

Neanderthal Origin of the Haplotypes Carrying the Functional Variant Val92Met in the *MC1R* in Modern Humans

Qiliang Ding,^{†,1} Ya Hu,^{†,1} Shuhua Xu,² Chuan-Chao Wang,¹ Hui Li,¹ Ruyue Zhang,¹ Shi Yan,¹ 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 Institute for Biological Sciences (SIBS), Chinese Academy of Sciences (CAS), Shanghai, China

[†]These authors contributed equally to this work.

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

Associate Editor: Joshua Akey

Abstract

Skin color is one of the most visible and important phenotypes of modern humans. Melanocyte-stimulating hormone and its receptor played an important role in regulating skin color. In this article, we present evidence of Neanderthal introgression encompassing the melanocyte-stimulating hormone receptor gene *MC1R*. The haplotypes from Neanderthal introgression diverged with the Altai Neanderthal 103.3 ka, which postdates the anatomically modern human–Neanderthal divergence. We further discovered that all of the putative Neanderthal introgressive haplotypes carry the Val92Met variant, a loss-of-function variant in *MC1R* that is associated with multiple dermatological traits including skin color and photoaging. Frequency of this Neanderthal introgression is low in Europeans (~5%), moderate in continental East Asians (~30%), and high in Taiwanese aborigines (60–70%). As the putative Neanderthal introgressive haplotypes carry a loss-of-function variant that could alter the function of *MC1R* and is associated with multiple traits related to skin color, we speculate that the Neanderthal introgression may have played an important role in the local adaptation of Eurasians to sunlight intensity.

Key words: Neanderthal, introgression, skin color, local adaptation, *MC1R*.

Introduction

Skin color is one of the most visible phenotypes of modern human. Skin color is important for its role in the adaptation of modern human to ultraviolet (UV) irradiation intensity (Jablonski 2004).

Skin pigmentation is mainly caused by the accumulation of pigment melanin in melanocytes in the bottom layer of the epidermis. Skin color is determined by the relative proportion of eumelanin (black pigment) and pheomelanin (red pigment), and this proportion is regulated in part by the melanocyte-stimulating hormone (MSH) and its receptor (Valverde et al. 1995). In human, MSH receptor is encoded by the *MC1R* (MIM 155555). Some loss-of-function variants in *MC1R* gene could alter the process of melanogenesis by reducing the production of eumelanin, which results in fair skin and red hair. So far, five loss-of-function variants in *MC1R* are found related to fair skin and red/blond hair: Val60Leu (rs1805005*T), Val92Met (rs2228479*A), Arg151Cys (rs1805007*T), Arg160Trp (rs1805008*T), and Asp294His (rs1805009*C) (Valverde et al. 1995, 1996; Box et al. 1997; Smith et al. 1998; Mogil et al. 2003). Further, a loss-of-function variant of *MC1R* (Arg307Gly) that is absent in modern human is identified in two Neanderthal individuals (Lalueza-Fox et al. 2007).

The Val92Met variant of *MC1R* is found as one of the most frequent *MC1R* mutations associated with red hair and Type

I/II skin (always burn, never or slightly tan; Valverde et al. 1995). The Val92Met is suspected to alter the α -helix structure of the second transmembrane domain of *MC1R*. In vitro study found that the Val92Met variant could reduce the affinity of *MC1R* to α -MSH by approximately 5-fold, suggesting that the Val92Met variant alters normal functions of the *MC1R* (Xu et al. 1996). Significant gene–gene interaction that could affect skin color of Tibetans is identified between the Val92Met of *MC1R* and the IVS13-15 of *OCA2* (Akey et al. 2001). Moreover, the Val92Met is also identified by a genome-wide association study of skin color in South Asians in one of its studied cohorts (Stokowski et al. 2007). Study on a Japanese population found the Val92Met is significantly associated with freckles and solar lentigines (Motokawa et al. 2008). Study on French also found Val92Met is associated with increased risk of photoaging (Elfakir et al. 2010). The Val92Met is found associated with malignant melanoma in a Spanish population (Fernandez et al. 2007). Furthermore, the Val92Met is also found associated with the effect of desipramine treatment on depression patients (Wu et al. 2011). All above functional and association studies implicate that the Val92Met variant is functional, and may play an important role in human response to sunlight.

Neanderthals are our close relatives that became extinct about 30,000 years ago (Finlayson et al. 2006). Y chromosome and mitochondrial DNA studies identified no evidence of

Neanderthal introgression in modern human (Hodgson and Disotell 2008). However, recent studies based on whole genome sequences of archaic hominin revealed Neanderthal and Denisovan introgression in modern Eurasians (Green et al. 2010; Reich et al. 2010; Prüfer et al. 2014). Segments from Neanderthal introgression account for less than 4% of the Eurasian genomes. Despite its small portion in the Eurasian genomes, several segments from Neanderthal are considered to play an important role in the local adaptation of modern Eurasians. Evidences of archaic introgression are found at the *HLA-A* (MIM 142800), *HLA-B* (MIM 142830), *HLA-C* (MIM 142840) (Abi-Rached et al. 2011), *HLA-DPB1* (MIM 142858) (Temme et al. 2014), *OAS1* (MIM 164350) (Mendez et al. 2012b), and *STAT2* (MIM 600556) (Mendez et al. 2012a). Further, Neanderthal introgression encompassing *HYAL2* (MIM 603551), a gene related to the cellular response to UV-B irradiation, is found under positive natural selection in East Asians (Ding, Hu, Xu, et al. 2014).

In this article, we present analyses of the genomic region encompassing the *MC1R* gene. We identified haplotypes from Neanderthal introgression in modern Eurasians (denoted as putative introgressive haplotypes) by comparing the Eurasian haplotypes with the Altai Neanderthal on the ancestry informative markers (AIMs) of the Altai Neanderthal. Neanderthal origin of those putative introgressive haplotypes was supported by further analyses. Remarkably, we found that the almost all of the derived alleles at the Val92Met variant of *MC1R* (rs2228479*A) in the human gene pool are carried by the putative introgressive haplotypes. As the Val92Met variant is reported associated with multiple dermatological traits including skin color and photoaging, we speculate that this Neanderthal introgression, together with the previously reported Neanderthal introgression at *HYAL2* (Ding, Hu, Xu, et al. 2014), may have played an important role in the local adaptation of modern Eurasians to sunlight intensity.

Results

Neanderthal Ancestry Encompassing *MC1R*

The *MC1R* gene is located on chromosome 16q24.3, spanning 3,099 bp. To interrogate possible Neanderthal ancestry encompassing the *MC1R*, we obtained sequence information of *MC1R* and its left and right 100-kb flanking regions (chr16: 89,884,286–90,087,384) for all the Eurasian and African haplotypes in the 1000 Genomes Project data set (denoted as the 1KG data set; see Materials and Methods). We then searched for Neanderthal AIMs encompassing the *MC1R*. Single nucleotide polymorphisms (SNPs) that resulted from mutations in Neanderthal lineage after the Neanderthal–anatomically modern human (AMH) divergence were considered as Neanderthal AIMs. In total, 59 Neanderthal AIMs were identified within the *MC1R* gene and its left and right 100-kb flanking regions (supplementary table S1, Supplementary Material online). It was observed that a number of non-African haplotypes carry segments with Neanderthal ancestry (see supplementary fig. S1, Supplementary Material online),

implicating the presence of Neanderthal introgression encompassing *MC1R*.

It was observed that some haplotypes are a mosaic of Neanderthal and AMH segments in supplementary figure S1, Supplementary Material online, indicating the existence of recombination between Neanderthal and non-Neanderthal segments. As recombination could interfere with phylogenetic analyses and time estimation, we shrank the analyzed region to chr16: 89,941,438–90,008,296 and removed those haplotypes that show evidence of recombination within the aforementioned region. In total, we included 97.5% (1,660/1,702) of the 1KG haplotypes in the new data set, and 98.7% (155/157) of the haplotypes with inferred Neanderthal ancestry encompassing *MC1R* were included in the new data set. For validation, we scanned the new data set using the Recombination Detection Program (RDP) v3.4 (Martin et al. 2010) with various algorithms (RDP, MaxChi, Chimaera, and GENECONV), and no recombination event was detected in the new data set. This new data set was used in all the following analyses.

Phylogenetic Evidence of Neanderthal Introgression

Given that preliminary evidence of Neanderthal introgression encompassing *MC1R* was demonstrated by the aforementioned AIM-based analysis, we further investigated the putative Neanderthal introgression using a phylogenetic approach.

We reconstructed a phylogenetic tree using all nonrecombinant haplotypes in the 1KG data set and the sequence of two archaic hominins (Altai Neanderthal and Denisovan) using a maximum-likelihood approach implemented in MEGA 6 (Tamura et al. 2013), with 100 bootstrap replications (fig. 1). Chimpanzee sequence (panTro3) was used to root the phylogenetic tree. It was observed that all haplotypes with an inferred Neanderthal origin by the AIM-based analysis (denoted as the “putative introgressive haplotypes” hereafter) coalesce first with the archaic hominins before they joining the rest of contemporary haplotypes. This conclusion was supported by the bootstrap value of the node distinguishing the archaic/introgressive haplotypes from other haplotypes. Further, the phylogenetic analysis also suggests that the putative introgressive haplotypes are closer to Altai Neanderthal than Denisovan. Thus, the phylogenetic analysis supports the Neanderthal introgression hypothesis, and also suggests that the putative introgressive haplotypes are closer to Altai Neanderthal than Denisovan.

Further Evidences of Neanderthal Introgression

Both AIM-based analysis and phylogenetic topology support a Neanderthal origin for the putative introgressive haplotypes. In addition to those analyses, we reconstructed the ancestral sequence for the putative introgressive haplotypes using a Bayesian Markov Chain Monte Carlo (MCMC) approach. It was observed that the Altai Neanderthal, the Denisovan, and the human reference sequence share 97.2% (1,060/1,090), 95.6% (1,033/1,081), and 85.5% (950/1,111) of the alleles with the ancestral sequence, respectively. This result is consistent with the earlier observation that the putative

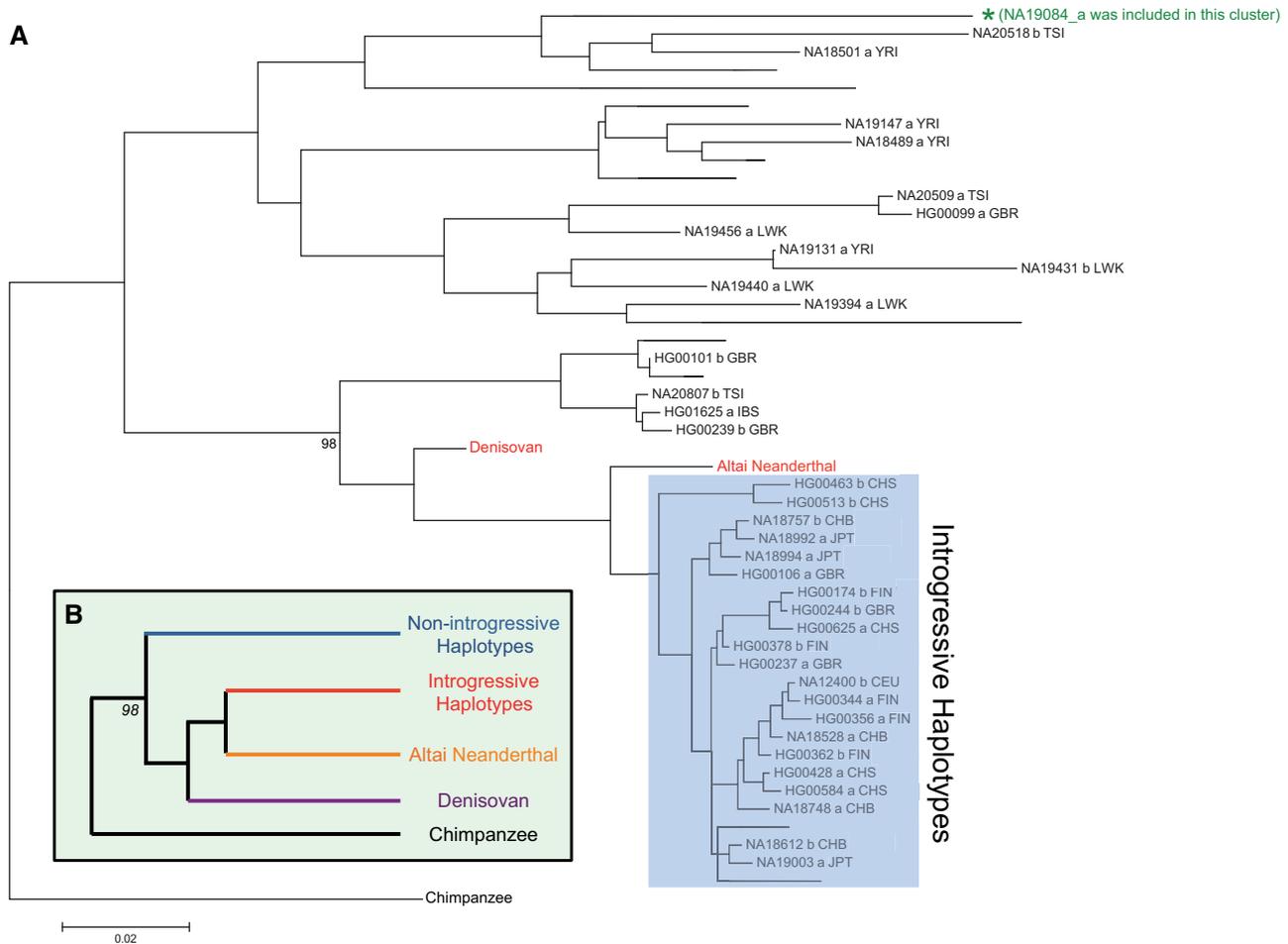


Fig. 1. Phylogenetic tree of haplotypes. (A) Phylogenetic tree of all nonrecombinant contemporary haplotypes, Altai Neanderthal, Denisovan, and chimpanzee reconstructed using the maximum-likelihood method in MEGA 6 with 100 bootstrap replications. Chimpanzee sequence was used to root the tree. Trivial monophyletic clusters were collapsed to enhance the display of the phylogenetic tree without losing the major topology. The collapsed monophyletic cluster that includes the NA19084_a haplotype was marked with a green asterisk (*). (B) A simplified illustration of the tree topology in panel (A). It was observed that the bootstrap value of the node distinguishing the archaic/introgressive haplotypes from other haplotypes is high (98).

introgressive haplotypes are closer to Altai Neanderthal than they are to Denisovan or human reference sequence.

Our previous study suggests that two types of derived alleles characterize an introgressive haplotype (Ding, Hu, Xu, et al. 2014). Specifically, the Type-1 derived alleles are from mutations that occurred before the archaic hominin–AMH divergence, whereas the Type-2 derived alleles are from mutations occurred in the archaic hominin lineage after the archaic hominin–AMH divergence. The Type-2 derived alleles could be considered as a signature of archaic introgression. We obtained a polymorphism table for contemporary and archaic hominin haplotypes (supplementary table S2, Supplementary Material online), and searched for Type-2 alleles in the putative introgressive haplotypes. In total, 21 Type-2 alleles were found within the analyzed region (supplementary table S3, Supplementary Material online). Frequencies of the putative introgressive variants (i.e., Type-2 alleles) were also listed in supplementary table S3, Supplementary Material online. It was observed that all of the putative introgressive variants were absent in Africans, further validating their archaic origin.

All above analyses suggest a fairly close relationship between the putative introgressive haplotypes and the Altai Neanderthal, which could be explained by recent Neanderthal introgression. However, an alternative model—the ancestral polymorphism model might also explain such observations (Ding, Hu, Xu, et al. 2014). According to this alternative model, the putative introgressive haplotypes existed in the human gene pool before the divergence between Neanderthal and AMH, but were subsequently lost in the contemporary Africans whereas preserved in Neanderthals and contemporary Eurasians. According to this model, the putative introgressive haplotypes existed in the human gene pool before 270–440 ka (Green et al. 2010), which is much earlier than the time of Neanderthal introgression (<100 ka). We examined this alternative model using a recombination-based method (see Materials and Methods; Mendez et al. 2012b; Ding, Hu, Xu, et al. 2014), and the ancestral polymorphism model was rejected as 96.9% (154/159) of the putative introgressive haplotypes have $P \leq 3.1 \times 10^{-4}$ (the significance threshold adjusted for 159 tests using Bonferroni correction).

Divergence Time and Time to Most Recent Common Ancestor Estimation

We then investigated the relationship between the putative introgressive haplotypes and other hominins by estimating the divergence time and time to most recent common ancestor (TMRCA) of the putative introgressive haplotypes. We employed a maximum likelihood-based method to estimate the divergence times between the putative introgressive haplotypes and the AMH or the Altai Neanderthal or the Denisovan haplotypes (Mendez et al. 2012a; Ding, Hu, Xu, et al. 2014). As required by this method, we computed K_1 , K_2 , C , and A for each haplotype. Averaged K_1 , K_2 , C , and A are shown in the [supplementary table S4, Supplementary Material online](#). Based on the values, it could be estimated that the putative introgressive haplotypes diverged with the AMH 552.5 ka, then diverged with the Denisovan 159.0 ka, and then diverged with the Altai Neanderthal 103.3 ka ([fig. 2](#)). It could be observed that the divergence time between the putative introgressive haplotypes and the Altai Neanderthal (103.3 ka) postdates the reported Neanderthal–AMH divergence time (>270 ka), which indicates the presence of postdivergence introgression.

We then employed a Bayesian MCMC analysis to estimate the TMRCA of all putative introgressive haplotypes (Drummond et al. 2012). Human–chimpanzee divergence time was set to 6,500 ka to calibrate molecular clock (Goodman 1999). The TMRCA of all putative introgressive haplotypes is 147.7 ka (95% confidence interval [CI]: 115.8–181.1 ka). This TMRCA predates the time of out-of-Africa migration of AMH (<100 ka; Li and Durbin 2011), which suggests that more than one copy of Neanderthal haplotype entered the modern human gene pool at the beginning.

The *MC1R* Val92Met Variant and the Putative Introgressive Haplotypes

To summarize, the above analyses provide convincing evidence that those putative introgressive haplotypes resulted from a recent Neanderthal introgression. Here, we evaluate the functional importance of this introgression. In particular, we investigated whether any of the five loss-of-function SNPs within the *MC1R* listed in OMIM (i.e., Val60Leu, Val92Met, Arg151Cys, Arg160Trp, and Asp294His) was exclusively or inclusively carried by the putative introgressive haplotypes.

Remarkably, it was observed that all putative introgressive haplotypes carry the derived allele of the Val92Met variant (rs2228479*A), and almost none (except the NA19084_a, as discussed below) of the nonintrogressive haplotypes carries the rs2228479*A. This suggests that the rs2228479*A carried by the putative introgressive haplotypes might be from archaic hominin lineage. However, rs2228479*A was not found in the Altai Neanderthal genome. Therefore, it is also possible that the mutation creating the rs2228479*A on the putative introgressive haplotypes occurred shortly after the putative introgressive haplotype entering modern human gene pool. We discussed this issue in Discussion that both possibilities might explain the origin of rs2228479*A on the putative introgressive haplotype.

As mentioned earlier, the rs2228479*A was also found carried by a nonintrogressive haplotype (from one Japanese individual in Tokyo, Japan, NA19084, denoted as NA19084_a). On Neanderthal AIMs, none of the alleles on NA19084_a matches Altai Neanderthal alleles, indicating that NA19084_a is not from Neanderthal introgression ([supplementary table S2, Supplementary Material online](#)). Further, NA19084_a shows a closer relationship to other nonintrogressive haplotypes than to the archaic hominin haplotypes and the putative introgressive haplotypes on phylogenetic tree ([fig. 1](#)), which also indicates that NA19084_a is not from archaic hominin introgression. Thus, we first suspect that the derived allele at rs2228479 on NA19084_a might be an SNP calling error in the 1000 Genomes Project. To validate the SNP calls for NA19084_a, we obtained DNA of the sample NA19084 from the Coriell Institute for Medical Research and genotyped the sample for rs2228479 and four flanking Type-2 SNPs (SNPs that harboring Type-2 derived alleles, rs11648433, rs62052212, rs62052213, and rs12600051, listed in [supplementary table S5, Supplementary Material online](#)). In results, our genotyping result is consistent with the 1000 Genomes Project ([supplementary fig. S2A, Supplementary Material online](#)), indicating that the derived allele at rs2228479 on NA19084_a is genuine, and might be from recurrent mutation, double recombination, or biased gene conversion. We further discussed the three possibilities in Discussion.

Furthermore, among all HapMap African individuals, it was observed that one YRI (Yoruba in Ibadan, Nigeria; NA18852) and two MKK (Massai in Kinyawa, Kenya; NA21339 and NA21574) carry heterozygous alleles at the rs2228479. We discussed this observation in Discussion and concluded that the observation did not affect our conclusion.

Global Distribution of the Putative Neanderthal Introgression

As suggested by the above analysis, the derived allele at Val92Met variant (rs2228479*A) in contemporary human populations may be of Neanderthal origin. However, the rs2228479*A was also carried by one nonintrogressive haplotype (NA19084_a) in the 1KG data set. To further search for cases of rs2228479*A carried by nonintrogressive haplotypes and to investigate the global distribution of archaic introgression encompassing *MC1R*, we genotyped a total of 757 individuals from 31 global populations for rs2228479 and four flanking Type-2 SNPs (i.e., rs11648433, rs62052212, rs62052213, and rs12600051; see [supplementary table S5, Supplementary Material online](#)). Frequencies of rs2228479 and the four flanking Type-2 SNPs in the populations genotyped by this study were listed in [supplementary table S6, Supplementary Material online](#).

We first searched for cases of rs2228479*A carried by nonintrogressive haplotypes. rs2228479*A will be considered carried by nonintrogressive haplotype if the given haplotype carrying the rs2228479*A contains none of the four Type-2 alleles on the four Type-2 SNPs. However, among all 757 individuals we genotyped, all haplotypes carrying rs2228479*A

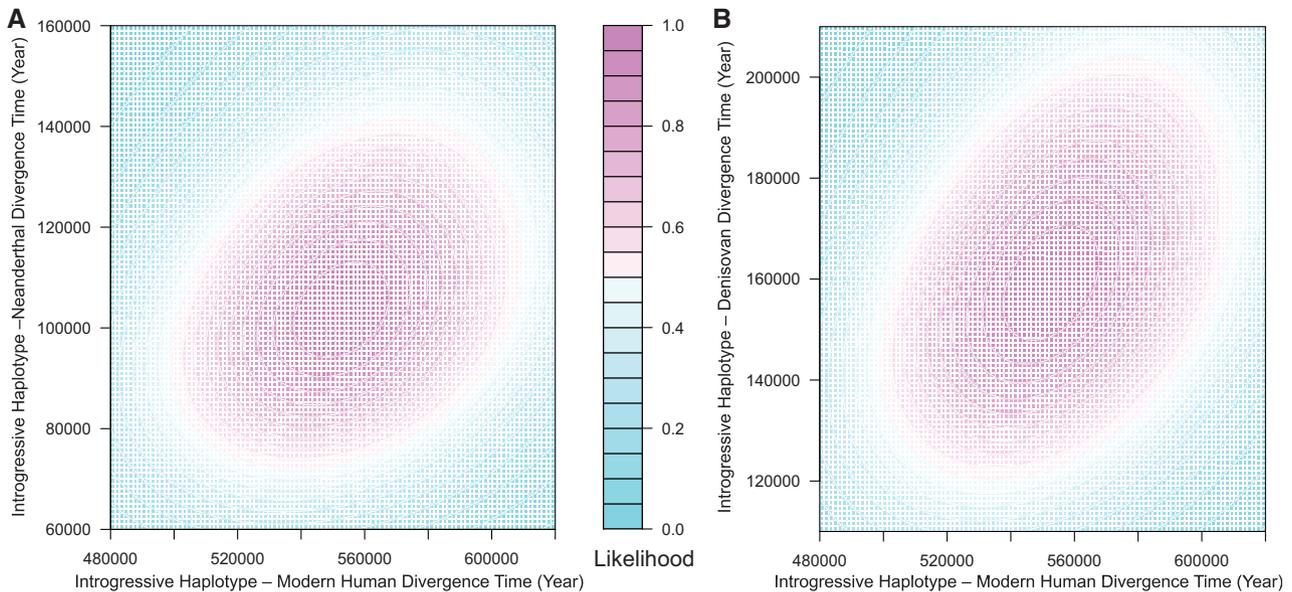


Fig. 2. Divergence estimation of the putative introgressive haplotypes. (A) Joint likelihood for the divergence time between the putative introgressive haplotypes and modern human (X axis) or the Altai Neanderthal (Y axis). (B) Joint likelihood for the divergence time between the putative introgressive haplotypes and modern human (X axis) or the Denisovan (Y axis).

contain more than one Type-2 allele on the four Type-2 SNPs (supplementary table S5, Supplementary Material online), which indicates that no further case of rs2224879*A carried by nonintrogressive haplotype could be found in our samples.

We then investigated the global distribution of the Neanderthal introgression encompassing *MC1R*. For the 1KG populations, a Neanderthal AIM (rs117406236*T) was used as diagnostic SNP of the Neanderthal introgression. For the HapMap 3 populations, we used rs2228479*A as diagnostic SNP of the Neanderthal introgression, as no nearby Neanderthal AIM was genotyped by HapMap. For the populations genotyped in this study, a Neanderthal AIM (rs11648433*A) was used as diagnostic SNP of the Neanderthal introgression. Frequencies of Neanderthal introgression in global populations were listed in supplementary tables S6–S8, Supplementary Material online, and plotted on fig. 3.

Furthermore, average frequencies of the rs2228479*A in East and Southeast Asians and Europeans are 31.82% and 5.52%, respectively. Moreover, frequency of rs2228479*A is significantly higher in the East/Southeast Asians comparing to the Europeans (Student's *t*-test, $P = 0.0003$). We also noticed that frequency of this allele is unusually high in the two populations of Taiwanese aborigines, including Atayal (71.05%) and Ami (59.26%). Sample sizes for the two populations are sufficiently large ($2n = 76$ and 54 , respectively).

Discussion

Cross Examination of Evidences

In this article, we present evidences of Neanderthal introgression encompassing the MSH receptor gene *MC1R*. Furthermore, our evidences support that the derived allele at the functional variant Val92Met of *MC1R* (i.e., rs2228479*A) is likely of two origins: The vast majority of haplotypes carrying this allele in the human gene pool is resulted from Neanderthal introgression, whereas one haplotype (NA19084_a) carrying

this allele may be from a recurrent mutation in the AMH lineage, double recombination, or biased gene conversion. Here, we cross-examined all evidences presented in this article.

Neanderthal ancestry encompassing *MC1R* was first revealed by Neanderthal AIMs. To facilitate further analyses, we shrank analyzed region and removed recombinant haplotypes. Next, we reconstructed a phylogeny for all contemporary nonrecombinant haplotypes and archaic hominin haplotypes, and found that those putative introgressive haplotypes identified by the AIM-based analysis coalesce with archaic hominins first. We then found that the ancestral sequence of putative introgressive haplotypes is closest to Neanderthal sequence among Neanderthal, Denisovan, and human reference sequences. Presence of Type-2 derived alleles also suggests archaic origin of the putative introgressive haplotypes. Moreover, we ruled out the alternative model of ancestral polymorphism by employing a recombination-based test. We further estimated that the putative introgressive haplotypes diverged with human reference sequence, Denisovan, and Altai Neanderthal 552.5, 159.0, and 103.3 ka, respectively. We noticed that the 103.3 ka postdates the AMH–Neanderthal divergence time (i.e., 270–440 ka; Green et al. 2010), which suggests that the putative introgressive haplotypes are from postdivergence introgression. Moreover, the putative introgressive haplotype is absent in Africans, further supporting our conclusion. To summarize, the above results form an evidence chain that suggests a recent Neanderthal origin of the putative introgressive haplotypes.

Proposing a Three-Step Approach for Demonstrating Archaic Introgression in AMH

After the publication of Neanderthal and Denisovan genomes (Green et al. 2010; Reich et al. 2010), several genomic regions have been identified to contain archaic hominin

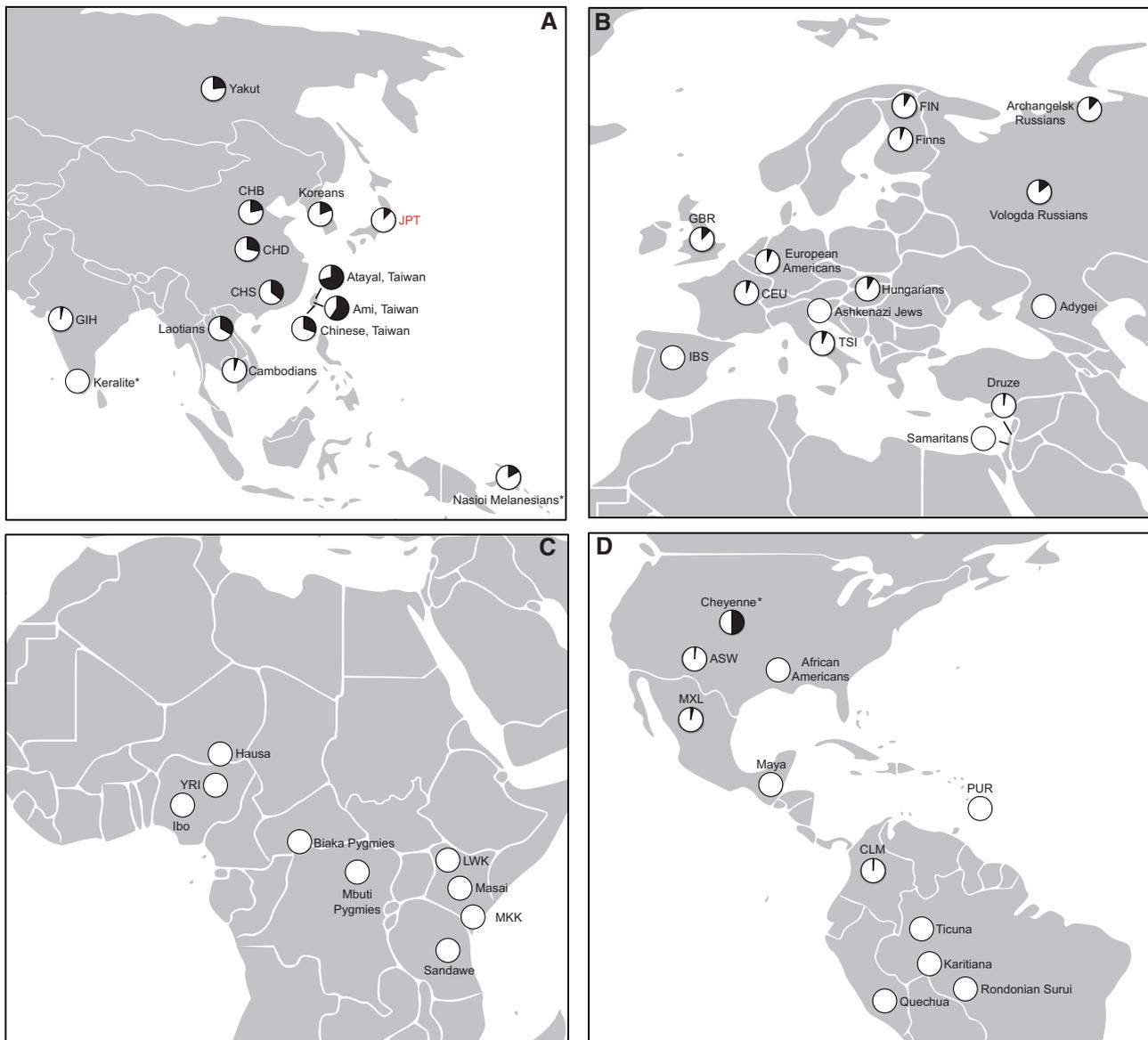


FIG. 3. Frequency of the Neanderthal introgression encompassing *MC1R* in global populations. Populations labeled with a three-letter code are 1KG or HapMap 3 populations. Other populations were genotyped in this study. The JPT population was labeled as one of the JPT haplotypes that carrying the rs2228479*A (rs2228479–chr16:89985940, NA19084_a) is not of Neanderthal origin. Populations with $2n \leq 20$ are marked with an asterisk (*). See also supplementary tables S6–S8, Supplementary Material online.

introgression, which include *HYAL2* (Ding, Hu, Xu, et al. 2014), *STAT2* (Mendez et al. 2012a), *OAS1* (Mendez et al. 2012b), *HLA Class I* (*HLA-A*, *HLA-B*, *HLA-C*; Abi-Rached et al. 2011), and *HLA-DPB1* (Ding, Hu, and Jin 2014; Temme et al. 2014). Based on the current and previous researches, we propose a three-step approach for demonstrating Neanderthal or Denisovan introgression in AMH. First, the putative introgressive segment should be close to Neanderthal or Denisovan sequence. This could be demonstrated by direct sequence comparison, phylogenetic reconstruction, or the presence of Type-2 derived alleles (or archaic hominin AIMs). Second, the ancestral polymorphism model should be rejected by applying the recombination-based test to the putative introgressive segment (Mendez et al. 2012b; Ding, Hu, Xu, et al. 2014). Third, the divergence time

between the putative introgressive segment and Neanderthal (or Denisovan) should postdate the AMH–Neanderthal (or Denisovan) divergence time. Further, other supportive evidences, such as global distribution of the putative introgressive segment, are preferred. We intend to make it convenient to distinguish regions with archaic introgression from false positives by implementing this three-step approach in future studies.

AIMs and Type-2 SNPs

We consider the SNPs resulted from mutations that occurred within the Neanderthal lineage after its divergence with AMH as Neanderthal AIMs. The principle of AIMs is the same as the principle of Type-2 SNPs. As the identification of AIMs relies on the Altai Neanderthal sequence, and the Altai Neanderthal

sequence may not be identical to the genome sequence of the actual Neanderthal population that admixed with AMH, it could be expected that the AIMs is a subset of Type-2 SNPs. As both the AIMs and the Type-2 SNPs could represent local Neanderthal ancestry, we genotyped four Type-2 SNPs (two of which are not identified as AIMs) flanking the rs2228479 to increase the resolution of local Neanderthal ancestry representation encompassing rs2228479 when investigating the two origins of the Val92Met variant of *MC1R*.

Origin of the Val92Met Variant Carried by the Putative Introgressive Haplotypes

It was observed that all introgressive haplotypes carry the Val92Met variant (i.e., rs2228479*A), and only one of the nonintrogressive haplotypes (NA19084_a) carries the rs2228479*A, which suggests that the rs2228479*A on the introgressive haplotypes might be of Neanderthal origin. However, the rs2228479*A was not observed in the published Altai Neanderthal genome. Further, among the four Type-2 alleles we used to “tag” Neanderthal ancestry, one of them (rs62052212*A) was also not observed in the Altai Neanderthal genome. The rs62052212 was found heterozygous in the low-coverage Vindija Neanderthal genome, which suggests that this SNP may be polymorphic in Neanderthal populations, and the sequenced high-coverage Altai Neanderthal does not reflect such polymorphism.

Here, we discuss the two possible scenarios could be used to explain the observation of rs2228479*A on the introgressive haplotypes and the absence of rs2228479*A in Altai Neanderthal: 1) The rs2228479 is polymorphic in Neanderthal populations, and 2) the mutation creating rs2228479*A occurred on the putative introgressive haplotype shortly after it entered modern human gene pool. If the above scenario #2 stands, the mutation creating rs2228479*A should occur when there is only one copy of Neanderthal introgressive haplotype in the modern human gene pool, as all of the putative introgressive haplotypes carry the rs2228479*A.

We then estimated the TMRCA for all nonrecombinant putative introgressive haplotypes using a Bayesian MCMC method. TMRCA was estimated as 147.7 ka (95% CI: 115.8–181.1 ka), which predates the time of out-of-Africa migration (< 100 ka; Li and Durbin 2011). A possible explanation for this is that there are two or more copies of Neanderthal haplotypes entering the modern human gene pool at the beginning. In this situation, the above scenario #2 is not likely to stand, and the above scenario #1 is still possible.

Further, we examined the BAM files of high-coverage Altai Neanderthal and Denisovan genomes. The rs2228479*A was found in one (out of 41) sequencing read of Altai Neanderthal and one (out of 20) sequencing read of Denisovan. All of the relevant sequencing reads and base calls passed quality control filter. Therefore, it could be possible that the Neanderthal and Denisovan are heterozygous at the rs2228479, and some unknown reason caused the imbalanced observation of ancestral and derived alleles of rs2228479 in sequencing reads.

To summarize, the above two possible scenarios could be used to explain the rs2228479*A carried by the putative introgressive haplotypes. Although our evidences based on TMRCA and raw sequencing reads in BAM files are in favor of the above scenario #1, we think that the above evidences are not sufficiently strong to reject the above scenario #2. Both scenarios did not affect our conclusion, which suggests that the putative introgressive haplotypes carrying the rs2228479*A are of Neanderthal origin.

Observation of Val92Met Variant in HapMap Africans

The Val92Met variant was absent in 1KG African populations and the African populations genotyped in this study, which is consistent with its putative Neanderthal origin. However, among HapMap African individuals, we found that one YRI sample (NA18852) and two MKK samples (NA21339, NA21574) carry heterozygous alleles at the Val92Met variant.

Presence of derived alleles at Val92Met (i.e., rs2228479*A) in Africans could be explained by three models: 1) The rs2228479*A existed in the human gene pool before the Neanderthal–AMH divergence and was drifted to very low frequency in modern Africans whereas preserved in the Neanderthals and modern Eurasians (i.e., the “ancestral polymorphism model”), 2) recent gene flow from modern Eurasians to Africans, or 3) independent origin (such as recurrent mutation).

For the first model, we rejected it using the aforementioned recombination-based test, as 96.9% of the putative introgressive haplotypes have $P \leq 3.1 \times 10^{-4}$ (the significance threshold adjusted for 159 tests using Bonferroni correction). Therefore, the second and third models could be used to explain the observation of Val92Met variant in Africans. Both models are consistent with our conclusion that the putative introgressive haplotypes are of Neanderthal origin.

Further, we obtained DNA of the two MKK samples from the Coriell Institute for Medical Research, and genotyped them on rs2228479 and four flanking Type-2 SNPs (see [supplementary fig. S2B and C, Supplementary Material](#) online). It was observed that all four Type-2 SNPs on both MKK samples indicate non-Neanderthal origin, which is in support of the previous model #3. Further, our genotyping result at the Val92Met SNP suggests that the two MKK samples are homozygotes of the ancestral allele, which is inconsistent with the HapMap data. This could be a genotyping error in HapMap and did not affect our conclusion.

To summarize, three models could be used to explain the observation of Val92Met variant in Africans, and the ancestral polymorphism model (above model #1) was rejected. The rest two models did not affect our conclusion. Furthermore, the observation of Val92Met variant in Africans could be due to genotyping error in HapMap. Due to the possible genotyping error, we removed the two MKK samples in obtaining global distribution of the putative introgressive haplotypes.

Possible Origins of the Val92Met Variant in NA19084_a

After we validated that the putative introgressive haplotypes are of Neanderthal origin, we found that all putative introgressive haplotypes carry the rs2228479*A (i.e., Val92Met variant of *MC1R*). However, one nonintrogressive haplotype (from 1KG JPT, NA19084_a) also carries this allele. Analyses based on AIMs and Type-2 SNPs suggest that NA19084_a is a nonrecombinant and nonintrogressive haplotype. Therefore, three possible explanations could be used to explain the allele observation at NA19084_a: 1) A recurrent mutation at rs2228479, 2) a double recombination between the two AIMs/Type-2 SNPs flanking rs2228479, or 3) a biased gene conversion event occurred nearby rs2228479. So far, we cannot rule out any of the three aforementioned possibilities.

Test for Natural Selection

Archaic introgressions in several regions have been reported under natural selection (Abi-Rached et al. 2011; Mendez et al. 2012a; Ding, Hu, Xu, et al. 2014). Here, we investigated whether the Neanderthal introgression at *MC1R* was under positive natural selection. We applied the Tajima's *D*, Fu and Li's test, and the integrated haplotype score (iHS) test to the European and East Asian populations (Tajima 1989; Fu and Li 1993; Voight et al. 2006). In results, Tajima's *D* and Fu and Li's test suggest that the region is neutral (supplementary table S9, Supplementary Material online). Further, iHS found no evidence of positive selection in either Europeans or East Asians (supplementary fig. S3, Supplementary Material online), suggesting that the Neanderthal introgression at *MC1R* could be neutral in Europeans and East Asians. However, the neutrality tests are based on the European (Utah Residents with Northern and Western European ancestry, known as CEPH [CEU]) and East Asian (Han Chinese in Beijing, China [CHB] + JPT) populations (HapMap 3), in which the frequency of rs2228479*A is ranging from 5.1% to 21.6% (supplementary table S6, Supplementary Material online). Due to lack of genome sequencing or genotyping data, we are not able to implement the neutrality tests to Taiwan Ami and Atayal populations. Therefore, we cannot rule out demographic effects such as population bottleneck as an explanation for the unusually high frequency of rs2228479*A in Atayal and Ami although positive natural selection is also a possibility (Nakayama et al. 2006).

Materials and Methods

Data Sources

Whole-genome sequencing data were obtained from the 1000 Genomes Project Phase 1 (1000 Genomes Project Consortium 2012). This data set (denoted as the 1KG data set) contains low-coverage whole-genome sequencing data of 1,092 individuals from 14 global populations. The populations in this data set are: YRI, LWK (Luhya in Wubuye, Kenya), CEU, FIN (Finnish in Finland), GBR (British in England and Scotland), TSI (Toscani in Italia), IBS (Iberian population in Spain), CHB, CHS (Southern Han Chinese), JPT, MXL (Americans with Mexican Ancestry in Los Angeles), CLM

(Colombians from Medellin, Columbia), PUR (Puerto Ricans from Puerto Rico), and ASW (Americans with African Ancestry in Southwest USA). The data set in the Variant Calling Format was downloaded from the 1000 Genomes Project FTP server. We also examined whole-genome genotyping data from the International HapMap Project Phase 3 (International HapMap 3 Consortium 2010). This data set, denoted as the HapMap data set, was used to extract allele frequency information and perform iHS test. The populations in this data set are: YRI, LWK, MKK, CEU, TSI, CHB, CHD (Chinese in Metropolitan Denver, CO), JPT, and GIH (Gujarati Indians in Houston, TX).

High-coverage genome of a Neanderthal from the Altai Mountains (denoted as the Altai Neanderthal) was from Prüfer et al. (2014). High-coverage genome of a Denisovan was from Meyer et al. (2012). BAM files were used to obtain sequence information. Criteria in Reich et al. (2010) were used when processing the BAM file. Chimpanzee sequences used in this study was obtained from the alignment of PanTro3 mapped to GRCh37. All genome coordinates in this article follow those in GRCh37.

Identification of Neanderthal AIMs

Common ancestors of Neanderthal and Denisovan diverged with ancestors of AMH about 800 ka. It could be expected that plethora of mutations accumulated in both lineages after their divergence. Those postdivergence mutations occurred in the Neanderthal lineage will not be observed in AMH unless in the case of an archaic hominin introgression. Thus, the postdivergence mutations in Neanderthal could be considered as AIMs for Neanderthal.

As the Sub-Saharan Africans are reported nearly free from Neanderthal introgression (Prüfer et al. 2014), we used the YRI and LWK as reference populations in the identification of postdivergence SNPs. In particular, if all YRI and LWK haplotypes are fixed at one allele at a given SNP, whereas the Altai Neanderthal is a homozygote of the other allele, this SNP will be considered as a candidate of Neanderthal AIM.

Phylogeny Reconstruction

We reconstructed a phylogenetic tree for all nonrecombinant contemporary haplotypes and archaic hominin sequences (Altai Neanderthal and Denisovan) using maximum-likelihood method by MEGA 6 with 100 bootstrap replications (Tamura et al. 2013). A chimpanzee sequence (PanTro3) was used to root the phylogenetic tree. We collapsed some trivial monophyletic clusters in the tree to enhance its display in the paper.

Ancestral Sequence Reconstruction

We reconstructed the ancestral sequence for all putative introgressive haplotypes using a Bayesian MCMC approach implemented in BEAST v1.8.0 (Drummond et al. 2012). We reconstructed the ancestral sequence for 10,000 times, and the sequence with highest probability was used as the ancestral sequence of all putative introgressive haplotypes for further comparison.

Type-1 and Type-2 Derived Alleles

Our previous study suggested that two types of derived alleles characterize an introgressive haplotype (Ding, Hu, Xu, et al. 2014). Type-1 derived alleles are resulted from mutations occurred before the divergence between archaic hominin and AMH, whereas Type-2 derived alleles are referred to those derived alleles resulted from mutations that occurred in the archaic hominin lineage after the AMH–archaic hominin divergence. SNPs that harbor the Type-2 derived alleles are denoted as Type-2 SNPs. We searched for Type-2 derived alleles in the putative introgressive haplotypes. A derived allele will be considered as a Type-2 derived allele if the given allele is inclusively and exclusively carried by the putative introgressive haplotypes. Principles of Neanderthal AIMs and Type-2 derived alleles are similar, and we will further discuss this issue in the Discussion section.

Test for Demographic Models

We applied a recombination-based test to examine the two mutually exclusive models that could explain the close relationship between the putative introgressive haplotypes and the Altai Neanderthal (see Results; Mendez et al. 2012b; Ding, Hu, Xu, et al. 2014). The alternative model (i.e., the “ancestral polymorphism model”) requires those putative introgressive haplotypes to exist in the human gene pool before the divergence of Neanderthal and AMH (>270 ka, Green et al. 2010). Therefore, we estimated the probability of persistence of the putative introgressive haplotypes over 270 ka, expressed as $P = e^{-\lambda \times \frac{T_p}{100 \times T_g}}$. In this equation, λ is the genetic length of the putative introgressive haplotypes (International HapMap 3 Consortium 2010), T_p is the divergence time between Neanderthal and AMH ($\geq 270,000$ years), and T_g is the time per generation (25 years). We estimated such probability for each and every putative introgressive haplotype. In total, we performed 159 tests, and the statistical significance threshold was adjusted to $P \leq 3.1 \times 10^{-4}$ using Bonferroni correction.

Divergence Time and TMRCA Estimations

We used a maximum-likelihood method as previously described to estimate the divergence time between the putative introgressive haplotypes and the Altai Neanderthal, Denisovan, or human reference sequence (Mendez et al. 2012a; Ding, Hu, Xu, et al. 2014). Given two sequences with complete coverage (denoted as sequence M and N) and a sequence with incomplete coverage (denoted as sequence R), this method could estimate the M–N and N–R divergence time. Technical detail of the method could be found in earlier studies (Mendez et al. 2012a; Ding, Hu, Xu, et al. 2014). In this study, we employed this time estimation method twice. In both estimations, the sequence M was the human reference sequence, and the sequence N was the introgressive haplotype. Altai Neanderthal sequence and Denisovan sequence were used as sequence R in the two estimations, respectively.

A Bayesian MCMC approach was employed in BEAST v1.8.0 to estimate the TMRCA of all putative introgressive haplotypes (Drummond et al. 2012). Markov chain length

was set to 10,000,000. Tracer v1.5 was used to analyze the outputs of BEAST. Human–chimpanzee divergence time of 6,500 ka was used to calibrate molecular clock in the divergence time and TMRCA estimations (Goodman 1999).

Genotyping of rs2228479 and Flanking Type-2 SNPs

To validate the genotyping calls of the 1KG sample NA19084 (JPT) and the HapMap samples NA21339 (MKK) and NA21574 (MKK), we obtained DNA of these samples from the Coriell Institute for Medical Research and genotyped the samples at the SNP of Val92Met variant (i.e., rs2228479) and four flanking Type-2 SNPs (rs11648433, rs62052212, rs62052213, and rs12600051; [supplementary table S5, Supplementary Material](#) online). To further investigate the origin of the rs2228479*A and obtain global distribution of the Val92Met variant, we genotyped additional 757 individuals from 31 global populations for the rs2228479 and the four flanking Type-2 SNPs (same as for the 1KG and HapMap samples). SNP genotyping was conducted using SNaPshot Multiplex System with protocols supplied by the Applied Biosystems. Primer sequences used in the SNaPshot was listed in the [supplementary table S10, Supplementary Material](#) online. All sampling and experimental procedures have been approved by the Ethics Review Board of the School of Life Sciences at Fudan University. All sampled individuals signed appropriate informed consent for sampling.

Detection of Selection

We used the Tajima’s D , Fu and Li’s test, and the iHS test to detect signal of positive selection encompassing *MC1R* (Tajima 1989; Fu and Li 1993; Voight et al. 2006). We computed Tajima’s D , Fu and Li’s D^* , and Fu and Li’s F^* using the 1KG data set. We computed iHS for Europeans (CEU) and East Asians (CHB + JPT) using the HapMap data set with the Whole-Genome Homozygosity Analysis and Mapping Machina.

Supplementary Material

Supplementary tables S1–S10 and figures S1–S3 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

This work was supported by the National Science Foundation of China (grant number 31271338, 31330038, 91331204, and 31171218) and the National Basic Research Program (grant number 2012CB944600). This work was also supported by the Fudan University Undergraduate Research Opportunities Program, FDUROP Wang Dao Grant 13075 to R.Z. The authors thank Dr Hua Zhong for technical assistance. They also thank the three anonymous reviewers for their constructive comments. They declare that there is no conflict of interest.

References

- 1000 Genomes Project 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.
- Akey JM, Wang H, Xiong M, Wu H, Liu W, Shriver MD, Jin L. 2001. Interaction between the melanocortin-1 receptor and P genes contributes to inter-individual variation in skin pigmentation phenotypes in a Tibetan population. *Hum Genet.* 108:516–520.
- Box NF, Wyeth JR, O’Gorman LE, Martin NG, Sturm RA. 1997. Characterization of melanocyte stimulating hormone receptor variant alleles in twins with red hair. *Hum Mol Genet.* 6:1891–1897.
- Ding Q, Hu Y, Jin L. 2014. Non-Neanderthal origin of the HLA-DPB1*0401. *J Biol Chem.* 289:10252–10252.
- Ding Q, Hu Y, Xu S, Wang J, Jin L. 2014. Neanderthal introgression at chromosome 3p21.31 was under positive natural selection in East Asians. *Mol Biol Evol.* 31:683–695.
- Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol.* 29:1969–1973.
- Elfakir A, Ezzedine K, Latreille J, Ambroisine L, Jdid R, Galan P, Hercberg S, Gruber F, Malvy D, Tschachler E, et al. 2010. Functional *MC1R*-gene variants are associated with increased risk for severe photoaging of facial skin. *J Invest Dermatol.* 130:1107–1115.
- Fernandez LP, Milne RL, Bravo J, Lopez JM, Avilés JA, Longo MI, Benitez J, Lazaro P, Ribas G. 2007. *MC1R*: three novel variants identified in a malignant melanoma association study in the Spanish population. *Carcinogenesis* 28:1659–1664.
- Finlayson C, Pacheco FG, Rodríguez-Vidal J, Fa DA, López JMG, Pérez AS, Finlayson G, Allue E, Preysler JB, Cáceres I, et al. 2006. Late survival of Neanderthals at the southernmost extreme of Europe. *Nature* 443:850–853.
- Fu YX, Li WH. 1993. Statistical tests of neutrality of mutations. *Genetics* 133:693–709.
- 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 Neanderthal genome. *Science* 328:710–722.
- Hodgson JA, Disotell TR. 2008. No evidence of a Neanderthal contribution to modern human diversity. *Genome Biol.* 9:206.
- International HapMap 3 Consortium. 2010. Integrating common and rare genetic variation in diverse human populations. *Nature* 467:52–58.
- Jablonski NG. 2004. The evolution of human skin and skin color. *Annu Rev Anthropol.* 33:585–623.
- Lalueza-Fox C, Römpler H, Caramelli D, Stäubert C, Catalano G, Hughes D, Rohland N, Pilli E, Longo L, Condemi S, et al. 2007. A melanocortin 1 receptor allele suggests varying pigmentation among Neanderthals. *Science* 318:1453–1455.
- Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature* 475:493–496.
- 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. 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. 2012b. Global genetic variation at *OAS1* provides evidence of archaic admixture in Melanesian populations. *Mol Biol Evol.* 29:1513–1520.
- Meyer M, Kircher M, Gansauge MT, Li H, Racimo F, Mallick S, Schraiber JG, Jay F, Prüfer K, Filippo CD, et al. 2012. A high-coverage genome sequence from an archaic Denisovan individual. *Science* 338:222–226.
- Mogil JS, Wilson SG, Chesler EJ, Rankin AL, Nemmani KVS, Lariviere WR, Groce MK, Wallace MR, Kaplan L, Staud R, et al. 2003. The melanocortin-1 receptor gene mediates female-specific mechanisms of analgesia in mice and humans. *Proc Natl Acad Sci U S A.* 100:4867–4872.
- Motokawa T, Kato T, Hashimoto Y, Takimoto H, Yamamoto H, Katagiri T. 2008. Polymorphism patterns in the promoter region of the *MC1R* gene are associated with development of freckles and solar lentiginos. *J Invest Dermatol.* 128:1588–1591.
- Nakayama K, Soemantri A, Jin F, Dashnyam B, Ohtsuka R, Duanchang P, Isa MN, Settheetham-Ishida W, Harihara S, Ishida T. 2006. Identification of novel functional variants of the melanocortin 1 receptor gene originated from Asians. *Hum Genet.* 119:322–330.
- Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, Heinze A, Renaud G, Sudmant PH, de Filippo C, et al. 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505:43–49.
- 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.
- Smith R, Healy E, Siddiqui S, Flanagan N, Steijlen PM, Rosdahl I, Jacques JP, Rogers S, Turner R, Jackson IJ, et al. 1998. Melanocortin 1 receptor variants in an Irish population. *J Invest Dermatol.* 111:119–122.
- Stokowski RP, Pant PV, Dadd T, Fereday A, Hinds DA, Jarman C, Filsell W, Ginger RS, Green MR, van der Ouderaa FJ, et al. 2007. A genome-wide association study of skin pigmentation in a South Asian population. *Am J Hum Genet.* 81:1119–1132.
- Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol Biol Evol.* 30:2725–2729.
- Temme S, Zacharias M, Neumann J, Wohlfromm S, König A, Temme N, Springer S, Trowsdale J, Koch N. 2014. A novel family of human leukocyte antigen class II receptors may have its origin in archaic human species. *J Biol Chem.* 289:639–653.
- Valverde P, Healy E, Jackson I, Rees JL, Thody AJ. 1995. Variants of the melanocyte-stimulating hormone receptor gene are associated with red hair and fair skin in humans. *Nat Genet.* 11:328–330.
- Valverde P, Healy E, Sikkink S, Haldane F, Thody AJ, Carothers A, Jackson IJ, Rees JL. 1996. The Asp84Glu variant of the melanocortin 1 receptor (*MC1R*) is associated with melanoma. *Hum Mol Genet.* 5:1663–1666.
- Voight BF, Kudaravalli S, Wen X, Pritchard JK. 2006. A map of recent positive selection in the human genome. *PLoS Biol.* 4:e72.
- Wu GS, Luo HR, Dong C, Mastronardi C, Licinio J, Wong ML. 2011. Sequence polymorphisms of *MC1R* gene and their association with depression and antidepressant response. *Psychiatr Genet.* 21:14–18.
- Xu X, Thörnwall M, Lundin LG, Chhajlani V. 1996. Val92Met variant of the melanocyte stimulating hormone receptor gene. *Nat Genet.* 14:384–384.