

RESEARCH ARTICLE

Lower promoter activity of the *ST8SIA2* gene has been favored in evolving human collective brains

Toshiyuki Hayakawa^{1,2*}, Masahiro Terahara¹, Naoko T. Fujito³, Takumi Matsunaga¹, Kosuke M. Teshima⁴, Masaya Hane⁵, Ken Kitajima⁶, Chihiro Sato⁶, Naoyuki Takahata³, Yoko Satta³

1 Graduate School of Systems Life Sciences, Kyushu University, Fukuoka, Japan, **2** Faculty of Arts and Science, Kyushu University, Fukuoka, Japan, **3** School of Advanced Sciences, SOKENDAI (The Graduate University for Advanced Studies), Hayama, Kanagawa, Japan, **4** Department of Biology, Kyushu University, Fukuoka, Japan, **5** Bioscience and Biotechnology Center, Nagoya University, Nagoya, Aichi, Japan, **6** Institute for Glyco-core Research (iGCORE), Nagoya University, Nagoya, Aichi, Japan

* thayakawa@artsci.kyushu-u.ac.jp



OPEN ACCESS

Citation: Hayakawa T, Terahara M, Fujito NT, Matsunaga T, Teshima KM, Hane M, et al. (2021) Lower promoter activity of the *ST8SIA2* gene has been favored in evolving human collective brains. PLoS ONE 16(12): e0259897. <https://doi.org/10.1371/journal.pone.0259897>

Editor: Iwamoto Kazuya, Kumamoto University Faculty of Life Sciences School of Medicine: Kumamoto Daigaku Daigakuin Seimei Kagaku Kenkyubu Igakubu, JAPAN

Received: May 24, 2021

Accepted: October 28, 2021

Published: December 16, 2021

Copyright: © 2021 Hayakawa et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its [Supporting Information](#) files.

Funding: This research was supported by the Japanese Society for Promotion of Science (JP16K07535 and JP19K06866 to T.H.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

ST8SIA2 is an important molecule regulating expression of the phenotype involved in schizophrenia. Lowered promoter activity of the *ST8SIA2* gene is considered to be protective against schizophrenia by conferring tolerance to psychosocial stress. Here, we examined the promoter-type composition of anatomically modern humans (AMHs) and archaic humans (AHs; Neanderthals and Denisovans), and compared the promoter activity at the population level (population promoter activity; PPA) between them. In AMHs, the TCT-type, showing the second lowest promoter activity, was most prevalent in the ancestral population of non-Africans. However, the detection of only the CGT-type from AH samples and recombination tracts in AH sequences showed that the CGT- and TGT-types, exhibiting the two highest promoter activities, were common in AH populations. Furthermore, interspecies gene flow occurred into AMHs from AHs and into Denisovans from Neanderthals, influencing promoter-type compositions independently in both AMHs and AHs. The difference of promoter-type composition makes PPA unique in each population. East and Southeast Asian populations show the lowest PPA. This results from the selective increase of the CGC-type, showing the lowest promoter activity, in these populations. Every non-African population shows significantly lower PPA than African populations, resulting from the TCT-type having the highest prevalence in the ancestral population of non-Africans. In addition, PPA reduction is also found among subpopulations within Africa via a slight increase of the TCT-type. These findings indicate a trend toward lower PPA in the spread of AMHs, interpreted as a continuous adaptation to psychosocial stress arising in migration. This trend is considered as genetic tuning for the evolution of collective brains. The inferred promoter-type composition of AHs differed markedly from that of AMHs, resulting in higher PPA in AHs than in AMHs. This suggests that the trend toward lower PPA is a unique feature in AMH spread.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Homo sapiens, anatomically modern humans (AMHs), emerged in Africa and expanded its habitat to almost every corner of the globe during the past 100 thousand years (kyr) [1]. During this spread across the globe, AMHs generated greater technological sophistication and larger bodies of adaptive knowhow to survive efficiently in new environments [2]. Once individuals evolve to learn from one other with sufficient accuracy (fidelity), social groups of individuals develop what might be called *collective brains* [2, 3]. Collective brains, in which individuals are regarded as neurons, generate technological sophistication and adaptive knowhow, and drive cumulative cultural evolution [2, 3].

The *ST8SIA2* gene encodes a sialyltransferase that synthesizes a linear homopolymer of sialic acid (polysialic acid; PSA) in the brain [4]. Sialic acids are a family of nine-carbon monosaccharides that are found at the outer end of glycan chains on the cell surface and are secreted molecules in the deuterostome lineage; they play important roles as ligands in cell–cell communication [5]. The major carrier of PSA is neural cell adhesion molecule (NCAM), and polysialylated NCAM (PSA-NCAM) plays an important role in neurite outgrowth, synapse formation and plasticity [4]. Thus, *ST8SIA2* contributes to these neuronal events in the brain by producing PSA.

Three SNPs (rs3759916, rs3759915, and rs3759914) in the promoter region of the *ST8SIA2* gene are associated with schizophrenia risk and involved in its promoter activity [6–9]. Since this association was detected in multiple populations including Asians and Europeans [6–9], these three SNPs can be regarded as a general regulator of the expression of the phenotype involved in schizophrenia. We previously reported that a non-risk haplotype of these three promoter SNPs [CGC promoter type (C in rs3759916, G in rs3759915, C in rs3759914)] has been under positive selection mainly in Asia during the past ~20–30 kyr [9]. Schizophrenia is an important mental disease because of the serious social impairment that it produces. Since environmental risk factors interact with genetic risk factors toward the onset of schizophrenia [10], the selective pressure in positive selection on the non-risk promoter type (i.e., CGC-type) should be one of the environmental risk factors. Considering that the *ST8SIA2* gene participates in brain functions, psychosocial stress, the sole major environmental risk factor that brain functions are responsible for dealing with [11–15], can be regarded as the selective pressure. Ultimately, positive selection on the CGC-type is considered as an adaptation to the environment exerting psychosocial stress [9]. Therefore, *ST8SIA2* is also a master regulator in the expression of the phenotype involved in psychosocial stress.

The CGC-type shows the lowest promoter activity among the four major promoter types (the rest are TGT-, TCT-, and CGT-types), suggesting that the lowered promoter activity of the *ST8SIA2* gene has been favored in adaptation to the environment causing psychosocial stress [9]. This is also supported by the finding that the TCT-type showing the second lowest promoter activity is identified as a promoter type not conferring a risk of schizophrenia in the population in which the frequency of the CGC-type is very low [8]. In addition to the CGC-type, the TCT-type can therefore be regarded as conferring tolerance of psychosocial stress.

Overcoming psychosocial stress by increasing the frequency of *ST8SIA2* promoter types conferring tolerance to such stress (i.e., tolerant types) at the population level would make collective brains bigger, associated with increases in the degree of social interaction and the population size. Therefore, it is expected that the *ST8SIA2* gene has been involved in the evolution of collective brains. To elucidate genetic tuning for the evolution of collective brains, we examined the change of the promoter-type composition of the *ST8SIA2* gene in AMH and archaic human (AH) lineages, and its functional consequence at the population level. This also provides insights into AMH-unique evolution of collective brains.

Materials and methods

DNA sequence data

Approximately 1 Mb of DNA sequences spanning the three promoter SNPs were retrieved from phase 3 of the 1000 Genomes Project database [16], covering 2,504 individuals (5,008 sequences). We classified them into five meta-populations: Africa (AFR), Europe (EUR), South Asia (SAS), East and Southeast Asia (EAS), and America (AMR) [216 for YRI (Yoruba in Ibadan, Nigeria), 198 for LWK (Luhya in Webuye, Kenya), 226 for GWD (Gambian in Western divisions in the Gambia), 170 for MSL (Mende in Sierra Leone), 198 for ESN (Esan in Nigeria), 122 for ASW (Americans of African ancestry in Southwest USA), and 192 for ACB (African Caribbeans in Barbados) in AFR (in total 1322); 198 for CEU (Utah residents with Northern and Western European ancestry), 214 for TSI (Toscani in Italy), 198 for FIN (Finnish in Finland), 182 for GBR (British in England and Scotland), and 214 for IBS (Iberian population in Spain) in EUR (in total 1006); 206 for GIH (Gujarati Indian from Houston, Texas, USA), 192 for PJI (Punjabi from Lahore, Pakistan), 172 for BEB (Bengali from Bangladesh), 204 for STU (Sri Lankan Tamil from the UK), and 204 for ITU (Indian Telugu from the UK) in SAS (in total 978); 206 for CHB (Han Chinese in Beijing, China), 208 for JPT (Japanese in Tokyo, Japan), 210 for CHS (Southern Han Chinese, China), 186 for CDX (Chinese Dai in Xishuangbanna, China), and 198 for KHV (Kinh in Ho Chi Minh City, Vietnam) in EAS (in total 1008); and 128 for MXL (Mexican ancestry in Los Angeles, California), 208 for PUR (Puerto Rican in Puerto Rico), 188 for CLM (Colombian in Medellin, Colombia), and 170 for PEL (Peruvian in Lima, Peru) in AMR (in total 694)] [9]. This sequence dataset is hereafter abbreviated as D_{1000} .

DNA sequences of AHs [three Neanderthals (Vindija, Altai, and Chagyrskaya) and one Denisovan] were obtained from the public database (<http://cdna.eva.mpg.de/neandertal/Vindija/>; <http://cdna.eva.mpg.de/neandertal/altai/>; <http://ftp.eva.mpg.de/neandertal/Chagyrskaya/VCF/>; <http://cdna.eva.mpg.de/neandertal/denisova/>). The chimpanzee sequence was obtained from the chimpanzee genome database (Clint_PTRv2/panTro6; <https://www.ncbi.nlm.nih.gov/genome>), and used as an out-group in tree constructions.

Schizophrenia-associated genetic loci

To examine the uniqueness of the promoter-type distribution of the *ST8SIA2* gene, we compared it with the allele frequencies of other schizophrenia-associated genetic loci identified from European and East Asian populations [17, 18]. A total of 308 index SNPs of which associations with schizophrenia risk were detected in meta-analyses using both European and East Asian populations were chosen because these are expected to link strongly to functional SNPs directly involved in schizophrenia risk.

Construction of haplotype sequences of archaic humans

The three promoter SNPs are located in an 18-kb region that is sandwiched in between recombination hot spots [9]. We therefore focused on this 18-kb region because the use of a longer region would cause ambiguity in the construction of archaic haplotypes by breaking the linkage of SNPs. The 18-kb region is homozygous in Altai Neanderthal (S1 Fig). Based on the linkage in the Altai Neanderthal sequence, we decided on the linkage of SNP sites in other AH sequences (S1 Fig).

Haplotype analysis

A phylogenetic tree of haplotype sequences was constructed using the Neighbor-Joining method [19]. Genetic distances were calculated with multiple hit corrections with MEGA X

software [20]. By following the work of Fujito et al. (2018) [9], homozygote tract lengths (HTLs) were obtained.

Detection of introgressed fragments

To detect introgressed fragments, we applied the Sprime method [21].

Ages of variants

Ages of variants were obtained from a website (<https://human.genome.dating/>; [22]).

Examination of archaic origin of promoter type not found in the AH database

To examine the archaic introgression of promoter types not found in the AH sequence database, we focused on the population that is considered as a direct descendant of the population in which the archaic introgression occurred, and used the information of the site differences between AMHs and AHs, the number of private haplotypes, recombination with other promoter types, and haplotypes shared with Africans. Based on these, we extracted candidates of descendant haplotypes of the archaic ancestry by following five steps.

Step 1: The number of site differences between AMHs and AHs (AMH–AH site differences) is obtained from comparison of the sequence between AMHs and AHs by ignoring recombinants.

Step 2: The pairwise comparison of site differences among haplotypes is performed to make a site difference matrix. Haplotypes showing the AMH–AH site differences in the matrix are selected for step 3 because these divergences correspond to the AMH–AH divergence.

Step 3: Phylogenetic trees of selected haplotypes are constructed. In the trees, it is expected that haplotypes can be divided into two groups corresponding to AMHs and AHs, if archaic introgression occurred.

Step 4: If haplotypes are distributed widely in AFR, such haplotypes and close relatives whose the number of site differences with them are less than the AMH–AH site differences are rejected.

Step 5: If there are variants shared only by the remaining haplotypes, the ages of such variants are examined. If not, the ages of variants uniquely found in the remaining haplotypes are examined. If the ages are not close to the time of archaic introgression, such haplotypes are rejected.

Promoter activity at the population level

Since the *ST8SIA2* gene is not an imprinting gene, it is assumed that the two promoter-type alleles contribute equally to the gene expression in individuals. Each promoter type has unique promoter activity (see also [9]), which was measured using the region whose sequence is the same as that from the most abundant haplotype of each promoter type (see Tables 1 and 2; [9]). Based on the relative difference in the promoter activity compared with that of the CGC-type (3.6, 2.9, and 2.1 for TGT-, CGT-, and TCT-types; [9]), we calculated the average of the relative differences (a_1 and a_2) of two promoter-type alleles as the individual promoter activity $[(a_1+a_2)/2]$, and then obtained the promoter activity at the population level (population

Table 1. Haplotypes* of the TGT-, TCT-, and CGC-types.

	Haplotype	AFR	EUR	SAS	EAS	AMR	Total
TGT	HG00132.1	12	26	78	79	69	264
	HG01271.0	76	2	0	0	3	81
	HG00728.0	62	1	0	1	2	66
	HG00553.0	59	0	0	0	4	63
	HG01886.1	55	0	0	0	1	56
	HG01241.0	48	1	0	0	3	52
	HG01110.1	45	0	0	0	2	47
	Others	255	3	31	53	42	384
	Total	612	33	109	133	126	1013
TCT	HG00097.1	70	417	188	279	192	1146
	HG00105.1	6	151	142	67	66	432
	HG01556.0	107	0	0	0	4	111
	HG01254.1	72	0	0	0	1	73
	HG01073.0	63	0	0	0	4	67
	Others	307	401	452	169	211	1540
		Total	625	969	782	515	478
CGC**	HG00356.0	6	1	67	304	77	455
	Others	5	1	9	45	5	65
		Total	11	2	76	349	82

* SNPs with minor allele frequency of 1% or over were used.

** Fujito et al. (2018) [9].

<https://doi.org/10.1371/journal.pone.0259897.t001>

Table 2. Distribution of CGT haplotypes.

Lineage	Haplotype	AFR	EUR	SAS	EAS	AMR	Total
CGT1	HG03105.0	1					1
	HG02095.1	1					1
	NA19908.0	4					4
	HG03572.1	1					1
	HG03311.0	8					8
	NA19116.1	1					1
	HG02433.0	8					8
	NA19468.0	1					1
	HG01073.1	12				1	13
	HG03100.1	3					3
	HG02095.0	1					1
	NA19472.1	1					1
	NA18909.0	1					1
	HG03437.1	1					1
	CGT2	NA19664.0					2
HG03667.1		2	1	5	8		16
HG03772.1				1			1
HG04042.1				4		3	7
HG02494.1				1		1	2
	HG01455.0					1	1
Total		46	1	11	8	8	74

<https://doi.org/10.1371/journal.pone.0259897.t002>

promoter activity; PPA) as the average of individual promoter activities in the population by

$$PPA = \frac{\sum \frac{a_1+a_2}{2}}{n}$$

where n is the number of individual samples in the population.

Results

Distribution of promoter types

In the present-day human population, over 99% of promoter types are the TGT-, TCT-, CGT-, and CGC-types ([9]; Fig 1; S1 Table). The TGT-type is the ancestral type [9] and is distributed globally. Its frequency is lower in non-AFR than in AFR [20.2% (1013/5008) in total, 46.3% (612/1322) in AFR, 3.3% (33/1006) in EUR, 11.1% (109/978) in SAS, 13.2% (133/1008) in EAS, and 18.2% (126/694) in AMR] (Fig 1; S1 Table; S2A Fig). In contrast to the TGT-type, the TCT-type is most prevalent in non-AFR, especially in EUR [67.3% (3369/5008) in total, 47.3% (625/1322) in AFR, 96.3% (969/1006) in EUR, 80.0% (782/978) in SAS, 51.1% (515/1008) in EAS, and 68.9% (478/694) in AMR] (Fig 1; S1 Table; S2B Fig). The CGT-type is the rarest among the four promoter types [1.5% (74/5008) in total, 3.5% (46/1322) in AFR, 0.1% (1/1006) in EUR, 1.1% (11/978) in SAS, 0.8% (8/1008) in EAS, and 1.2% (8/694) in AMR], and its frequency is lower outside of Africa (Fig 1; S1 Table; S2C Fig). The CGC-type is relatively common in SAS, EAS, and AMR [10.4% (520/5008) in total, 0.8% (11/1322) in AFR, 0.2% (2/1006) in EUR, 7.8% (76/978) in SAS, 34.6% (349/1008) in EAS, and 11.8% (82/694) in AMR] (Fig 1; S1 Table; S2D Fig). We previously reported that this distribution of the CGC-type has been influenced by positive selection [9].

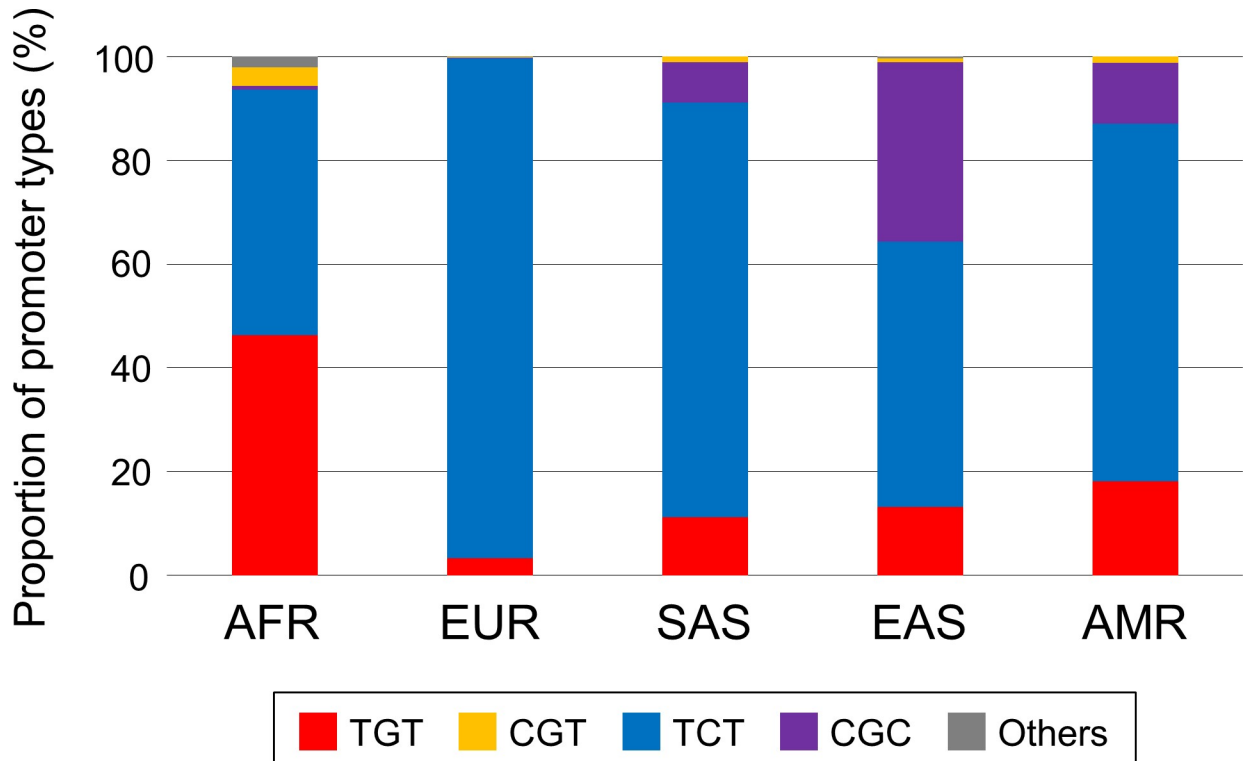


Fig 1. Promoter-type compositions in five meta-populations. For compositions in subpopulations of each meta-population, please see our previous report [9].

<https://doi.org/10.1371/journal.pone.0259897.g001>

Composition of haplotypes in meta-populations

Since the 18-kb region containing the three promoter SNPs is sandwiched between recombination hot spots [9], we focused on it to examine the haplotype composition using SNPs whose minor allele frequency (MAF) is >1%. Eighty-one TGT haplotypes were identified from D_{1000} (S2 Table). The number of haplotypes is 58, 7, 18, 7, and 25, for AFR, EUR, SAS, EAS, and AMR, respectively. In AFR, over half of TGT sequences are occupied by six haplotypes (HG01271.0, HG00728.0, HG00553.0, HG01886.1, HG01241.0, and HG01110.1) (Table 1; S3A Fig). In contrast, HG00132.1, which is minor in AFR (2%), occupies over half of haplotypes in every non-AFR meta-population (Table 1; S3A Fig).

The TCT-type sequences in D_{1000} are composed of 163 haplotypes (S3 Table). AFR, EUR, SAS, EAS, and AMR contain 62, 62, 58, 31, and 51 haplotypes, respectively. In AFR, HG01556.0 is most abundant and occupied around half of sequences with three haplotypes (HG00097.1, HG01254.1, and HG01073.0) (Table 1; S3B Fig). In contrast, HG00097.1 is the most prevalent type in every non-AFR meta-population and occupies around half of sequences with or without HG00105.1, which are rare in AFR (<1%), in each non-AFR meta-population (Table 1; S3B Fig).

Compared with the TGT- and TCT-types, the number of haplotypes is smaller in both CGC- and CGT-types (31 CGC haplotypes and 20 CGT haplotypes; [9]; Table 2; S4 Table). As for the CGC-type, AFR, EUR, SAS, EAS, and AMR contain 3, 2, 7, 21, and 5 haplotypes, respectively [9]. The excess in EAS can be explained by the increase of CGC sequences due to positive selection [9]. HG00356.0 is most prevalent in every meta-population ([9]; S3C Fig). As for the CGT-type, the number of haplotypes is 15, 1, 4, 1, and 5 in AFR, EUR, SAS, EAS, and AMR, respectively (S4 Table). Two haplotypes (HG01073.1 and HG03667.1) are shared by AFR and non-AFR. However, almost all HG01073.1 sequences are found in AFR (12/13), and the distribution of HG03667.1 is biased to non-AFR (14/16) (Table 2; S4 Table). In the remaining 18 CGT haplotypes, 13 haplotypes are found only in AFR and the others are identified only in non-AFR (Table 2; S4 Table). Based on these distributions, the CGT haplotypes can be divided into two groups: CGT1 (AFR unique haplotypes and HG01073.1) and CGT2 (non-AFR unique haplotypes and HG03667.1) (Table 2; S4 Fig; S4 Table).

High prevalence of the TCT-type in the ancestral population of non-AFR

The TCT-type is most prevalent in every non-AFR meta-population, and has an impact on the promoter-type composition of non-AFR. The positive selection on the CGC-type has influenced the promoter-type composition in Asian and American populations [9]. However, the very low frequency of the CGC-type (0.2%) in EUR indicates that the impact of such positive selection on the promoter-type composition can be ignored in EUR. Thus, EUR is regarded as a meta-population in which promoter-type composition has not been influenced by local events. Since the TCT-type can be divided from the others by rs3759915, the TCT-type and the others can be treated as biallelic. To determine the uniqueness of the high prevalence of the TCT-type outside of Africa, we therefore examined the change of allele frequency of 308 schizophrenia-associated SNPs between AFR and EUR by dividing the SNPs into frequency bins.

Based on the TCT-type frequency in AFR (47.3%), 45 schizophrenia-associated SNPs showing 40.1%–49.2% frequency were selected. By comparison with the frequencies of these SNPs in EUR, the average value of fold difference was obtained as 1.20 (1.02–1.38; 99% confidence interval of the normal distribution accepted by Kolmogorov–Smirnov test). Since the fold difference of the TCT-type frequency from AFR is 2.04 in EUR (Fig 2), the TCT-type frequency in EUR (96.3%) is much higher as expected from the frequency in AFR (48.4%–65.1%). Even

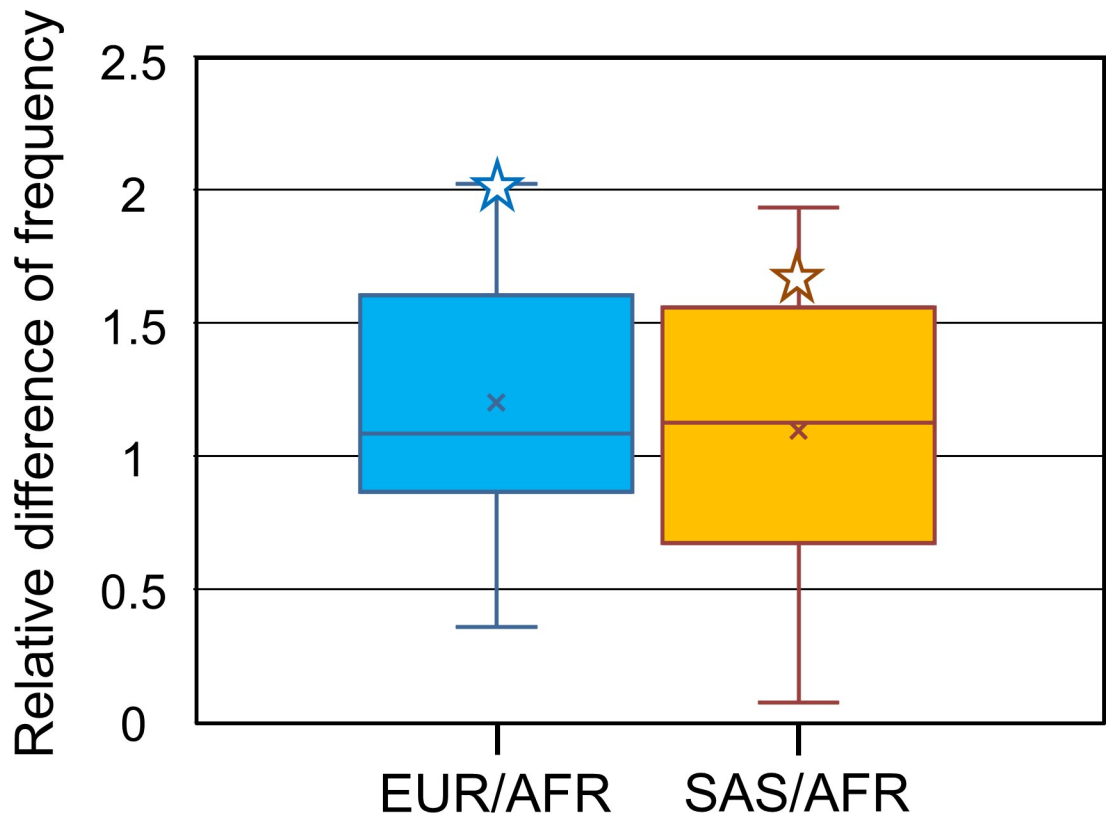


Fig 2. Change of frequencies in schizophrenia-associated genetic loci. Change of frequency is shown as difference of frequency in EUR and SAS relative to AFR. Based on the TCT-type frequency in AFR (47.3%), relative differences were calculated using the schizophrenia-associated genetic loci that show frequencies of 40.1%–49.2% in AFR. Relative difference of TCT-type is represented by an open star.

<https://doi.org/10.1371/journal.pone.0259897.g002>

though the TCT-type frequency is very high in EUR, we did not detect any signal of positive selection on the TCT-type in EUR using F_{ST} , extended haplotype homozygosity, or F_c ([9]; very small XP-EHH scores of 0.45 and 0.55 between EUR and AFR are not significant [23]). Furthermore, the nucleotide diversity (π) of the TCT-type in EUR (0.07%) is similar to that in AFR (0.09%). These findings indicate that the very high frequency of the TCT-type in EUR does not result from the recent local selection, consistent with the high prevalence of the TCT-type in the ancestral population of non-AFR. The median value of fold difference from AFR in EUR is 0.97 using 41 schizophrenia-associated SNPs showing 90.1%–100.0% frequency in EUR. Using this median value, the TCT-type frequency in the ancestral population of non-AFR can be inferred to be around 90%. Taken together, it is suggested that selected increase of the TCT-type was occurred in the ancestral population of non-AFR.

We also found similar results even if focus was placed on SAS. The average value of fold difference (SAS/AFR) is 1.10 (0.91–1.28; 99% confidence interval of the normal distribution accepted by Kolmogorov–Smirnov test) (Fig 2). The TCT-type frequency in SAS (80.0%) is much higher than expected from the frequency in AFR (43.0%–60.6%) because the fold difference of the TCT-type frequency from AFR is 1.69 in SAS. The frequency of the TGT-type is higher in SAS than in EUR. According to the phylogenetic tree of the TGT-type haplotypes from SAS, this higher frequency results from the recent diversification of one TGT-type lineage (S5 Fig). The timing of this diversification was obtained as 31 kyr [from ~11 kb of mean upstream homozygosity tract length (HTL)] using a published equation [9]. It is therefore

concluded that the increase of the TGT-type frequency occurred after the migration out of Africa, compatible with the high prevalence of the TCT-type in the ancestral population of non-AFR.

Unique TCT haplotype composition in LWK

In addition to the high prevalence of the TCT-type, the TCT haplotype composition in non-AFR is different from that in AFR (Table 1; S3B Fig; S3 Table) as mentioned above. HG00097.1 is just one of the common haplotypes in AFR, but is markedly prevalent in non-AFR. In contrast, HG01556.0 is the most prevalent type in AFR, but very rare in non-AFR. In AFR, the frequencies of HG00097.1 in ACB (12.6%), ASW (23.4%), and LWK (18.3%) are higher than the average value of AFR (11.2%). Among these three subpopulations, ACB can be grouped with others (ESN, GWD, MSL, and YRI) because of the lack of significance of the difference in frequency between them ($P > 0.01$). As for ASW, admixture with Europeans occurred during its history [24], and the high frequency of HG00097.1 would result from the introgression by this admixture. By excluding ASW, AFR sub-populations can thus be divided into two groups: LWK and the others. LWK shows that the HG00097.1 frequency (18.3%) is not significantly different from that (24.0%) in SAS, the non-AFR meta-population showing the lowest TCT-type frequency ($P = 0.258$). In addition to this high frequency of HG00097.1, the frequency of HG00105.1, the second major haplotype in non-AFR, is highest in LWK (4.2%). Therefore, it is concluded that LWK is most closely related to non-AFR in AFR subpopulations.

To examine the effect of recent admixture on the close relationship between LWK and non-AFR, we focused on HG01556.0, the most prevalent haplotype in AFR. ASW shows the lowest frequency (7.8%) in AFR subpopulations. Since ASW has admixed with Europeans, this lowest frequency would result from the reduction of HG01556.0 by the admixture because of very low frequency of HG01556.0 in non-AFR (0.8% in AMR). In contrast to ASW, LWK shows the highest frequency (22.5%) in AFR subpopulations (S3 Table). This significant difference between ASW and LWK ($P < 0.01$) indicates that LWK has not undergone significant admixture with non-AFR.

In AFR, LWK shows the highest frequencies of both HG00097.1, the most prevalent type in non-AFR, and HG01556.0, the most prevalent type in AFR. In line with this, this hybrid composition in LWK reflects the TCT haplotype composition of the common ancestor of non-AFR and AFR. Interestingly, the TGT-type, the ancestral promoter type, has the highest frequency in LWK (58.1%; see S2 Table). Taking these findings together, it is likely that the promoter-type composition of LWK is close to those of AMH ancestors more than other subpopulations.

Promoter types in AHs

We extracted genomic sequences from three Neanderthals (Vindija, Altai, and Chagyrskaya) and one Denisovan, and found that they possess only CGT at the three promoter SNPs (S1 and S6 Figs). These AH samples came from Croatia (Vindija Neanderthal) and Siberia (Altai Neanderthal, Chagyrskaya Neanderthal, and Denisovan), and thus are from locations distant from each other [25, 26]. This shows that the CGT haplotype was widely dispersed in AH species. Only the CGT-type was identified from the AH samples, indicating that the CGT/CGT homozygote was common in AH population. Based on the identification of only CGT/CGT homozygosity in all four AH individuals, the CGT-type frequency in AH populations was estimated as over 71.7% (95% credible interval from the β -distribution). To examine the existence of non-CGT-types in AH populations, we detected recombination tracts because their

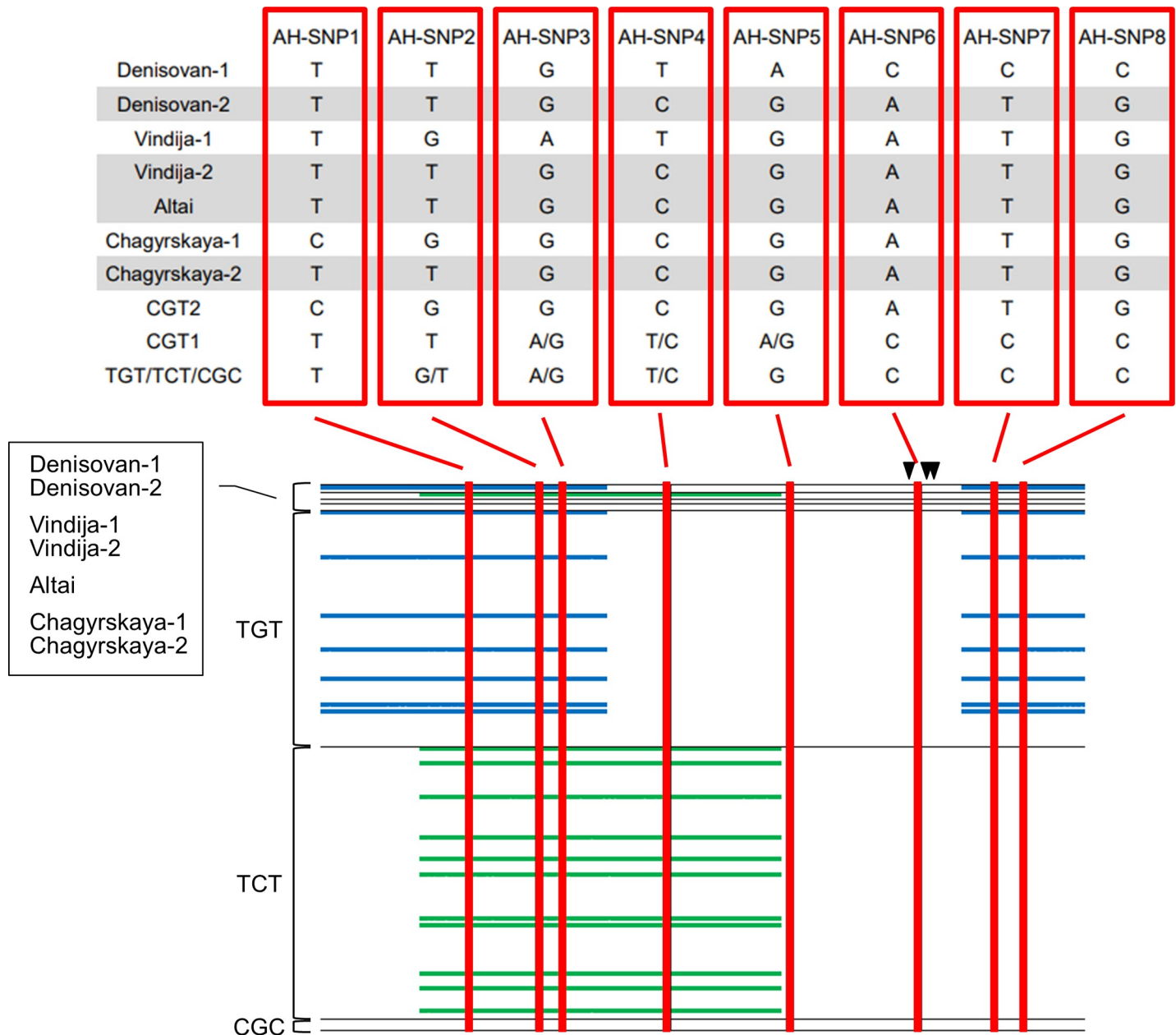


Fig 3. Comparison between AH haplotypes and AFR haplotypes. SNPs with minor allele frequency of 0.5% or over in the total AMH population were used, with the exception of AH-SNP5 (rs144258052; 0.34% in total). As for non-CGT-types, all AFR haplotypes were used. AH haplotypes of eight AH-SNP sites are represented in the upper panel. In the lower panel, the longest identical tracts detected between AH sequences and non-CGT haplotypes are represented by blue bars (Denisovan-1) and green bars (Vindija-1). Arrowheads represent positions of three promoter SNPs. Uncertain sites ("N") in AH sequences were ignored in the comparison.

<https://doi.org/10.1371/journal.pone.0259897.g003>

presence reflects the combination of promoter types. In this analysis, we used SNPs with MAF of 0.5% or more because some AH SNPs have an MAF in AMHs of less than 1%.

Among AH SNPs, eight SNP sites [rs145606077 (AH-SNP1), rs13379489 (AH-SNP2), rs4777968 (AH-SNP3), rs4570781 (AH-SNP4), rs144258052 (AH-SNP5), rs150882587 (AH-SNP6), rs192224395 (AH-SNP7), and rs373352748 (AH-SNP8)] are also polymorphic in AMHs (see Fig 3 and S6 Fig). The Altai haplotype of these SNPs, T-T-G-C-G-A-T-G, is shared with Denisovan-2, Vindija-2, and Chagyrskaya-2. Denisovan-1, Vindija-1, and Chagyrskaya-1

are different from Altai haplotype at five, three, and two AH-SNP sites, respectively. Chagurskaya-1 and Vindija-1 have one unique allele, “C” at AH-SNP1 and “A” at AH-SNP3, respectively. In contrast, “A,” “C,” “C,” and “C” at AH-SNP5–8 are unique to Denisovan-1. To examine the recombination with non-CGT-type in the emergence of these unique alleles, we compared the lengths of tracts identical to Denisovan-1, Vindija-1, and Chagurskaya-1 among non-CGT haplotypes (Fig 3; S6 Fig). Uncertain sites (“N”) in AH sequences were ignored in this comparison.

The tract identical to Denisovan-1 in the 3′ end region containing AH-SNP7 and AH-SNP8 was longest in some TGT haplotypes (Fig 3; S6 Fig), which suggests that “C”s in AH-SNP7 and AH-SNP8 were derived from a TGT-type by recombination. In addition, some of these TGT haplotypes show a tract identical to Denisovan-1 in the 5′ end region containing AH-SNP1–3. These findings indicate that the ~8-kb part containing the three promoter SNPs is sandwiched between the TGT-type sequences in the 18-kb region of Denisovan-1. It is therefore concluded that Denisovan-1 was established by recombination in which the ~8-kb middle part of the TGT-type was exchanged for CGT-type. Using a published equation [27] with the recombination rate of the 18-kb region (3.3 cM/Mb; [9]), the Neanderthal–Denisovan split time of 450 kyr [25], introgression time of 50 kyr, and generation time of 29 years, the length of the part containing the three promoter SNPs rules out that the Denisovan-1 sequence was derived from the AH common ancestor ($P = 0.006$). In other words, the recombination between the CGT- and TGT-types occurred in the AH population.

The tract identical to Vindija-1 in the region surrounding AH-SNP3 was longest in some TCT haplotypes (Fig 3; S6 Fig), which suggests that “A” at AH-SNP3 was derived from a TCT-type by recombination. The length of the identical tract is ~9 kb, which refutes that the Vindija-1 sequence is derived from the AH common ancestor ($P = 0.002$). In addition to the TGT-type, the TCT-type anticipated the recombination with the CGT-type in the AH population.

According to the position of the recombination tracts, Denisovan-1 has the TGT part at both ends of the 18-kb region (i.e., TGT-CGT-TGT), and Vindija-1 contains the TCT part in the middle (i.e., CGT-TCT-CGT). Since three CGT-types, two TGT-types (or recombinants with the TGT-type), and one TCT-type (or recombinant with the TCT-type) are needed to make these haplotype structures (S7 Fig), the detected structural difference suggests that the frequency of the TGT-type was higher than that of the TCT-type in the AH population.

Archaic introgression of the CGT-type from AHs into AMHs

To examine the phylogenetic relationship of the CGT haplotypes of AMHs and AHs, we constructed a phylogenetic tree. Interestingly, the obtained tree shows that all CGT2 haplotypes are clustered with AH sequences after deep divergence from the CGT1 haplotypes (Fig 4). To explain the topology of this tree, two scenarios are proposed. One is ancestral polymorphism, meaning that the topology of the obtained phylogenetic tree already existed in the common ancestor of AMHs and AHs (incomplete lineage sorting). The other is archaic introgression after the split of AMHs and AHs [25, 28, 29]. The mean value of HTLs was 28,042 bp for the CGT2 sequences. Using a published equation [27] and conditions [30] with the exception of the recombination rate of the 18-kb region (3.3 cM/Mb; [9]), this mean HTL rules out the former scenario ($P = 9.7 \times 10^{-14}$). Unlike AH sequences, all of the CGT2 sequences form a single cluster in the phylogenetic tree after the divergence from AH sequences (Fig 4), which indicates the archaic introgression into AMHs from AHs.

Based on the number of accumulated mutations within the CGT2 cluster, the maximum likelihood estimation [31] shows that the CGT2 lineage began to diversify into sub-lineages 46 kyr ago (33–59 kyr ago; mutation rate of 0.5×10^{-9} per site per year; [32]). Furthermore,

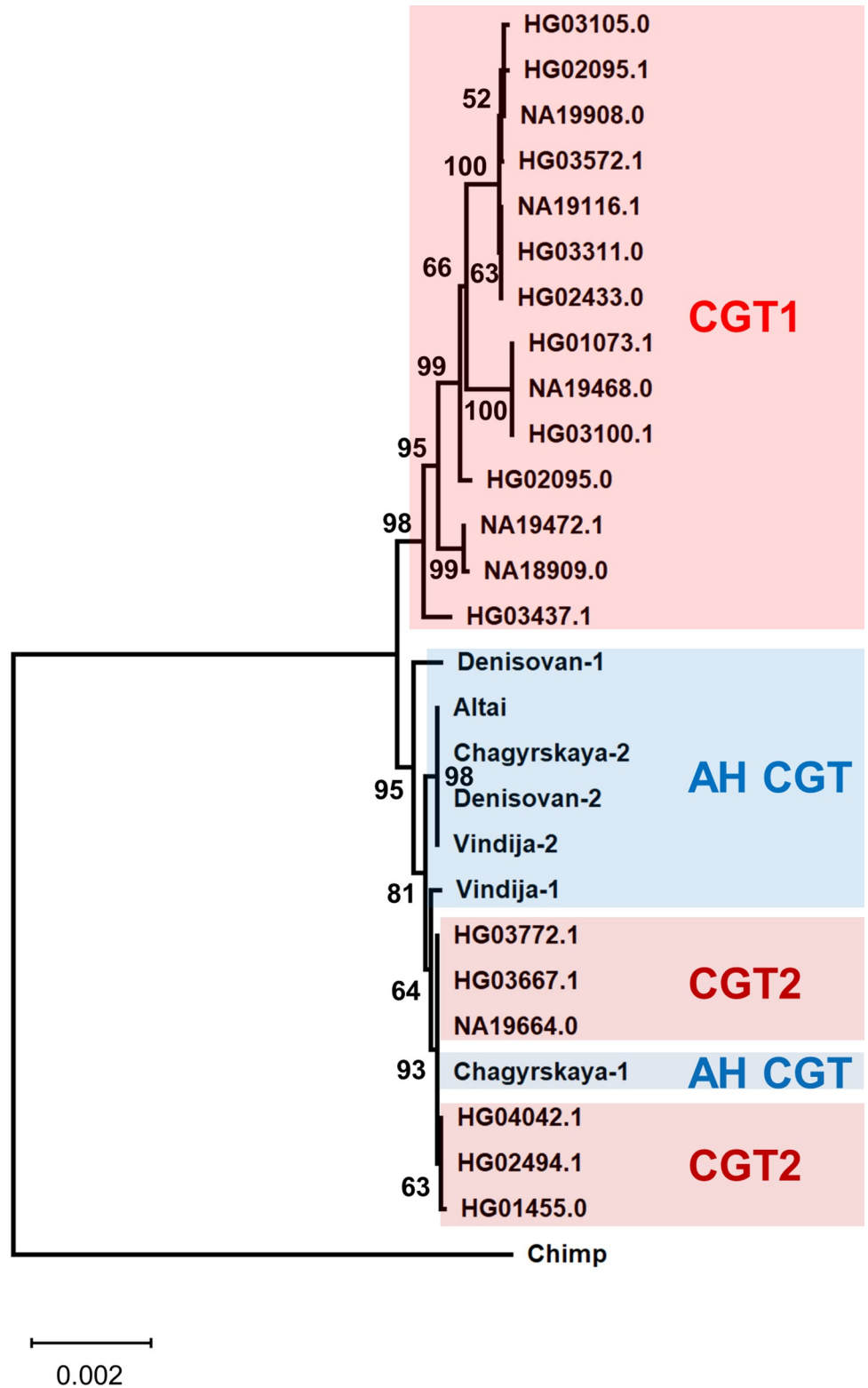


Fig 4. Phylogenetic relationship of the CGT-type sequences from AMHs and AHs. Phylogenetic tree is constructed using the Neighbor-Joining method with Jukes–Cantor correction. Bootstrap values of more than 50% from 1,000 replications are shown on the tree branches.

<https://doi.org/10.1371/journal.pone.0259897.g004>

derived alleles of six SNP sites (rs139630787, rs186699149, rs150882587, rs138393846, rs192224395, and rs373352748) are shared only between AH and CGT2 sequences (see [S1](#) and [S6](#) Figs), and the ages of these variants were obtained as 71, 69, 66, 63, 63, and 82 kyr, respectively (joint clock; combined dataset; assuming 29 years per generation; [22]). These estimated times are much younger than the split time of AMHs and AHs (550–765 kyr) [25] and consistent with archaic introgression.

We detected 29 CGT2 sequences in D_{1000} (Table 2). Sixteen sequences are classified into the HG03667.1 haplotype found widely in AFR, EUR, EAS, and SAS (Table 2). The HG03667.1 haplotype is also an ancestral type because it is most closely related to the AH sequences (Fig 4). The second major haplotype to which seven sequences belong is HG04042.1 found in SAS and AMR, and other haplotypes contain only one or two sequences from SAS and/or AMR (Table 2). SAS contains nearly half of the sequences (11 sequences) that belong to four haplotypes, including the ancestral type. It is therefore considered that the archaic introgression of the CGT-type occurred in an ancestral population of SAS.

Archaic introgression from Neanderthals to Denisovans

We also examined the archaic introgression without determining the allelic linkage of AH sequences. To detect introgressed segments in the 978 SAS genomes, we applied the Sprime method [21]. Similar to the S^* approach [33], it is reference-free and therefore advantageous when introgressed archaics are unknown. Chen et al. (2020) [34] pointed out false-negatives of these approaches where a modern reference (African) population is not completely free from archaic segments by introgression and/or gene flow. Among 116 possible introgressed segments identified in chromosome 15, one is about 250 kb long and spans the *ST8SIA2* gene. Curiously, however, the match proportion in this particular segment is ~83% both to the Altai Neanderthal genome [25] and to the Denisovan genome [28], suggesting introgression not only between modern humans and an archaic but also between archaics themselves. Examining the unphased genotype data of the Denisovan and Neanderthal genomes [26, 35], we found heterogeneous allele sharing and/or introgression in the 250 kb segment spanning the *ST8SIA2* gene. Focusing on one subsegment (22 kb long) ranging from 92913845 to 92936297 that includes the promoter SNPs, we computed the number of nucleotide differences (distance) in all pairwise comparisons of the four archaic diploid and one CGT2 (HG03667.1; Table 2; S4 Table) haploid sequences. The Altai sequence turned out to be homozygous for the entire subsegment, so that the distance at a nucleotide site was defined as 1 and 1/2 where an Altai Neanderthal site differs from a homozygous and heterozygous site, respectively, of a diploid in comparison. Similarly, we computed the distance between two unphased diploid sequences. The distance matrix and the resulting model of the relatedness of the segments are given in Fig 5. They suggest that one of the Denisovan segments was replaced by a Neanderthal one. Taken together with widespread admixture between Altai Neanderthals and Denisovans [36], multiple interspecies gene flows in the *ST8SIA2* gene make it difficult to know which AH provided archaic ancestry of CGT2 in the admixture with AMHs. It is noteworthy that the topology shown in Fig 5 is compatible with that in Fig 4. This suggests that our linkage determination in constructing the AH haplotype is reasonable.

No detection of the archaic ancestry of non-CGT-type

Multiple interspecies introgression in the *ST8SIA2* gene raised a possibility that introgression of non-CTG-types from archaic humans was also occurred in AMHs. Thus, we tried to detect the archaic introgression of non-CGT-types by focusing on SAS sequences. This provides further insight into the promoter-type composition in AH population. According to the

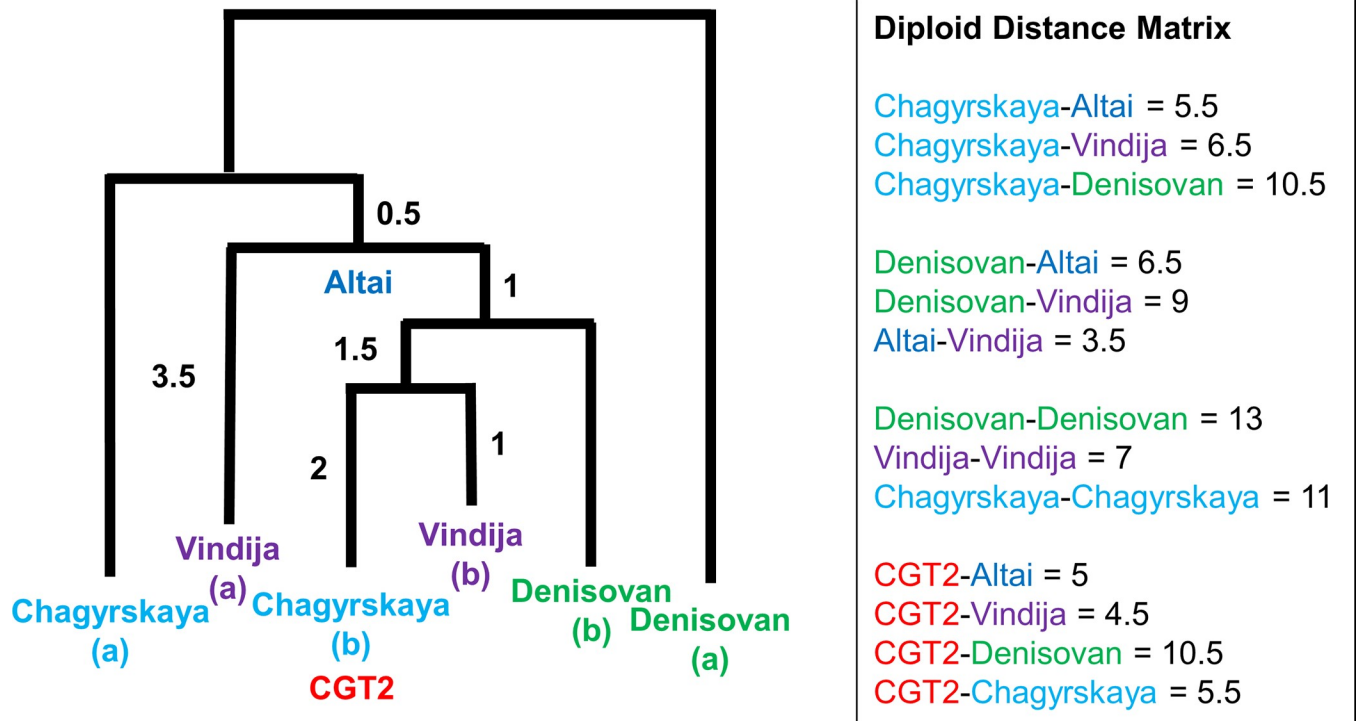


Fig 5. Diploid distance matrix and resulting tree of 22-kb region encompassing the three promoter SNPs. The tree topology was obtained by realizing compatibility of diploid distances among sequences.

<https://doi.org/10.1371/journal.pone.0259897.g005>

phylogenetic tree of the CGT-type (Fig 4), the CGT1 and CGT2 lineages correspond to the AMH and AH lineages. The numbers of different sites between the CGT1 and CGT2 haplotypes can be regarded as those expected in the AMH-AH divergence. Since parts of the HG03437.1, NA18909.0, and NA19472.1 sequences were influenced by recombination with the TGT- and TCT-types (S6 Fig), we counted the number of different sites between the CGT1 and CGT2 haplotypes by eliminating these three haplotypes. The obtained numbers of different sites were 12–15 (S5 Table), which can be regarded as the AMH–AH site differences. This is consistent with the site differences (10–14 site differences) expected from the divergence time between AMHs and AHs (550–765 kyr; [25]).

Among the TGT haplotypes from SAS, only one haplotype, HG03779.1, represents 12–15 site differences from 15 haplotypes (S8A Fig; S6 Table). These 16 haplotypes were selected. However, a member of the 15 haplotypes, HG00132.1, is found widely in AFR (S8A Fig; S2 Table), and estimated ages of variants unique to HG03779.1 were much older than the time of the archaic introgression [35, 37, 38] [754 kyr (combined) of rs10852173, 828 kyr (TGP) of rs56027313, 123 kyr (combined) of rs55649570, 1010 kyr (combined) of rs13379489, and 620 kyr (combined) of rs17646351; 29 years of generation time; [22]]. Thus, it is concluded that there is no TGT haplotype derived from archaic ancestry.

Among the TCT haplotypes from SAS, 12–15 site differences from other haplotypes are found multiply in two haplotypes, HG02127.1 and HG03729.1 (S7 Table), and 12 haplotypes showed 12–15 site differences from both HG02127.1 and HG03729.1. Thus, we selected these 14 haplotypes. In the tree (S8B Fig), the 12 haplotypes form a single cluster and are distantly related to HG02127.1 and HG03729.1, which are deeply divergent from one other. Seven of the 12 haplotypes are found in AFR (S8B Fig; S3 Table). In addition, HG02127.1 and HG03729.1 showed that the estimated ages of unique variants were much older than the time

of archaic introgression [35, 37, 38] [509 kyr (combined) of rs2129796, 1360 kyr (combined) of rs6416575, 454 kyr (combined) of rs12913269, 1321 kyr (combined) rs4777968, 828 kyr (combined) of rs7175280, 1322 kyr (combined) of rs4583192, 1343 kyr (combined) of rs8035799, 1304 kyr (combined) of rs4570781, 1432 (TGP) of rs4238483, 669 kyr (TGP) of rs7172135, 617 kyr (combined) of rs7171140, 1112 kyr (combined) of rs7173401, 1362 kyr (combined) of rs6496930, 890 kyr (combined) of rs3759917, 880 kyr (combined) of rs3784746, and 643 kyr (combined) of rs3784745; 29 years of generation time; [22]]. Thus, the selected haplotypes were not derived from the archaic ancestry.

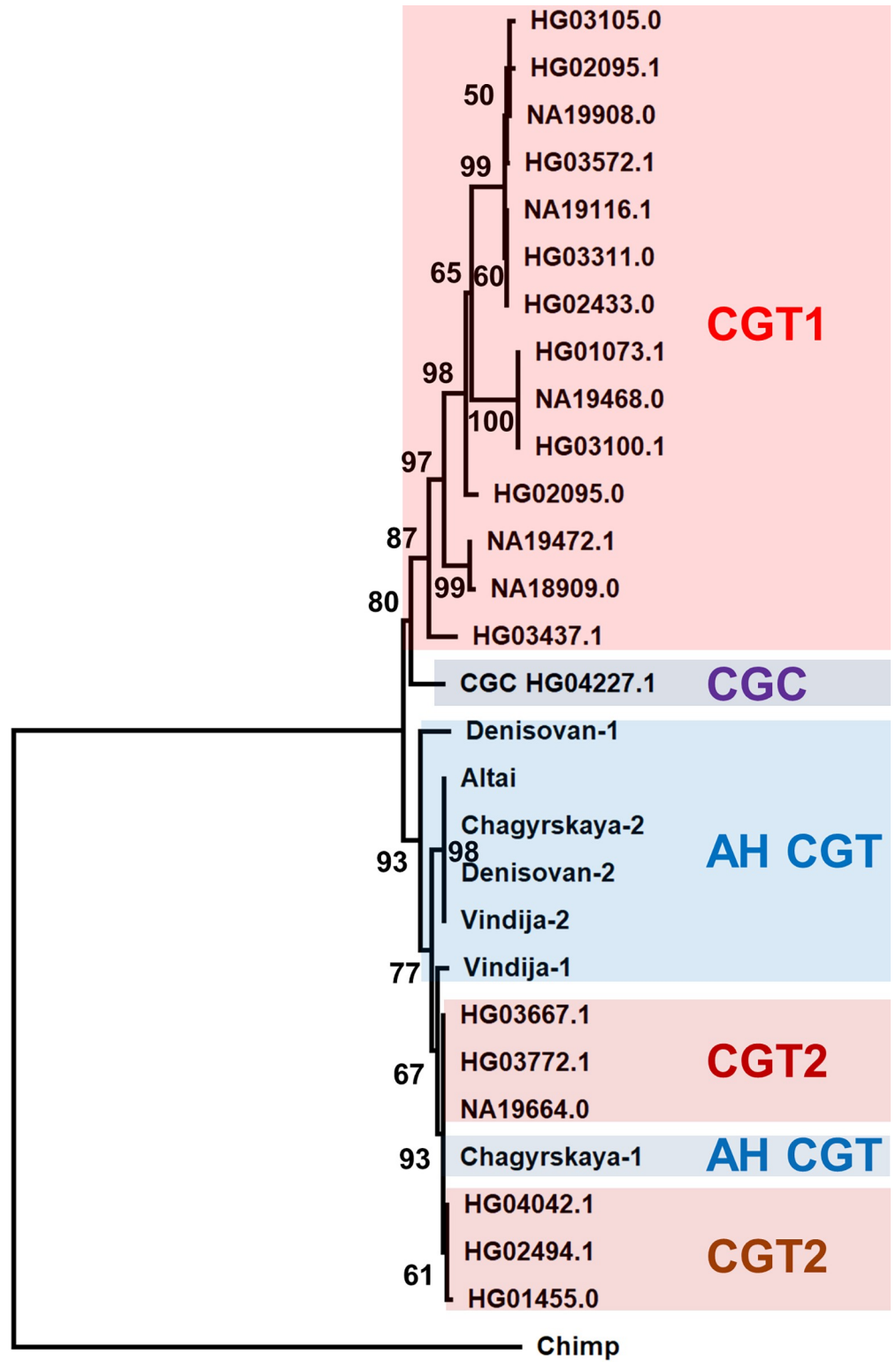
The CGC haplotypes from SAS did not show any 12–15 site differences in the pairwise comparison (S8 Table). According to our previous work, the CGC-type emerged in the AMH lineage [9]. This is supported by the tree topology in which the CGC-type sequence belonging to the most prevalent haplotype is closely related to the CGT1 sequences because the archaic origin of the CGT2 sequences suggests that the CGT1 and CGT2 lineages correspond to the AMH and AH lineage, respectively (Figs 4 and 6). The most prevalent CGC haplotype is widely distributed in AFR sub-populations, and Mbuti Pygmy, a deeply divergent population in the AMH lineage, has a CGC haplotype [9]. These findings support the AMH origin of the CGC-type. Taking these results together, it is concluded that the absence of haplotypes showing the AMH–AH site difference in the CGC haplotypes from SAS results from the AMH origin of the CGC-type.

Since the frequency of the CGT2 sequences is very low in AMHs (0.6% in total), the detection of only CGT-type introgression indicates that the impact of the archaic introgression on the promoter-type composition in AMHs is very small. Considering that the chance of introgression of promoter types would have been decided by their frequencies in AH populations, it is concluded that the frequency of the CGT-type was very high in AH populations, consistent with the inferred CGT-type frequency (>71.7%) in AH populations. The promoter-type composition is markedly different between AMHs and AHs (S9 Fig).

Distribution of promoter activities

We previously measured the promoter activity of each promoter type, and found that the CGC-type shows significantly lower activity than the others [9]. This comparison also indicates that promoter activity is significantly different among promoter types ($P < 0.005$; see also [9]), which indicates that each promoter type has unique promoter activity. Since the composition of promoter types is diversified among meta-populations, this raised the possibility that the promoter activity differs at the population level. To examine this possibility, we established a concept representing promoter activity at the population level (population promoter activity: PPA) as a population phenotype by using relative promoter activities measured experimentally [9]. As shown in Fig 7A, EAS shows significantly lower PPA than the others ($P < 0.01$; Kolmogorov–Smirnov test). This is consistent with the highest frequency of the CGC-type in EAS ([9]; S2D Fig). Outside of Africa, a significant difference was also found among other meta-populations ($P < 0.01$; Fig 7A). Furthermore, every non-AFR (EUR, SAS, EAS, and AMR) shows significantly lower PPA than AFR ($P < 0.01$; Fig 7A). This is compatible with the highest frequency of the TCT-type showing the second lowest promoter activity in every non-AFR meta-population.

LWK shows the highest PPA (Fig 7B). This is compatible with the highest frequency (58.1%) of the TGT-type, which shows the highest promoter activity, in LWK. As mentioned previously, LWK has retained the ancestral promoter-type composition more than other sub-populations. Thus, the PPA has become lower even within Africa. Taking these findings together with the lower PPA in non-AFR than in AFR, we can see the trend toward lowered PPA in the spread of AMHs (Fig 7B).



0.002

Fig 6. Tree of the CGT-type sequences with a CGC-type sequence belonging to the most prevalent CGC haplotype. The phylogenetic trees were constructed by the Neighbor-Joining method with Jukes–Cantor correction. Bootstrap values of more than 50% from 1,000 replications are shown on the tree branches.

<https://doi.org/10.1371/journal.pone.0259897.g006>

PPA of AHs

As mentioned above, AH populations had the CGT-, TGT-, and TCT-types. Their CGT-type frequency was inferred to be over 71.7%, and the frequency of the TGT-type was assumed to be higher than that of the TCT-type. Based on these findings, the promoter-type combination of individuals was generated by random mating with no difference in the promoter-type frequency between males and females, and used for PPA calculation (S9 Table). All calculated PPAs were slightly higher than that of AFR (1.01- to 1.08-fold difference; S9 Table). Furthermore, even if we looked closely at the PPAs of AFR subpopulations, the inferred lowest PPA of AHs is higher than those of AFR subpopulations, with the exceptions of LWK and ESN, which show the top two PPAs (Fig 7B; S9 Table).

Although AHs lived outside of Africa, their PPA was inferred to be higher than those of non-AFR subpopulations (see Fig 7B). If the PPA of LWK is assumed to be close to that of the ancestor of AMHs and AHs, it is considered that the out-of-Africa migration did not significantly change the PPA at the origin of AHs. It is therefore likely that the trend toward lower PPA is unique to AMHs.

Discussion

The numbers of heterozygous sites in Denisovan, Vindija, and Chagyrskaya sequences (12, 8, and 6 sites; S1 Fig) were greater than those (~2 sites) detected in the whole-genome analysis [25]. As mentioned above, recombination with non-CGT-type was detected in Denisovan-1 and Vindija-1 (Fig 3; S6 Fig). This is consistent with the higher recombination rate (3.3 cM/Mb; [9]) in the 18-kb region as compared with the standard value [32]. In addition, the introgression from Neanderthals influenced Denisovan heterozygosity in the 18-kb region, consistent with the largest number of heterozygous sites in Denisovan sequences. A high recombination rate and introgression may be involved in the unexpectedly large number of heterozygous sites.

We did not detect any non-CGT haplotype derived from archaic ancestry. In cases where a promoter type that emerged uniquely in the AH lineage was introgressed, our method is useless because the AMH–AH site difference is used to identify candidates. In addition, if only sequences that underwent recombination with different promoter types in the AH lineage were introgressed, our method cannot detect haplotypes derived from such archaic ancestry as candidates because such haplotypes do not show the AMH–AH site difference with any haplotype. These are limitations of our method. As for the former, we did not encounter such a case in this study. However, even if it occurred, the distribution inside of Africa may be helpful to detect it because a wide distribution of haplotypes from the archaic ancestry cannot be expected in Africans. As for the latter, the use of unrecombined region would be important to avoid encountering such a case. However, the use of many haplotypes makes the length of an unrecombined region short, resulting in a low resolution. In this study, we used the 18-kb region containing no recombination hotspots. We therefore think that the resolution of our method was well maintained by the minimizing the effect of recombination in this study.

We found a difference in PPA among meta-populations (Fig 7A). Since each promoter type has unique promoter activity (see also [9]), this difference should depend on the composition of promoter types. The CGC-type has the lowest promoter activity, and its frequency is highest in EAS [9] (Fig 1; S1 Table). It is therefore considered that the lowest PPA in EAS results from

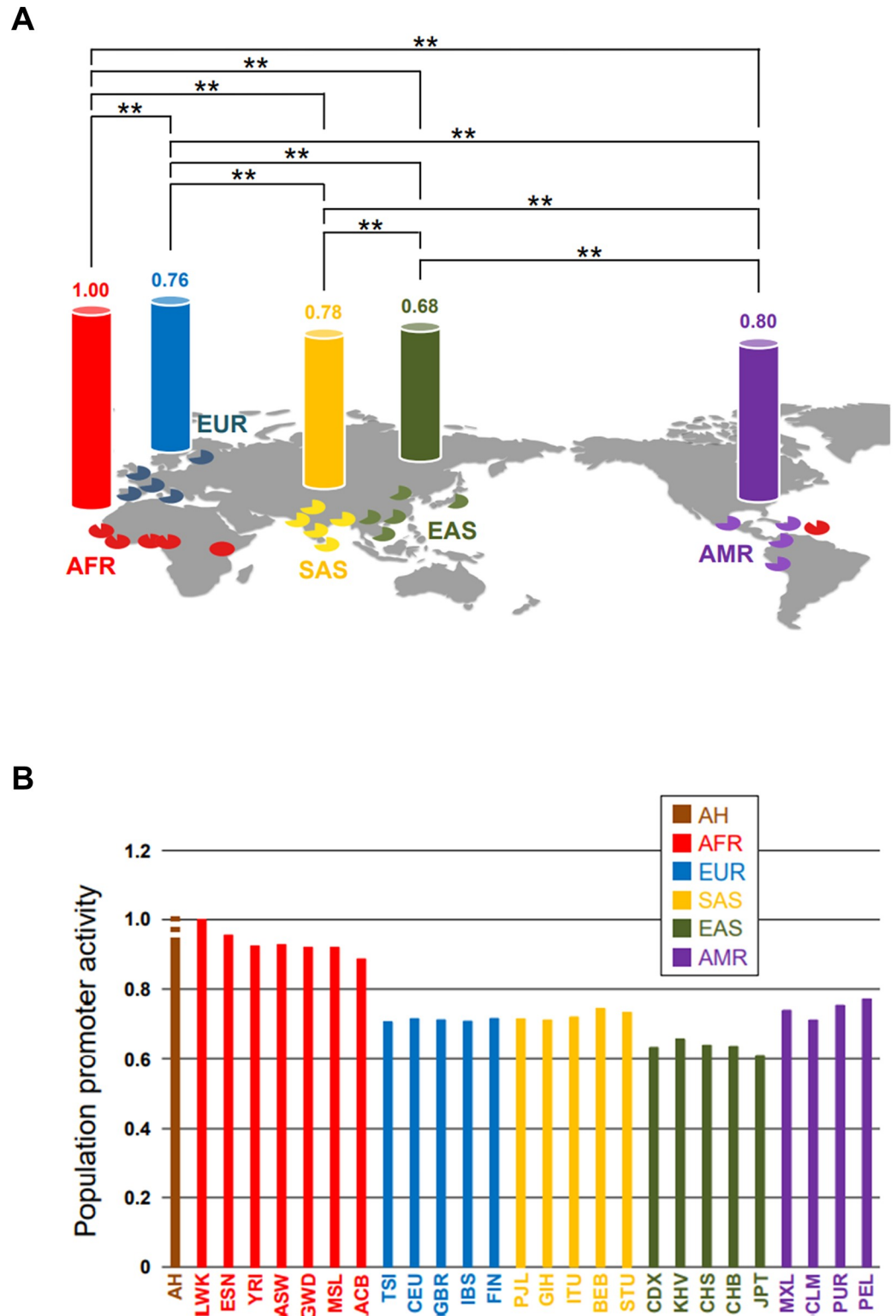


Fig 7. Distribution of PPA. (A) PPA value of each meta-population is represented by relative activity as compared with AFR. Pie charts represent relative activities as compared with the highest activity (LWK) among subpopulations. ** $P < 0.01$. (B) Relative difference of PPA as compared with LWK. For AH, the inferred lowest PPA is represented by a filled line, and the top of the additional dashed line shows the highest inferred PPA value.

<https://doi.org/10.1371/journal.pone.0259897.g007>

the increase in the frequency of the CGC-type under selection. This indicates that the selective increase of one functional genotype can influence a biological trend at the population level (population phenotype), even if the selected allele is still segregating at intermediate frequency.

The CGC-type frequency has been influenced by the recent positive selection in SAS, EAS, and AMR, but is very low in EUR and AFR. This suggests that the frequency of the CGC-type was very low in the ancestral population of non-AFR. However, the TCT-type shows the highest frequency in every non-AFR meta-population (Fig 1; S1 Table). Furthermore, we found that the TCT-type was highly prevalent in the ancestral population of non-AFR (~90%). Since the TCT-type shows the lowest promoter activity with the exception of the CGC-type [9], it is reasonable that the lowered PPA was already established in the ancestral population of non-AFR, which is compatible with the lowered PPA shared by every non-AFR population.

Based on the TCT haplotype composition (Table 1; S3B Fig; S3 Table), it is likely that the ancestral population of non-AFR (high frequency of HG00097.1 and low frequency of HG01556.0) emerged directly from the ancestor represented by LWK (high frequencies of both HG00097.1 and HG01556.0), accompanying the reduction of HG01556.0 frequency. While the TCT-type frequency is lowest in LWK (35.9%), it was inferred to be very high in the ancestral population of non-AFR (~90%). Therefore, the increase of the TCT-type would have occurred in the ancestral population of non-AFR, suggesting positive selection on the TCT-type in the non-AFR ancestor. The similar π values of the TCT-type between AFR and EUR indicate the concentration of the TCT-type in the non-AFR ancestor, and the out-of-Africa migration would have acted as a filter for this concentration.

The selective increase of the TCT-type in the ancestral population of non-AFR was suggested in the comparison with 308 schizophrenia-associated index SNPs (Fig 2). The linkage disequilibrium between these index SNPs and functional SNPs directly involved in schizophrenia risk is not guaranteed to be strong in every population. However, since the index SNPs were identified in meta-analyses using European populations [17, 18], the comparison using allele frequencies in EUR (Fig 2) may reflect the change of frequencies of functional alleles. This implies uniqueness of the *ST8SIA2* gene in the schizophrenia-associated genes.

Since schizophrenia develops through the interaction between genetic risk factors and environmental risk factors, an increase of the non-risk type of brain-related genes can be regarded as a result of adaptation to psychosocial stress [9, 11–15]. In terms of stress resistance, risk and non-risk types therefore confer sensitivity and tolerance to psychosocial stress and are regarded as sensitive and tolerant types, respectively. In addition to the CGC-type showing the lowest promoter activity, the TCT-type showing the second lowest promoter activity has been identified as a non-risk type in populations in which the frequency of the CGC-type is very low [8]. It is therefore considered that the lowering of PPA resulting from the increase of the frequency of tolerant types, namely, the TCT- and CGC-types, occurred due to selection as an adaptation to emerging stressful social environments. The lower PPA in non-AFR than in AFR (see Fig 7) suggests that migration was a factor raising psychosocial stress, consistent with psychosocial stress in migration (immigration) having been identified as an environmental risk factor [13, 15, 39–41]. In the case of the spread of AMHs, it is considered that such stress was caused by the close encounters with earlier migrants including AHs. Therefore, it is concluded that the facilitation of social interaction (cooperation) by overcoming psychosocial stress has been advantageous in the spread of AMHs (S10 Fig).

Societies and social networks act as collective brains in which individuals are regarded as neurons and drive cumulative cultural evolution [2, 3]. It is likely that overcoming psychosocial stress in close encounters with others makes collective brains bigger by increasing the degree of social interaction and the population size, and accelerates cumulative cultural evolution. Therefore, the trend toward lower PPA in the *ST8SIA2* gene would be an example of

genetic tuning needed for the evolution of collective brains. This is also supported by significant negative correlation between PPA and the effective population size change of each sub-population during the past 60 kyr [$r = -0.6$ ($P = 0.001$) using values from the smoothed PSMC curve [16]; $r = -0.7$ ($P = 6.2 \times 10^{-5}$) using value from the unsmoothed PSMC curve [16]; S11 Fig].

Since the PPA of AHs was inferred to be high (Fig 7B), AHs might have been under stress when encountering AMHs. Because of the near monomorphism of the sensitive type (over 71.7% of the CGT-type) in AHs (S9 Fig), it is assumed that AHs could not achieve a rapid change in their high PPA. Considering that AHs became extinct after the encounter with AMHs, their stable high PPA might have been disadvantageous for surviving the encounter with AMHs. Furthermore, the stable high PPA in AHs, resulting from near monomorphism of the sensitive type, also raises the possibility that AHs had stable small collective brains because they could not undergo increases of the degree of social interaction and the population size. This is consistent with the smaller population size in AHs than in AMHs [25] and supports Henrich's hypothesis that AMHs have bigger collective brains than AHs [2].

Each meta-population represents a unique PPA (Fig 7A). Diamond (1997) argued that the rate of spread of food production and technological innovations was higher in the Eurasian continent than in the African and American continents, resulting from their geographical and biogeographical differences (i.e., east–west major axis of the Eurasian continent, and north–south major axes of the African and American continents) [42]. Considering that the rate of spread depends on the degree of social interactions through migration, this is consistent with the finding that AFR and AMR show the two highest PPAs (Fig 7A). The difference in AMR from other non-AFR meta-populations is smaller than that in AFR from AMR (Fig 7A). However, this smaller difference might result from the short history of human settlement of the Americas (i.e., a recent reduction of psychosocial stress). Within the Eurasian continent, Diamond argued that geographical connectedness is high, moderate, and low in East Asia (China), Europe, and the Indian subcontinent, respectively [42]. Since the degree of social interactions through migration is proportional to that of geographical connectedness, this is compatible with the PPAs of EAS, EUR, and SAS (see Fig 7A). Taking these findings together, the trend toward lower PPA might roughly reflect the local differences of the degree of social interaction resulting from the geographical and biogeographical differences.

Although the CGT-type is a sensitive type, archaic introgression of the CGT-type was accepted in the ancestral population of SAS. In addition, a local increase of the other sensitive type, the TGT-type, is found in SAS. The increase of these sensitive types in SAS ancestors might have been accepted because of low psychosocial stress resulting from low geographical connectedness in the Indian subcontinent [42].

The *ST8SIA2* gene was identified as a schizophrenia-related gene because it is functionally involved in sensitivity to psychosocial stress, an environmental risk factor of schizophrenia. In the adaptive increase of its tolerant types, the onset of schizophrenia is not necessarily required as a selective target (see S12 Fig). Therefore, the trend toward lower PPA should not be hastily interpreted as a result of negative selection against the schizophrenia phenotype (i.e., positive and negative symptoms) or schizophrenia onset. Schizophrenia might be a modern phenotype induced by accidental exposure to genetically intolerable levels of environmental risk factors. Archaic admixture influences disease risk in AMHs [30, 43]. The introgression of the CGT-type from AHs is an example of an introgressed allele negatively affecting present-day humans in terms of the risk of schizophrenia.

One may argue that the trend toward lower PPA results in a lowering of the prevalence of schizophrenia from AFR to non-AFR. However, this is not guaranteed because of the variety of environmental risk factors. Although it is considered that *ST8SIA2* played a central role in

the adaptation to psychosocial stress associated with the spread of AMHs, its current role in the onset of schizophrenia should be limited because psychosocial stress caused by migration is only one of the environmental risk factors.

In conclusion, we detected a difference of the promoter-type composition between AMHs and AHs as well as within AMHs, and found that PPA has changed in the spread of AMHs by the compositional alteration of promoter types. In AMHs, the TCT-type was most prevalent in the ancestral population of non-AFR in addition to the selective increase of the CGC-type in Asia. In contrast, the frequency of the CGT-type was inferred to be very high in AHs. Promoter-type compositions were influenced independently in AMHs and AHs by interspecies gene flows: one into AMHs from AHs, and the other into Denisovans from Neanderthals. The difference of the promoter-type composition confers a unique PPA to each AMH population. PPA has become lower inside of Africa via the increase of the TCT-type. Furthermore, every non-AFR meta-population shows significantly lower PPA than AFR, which resulted from the high prevalence of the TCT-type in the ancestral population of non-AFR. In addition, EAS shows the lowest PPA by the selective increase of the CGC-type. These findings indicate that the increase of the TCT- and CGC-types, which show the two lowest promoter activities, has induced the trend toward lower PPA in the spread of AMHs. The inference of higher PPA in AHs than in AMHs suggests that this trend toward lower PPA is unique to AMHs. Since the TCT- and CGC-types can be regarded as tolerant types to psychosocial stress associated with the migration, it is concluded that the trend toward lower PPA results from the adaptation to the increased psychosocial stress associated with the spread of AMHs. The trend toward lower PPA is also considered as a form of genetic tuning for the evolution of collective brains in AMHs.

Supporting information

S1 Fig. Sequence alignment of haplotypes containing three promoter SNPs of the *ST8SIA2* gene. The three promoter SNPs are highlighted in yellow. The “N”s represent uncertain sites. The name of the AMH haplotype is shown by the ID of the individual who has the sequence (phase 3 of the 1000 Genomes Project database [16]). The six SNPs used for the estimation of variant age are highlighted in orange.

(PDF)

S2 Fig. Global distribution of the TGT (A), TCT (B), CGT (C), and CGC (D) sequences. Pie charts represent relative frequencies as compared with the highest frequency among subpopulations [LWK in (A), TSI in (B), YRI in (C), and JPT in (D)]. ASW is not shown because of a lack of information about homelands in Africa.

(PDF)

S3 Fig. Haplotype composition of the TGT-, TCT-, and CGC-types in meta-populations. (A) Proportions of major TGT haplotypes (see also Table 1 and S2 Table). (B) Proportions of major TCT haplotypes (see also Table 1 and S3 Table). (C) Proportions of major CGC haplotypes (see also Table 1 and [9]).

(PDF)

S4 Fig. Global distribution of the CGT1 (A) and CGT2 (B) sequences. Pie charts represent relative frequencies as compared with the highest frequency among subpopulations [YRI in (A) and GIH in (B)]. ASW is not shown because of a lack of information about homelands in Africa.

(PDF)

S5 Fig. Tree of the TGT haplotypes from SAS. The phylogenetic tree was constructed by the Neighbor-Joining method with the number of differences. Bootstrap values of more than 50% from 1,000 replications are shown on the tree branches.

(PDF)

S6 Fig. Sequence alignment of AH and AFR haplotypes. Except for rs144258052 (0.34% in total; AH-SNP5), SNPs with minor allele frequency of 0.5% or over in the total AMH population are shown. Open boxes represent AH-SNP sites. As for the non-CGT-types, all AFR haplotypes are represented. The three promoter SNPs are represented by asterisks. Longest identical tracts between AH sequences and non-CGT haplotypes are highlighted in blue (Denisovan-1) and green (Vindija-1), and those between CGT1 haplotypes and non-CGT haplotypes are highlighted in orange and yellow. The positions of six SNPs used for the estimation of variant age are indicated in red (see [S1 Fig](#)). AH-SNP1 was not used for variant age estimation because the C allele is shared with a TCT haplotype found only in EUR (HG00145.0; two sequences; [S3 Table](#)). It is considered that HG00145.0 containing the C allele at AH-SNP1 is a recombinant with a CGT2 haplotype (HG03667.1) found in EUR because of sequence identity in the region surrounding AH-SNP1 between them ([S13 Fig](#)).

(PDF)

S7 Fig. Recombination process producing Denisovan-1 and Vindija-1. Star represents the position of the three promoter SNPs.

(PDF)

S8 Fig. Haplotype trees of promoter types from SAS. (A) Tree of the TGT haplotypes. (B) Tree of the TCT haplotypes. These haplotypes were selected based on the analyses using site differences (see the “Materials and Methods” section). The phylogenetic trees were constructed by the Neighbor-Joining method with the number of differences. Bootstrap values of more than 50% from 1,000 replications are shown on the tree branches. Red star represents the haplotype found in AFR.

(PDF)

S9 Fig. Evolution of promoter types in AMHs and AHs. Pie chart of AHs was drawn by promoter-type composition producing the lowest inferred PPA (see [S9 Table](#)).

(PDF)

S10 Fig. Hypothetical process of adaptive increase of tolerant types in the *ST8SIA2* gene. For details of the functional consequence on social interaction, please see our previous report [9].

(PDF)

S11 Fig. PPA and effective population size change during the past 60 kyr. (A) PPAs and effective population size changes obtained roughly from the PSMC smoothed curve [16]. (B) PPAs and effective population size changes obtained roughly from the PSMC unsmoothed curve [16]. As for PPA, relative difference as compared with LWK is shown. As for effective population size change, relative difference as compared with BEB is shown.

(PDF)

S12 Fig. Conditional difference between adaptation to psychosocial stress and the onset of schizophrenia. In the onset of schizophrenia, inevitable stay under the environmental risk factors is essential. However, this inevitability of staying is not necessarily assumed in the adaptation to psychosocial stress.

(PDF)

S13 Fig. Sequence alignment of AH, CGT and EUR TCT haplotypes. Except for rs144258052 (0.34% in total; AH-SNP5), SNPs with minor allele frequency of 0.5% or over in the total AMH population are shown. Open boxes represent AH-SNP sites. The three promoter SNPs are represented by asterisks. Longest identical tract surrounding AH-SNP1 between a CGT2 sequence (HG03667.1) and an EUR-unique TCT haplotype (HG00145.0) is highlighted in purple. This sequence identity suggests that HG00145.0 containing the C allele at AH-SNP1 is a recombinant with a CGT2 haplotype (HG03667.1) found in EUR.
(PDF)

S1 Table. Distribution of promoter types.
(PDF)

S2 Table. Distribution of the TGT haplotypes.
(PDF)

S3 Table. Distribution of the TCT haplotypes.
(PDF)

S4 Table. Distribution of the CGT haplotypes.
(PDF)

S5 Table. Matrix of site differences among the CGT haplotypes.
(PDF)

S6 Table. Matrix of site differences among the TGT haplotypes from SAS.
(PDF)

S7 Table. Matrix of site differences among the TCT haplotypes from SAS.
(PDF)

S8 Table. Matrix of site differences among the CGC haplotypes from SAS.
(PDF)

S9 Table. Inference of promoter type composition and PPA of AH population.
(PDF)

Acknowledgments

We thank Edanz (<https://en-author-services.edanz.com/ac>) for editing a draft of this manuscript.

Author Contributions

Conceptualization: Toshiyuki Hayakawa.

Data curation: Toshiyuki Hayakawa, Masahiro Terahara, Naoko T. Fujito, Kosuke M. Teshima, Yoko Satta.

Formal analysis: Toshiyuki Hayakawa, Masahiro Terahara, Takumi Matsunaga, Naoyuki Takahata, Yoko Satta.

Funding acquisition: Toshiyuki Hayakawa.

Investigation: Toshiyuki Hayakawa, Masaya Hane, Ken Kitajima, Chihiro Sato.

Project administration: Toshiyuki Hayakawa, Yoko Satta.

Writing – original draft: Toshiyuki Hayakawa, Yoko Satta.

Writing – review & editing: Toshiyuki Hayakawa, Naoko T. Fujito, Kosuke M. Teshima, Masaya Hane, Ken Kitajima, Chihiro Sato, Naoyuki Takahata, Yoko Satta.

References

1. Nielsen R, Akey JM, Jakobsson M, Pritchard JK, Tishkoff S, Willerslev E. Tracing the peopling of the world through genomics. *Nature*. 2017; 541(7637):302–10. <https://doi.org/10.1038/nature21347> PMID: 28102248
2. Henrich J. *The secret of our success: how culture is driving human evolution, domesticating our species, and making us smarter*. Princeton, NJ: Princeton University Press; 2016.
3. Muthukrishna M, Henrich J. Innovation in the collective brain. *Philos Trans R Soc Lond B Biol Sci*. 2016; 371(1690):20150192. <https://doi.org/10.1098/rstb.2015.0192> PMID: 26926282
4. Sato C, Kitajima K. Impact of structural aberrancy of polysialic acid and its synthetic enzyme ST8SIA2 in schizophrenia. *Front Cell Neurosci*. 2013; 7:61. <https://doi.org/10.3389/fncel.2013.00061> PMID: 23675315
5. Hayakawa T, Varki A. Human-specific changes in sialic acid biology. In: Hirai H, Imai H, Go Y, editors. *Post-genome biology of primates*. Primatology monographs: Springer; 2012. p123–48.
6. Arai M, Yamada K, Toyota T, Obata N, Haga S, Yoshida Y, et al. Association between polymorphisms in the promoter region of the sialyltransferase 8B (SIAT8B) gene and schizophrenia. *Biol Psychiatry*. 2006; 59(7):652–9. <https://doi.org/10.1016/j.biopsych.2005.08.016> PMID: 16229822
7. Tao R, Li C, Zheng Y, Qin W, Zhang J, Li X, et al. Positive association between SIAT8B and schizophrenia in the Chinese Han population. *Schizophr Res*. 2007; 90(1–3):108–14. <https://doi.org/10.1016/j.schres.2006.09.029> PMID: 17126533
8. Gilabert-Juan J, Nacher J, Sanjuán J, Molto MD. Sex-specific association of the ST8SIAII gene with schizophrenia in a Spanish population. *Psychiatry Res*. 2013; 210(3):1293–5. <https://doi.org/10.1016/j.psychres.2013.09.001> PMID: 24070986
9. Fujito NT, Satta Y, Hane M, Matsui A, Yashima K, Kitajima K, et al. Positive selection on schizophrenia-associated ST8SIA2 gene in post-glacial Asia. *PLoS One*. 2018; 13(7):e0200278. <https://doi.org/10.1371/journal.pone.0200278> PMID: 30044798
10. Sato C, Hane M, Kitajima K. Relationship between ST8SIA2, polysialic acid and its binding molecules, and psychiatric disorders. *Biochim Biophys Acta*. 2016; 1860(8):1739–52. <https://doi.org/10.1016/j.bbagen.2016.04.015> PMID: 27105834
11. van Os J, Kenis G, Rutten BP. The environment and schizophrenia. *Nature*. 2010; 468(7321):203–12. <https://doi.org/10.1038/nature09563> PMID: 21068828
12. Howes OD, Murray RM. Schizophrenia: an integrated sociodevelopmental-cognitive model. *Lancet*. 2014; 383(9929):1677–87. [https://doi.org/10.1016/S0140-6736\(13\)62036-X](https://doi.org/10.1016/S0140-6736(13)62036-X) PMID: 24315522
13. Howes OD, McCutcheon R, Owen MJ, Murray RM. The Role of Genes, Stress, and Dopamine in the Development of Schizophrenia. *Biol Psychiatry*. 2017; 81(1):9–20. <https://doi.org/10.1016/j.biopsych.2016.07.014> PMID: 27720198
14. Devylder JE, Oh HY, Yang LH, Cabassa LJ, Chen FP, Lukens EP. Acculturative stress and psychotic-like experiences among Asian and Latino immigrants to the United States. *Schizophr Res*. 2013; 150(1):223–8. <https://doi.org/10.1016/j.schres.2013.07.040> PMID: 23932446
15. Brown AS. The environment and susceptibility to schizophrenia. *Prog Neurobiol*. 2011; 93(1):23–58. <https://doi.org/10.1016/j.pneurobio.2010.09.003> PMID: 20955757
16. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*. 2015; 526(7571):68–74. <https://doi.org/10.1038/nature15393> PMID: 26432245
17. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014; 511(7510):421–7. <https://doi.org/10.1038/nature13595> PMID: 25056061
18. Lam M, Chen CY, Li Z, Martin AR, Bryois J, Ma X, et al. Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nat Genet*. 2019; 51(12):1670–8. <https://doi.org/10.1038/s41588-019-0512-x> PMID: 31740837
19. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*. 1987; 4(4):406–25. <https://doi.org/10.1093/oxfordjournals.molbev.a040454> PMID: 3447015
20. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol*. 2018; 35(6):1547–9. <https://doi.org/10.1093/molbev/msy096> PMID: 29722887

21. Browning SR, Browning BL, Zhou Y, Tucci S, Akey JM. Analysis of Human Sequence Data Reveals Two Pulses of Archaic Denisovan Admixture. *Cell*. 2018; 173(1):53–61. <https://doi.org/10.1016/j.cell.2018.02.031> PMID: 29551270
22. Albers PK, McVean G. Dating genomic variants and shared ancestry in population-scale sequencing data. *PLoS Biol*. 2020; 18(1):e3000586. <https://doi.org/10.1371/journal.pbio.3000586> PMID: 31951611
23. Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, et al. Genome-wide detection and characterization of positive selection in human populations. *Nature*. 2007; 449(7164):913–8. <https://doi.org/10.1038/nature06250> PMID: 17943131
24. McHugh C, Brown L, Thornton TA. Detecting Heterogeneity in Population Structure Across the Genome in Admixed Populations. *Genetics*. 2016; 204(1):43–56. <https://doi.org/10.1534/genetics.115.184184> PMID: 27440868
25. Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, et al. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*. 2014; 505(7481):43–9. <https://doi.org/10.1038/nature12886> PMID: 24352235
26. Mafessoni F, Grote S, de Filippo C, Slon V, Kolobova KA, Viola B, et al. A high-coverage Neandertal genome from Chagyrskaya Cave. *Proc Natl Acad Sci USA*. 2020; 117(26):15132–6. <https://doi.org/10.1073/pnas.2004944117> PMID: 32546518
27. Huerta-Sánchez E, Jin X, Asan, Bianba Z, Peter BM, Vinckenbosch N, Liang Y, et al. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature*. 2014; 512(7513):194–7. <https://doi.org/10.1038/nature13408> PMID: 25043035
28. Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, et al. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*. 2010; 468(7327):1053–60. <https://doi.org/10.1038/nature09710> PMID: 21179161
29. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, et al. A draft sequence of the Neandertal genome. *Science*. 2010; 328(5979):710–22. <https://doi.org/10.1126/science.1188021> PMID: 20448178
30. Zeberg H, Pääbo S. The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. *Nature*. 2020; 587(7835):610–2. <https://doi.org/10.1038/s41586-020-2818-3> PMID: 32998156
31. Weir BS. *Genetic Data Analysis II*. Sunderland, MA, USA: Sinauer Associates, Inc; 1996.
32. Scally A, Durbin R. Revising the human mutation rate: implications for understanding human evolution. *Nat Rev Genet*. 2012; 13(10):745–53. <https://doi.org/10.1038/nrg3295> PMID: 22965354
33. Plagnol V, Wall JD. Possible ancestral structure in human populations. *PLoS Genet*. 2006; 2(7):e105. <https://doi.org/10.1371/journal.pgen.0020105> PMID: 16895447
34. Chen L, Wolf AB, Fu W, Li L, Akey JM. Identifying and Interpreting Apparent Neanderthal Ancestry in African Individuals. *Cell*. 2020; 180(4):677–87. <https://doi.org/10.1016/j.cell.2020.01.012> PMID: 32004458
35. Prüfer K, de Filippo C, Grote S, Mafessoni F, Korlevic P, Hajdinjak M, et al. A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science*. 2017; 358(6363):655–8. <https://doi.org/10.1126/science.aao1887> PMID: 28982794
36. Peter BM. 100,000 years of gene flow between Neandertals and Denisovans in the Altai mountains. *bioRxiv*. 2020; <https://doi.org/10.1101/2020.03.13.990523>
37. Jacobs GS, Hudjashov G, Saag L, Kusuma P, Darusallam CC, Lawson DJ, et al. Multiple Deeply Divergent Denisovan Ancestries in Papuans. *Cell*. 2019; 177(4):1010–21. <https://doi.org/10.1016/j.cell.2019.02.035> PMID: 30981557
38. Sankararaman S, Patterson N, Li H, Paabo S, Reich D. The Date of Interbreeding between Neandertals and Modern Humans. *PLoS Genet*. 2012; 8(10):e1002947. <https://doi.org/10.1371/journal.pgen.1002947> PMID: 23055938
39. Cantor-Graae E, Selten JP. Schizophrenia and migration: a meta-analysis and review. *Am J Psychiatry*. 2005; 162(1):12–24. <https://doi.org/10.1176/appi.ajp.162.1.12> PMID: 15625195
40. van Os J, Kapur S. Schizophrenia. *Lancet*. 2009; 374(9690):635–45. [https://doi.org/10.1016/S0140-6736\(09\)60995-8](https://doi.org/10.1016/S0140-6736(09)60995-8) PMID: 19700006
41. Morgan C, Charalambides M, Hutchinson G, Murray RM. Migration, ethnicity, and psychosis: toward a sociodevelopmental model. *Schizophr Bull* 2010; 36(4):655–64. <https://doi.org/10.1093/schbul/sbq051> PMID: 20513653
42. Diamond J. *GUNS, GERMS, AND STEEL. The fates of human societies*. New York: W. W. Norton & Company, Inc; 1997.
43. Simonti CN, Vernot B, Bastarache L, Bottinger E, Carrell DS, Chisholm RL, et al. The phenotypic legacy of admixture between modern humans and Neandertals. *Science*. 2016; 351(6274):737–41. <https://doi.org/10.1126/science.aad2149> PMID: 26912863