

Neanderthal Introgression at Chromosome 3p21.31 Was Under Positive Natural Selection in East Asians

Qiliang Ding,^{†,1} Ya Hu,^{†,1} Shuhua Xu,² Jiucun Wang,¹ and Li Jin^{*,1,2}

¹State Key Laboratory of Genetic Engineering and Ministry of Education Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai, China

²CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences (SIBS), Chinese Academy of Sciences (CAS), Shanghai, China

[†]These authors contributed equally to this work.

*Corresponding author: E-mail: lijn.fudan@gmail.com.

Associate editor: Joshua Akey

Abstract

Studies of the Neanderthal and Denisovan genomes demonstrate archaic hominin introgression in Eurasians. Here, we present evidence of Neanderthal introgression within the chromosome 3p21.31 region, occurring with a high frequency in East Asians (ranging from 49.4% to 66.5%) and at a low frequency in Europeans. We also detected a signal of strong positive selection in this region only in East Asians. Our data indicate that likely candidate targets of selection include rs12488302-T and its associated alleles—among which four are nonsynonymous, including rs35455589-G in *HYAL2*, a gene related to the cellular response to ultraviolet-B irradiation. Furthermore, suggestive evidence supports latitude-dependent selection, implicating a role of ultraviolet-B. Interestingly, the distribution of rs35455589-G suggests that this allele was lost during the exodus of ancestors of modern Eurasians from Africa and reintroduced to Eurasians from Neanderthals.

Key words: Neanderthals, introgression, positive selection, East Asians, *HYAL*.

Introduction

Studies of mitochondrial and Y chromosome DNA have provided solid evidence supporting an African origin for all modern Eurasians—known as the out-of-Africa theory of modern human origins (Cann et al. 1987; Hammer 1995). The recent publication of the Neanderthal and Denisovan genomes enables direct comparison between modern human genomes and archaic hominin genomes. Examinations of archaic hominin genomes reveal Neanderthal introgression among modern Eurasians (Green et al. 2010) and Denisovan introgression in modern Oceanians (Reich et al. 2010), but their contributions to modern humans are very limited. *Abi-Rached et al. (2011)* identified Neanderthal and Denisovan introgression at the *HLA* loci (MIM 142800, 142830, 142840) and suggested that the introgressed alleles significantly reshaped the immune system of modern humans, especially those outside Africa. *Mendez et al. (2013)* found both Neanderthal and Denisovan introgression around the *OAS* gene cluster (MIM 164350, 603350, and 603351). *Mendez et al. (2012b)* also found both Neanderthal and Denisovan introgression in the region surrounding *STAT2* (MIM 600556). Notably, the results of a neutrality test on the *STAT2* region suggest that Neanderthal introgression is under positive natural selection among Papua New Guineans.

In this study, we analyzed a region of the chromosome 3p21.31 (denoted as the *HYAL* region), which encompasses 18 genes, including *HYAL1* (MIM 607071), *HYAL2* (MIM 603551), and *HYAL3* (MIM 604038). These three *HYAL* genes encode

hyaluronoglucosaminidases, which are involved in the cellular response to ultraviolet-B (UV-B). We analyzed data from the 1000 Genomes Project (1KG) Phase 1 (1000 Genomes Consortium 2012) and found archaic hominin introgression in the genomes of contemporary Eurasians, with East Asians showing a high frequency of introgressive haplotypes, ranging from 49.4% in a Japanese population to 66.5% in a Southern Han Chinese population. Both the integrated haplotype score (iHS) and the composite of multiple signals (CMS) tests also revealed a signal of strong positive selection among East Asians. The target of this selection appears to be rs12488402-T—which is exclusively carried by the introgressive haplotypes—or one of the four non-synonymous single-nucleotide polymorphisms (SNPs) found to be in strong linkage disequilibrium (LD) with rs12488302. To conclude, the *HYAL* region is identified as showing archaic hominin introgression with the introgressive haplotypes under positive selection in East Asian populations.

Results

LD Structure and Haplotypes in the *HYAL* Region

Investigation of the LD structure in the *HYAL* region revealed a large LD block spanning 200 kb in East Asian populations but not in other populations. Specifically, the length of the LD block is 196 kb in the Han Chinese population in Beijing, China (CHB; [supplementary fig. S1A, Supplementary Material online](#)), 216 kb in the Chinese population in Metropolitan Denver, CO (CHD; [supplementary fig. S1B, Supplementary Material online](#)), 195 kb in the Japanese population in

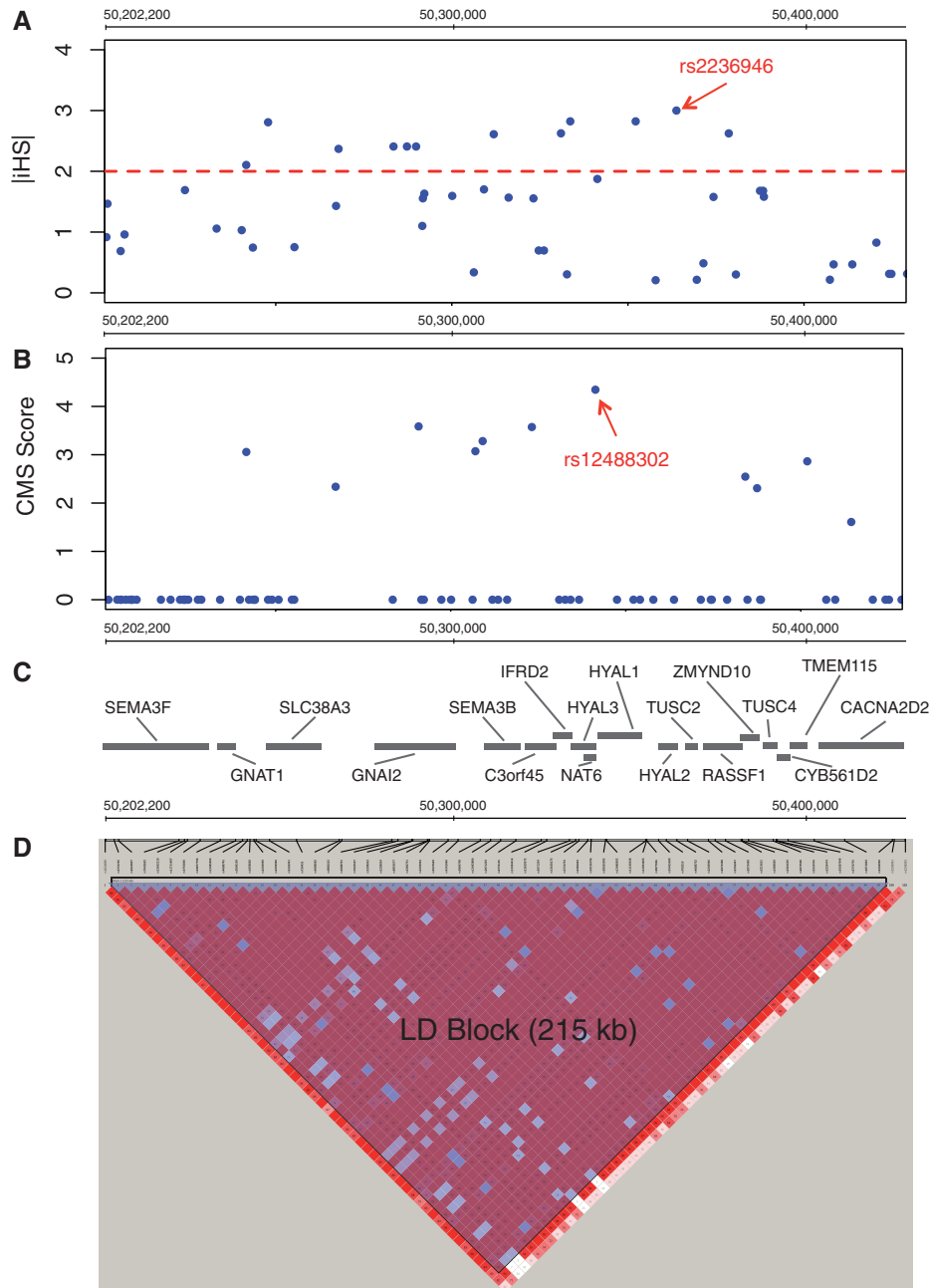


Fig. 1. Overview of the studied region. All panels are to the same scale. (A) Standardized $|iHS|$ plot of the East Asian populations (CHB + JPT, HapMap 3). The red horizontal line indicates the threshold of positive selection signal ($|iHS| = 2.0$). (B) CMS score of East Asians populations (CHB + JPT, HapMap 2). (C) Physical location of genes surrounding the studied region. (D) LD pattern of the studied region in East Asian populations (CHB + JPT, HapMap 3). SNPs with MAF >0.05 are shown.

Tokyo, Japan (JPT; [supplementary fig. S1C, Supplementary Material online](#)), and 215 kb in CHB + JPT ([fig. 1D](#)). In contrast, this region is more fragmented in Europeans and Africans, including the population of Northern and Western European ancestry (known as CEPH) in Utah, USA (CEU) and the population of Toscani in Italy (TSI; [supplementary fig. S1D, Supplementary Material online](#)), and among the Yoruba in Ibadan, Nigeria (YRI; [supplementary fig. S1E, Supplementary Material online](#)). For subsequent analyses, we chose the boundaries of the segment that were shared by all East Asian populations (chr3: 50,225,029–50,420,554,

hg19) as the preliminary boundaries. All human genome coordinates in this article are in GRCh37 (hg19) unless otherwise noted.

We reconstructed a phylogenetic tree of the haplotypes within the preliminary boundaries, including those from the 1KG data set, Altai Neanderthal, and Denisovan, using chimpanzee as outgroup. We observed that some haplotypes of contemporary populations formed a cluster first with Neanderthal and Denisovan before joining together with the rest of the haplotypes of contemporary populations, suggesting an influence of archaic hominin introgression.

Such haplotypes resulting from archaic introgression will be hereafter referred to as “introgressive haplotypes.”

To eliminate possible topological distortion of the phylogeny in the presence of recombination and to further define the boundaries of the tentative introgressed segment, we identified and removed 19 haplotypes that were recombinants of non-introgressive and introgressive haplotypes (supplementary fig. S3 and table S1, Supplementary Material online). After removal of recombinants, the redefined boundaries of the introgressive segment became 50,238,376 and 50,401,841. This data set was used in all following analyses.

Evidence of Introgression of the Putative Introgressive Segments

A phylogenetic tree was reconstructed using all non-recombinant haplotypes of contemporary populations, Altai Neanderthal, and Denisovan, using chimpanzee as outgroup (fig. 2). Trivial monophyletic clusters were collapsed to simplify presentation without losing major structures of the tree. The phylogenetic tree shows that some haplotypes of contemporary populations coalesce first with archaic hominins (Altai Neanderthal and Denisovan) before joining with the rest of the haplotypes of contemporary populations, suggesting archaic hominin introgression. This is supported by the bootstrap value of 98 at Node A in figure 2, which separates all archaic/introgressive haplotypes from other haplotypes of contemporary populations. The introgressive haplotypes first coalesce with Altai Neanderthals (bootstrap value = 99), then with Denisovan (bootstrap value = 96), indicating that Neanderthal is more likely to be the source of introgression than Denisovan. We inferred the ancestral sequence of all introgressive haplotypes using a maximum likelihood procedure implemented in MEGA 5 (Tamura et al. 2011) and further compared it with archaic genomes. The ancestral sequence of introgressive haplotypes shares 99.3% (731/736) of alleles with Altai Neanderthal, 98.6% (726/736) with Vindija Neanderthal, and 97.2% (716/736) with Denisovan. This result also supports that the introgressive haplotypes observed in contemporary populations are closer to Neanderthal than Denisovan.

In addition to this phylogenetic approach, we also searched for Type 2 derived alleles on the putative introgressive haplotypes (see Material and Methods). Type 2 derived alleles were originated from mutations in the archaic hominin lineage after their divergence with ancestors of modern humans. We obtained the polymorphism table for all non-recombinant haplotypes (supplementary table S5, Supplementary Material online). We found that the Type 2 derived alleles existed in the putative introgressive haplotypes, which fit the allele pattern of introgressive haplotypes. Furthermore, of the Type 2 alleles, 90.3% (28/31) are identical to the Altai Neanderthal alleles and 16.1% (5/31) to the Denisovan alleles. This suggests that the putative introgressive haplotypes are not false positives originated inside the modern human lineage, because the haplotypes of contemporary populations would have similar sequence divergence with Neanderthal and Denisovan. In addition, the high

sequence similarity to Neanderthal indicates a likelihood that the putative introgressive haplotypes are of Neanderthal origin.

It is also possible that the observed high similarity of the introgressive haplotypes to the Neanderthal sequence could be a result of ancestral shared polymorphism. The “ancestral shared polymorphism” hypothesis suggests that the putative introgressive haplotypes existed before the divergence of ancestors of modern humans and Neanderthals and were drifted to almost zero in Europeans and Africans, while they were preserved in East Asians. Using the method described by Mendez et al. (2012a), we analyzed the haplotype structure at the *HYAL* region. The “recent introgression” hypothesis would predict a longer LD block than that predicted under the ancestral shared polymorphism hypothesis. The LD block shared by all East Asian populations is 0.0457 cM in length (Hinch et al. 2011); thus, the probability of persistence of such an LD block follows an exponential distribution with $\lambda = 0.000457$: $P = e^{-\lambda \times \frac{\text{population divergence time}}{\text{time per generation}}}$. The time per generation was set to 25 years. Following the ancestral shared polymorphism hypothesis, the LD block structure should exist before the divergence of Neanderthals from ancestors of modern humans (i.e., 270–440 thousand years ago [KYA], estimated by Green et al. 2010). This probability is 7.186×10^{-3} (270 KYA); therefore, the ancestral shared polymorphism hypothesis could be rejected with a *P* value of $\leq 7.186 \times 10^{-3}$.

Overall, both the results of phylogenetic reconstruction and analyses of derived allele patterns demonstrate a high similarity of the introgressive haplotypes to Neanderthal. Further, we reject that these observations could result from ancestral shared polymorphism, leading to the conclusion that the introgressive haplotypes were from a recent Neanderthal introgression.

Global Distribution of the Introgressive Haplotypes

The 1KG and the International Haplotype Map Project Phase 3 (HapMap 3; International HapMap 3 Consortium 2010) data sets were used to investigate the global distribution of introgressive haplotypes. We observed high frequencies of the introgressive haplotypes in East Asian populations, ranging from 49.4% in JPT to 66.5% in the Han Chinese from Southern China (CHS; fig. 3 and supplementary table S2, Supplementary Material online), consistent with the pattern observed in figure 2. The introgressive haplotypes are also present at low frequencies in CEU (0.57%, a singleton), in the British population in England and Scotland (GBR; 0.56%, a singleton), and in the Finnish population in Finland (FIN; 3.7%, 7 haplotypes), as well as in four New World populations, including the population with Mexican Ancestry in Los Angeles, CA (MXL; 34.8%), Puerto Ricans in Puerto Rico (PUR; 11.8%), Columbians in Medellin, Columbia (CLM; 20.0%), and the population of Americans with African ancestry in the Southwest USA (ASW; 1.64%).

We also obtained frequency of introgressive haplotypes in more populations from the Allele Frequency Database (ALFRED; Rajeevan et al. 2012) (supplementary table S6,

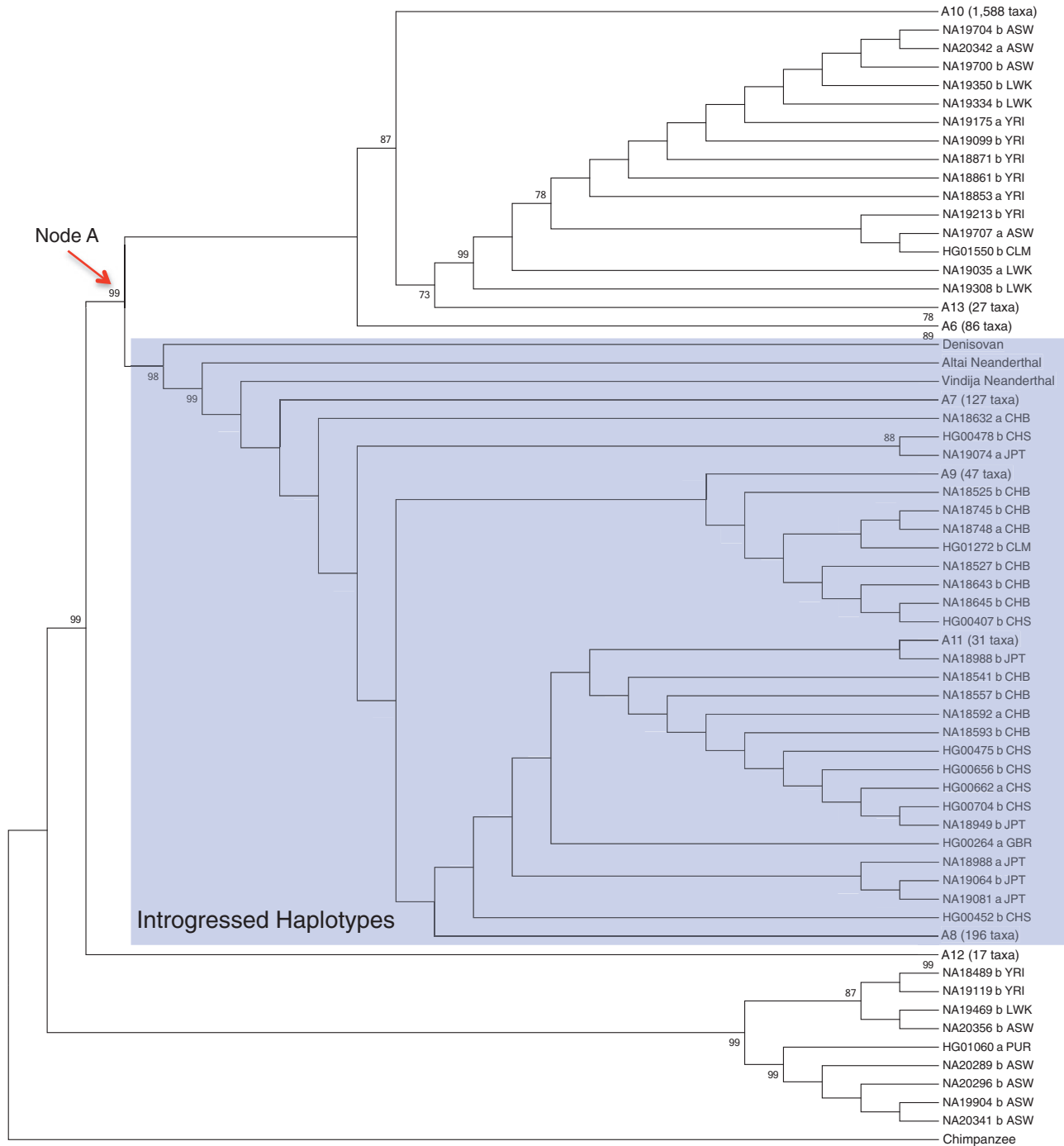


Fig. 2. Phylogenetic tree of all non-recombinant haplotypes of contemporary populations, Altai Neanderthal, Denisovan, and chimpanzee reconstructed using the maximum likelihood method in MEGA with 100 bootstrap replications. Some trivial monophyletic clusters were collapsed, and the “topology only” option in MEGA was used to enhance display. Bootstrap values larger than 50 are shown.

Supplementary Material online). The diagnostic SNP used here is rs11130248-G. Notably, the G allele is the ancestral allele. Therefore, this diagnostic SNP could not be used in African populations. Furthermore, many ALFRED populations represent a very small sample size (e.g., $2n$ less than 20), which could lead to imprecise estimation of allele frequency. For these reasons, we did not incorporate the ALFRED data into figure 3.

As the frequency of introgressive haplotypes is higher in East Asians than in Europeans, it is possible that archaic

introgression occurred in the ancestral population shared by East Asians and Native Americans, while the introgressive haplotypes in Europeans might be the result of recent gene flow from East Asia to Europe. To examine this hypothesis, we obtained the sequence of the 500-kb regions flanking the *HYAL* region to the left and right (chr3: 49,725,029–50,225,028 and chr3: 50,420,555–50,920,554) for Eurasians. We divided the Eurasian haplotypes into four groups: 1) European non-introgressive haplotypes (EUR-N); 2) European introgressive haplotypes (EUR-I); 3) East Asian

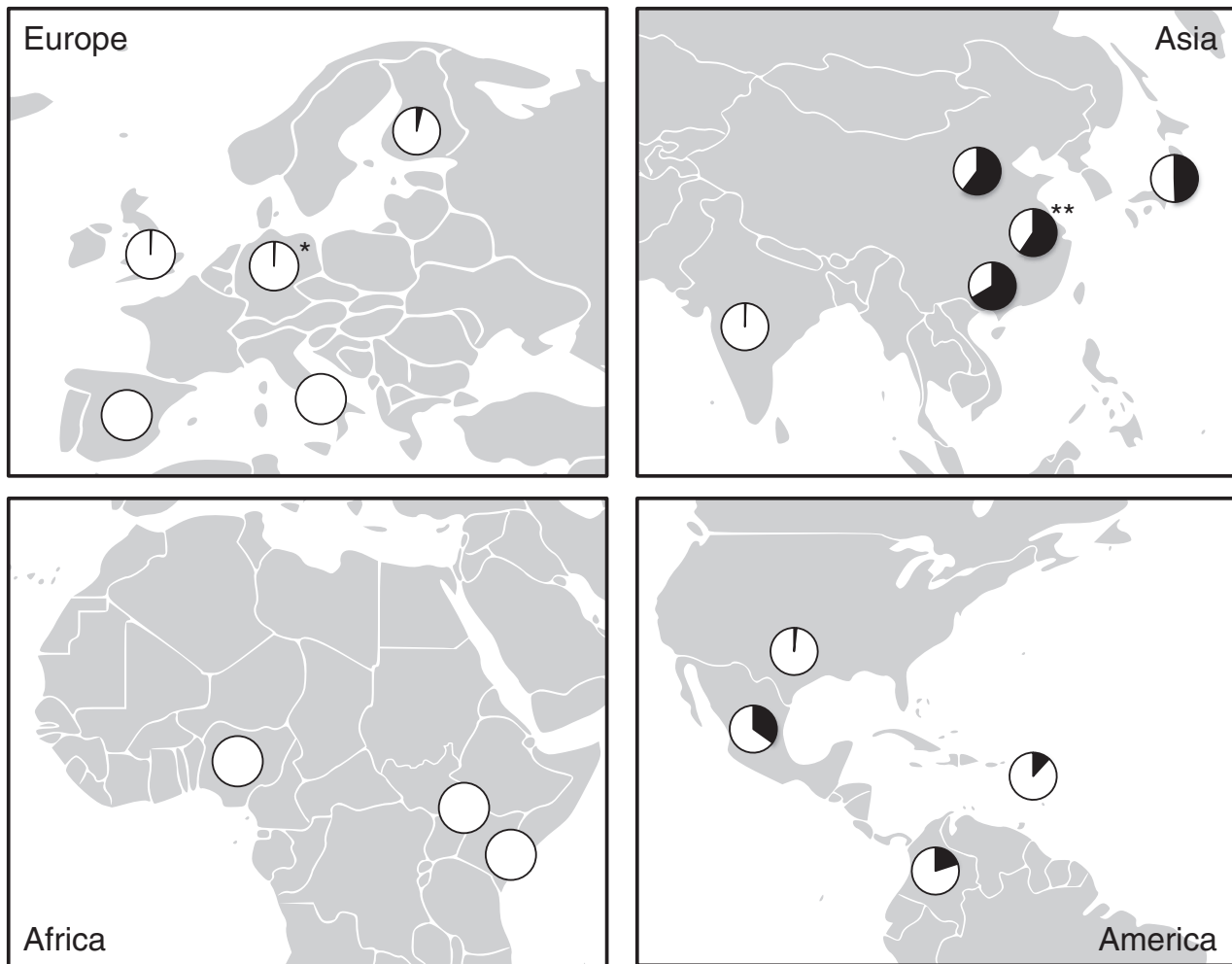


Fig. 3. Global distribution of the introgressive haplotypes, shown in the filled region of the pie chart. *CEU (1KG Phase 1); **CHD (HapMap 3).

non-introgressive haplotypes (ASN-N); and 4) East Asian introgressive haplotypes (ASN-I). If the introgressive haplotypes in Europeans resulted from East Asian gene flow, the flanking regions should be more similar to East Asians than to Europeans. We calculated the average sequence similarities among the four groups (supplementary table S7, Supplementary Material online), analyzing a total of 9,180 polymorphic sites. Interestingly, EUR-I shows significantly higher sequence similarity with ASN-N (0.9611) and ASN-I (0.9683) than with EUR-N (0.9538), with a P value of 0.0000 (right-tailed Student's t -test). This observation is consistent with our hypothesis that the introgressive haplotypes in Europe resulted from recent East Asian gene flow.

Positive Selection in the *HYAL* Region

To explore the extent of positive selection in the *HYAL* region, we calculated the standardized iHS based on the HapMap 3 data set and obtained a CMS score based on HapMap 2 with the CMS Viewer. A positive selection signal, revealed by $|iHS| > 2.0$, is observed in East Asians (fig. 1A) but not in Europeans or Africans (supplementary fig. S2A and B, Supplementary Material online). Similarly, a high CMS score

is also found in East Asians (fig. 1B) but not in Europeans or Africans (supplementary fig. S2C and D, Supplementary Material online). Remarkably, the SNP with the highest CMS score (rs12488302; CMS score = 4.35; chr3: 50,341,326) and the SNP with the highest $|iHS|$ (rs2236946; $|iHS| = 3.0$; chr3: 50,363,771) are in strong LD ($r^2 = 0.97$, $D' = 0.99$ using 1KG Phase 1 data) in East Asians. These findings indicate that the *HYAL* region was subjected to positive selection, and we further speculate that the SNPs in strong LD with rs12488302 and rs2236946 could be the candidate targeting sites of this positive selection.

Next, we explored the allele that was positively selected or hitchhiked at rs12488302. The derived allele (T) of rs12488302 is inclusively and exclusively carried by the introgressive haplotypes. We calculated the extended haplotype homozygosity (EHH, Sabeti et al. 2002) and drew a haplotype bifurcation graph for both alleles. The decay of haplotype homozygosity is much slower for the derived allele (T) compared with the ancestral allele (C) on the EHH plot (fig. 4A). Bifurcation of haplotypes is also slower for the derived allele (fig. 4C) compared with the ancestral allele (fig. 4B). Results from the EHH plot and haplotype bifurcation graph both support that the

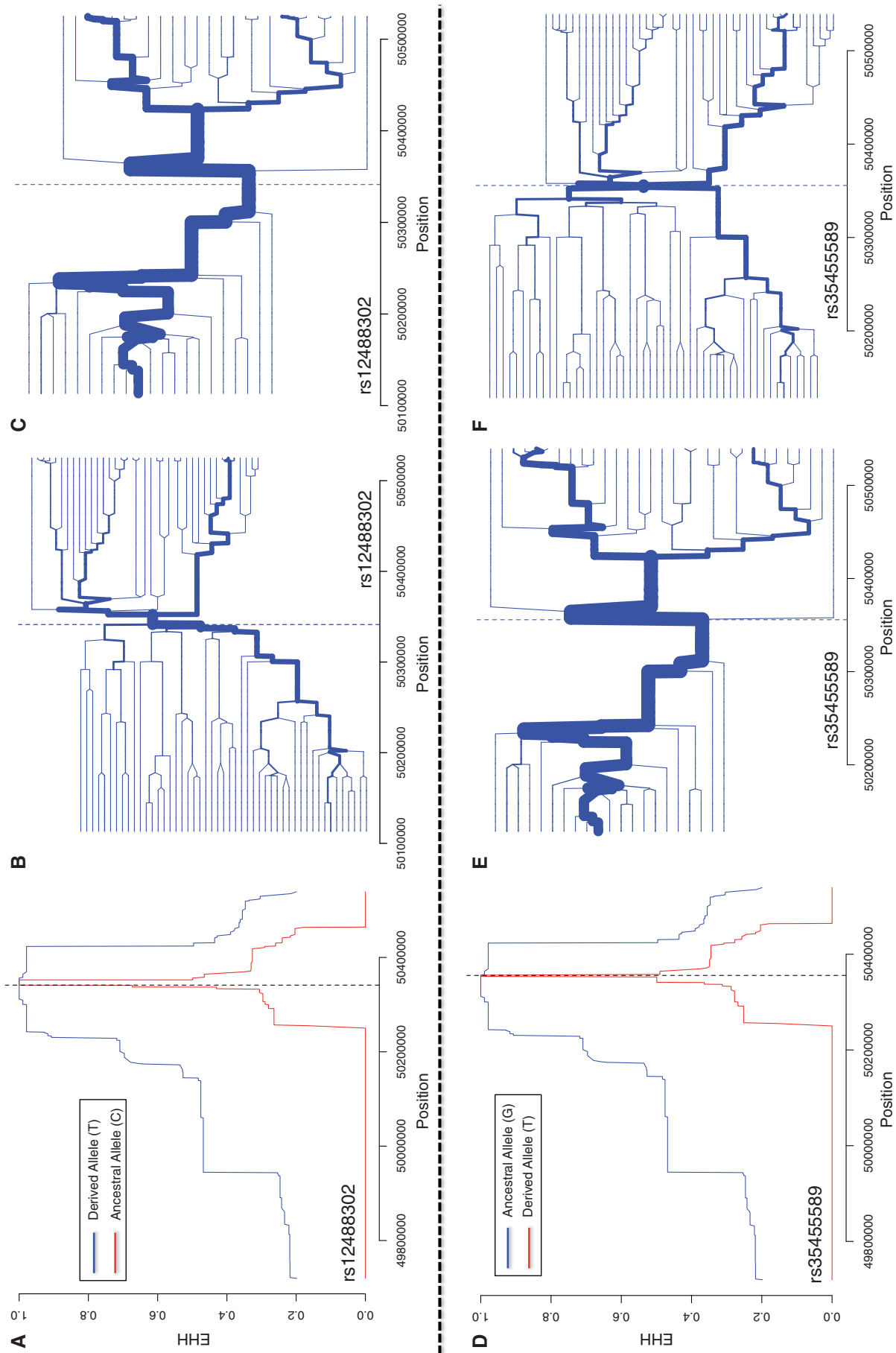


Fig. 4. EHH and haplotype bifurcation graphs (HBGs). (A) EHH graph for both alleles at rs12488302. (B) HBG for rs12488302-C (the ancestral allele). (C) HBG for rs12488302-T (the derived allele). (D) EHH graph for both alleles at rs35455589. (E) HBG for rs35455589-G (the ancestral allele). (F) HBG for rs35455589-T (the derived allele). (G) EHH graph for rs12488302-T (the derived allele). (H) EHH graph for rs12488302-C (the ancestral allele). The haplotype thicknesses in a HBG represent the haplotype frequency. Long and thick haplotypes, indicating long-range homozygosity, are observed for rs12488302-T (panel C) and rs35455589-G (panel E) but not for rs12488302-C (panel B) and rs35455589-T (panel F).

Table 1. Non-Synonymous SNPs Exclusively Carried by the Introgressive Haplotypes.

dbSNP ID	LD ^a	Conservation ^b	Regulatory Chromatin States ^c	No. Altered Motifs	Gene
rs2071203	0.95	SiPhy	Poised promoter in NHEK	9	SEMA3B
rs2269432	1	GERP	—	5	NAT6
rs35455589	0.97	SiPhy and GERP	Strong enhancer in HMEC, NHEK	4	HYAL2
rs4688725	0.96	SiPhy	Active promoter in NHEK	9	RASSF1

NOTE.—See also [supplementary table S3, Supplementary Material](#) online. NHEK, epidermal keratinocytes; HMEC, mammary epithelial cells.

^aLD (r^2 score) with rs12488302.

^bLocated in conserved regions predicted by SiPhy and/or GERP algorithms.

^cAnnotated by the ENCODE project. Only active, poised, or strong promoters and/or enhancers are listed.

derived allele (T) at rs12488302 or the alleles at nearby sites carried by the introgressive haplotypes could be the candidate targets of positive selection.

Search for Candidate Targets of Positive Selection in the Introgressive Haplotypes

As the derived allele (T) of the SNP with the highest CMS score (rs12488302) was a marker representing introgressive haplotypes, we focused on the loci harboring alleles that were exclusively carried by the introgressive haplotypes ([supplementary table S3, Supplementary Material](#) online). LD (r^2 score) with rs12488302 was estimated for the loci harboring those alleles (based on 1KG East Asian populations), and the r^2 scores are larger than 0.8. Functional annotations for those alleles were obtained from HaploReg ([supplementary table S3, Supplementary Material](#) online). Among them, four non-synonymous SNPs are found in *SEMA3B* (MIM 601281), *NAT6* (MIM 607073), *HYAL2* (MIM 603551), and *RASSF1* (MIM 605082) (see [table 1](#)). Interestingly, the alleles carried by the introgressive haplotypes at the four non-synonymous SNPs are all ancestral alleles, which are also observed among contemporary Africans where selection was absent.

Adaptation by Acquiring the Lost Allele(s) Through Introgression: Hypothesis and Evidence

Alleles carried by the introgressive haplotypes at the sites in strong LD with rs12488302 are either derived or ancestral ([supplementary table S3, Supplementary Material](#) online). For example, rs12494414-T is not shared with chimpanzee and is only found in the introgressive haplotypes, suggesting that it might have emerged after the divergence of Neanderthals and ancestors of modern humans, and is therefore derived. However, other alleles, represented by rs35455589-G, are shared with chimpanzee and are therefore ancestral. Those ancestral alleles are also found in African populations with moderate frequencies, ranging from 1% to 57%, raising the question of whether those ancestral alleles on the introgressive haplotypes are the result of archaic hominin introgression or the direct inheritance from ancestors of modern humans from Africa.

To examine the two mutually exclusive hypotheses, for each allele that was both ancestral in the introgressive haplotypes and carried by the site in strong LD with rs12488302, we reconstructed a phylogenetic tree using haplotypes of 1KG individuals, Altai Neanderthal, Denisovans, and

chimpanzee. For all such alleles, the topologies of these trees are similar to each other. As an example, [figure 5](#) shows the tree for one such allele, rs35455589-G. The haplotypes carrying rs35455589-G form two clusters in the phylogenetic tree, one consisting of East Asian haplotypes, which cluster first with archaic hominin haplotypes and then join with the other haplotypes that are mostly of African origin. If rs35455589-G was directly inherited from African populations, one would expect a direct derivation of East Asian haplotypes from African ones. We therefore hypothesize that the haplotypes carrying rs35455589-G in East Asians were acquired from Neanderthal introgression rather than inherited directly from their ancestors in Africa. The African haplotypes carrying rs35455589-G were lost, probably due to genetic drift during the exodus of ancestors of modern Eurasians from Africa. Similar pattern is also observed for all aforementioned alleles at other sites (data not shown).

Divergence and Expansion Time Estimation for the Introgressive Haplotypes

Following the maximum likelihood-based method proposed by Mendez et al. (2012b), we estimated the divergence time between the introgressive haplotypes and the contemporary populations or the Altai Neanderthal. As required by the method, we calculated K_1 , K_2 , C , and A for each introgressive haplotype (see Materials and Methods), finding average values of 39, 27.19, 27.19, and 24.98, respectively. Based on these average values, we determined that the t and t_d with maximum likelihood were 429.75 and 34.84 KYA, respectively. [Figure 6A](#) shows the joint likelihood for M–N and N–R divergence. Our results show that the introgressive haplotypes diverged 429.75 KYA from ancestors of modern humans and 34.84 KYA from Altai Neanderthals, which is consistent with previous reports (Green et al. 2010; Mendez et al. 2012b).

We further employed Bayesian Markov chain Monte Carlo (MCMC) analysis and drew a Bayesian skyline plot (BSP) to interrogate the change of effective population size of the introgressive haplotypes in contemporary populations ([fig. 6B](#)). The BSP suggests that the introgressive haplotypes coalesced at approximately 45,000 years ago (Drummond et al. 2005). From 45,000 years to 5,000 years before present, the effective population size of the introgressive haplotypes increased at a steady rate. Notably, the growth rate of the effective population size increased at around 5,000 to 3,500 years before the present, suggesting a population expansion event.

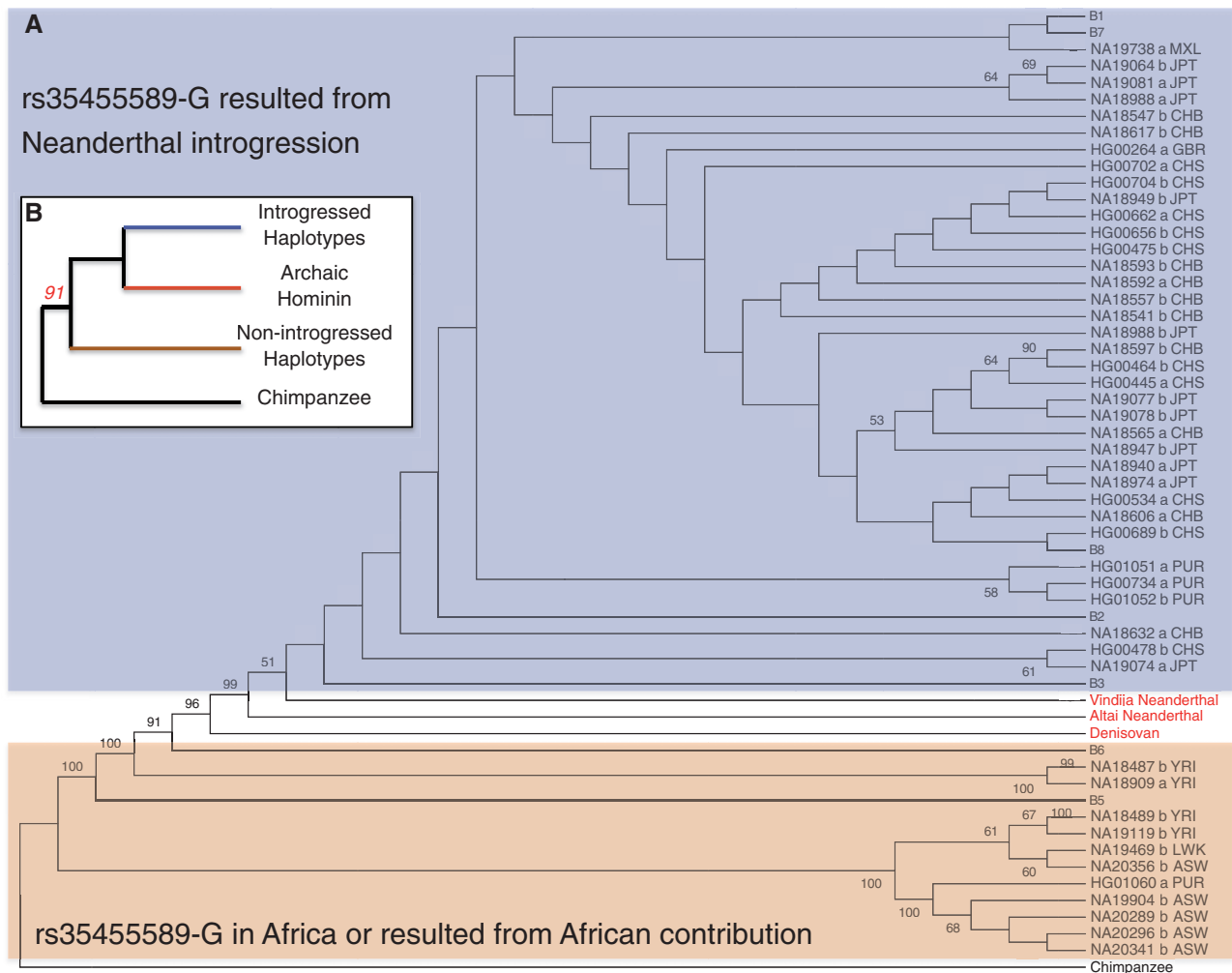


Fig. 5. Phylogenetic tree of the haplotypes carrying rs35455589-G. (A) Phylogenetic tree of non-recombinant haplotypes of contemporary populations carrying rs35455589-G, Altai Neanderthal, Denisovan, and chimpanzee with 1,000 bootstrap replications, reconstructed using the maximum likelihood method. A cluster of haplotypes (in blue) comprising the introgressive haplotypes, coalescing first with Neanderthal, rather than coalescing first with haplotypes of contemporary populations from Africa. Bootstrap values larger than 50 are shown. (B) A simplified illustration of the tree topology from panel (A). The italicized number in red is the bootstrap value at the node separating the introgressive/archaic haplotypes from other haplotypes of contemporary populations.

This observation could either result from demographic expansions during Neolithic time or a positive selection, which requires further investigation.

Discussion

Here, we present evidence of the introgression of a Neanderthal segment encompassing 18 genes into the genomes of contemporary populations. We further find evidence that the introgressive haplotypes are subjected to strong positive selection in East Asian populations. Interestingly, the introgressive haplotypes are almost East Asian specific, except for some sporadic distribution in Europe and South Asia. Our findings suggest that the introgression occurred within the ancestral population shared by East Asians and Native Americans, and the introgressive haplotypes in Europe resulted from recent East Asian gene flow. This “Asian-specific” Neanderthal introgression is consistent

with a previous report of higher levels of Neanderthal ancestry in East Asians than in Europeans (Wall et al. 2013).

Genes in the *HYAL* Region

There are 18 genes in the *HYAL* region, 15 of which are present in a smaller block defined by LD $r^2 > 0.80$ with rs12488302 in East Asian populations using 1KG Phase 1 data. These include *SLC38A3* (MIM 604437), *GNAI2* (MIM 139360), *SEMA3B*, *C3orf45*, *IFRD2* (MIM 602725), *HYAL3*, *NAT6*, *HYAL1*, *HYAL2*, *TUSC2* (MIM 607052), *RASSF1*, *ZMYND10* (MIM 607070), *NPRL2* (MIM 607072), *CYB561D2* (MIM 607078), and *CACNA2D2* (MIM 607082). Notably, the 3p21.3 region contains a cluster of well-studied tumor-suppressor genes (Hesson et al. 2007), which includes *HYAL2*, *RASSF1*, and *CACNA2D2* in the *HYAL* region.

SEMA3B encodes semaphorin 3B. The semaphorin protein family plays an important role in axon guidance during

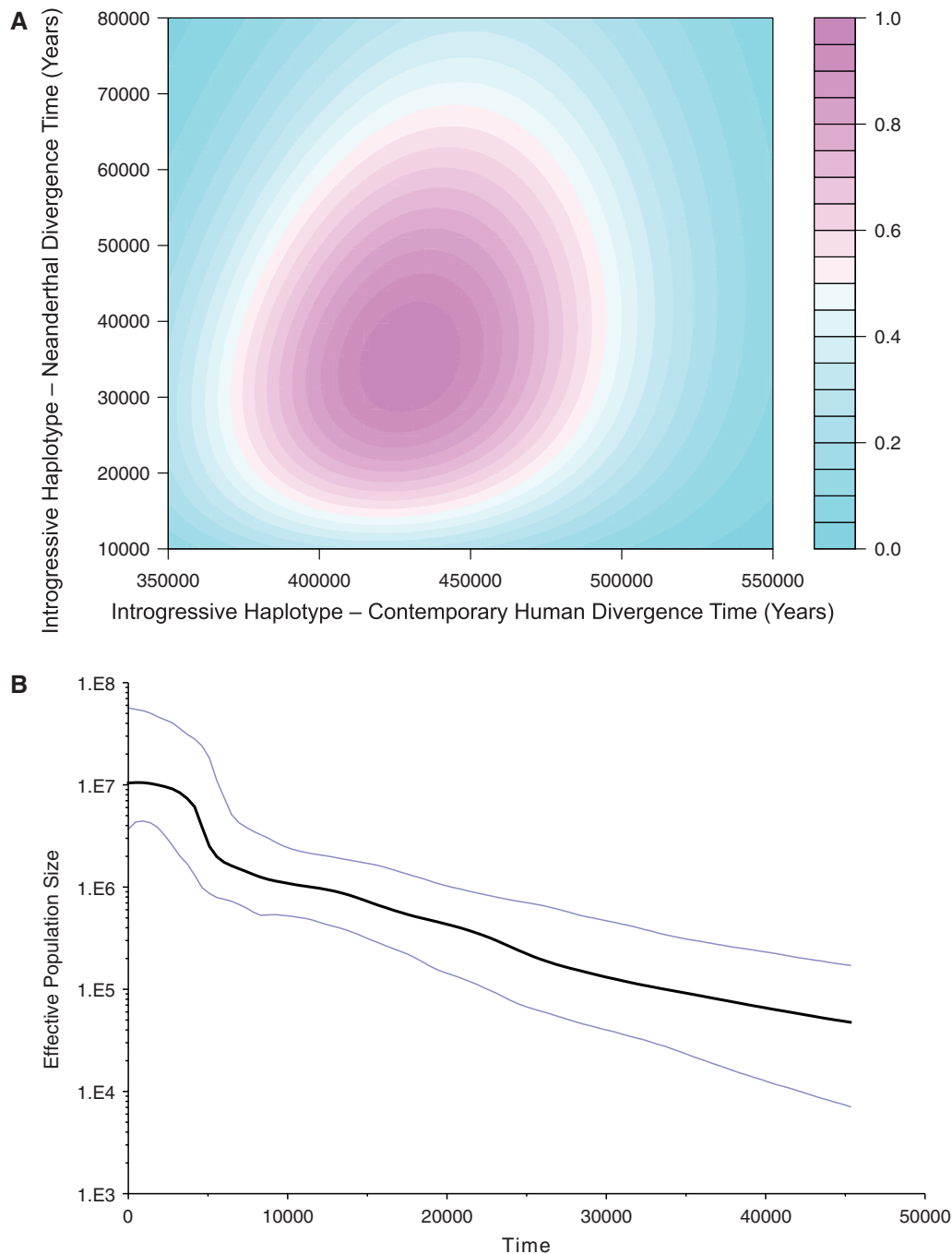


Fig. 6. Divergence and expansion time estimation for the introgressive haplotypes. (A) Joint likelihood for the divergence time between the introgressive haplotypes and contemporary populations or the Altai Neanderthal. (B) BSP of the introgressive haplotypes. The black line is the median estimate, and the blue lines indicate the 95% highest posterior density limits. The lines start at about 45,000 years ago, which is the coalescent time for all introgressive haplotypes in the analysis. Since 45,000 years ago, the growth rate of the effective population size remained steady until about 5,000 years ago, at which time the growth rate increased.

neuronal development, and *SEMA3B* can induce apoptosis in tumor cells and serves as a tumor suppressor (Castro-Rivera et al. 2004). Interestingly, a missense SNP at *SEMA3B* in strong LD with rs12488302 (rs2071203; $r^2 = 0.95$; table 1) is reportedly associated with prostate cancer risk. Homozygosity of the allele introgressed from Neanderthals (TT) increases the risk of prostate cancer by 2-fold in a Hispanic population (Beuten et al. 2009).

HYAL genes have attracted special attention due to their functional importance. Mutation of *HYAL1* in humans can result in mucopolysaccharidosis IX (Triggs-Raine et al. 1999). Deficiencies of *HYAL2* and *HYAL3* have not been reported in humans. *Hyal3* knockout mice show no obvious signs of defect compared with wild-type mice, suggesting that *Hyal3* does not play a major role in hyaluronan degradation (Atmuri et al. 2008). *Hyal2* knockout (*Hyal2*^{-/-}) mice are viable and

fertile but exhibit thrombocytopenia, hemolysis, birth defects of frontonasal and vertebral bone formation, and a 10-fold increase in plasma hyaluronan level (Chowdhury et al. 2013), suggesting that *Hyal2* plays an important role in mice, especially in hyaluronan metabolism. In naked mole rats, *HYAL2* was also found to play an important role in hyaluronan metabolism. High-molecular-mass hyaluronan reportedly accumulates in naked mole-rat tissues and is responsible for the naked mole rat's cancer resistance (Tian et al. 2013).

Latitudinal Gradient of Frequency of the Introgressive Haplotypes in East Asia

Analysis of the HapMap 3 and 1KG Phase 1 data shows a high frequency of introgressive haplotypes in East Asian populations. Notably, this frequency is higher in the southern Chinese population (CHS, 66.5%) than the Japanese population (JPT, 49.4%), raising questions about the existence of a correlation between latitude and frequency of introgressive haplotypes among East Asian populations. To further investigate this issue, we turned our attention to ALFRED for allele frequency information of rs12488302-T in East Asian populations. However, the ALFRED does not contain genotype information for rs12488302. As rs11130248 was in strong LD with rs12488302 ($r^2 = 0.97$), we used rs11130248-G as a proxy for rs12488302-T in East Asians when calculating frequency of the introgressive haplotypes. The location information of the samples was obtained from ALFRED.

Allele frequencies of rs11130248-G in East Asian populations are listed in [supplementary table S4, Supplementary Material](#) online. We conducted a linear regression analysis between the frequency of introgressive haplotypes and latitude of sampling location and observed a relatively weak negative correlation between the two variables ([supplementary fig. S4A, Supplementary Material](#) online), with a regression coefficient of -0.5028 . The correlation becomes more evident after we removed the Tibeto-Burman speaking populations with a well-documented recent migration from the north (Yi, Naxi, and Lahu; [supplementary fig. S4B, Supplementary Material](#) online; $r = -0.7508$). This correlation could have resulted from latitude-dependent positive natural selection. However, the population size in the ALFRED was relatively small, and thus the allele frequency information in these populations may be imprecise. A future more detailed sampling could resolve this issue.

UV Irradiation and Positive Selection

Considering the latitudinal gradient of frequency of the introgressive haplotypes, the selection force for the *HYAL* region might be latitude-dependent. Gene Ontology reports that *HYAL1*, *HYAL2*, and *HYAL3* are important in the cellular response to UV-B. We suspect that selection force at the *HYAL* region might be related to intensity of UV-B irradiation, which is also latitude-dependent (Shi et al. 2009). Previous evidence shows increased hyaluronan metabolism after exposure to UV-B (Hasova et al. 2011). Various cell types have been used to study hyaluronan metabolism after UV-B exposure, including fibroblasts and HaCaT cells (immortalized human

keratinocytes). In HaCaT cells, decreased *HYAL2* expression has been observed at the mRNA and protein levels after UV-B exposure (Hasova et al. 2011; Kurdykowski et al. 2011). However, Averbek et al. (2007) reported increased *HYAL2* mRNA expression level in HaCaT cells at 3 h after UV-B exposure. This study also reported that the *HYAL2* mRNA level in fibroblasts decreased at 3 h after UV-B exposure and increased at 24 h after UV-B exposure. Despite this remaining disagreement regarding the direction of the regulation, it is clear that the expression level of *HYAL2* is regulated by UV-B irradiation.

Interestingly, rs35455589, a non-synonymous SNP located in the fourth exon of *HYAL2*, is found to be in strong LD with rs12488302 ($r^2 = 0.97$) in East Asian populations. Changing from ancestral allele (G) to derived allele (T) has resulted in the replacement of leucine with isoleucine at the 418th amino acid of *HYAL2*. In mice, all *Hyal* genes have four exons, and exons 2–4 show highest level of sequence conservation (Shuttleworth et al. 2002). Both SiPhy and GERP algorithms predicted rs35455589 as conservative. Furthermore, EHH analyses ([fig. 4D–F](#)) revealed that rs35455589-G might be the target of positive selection.

rs11130248-G (in strong LD with rs35455589-G, $r^2 = 1.00$) has also been identified by a genome-wide association study as being related to keloid susceptibility in Japanese populations (Nakashima et al. 2010). Keloid results from overgrowth of granulation tissue and is most prevalent in Africa and Asia. Reduced hyaluronan levels have been observed in keloid tissue (Meyer et al. 2000). As *HYAL2* encodes an enzyme that degrades hyaluronan, we suspect that the increased keloid risk in African and Asian populations might be related to the polymorphism at rs35455589, although the same allele had different evolutionary history in these two groups. Further studies should be carried out to focus on the effect of the allele change at rs35455589 on *HYAL2* expression, hyaluronan metabolism, and keloid pathogenesis.

Material and Methods

Data Sources

Whole-genome sequencing and genotyping data were obtained from 1KG Phase 1 and HapMap 3, respectively. The 1KG Phase 1 data set contains low-coverage whole-genome sequencing data for 1,092 individuals from the following 14 populations: YRI, LWK (Luhya in Wubuye, Kenya), CEU, FIN, GBR, TSI, IBS (Iberian population in Spain), CHB, CHS, JPT, ASW, MXL, CLM, and PUR. This data set was used for reconstructing phylogeny and extracting frequency distribution. The HapMap 3 data set contains whole-genome genotyping data for 647 unrelated individuals from the following 10 populations: YRI, LWK, MKK (Massai in Kinyawa, Kenya), CEU, TSI, CHB, CHD, JPT, GIH (Gujarati Indians in Houston, TX), and ASW. This data set was used for detecting positive selection and extracting frequency distribution.

We obtained a draft of the Neanderthal genome from the Vindija Cave in Croatia from Green et al. (2010), a high-coverage Altai Neanderthal genome from the Max Planck Institute for Evolutionary Anthropology, and a high-coverage

Denisovan genome from Meyer et al. (2012). The BAM files were used to obtain sequence information. We used the same criteria as used in Reich et al. (2010) for the filtering process. In cases where multiple reads were observed on a given locus, one read was randomly selected. The chimpanzee genome (PanTro3) was obtained from the alignment of PanTro3 mapped to GRCh37 (hg19).

Detection of LD

Haploview (Barrett et al. 2005) was used to identify LD blocks in the *HYAL* region using SNPs with a minor allele frequency (MAF) >0.05 in HapMap 3 Release 27. We used an algorithm described by Gabriel et al. (2002) to identify LD blocks.

Removal of Recombinants and Identification of Haplotypes

Sequence information was obtained for all individuals in the 1KG Phase 1 data set using the preliminary boundaries. The phylogenetic relationship was reconstructed for the haplotypes using the maximum likelihood method implemented in MEGA 5.05 (Tamura et al. 2011).

An introgressive haplotype is assumed to carry two types of derived alleles: Type 1, those originated from mutations before the divergence of ancestors of modern humans and archaic hominins; and Type 2, those originated from mutations in the archaic hominin lineage after the divergence of ancestors of modern humans and archaic hominins. Type 2 alleles were considered to be a signature of archaic hominin introgression. Furthermore, allele statuses at SNPs harboring Type 2 alleles (denoted as Type 2 SNPs) could reflect the origin of haplotypes, with ancestral alleles reflecting non-introgressive haplotype and derived alleles reflecting introgressive haplotype.

As the presence of recombination may distort the topological structure of a phylogenetic tree, it is necessary to remove recombinant haplotypes prior to further analyses. Recombinants were identified based on Type 2 alleles. Ideally, if no recombination occurs between non-introgressive and introgressive haplotypes, allele observations at the Type 2 SNPs would be either all ancestral (for non-introgressive haplotypes) or all derived (for introgressive haplotypes). Haplotypes in which some Type 2 SNPs were of ancestral status while others were of derived status were considered recombinants and were removed. As this approach could only detect recombination events that occurred between the first and last Type 2 SNPs, we also redefined the boundaries of the analyzed region using the genome coordinates of these SNPs. The resulting data set without recombinant haplotypes and with redefined boundaries was used in all following analyses.

The identified Type 2 alleles were also used to demonstrate the archaic origin of the introgressive haplotypes. As mentioned, Type 2 SNPs can be considered a signature of archaic introgression. We observed that the Type 2 alleles existed in the introgressive haplotypes (supplementary table S5, Supplementary Material online). We further compared the allele observations at the Type 2 alleles with the Neanderthal

and Denisovan genomes to examine the relationship between the introgressive haplotypes and archaic hominins.

Phylogenetic Analysis of Haplotypes

We reconstructed a phylogenetic tree with all non-recombinant haplotypes of contemporary populations, Altai Neanderthal, Denisovan, and chimpanzee, using the maximum likelihood method in MEGA 5.05 with 100 bootstrap replications. For each allele that was both ancestral in the introgressive haplotypes and at a site in strong LD with rs12488302, a phylogenetic tree was reconstructed using haplotypes of 1KG individuals, Altai Neanderthal, Denisovans, and chimpanzee, with 1,000 bootstrap replications. The Vindija Neanderthal sequence was excluded from all phylogenetic analyses, because its low coverage led to numerous missing calls and ambiguous tree topology. To enhance the display of phylogenetic trees, we used the “topology only” option in MEGA, and trivial monophyletic clusters were collapsed. Chimpanzee was used as the outgroup in all phylogenetic analyses.

Test for Positive Selection

To detect positive selection, we used the *iHS* (Voight et al. 2006), which detects extensive LD and long-range haplotypes resulting from positive selection, and the CMS test (Grossman et al. 2010), which combines information from multiple tests, including F_{ST} , XP-EHH, and *iHS*. We used Whole-Genome Homozygosity Analysis and Mapping Machina (WHAMM) to calculate a standardized *iHS* based on the HapMap 3 data set for Africans (YRI), Europeans (CEU and TSI), and East Asians (CHB, JPT, and CHD). From the CMS Viewer, we obtained CMS scores based on HapMap Phase 2 for YRI, CEU, and CHB + JPT. Standardized $|iHS|$ and CMS scores were plotted for East Asian, European, and African populations.

The SNP with the highest CMS score was considered to be the SNP under positive natural selection. We used the *rehh* package (Gautier and Vitalis 2012) in R to draw the haplotype bifurcation graph and calculate the EHH (Sabeti et al. 2002) for both ancestral and derived alleles at the SNP. The LD r^2 score of the SNP with the highest CMS score and the SNP with the highest $|iHS|$ score was calculated based on the 1KG Phase 1 East Asian populations.

Global Distribution of the Introgressive Haplotypes

To obtain the global distribution of the introgressive haplotypes, we combined data from the 1KG and HapMap 3 data sets. For populations included in both data sets (YRI and LWK in Africa, CEU and TSI in Europe, and CHB and JPT in East Asia), data from the 1KG data set were used in frequency calculations. One of the Type 2-derived alleles (rs12488302-T) was used to define introgressive haplotypes. This definition was confirmed by phylogenetic tree, as all haplotypes carrying rs12488302-T coalesced with Neanderthal first, while the other haplotypes coalesced first with Africans.

To further investigate the global distribution, we incorporated data from the Allele Frequency Database

(Rajeevan et al. 2012), which did not include rs12488302. As rs11130248 was in strong LD ($r^2 = 0.97$) with rs12488302, we used rs11130248-G to define introgressive haplotypes in ALFRED. However, many of the ALFRED populations had only a small sample size; therefore, the allele frequencies in these populations might be imprecise and should be used with caution.

Searching for Targets of Positive Selection

We searched for SNPs that were linked to the SNP with the highest CMS score (rs12488302) with LD $r^2 > 0.8$ in 1KG Phase 1. HaploReg (Ward and Kellis 2012) was used to investigate the potential functions of the SNPs. We used the National Human Genome Research Institute's Catalog of Published Genome-Wide Association Studies database (Hindorf et al. 2009) to obtain additional information about SNPs that were identified by genome-wide association studies in this region. To further investigate gene functions in the region, we referred to their gene ontology annotations (Ashburner et al. 2000).

Divergence and Expansion Time Estimation for the Introgressive Haplotypes

To estimate the divergence time between the introgressive haplotypes and contemporary populations or the Altai Neanderthals, we used the method described by Mendez et al. (2012b). Given two lineages with complete sequence coverage (denoted as M and N, where N is closer to R than M) and one lineage with incomplete sequence coverage (denoted as R), the maximum likelihood-based method can be used to estimate the M–N divergence time and the N–R divergence time. In our analysis, lineage M was the human reference sequence (GRCh37), lineage N was the introgressive haplotype, and lineage R was the Altai Neanderthal. The method defined four variables: K_1 , K_2 , C, and A. K_1 and K_2 , respectively, were the numbers of mutations in the lineages M and N since their divergence, C was the number of K_2 mutations with alleles observed in the lineage R, and A was the number of K_2 mutations carried by the lineage R at those sites with alleles observed. The likelihood function reached maximum at $t = K_1 + K_2/2 \times \mu$ and $t_d = t \times (1 - (A/C))$, where t is the M–N divergence time, t_d is the N–R divergence time, and μ is the mutation rate for the entire region (in nucleotides/year). We calibrated μ using a human–chimpanzee divergence time of 6.5 Ma (Goodman 1999) and obtained a value of 7.70×10^{-5} /year.

Recombination event detection for the introgressive haplotypes was performed using four algorithms, including the RDP, MaxChi, GENECONV, and Chimaera via the Recombination Detection Program version 3 (Martin et al. 2010). If no recombination event was detected, MCMC analysis was employed to estimate the expansion time of the introgressive haplotypes with BEAST v1.7.5 (Drummond et al. 2012). Markov chain length was set to 30,000,000. Based on the estimation of Roach et al. (2010), we used a mutation rate of $4.4e-10$ per nucleotide per year. Tracer v1.5 was used to analyze the data generated by BEAST. To obtain

an effective sample size of 200, two independent runs were combined in Tracer. The BSP was drawn to reflect the change of the effective population size of the introgressive haplotypes over time.

Supplementary Material

Supplementary figures S1–S4 and tables S1–S7 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

The authors are grateful to two anonymous reviewers for their constructive comments. They thank Dr Hua Zhong for technical assistance and editors from the San Francisco Edit for editing and proofreading the manuscript. This work was supported by the National Science Foundation of China (31271338, 31330038) and the National Basic Research Program (2012CB944600). Q.D. was also supported by the Chinese Ministry of Education Top Talent Undergraduate Training Program and the National Science Foundation for Fostering Talents in Basic Research of the National Natural Science Foundation of China.

References

- 1000 Genomes Consortium. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* 491:56–65.
- Abi-Rached L, Jobin MJ, Kulkarni S, McWhinnie A, Dalva K, Gragert L, Babrzadeh F, Gharizadeh B, Luo M, Plummer FA, et al. 2011. The shaping of modern human immune systems by multiregional admixture with archaic humans. *Science* 334:89–94.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. 2000. Gene Ontology: tool for the unification of biology. *Nat Genet.* 25:25–29.
- Atmuri V, Martin DC, Hemming R, Gutsol A, Byers S, Sahebjam S, Thliveris JA, Mort JS, Carmona E, Anderson JE, et al. 2008. Hyaluronidase 3 (HYAL3) knockout mice do not display evidence of hyaluronan accumulation. *Matrix Biol.* 27:653–660.
- Averbeck M, Gebhardt CA, Voigt S, Beilharz S, Anderegg U, Termeer CC, Sleeman JP, Simon JC. 2007. Differential regulation of hyaluronan metabolism in the epidermal and dermal compartments of human skin by UVB irradiation. *J Invest Dermatol.* 127:687–697.
- Barrett JC, Fry B, Maller J, Daly MJ. 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263–265.
- Beuten J, Garcia D, Brand TC, He X, Balic I, Canby-Hagino E, Troyer DA, Baillargeon J, Hernandez J, Thompson IM, et al. 2009. Semaphorin 3B and 3F single nucleotide polymorphisms are associated with prostate cancer risk and poor prognosis. *J Urol.* 182:1614–1620.
- Cann RL, Stoneking M, Wilson AC. 1987. Mitochondrial DNA and human evolution. *Nature* 325:31–36.
- Castro-Rivera E, Ran S, Thorpe P, Minna JD. 2004. Semaphorin 3B (SEMA3B) induces apoptosis in lung and breast cancer, whereas VEGF165 antagonizes this effect. *Proc Natl Acad Sci U S A.* 101: 11432–11437.
- Chowdhury B, Hemming R, Hombach-Klonisch S, Flamion B, Triggs-Raine B. 2013. Murine hyaluronidase 2 deficiency results in extracellular hyaluronan accumulation and severe cardiopulmonary dysfunction. *J Biol Chem.* 288:520–528.
- Drummond AJ, Rambaut A, Shapiro B, Pybus OG. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol.* 22:1185–1192.
- Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol.* 29: 1969–1973.

- Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, et al. 2002. The structure of haplotype blocks in the human genome. *Science* 296: 2225–2229.
- Gautier M, Vitalis R. 2012. rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics* 28:1176–1177.
- Goodman M. 1999. The genomic record of humankind's evolutionary roots. *Am J Hum Genet.* 64:31–39.
- Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH, et al. 2010. A draft sequence of the Neandertal genome. *Science* 328:710–722.
- Grossman SR, Shlyakhter I, Karlsson EK, Byrne EH, Morales S, Frieden G, Hostetter E, Angelino E, Garber M, Zuk O, et al. 2010. A composite of multiple signals distinguishes casual variants in regions of positive selection. *Science* 327:883–886.
- Hammer MF. 1995. A recent common ancestry for human Y chromosomes. *Nature* 378:376–378.
- Hasova M, Crhak T, Safrankova B, Dvorakova J, Muthny T, Velebny V, Kubala L. 2011. Hyaluronan minimizes effects of UV irradiation on human keratinocytes. *Arch Dermatol Res.* 303:277–284.
- Hesson LB, Cooper WN, Latif F. 2007. Evaluation of the 3p21.3 tumour-suppressor gene cluster. *Oncogene* 26:7283–7301.
- Hinch AG, Tandon A, Patterson N, Song Y, Rohland N, Palmer CD, Chen GK, Wang K, Buxbaum SG, Akyzbekova EL, et al. 2011. The landscape of recombination in African Americans. *Nature* 476:170–175.
- Hindorf LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A.* 106:9362–9367.
- International HapMap 3 Consortium. 2010. Integrating common and rare genetic variation in diverse human populations. *Nature* 467: 52–58.
- Kurdykowski S, Mine S, Bardey V, Danoux L, Jeanmaire C, Pauly G, Brabencova E, Wegrowski Y, Maquart FX. 2011. Ultraviolet-B irradiation induces differential regulations of hyaluronidase expression and activity in normal human keratinocytes. *Photochem Photobiol.* 87:1105–1112.
- Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefevre P. 2010. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* 26:2462–2463.
- Mendez FL, Watkins JC, Hammer MF. 2012a. Global genetic variation at OAS1 provides evidence of archaic admixture in Melanesian populations. *Mol Biol Evol.* 29:1513–1520.
- Mendez FL, Watkins JC, Hammer MF. 2012b. A haplotype at STAT2 introgressed from Neanderthals and serves as a candidate of positive selection in Papua New Guinea. *Am J Hum Genet.* 91:265–274.
- Mendez FL, Watkins JC, Hammer MF. 2013. Neandertal origin of genetic variation at the cluster of OAS immunity genes. *Mol Biol Evol.* 30: 798–801.
- Meyer LJM, Russell SB, Russell JD, Trupin JS, Egbert BM, Shuster S, Stern R. 2000. Reduced hyaluronan in keloid tissue and cultured keloid fibroblasts. *J Invest Dermatol.* 114:953–959.
- Meyer M, Kircher M, Gansauge MT, Li H, Racimo F, Mallick S, Schraiber JG, Jay F, Prüfer K, de Filippo C, et al. 2012. A high-coverage genome sequence from an archaic Denisovan individual. *Science* 338: 222–226.
- Nakashima M, Chung S, Takahashi A, Kamatani N, Kawaguchi T, Tsunoda T, Hosono N, Kubo M, Nakamura Y, Zembutsu H. 2010. A genome-wide association study identifies four susceptibility loci for keloid in the Japanese population. *Nat Genet.* 42:768–771.
- Rajeevan H, Soundararajan U, Kidd JR, Pakstis AJ, Kidd KK. 2012. ALFRED: an allele frequency resource for research and teaching. *Nucleic Acids Res.* 40:D1010–D1015.
- Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, Viola B, Briggs AW, Stenzel U, Johnson PL, et al. 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468: 1053–1060.
- Roach JC, Glusman G, Smit AFA, Huff CD, Hubley R, Shannon PT, Rowen L, Pant KP, Goodman N, Bamshad M, et al. 2010. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* 328:636–639.
- Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, et al. 2002. Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419:832–837.
- Shi H, Tan SJ, Zhong H, Hu W, Levine A, Xiao CJ, Peng Y, Qi XB, Shou WH, Ma RL, et al. 2009. Winter temperature and UV are tightly linked to genetic changes in the p53 tumor suppressor pathway in Eastern Asia. *Am J Hum Genet.* 84:534–541.
- Shuttleworth TL, Wilson MD, Wicklow BA, Willkins JA, Triggs-Raine BL. 2002. Characterization of the murine hyaluronidase gene region reveals complex organization and cotranscription of Hyal1 with downstream genes, Fus2 and Hyal3. *J Biol Chem.* 277:23008–23018.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 28:2731–2739.
- Tian X, Azpurua J, Hine C, Vaidya A, Myakishev-Rempel M, Ablaeva J, Mao Z, Nevo E, Gorbunova V, Seluanov A. 2013. High-molecular-mass hyaluronan mediates the cancer resistance of the naked mole rat. *Nature* 499:346–349.
- Triggs-Raine B, Salo TJ, Zhang H, Wicklow BA, Natowicz MR. 1999. Mutations in HYAL1, a member of a tandemly distributed multi-gene family encoding disparate hyaluronidase activities, cause a newly described lysosomal disorder, mucopolysaccharidosis IX. *Proc Natl Acad Sci U S A.* 96:6296–6300.
- Voight BF, Kudravalli S, Wen X, Pritchard JK. 2006. A map of recent positive selection in the human genome. *PLoS Biol.* 4:e72.
- Wall JD, Yang MA, Jay F, Kim SK, Durand EY, Stevison LS, Gignoux C, Woerner A, Hammer MF, Slatkin M. 2013. Higher levels of Neandertal ancestry in East Asians than in Europeans. *Genetics* 194:199–209.
- Ward LD, Kellis M. 2012. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* 40:D930–D934.