PROJECT IN COLUMN	Desirent and	1001031000	1001031001	011101001	134121154	Contract	Decision in	130131200	SURVER	000103001	NUMBER	ZZUSIE	THURSDAY	Contrasts of	Part of the local data	CONTRACT	136131410	62461661	ABNUNES	100101401	120100	130 LANATS	Control and	SCHOOL ST	000112000	1301111570	apsigues.	100112100	PRODUCTION 1	CREATING	BANCINCI	1201231021	DISKNEY	DONUSIES	10010100	PONCI BCL	NUISING	WALKSHOP	134125485	Condiciance.	Cymerico I	CARCUNCI	1000011001	BOHCCI BCI	APPECT OF L	10010000	Tal 136238	100100237	BCORT NO.	AT A REAL OF A	CLASSING .	(PERCENCE)	ASSECTANCE	130137564	138131360	1 Million (The second se	ABO match
0.0	a G	1	a a	a 0	0.0	1	0.0	0.0	c 1	1 0	10	2 2	0.0	0.0	0	9	5 1		5 0	0.0	00	0.0	0.0	0	A	0.0		-0	e	e	6	c 6	1	0 6	0.0	60	0.0		0 J	0	0 0	Å	0 0	T I		1	Â	6	9	c c	0.0	0.0	0 0			G	0.0	9.0	6	£	02 01
00	e G	1	0 1 0 1		0 0	÷	0.0	e e	0		1 0	11.11	00	0.0	6 5	0 G	5 6		5 0 5 0	0.0		0 0	6 5	0.0	2	5 0		0	ii T	5	6 1	6 6		0	0.0	5	0 0	6 1 6	0 4	0		A A	0 0	r e		8	â	6	0	5 11	0 0	0.0	6 6	0 0		6	5	a 1 6	8	1 N 1 N	áo multich Ao multich
00	G	į.	6 1 6 1	3 0 5 0	c c	7	0.0	а с	0.0	r 0	5 0	5	0.0	00	0.0	a a	5 (5 (s 0	0	c c	0.0	g n	0.0	A .			0	a	E.	G 3		1	0	0	0 6	0.0	с і с і	0 4		6	*	0.0	T I		8	\$		0	5	0.0	c c	0 0	0 0		a a	e c	0 0	0 0	1.	CZ AD miatch
00	9	1	6 i	3 0 5 0	0	ž	00	5	6	r o	1 0	10	0 0	0.0	0.0	9	5 6 5 6		3 0	0 0	5	-C -C	ц 5	0.0	A A			-0	U.	c	0.1	c . e		0	0.0	u u	0.0	5	0 4	0	0	A A	0.0	1.1		8	â	6 4	0	с с	0.0	0.0	0 0			6	5	8	6	1 N 1 N	io match io match

Figure 3: Possible configurations of unphased heterozygosity blocks for the two Chagyrskaya Neanderthal chromosomes, including closest match in modern ABO haplotypes

Supplementary Figure 4



Figure 4: Possible configurations of unphased heterozygosity blocks for the two Vindija Neanderthal chromosomes, including closest match in modern ABO haplotypes

Supplementary Figure 5

Figure denisova phased.pdf 1361315 1361314 1361314 1361314 1361314 1361314 1361314 1361314 1361314 1361314 1361313 1361313 1361313 C G T G G G C T C C C T G с д т д д д с т с с С G G C G C C C C C A G A C CGTGGGCTC ATCC C C C G G A C C C G AC C G T G G G C T C C C C G T G G G C T C C T C G G A C C A G A C G C G G G G C G A C G ATCA CGTGGGCTC GGAO G C G GCGCCC G G A C G A G C C G C G T G G G C T C C C AGAC C C C G G G G C G GTGCDAAGG C G T G G G C T C C C T G C C G T G G G C T C C T T G

Figure 5: Possible configurations of unphased heterozygosity blocks for the two Denisovan chromosomes, including closest match in modern ABO haplotypes

7 Introgressed genome fragments containing ABO

Supplementary Figure 6

																				ł	ľ	ie	ςι	u	r	е	ŧ	a	r	cl	h	a	i	С	а	l	iĮ	gı	10	е	d	·]	þ	ł	f																
138131058	136131059	136131050	136121064	136131069	136131086	138131109	136131118	136131164	136131192	136131240	136131289	138131315	136131316	135131319	136131322	138131347	136131350	136131375	136131389	126121607	136131415	100101007	136131451	138131459	138131472	136131490	136131523	138131539	138131558	136131576	136131589	138131590	200101001	136131593	136131616	136131621	136131630	136131635	106131636	138131854	136131704	135131718	136132845	136132852	136132853	136132954	138132908	136133466	136133506	138135226	136136232	138135237	136135238	136136728	136136770	139196773	136137551	136137554	136137555	attordeu	Linebatros
С	G	T (G (GG	C	т	С	C	СТ	G	C	С	С	С	G	G	G	С	G	G	C	0	G	G	i A	C	G	Α	G	С	С	G	C (C (G	G	G	G	CO	3 A	C	G	A	G	Т	G (c	A	A	G	G	CG	C	С	С	C /	١G	С	G	Alta	i 02
C	G	T I	G d	GG	С	T	С	C I	СТ	G	c	C	C	C	G	G	G	C	G	G	C C	> 0	G	; G	i A	C	G	Α	G	С	C	G	C 4	C C	G	G	G	G	C C	λ	C C	G	Α	G	Т	G	; c	A	A	G	G i	C Q	; C	С	С	c,	4 G	¢	G	i Alta	i 02
С	G	T I	G (GG	С	т	С	C	СТ	G	C	C	С	С	G	G	G	С	G	G	C	2 0	G	G	iΛ	C	G	٨	G	С	С	G	C (C	; G	G	G	G	CO	3 A	C	G	A	G	Т	G	C	A	A	G	G	CG	i C	С	С	CI	G	С	G	6 Chagyrs	kaya O2
¢	G	TI	G 4	GG	C	Т	C	C I	С Т	G	C	C	C	C	G	G	G	C	G	G	CO	2 0	; G	; G	: A	C	G	G	G	С	C	G	C (G	G	G	G	G	CO	a A	C	: G	A	G	т	G	0	A	G	G	G i	CQ	; C	С	С	C I	; G	С	G	Chagyrs	kaya O1
С	G	T I	G 4	GG	C	т	С	C	СТ	G	C	C	C	C	G	G	G	С	G	G	CO	0	G	G	5 A	C	G	G	G	С	С	G	C (G	G	G	G	G	CO	3 A	C	G	A	G	т	G	0	A	G	G	G	CG	i C	С	C	CI	; G	С	G	5 Vind	a 01
С	G	TI	6 0	G G	С	Т	C	C .	ΤT	G	C	G	C	C	Т	G	G	С	G	G	C C	2 0	A	6	i A	C	G	A	G	C	С	G	C (C	G	G	G	G	CC	3 A	C	G	Α	G	Т	6 (A	G	G	G I	C C	C C	С	С	C I	C 6	С	G	S Vindija	a B(A)
C	G	TI	G d	GG	C	т	С	С	СТ	G	C	C	C	C	G	G	G	С	G	G	C			G	i A	C	G	A	G	С	С	G	C C	G	G	G	G	G	C C	3 A	C	G	A	G	т	G	D	A	G	G	G	c [C	C	C	C I	; G	C	G	Denisor	van 01
С	G	TI	6 0	G G	С	т	С	C .	ΤT	G	C	G	C	C		G	G	С	6	G	τ¢		A	(G	A	C	G	Α	G	С	C	G	C (CO	6 G	G	G	G	CO	3 A	C	G	A	G	т	G	D	A	Α	G	G	c 🖊	ιT	С	A	C I	G	С	6 0	B Denisov	ran O(B)

Figure 6: Phasing of ABO functional sites based on archaic variation only

Supplementary Figure 7



Figure 7: Phasing of ABO functional sites using surviving introgressed haplotypes as a reference

References

- Sharon R Browning, Brian L Browning, Ying Zhou, Serena Tucci, and Joshua M Akey. Analysis of human sequence data reveals two pulses of archaic denisovan admixture. *Cell*, 173(1):53–61, 2018.
- International HapMap Consortium et al. A second generation human haplotype map of over 3.1 million snps. *Nature*, 449(7164):851, 2007.
- Katerina Douka, Viviane Slon, Zenobia Jacobs, Christopher Bronk Ramsey, Michael V Shunkov, Anatoly P Derevianko, Fabrizio Mafessoni, Maxim B Kozlikin, Bo Li, Rainer Grün, et al. Age estimates for hominin fossils and the onset of the upper palaeolithic at denisova cave. *Nature*, 565(7741): 640–644, 2019.
- Emilia Huerta-Sánchez, Xin Jin, Zhuoma Bianba, Benjamin M Peter, Nicolas Vinckenbosch, Yu Liang, Xin Yi, Mingze He, Mehmet Somel, Peixiang Ni, et al. Altitude adaptation in tibetans caused by introgression of denisovan-like dna. *Nature*, 512(7513):194, 2014.
- Santosh Kumar Patnaik, Wolfgang Helmberg, and Olga O Blumenfeld. Bgmut: Ncbi dbrbc database of allelic variations of genes encoding antigens of blood group systems. *Nucleic acids research*, 40(D1):D1023–D1029, 2012.
- Kay Prüfer, Fernando Racimo, Nick Patterson, Flora Jay, Sriram Sankararaman, Susanna Sawyer, Anja Heinze, Gabriel Renaud, Peter H Sudmant, Cesare De Filippo, et al. The complete genome sequence of a neanderthal from the altai mountains. *Nature*, 505(7481):43, 2014.
- Kay Prüfer, Cesare de Filippo, Steffi Grote, Fabrizio Mafessoni, Petra Korlević, Mateja Hajdinjak, Benjamin Vernot, Laurits Skov, Pinghsun Hsieh, Stéphane Peyrégne, et al. A high-coverage neandertal genome from vindija cave in croatia. *Science*, 358(6363):655–658, 2017.
- SP Yip. Sequence variation at the human abo locus. Annals of human genetics, 66(1):1–27, 2002.