Reintroduction of a Homocysteine Level-Associated Allele into East Asians by Neanderthal Introgession

Ya Hu,1,2,† Qiangt Ding,1,2,3,† Yungang He,2 Shuhua Xu,2 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
3Department of Molecular Biology and Genetics, Cornell University
†These authors contributed equally to this work.
*Corresponding author: E-mail: lijin@fudan.edu.cn.
Associate editor: Xun Gu

Abstract

In this study, we present an analysis of Neanderthal introgression at the dipeptidase 1 gene, DPEP1. A Neanderthal origin for the putative introgressive haplotypes was demonstrated using an established three-step approach. This introgression was under positive natural selection, reached a frequency of >50%, and introduced a homocysteine level- and pigmentation-associated allele (rs460879-T) into East Asians. However, the same allele was also found in non-East Asians, but not from Neanderthal introgression. It is likely that rs460879-T was lost in East Asians and was reintroduced subsequently through Neanderthal introgression. Our findings suggest that Neanderthal introgression could reintroduce an important previously existing allele into populations where the allele had been lost. This study sheds new light on understanding the contribution of Neanderthal introgression to the adaptation of non-Africans.

Key words: Neanderthal, archaic introgression, positive selection, DPEP1, plasma homocysteine level, pigmentation.

Introduction

Neanderthals are our extinct relatives that diverged from modern Africans ~550 thousand years ago (KYA) (Prüfer et al. 2014). Non-Africans exhibit approximately 1% Neanderthal ancestry. Neanderthal introgressions involving skin color and immunity loci were identified (Abi-Rached et al. 2011; Ding, Hu, Xu, Wang, Jin 2014, Ding, Hu, Xu, Wang, Li, et al. 2014).

A previous study proposed that putative introgressive haplotypes should satisfy the following three criteria to be considered as recent Neanderthal introgressions (Ding, Hu, Xu, Wang, Li, et al. 2014): 1) a close phylogenetic relationship between Neanderthal and putative introgressive haplotypes; 2) putative introgressive haplotype–Neanderthal divergence time postdates the anatomically modern human (AMH)–Neanderthal divergence time; and (3) rejection of the incomplete lineage sorting model.

Here, we report that Neanderthal introgression at DPEP1 was under positive selection in East Asians. First, we identified putative introgressive haplotypes at DPEP1 using Neanderthal ancestral informative markers (AIMs). We then demonstrated a Neanderthal origin for the putative introgressive haplotypes using the aforementioned three criteria. This introgression reached a high frequency in East Asians and was found to be under positive selection in East Asians by the Tajima’s D statistic, composite of multiple signals (CMS), and integrated haplotype score (iHS) tests. We also explored the possible evolutionary importance of this introgression.

Results

Identification of Putative Introgressive Haplotypes

To investigate putative Neanderthal introgression at DPEP1, we obtained sequence information for all haplotypes in the 1000 Genomes Project Phase 3 data set for DPEP1 and its up- and downstream 50-kb flanking regions (chr16: 89629716–89754839, all coordinates in GRCh37) (1000 Genomes Consortium 2012). Using a previous approach (Ding, Hu, Xu, Wang, Li, et al. 2014), we identified 27 Neanderthal AIMs in this region (supplementary table S1, Supplementary Material online). The presence of putative Neanderthal introgression at DPEP1 was suggested by haplotypes that carry Neanderthal alleles on the Neanderthal AIMs. As recombination could interfere with further analyses (Ding, Hu, Xu, Wang, Jin 2014; Ding, Hu, Xu, Wang, Li, et al. 2014), we refined the boundaries of the analyzed region to chr16: 89700677–89746126 and removed haplotypes that were recombinants of the putative introgressive and nonintrogressive haplotypes. In total, 98.1% (4,974/5,070) of the haplotypes in the 1000 Genomes Project data set were included in the refined data set for further analyses. We used a Neanderthal AIM (rs76607656-C) as the diagnostic single nucleotide polymorphism (SNP) for the putative introgressive haplotypes.

Neanderthal Origin of Putative Intrtractive Haplotypes

We followed a previously established three-step approach (as described in Introduction section) to demonstrate a
Neanderthal origin for the putative introgressive haplotypes. First, we reconstructed a phylogenetic tree for the putative introgressive, nonintrogressive, Altai Neanderthal, and Denisovan haplotypes using the neighbor-joining method (Tamura et al. 2013). It was observed that all putative introgressive haplotypes coalesced with Altai Neanderthal first, supporting their Neanderthal origin. Second, we estimated the divergence times between the putative introgressive haplotypes and Altai Neanderthal and Denisovan haplotypes as 91.50 ± 16.29 and 495.15 ± 11.07 KYA, respectively (supplementary table S2, Supplementary Material online). The putative introgressive haplotype–Altai Neanderthal haplotype divergence time (91.50 KYA) far post-dates the AMH–Neanderthal divergence time (550 KYA), which supports a recent Neanderthal origin for the putative introgressive haplotypes.

Third, we applied a recombination-based test (Ding, Hu, Xu, Wang, Jin 2014) to reject the incomplete lineage sorting model. We need to reject this alternative model, as it could also explain the close relationship between the putative introgressive and Altai Neanderthal haplotypes, shown in figure 1A (Ding, Hu, Jin 2014). This alternative model assumed that the putative introgressive haplotypes existed in the human gene pool before the Neanderthal–AMH divergence (550 KYA). Thus we computed the probability of the earlier assumption of the incomplete lineage sorting model being true,

\[ \text{Probability} = \frac{1}{n} \]

where \( n \) is the number of haplotypes. The probability was found to be 0.01, which supports the recent Neanderthal origin for the putative introgressive haplotypes.

![Phylogenetic trees. The two trees were reconstructed using the neighbor-joining method implemented in MEGA 6. Chimpanzee sequence was used to root the trees. (A) Tree for putative introgressive, nonintrogressive, Altai Neanderthal, and Denisovan haplotypes. All putative introgressive haplotypes coalesced with Altai Neanderthal first, supporting their Neanderthal origin. (B) Tree for haplotypes that carry the rs460879-T and Altai Neanderthal haplotypes. In the tree, most (96.9%) East Asian haplotypes coalesced with Neanderthal first, while non-East Asian haplotypes rarely (6.2% in South Asians, 0.6% in Europeans, 0.6% in Americans, and 0% in Africans) coalesced with Neanderthal first, which suggests that the rs460879-T is of independent origins in East Asians (Neanderthal origin) and non-East Asians (non-Neanderthal origin).]

**Fig. 1.** Phylogenetic trees.
i.e., the putative introgressive haplotypes persisted over 550 KYA, using a recombination-based test. Because the genetic length of the analyzed region was 0.026 cM, the incomplete lineage sorting model was thus rejected ($P < 7.85/10^{4}$).

In conclusion, the earlier analyses suggest that the putative introgressive haplotypes are of Neanderthal origin. We also noted that the origin of the putative introgressive haplotypes is more likely to be Neanderthal than Denisovan, because the putative introgressive haplotypes are closer to Neanderthal than Denisovan in both the phylogenetic analysis and divergence time estimation. We denote the putative introgressive haplotypes as introgressive haplotypes hereafter.

Global Distribution

We then investigated the global distribution of the introgressive haplotypes, using populations from the 1000 Genomes Project Phase 3 data set (fig. 2A). Consistent with its Neanderthal origin, the introgressive haplotype was absent in Africans. The frequency of the introgressive haplotype was low in south Asians, Europeans, and Americans (<5%). However, the frequency of the introgressive haplotypes was strikingly high in East Asians, ranging from 50% in Kinh in Ho Chi Minh City, Vietnam (KHV) to 58% in Chinese Dai in Xishuangbanna (CDX). This high frequency in East Asians could be explained by positive selection.

Tests for Positive Selection

To test for possible positive selection, we first employed Tajima’s $D$ statistic (Tajima 1989). Tajima’s $D$ for the introgressive haplotypes was $-2.57$ ($P < 0.001$), indicating a significant departure from neutrality. Then we obtained the CMS scores of SNPs in the analyzed region (chr16: 89700677–89746126) in East Asians (Chinese Han in Beijing [CHB] and Japanese in Tokyo, Japan [JPT]), Europeans (Utah residents with Northern and Western European ancestry [CEU]), and Africans (Yoruba in Ibadan, Nigeria [YRI]) from the CMS Viewer (Grossman et al. 2010). Two SNPs were observed showing high CMS scores (rs3751680 [3.04] and rs35414122 [2.87]) in East Asians, indicating a signature of positive selection (fig. 3A). The two SNPs were in strong linkage disequilibrium (LD) with the diagnostic SNP of introgressive haplotypes ($D’ = 0.99$ and $0.94$, respectively), suggesting that
the introgressive haplotypes were the target of selection. This pattern was not observed in Europeans and Africans (supplementary fig. S1, Supplementary Material online).

To provide additional support for the above conclusion, we computed the iHS and extended haplotype homozygosity (EHH) plots and haplotype bifurcation graphs for the SNPs in the aforementioned analyzed region (Sabeti et al. 2002; Voight et al. 2006; Gautier and Vitalis 2012), based on East Asian haplotypes in the 1000 Genomes Project Phase 3 data set (CHB, Southern Han Chinese [CHS], KHV, CDX, and JPT). It was observed that the standardized $|\text{iHS}|$ for four SNPs (rs75670182 [2.57], rs3794635 [2.55], rs76878855 [2.16], and rs79799075 [2.13]) passed the threshold for positive selection ($|\text{iHS}| > 2.00$) (fig. 3A). The four SNPs were also in strong LD with the diagnostic SNP of introgressive haplotypes ($D' > 0.80$). Furthermore, EHH plots and haplotype bifurcation graphs of the diagnostic SNP of introgressive haplotypes (rs76607656) showed that the introgressive haplotypes were the target of positive selection (fig. 3B and C).

Using simulations, we showed that the false-positive rates of Tajima’s $D$, CMS, and iHS tests were not likely to be affected by Neanderthal introgression (for details, see supplementary text, supplementary tables S4 and S5, Supplementary Material online).
Moreover, we noted that Neanderthal introgression under positive selection did not reach fixation in any East Asian population in this study. Because positive selection may cause the fixation of the positively selected haplotype, several possibilities could explain the observation that Neanderthal introgression did not reach fixation. First, positive selection could still be ongoing. Second, the selective pressure might have been relaxed before the Neanderthal introgression reached fixation. Third, this could be the result of a stabilizing selective pressure. This stabilizing selective pressure is likely an interplay between a positive selective pressure (as shown by Tajima’s $D$, CMS, and iHS) and a potentially negative selective pressure.

In summary, we found that the Neanderthal introgression at $DPEP1$ was under positive selection and reached a high frequency in East Asians. This selective pressure would require further investigation.

**Discussion**

**Exploring Possible Selective Pressure**

We found that Neanderthal introgression at $DPEP1$ was under positive selection in East Asians. $DPEP1$ encodes dipeptidase 1, which is involved in the metabolism of dipeptides, glutathione, and chemicals containing beta-lactam rings (Kozak and Tate 1982; Campbell et al. 1984).

We explored possible function of the Neanderthal introgression by investigating SNPs in strong LD with the diagnostic SNP of introgressive haplotypes in East Asians. rs460879-T, an SNP in strong LD ($D' = 0.92$) with the diagnostic SNP, was reported in two previous association studies as being associated with plasma homocysteine (Hcy) levels in Europeans (Paré et al. 2009) and East Asians (Filipinos) (Lange et al. 2010). There were four SNPs in strong LD ($D' > 0.90$) with rs460879-T in both Europeans and East Asians (rs460879-T, rs352939-A, rs2460448-A, and rs164749-T). rs164749-T is an expression quantitative trait locus (eQTL), which suggests that this Neanderthal introgression may play a role in regulating gene expression (Schadt et al. 2008; Ward and Kellis 2012). Furthermore, we showed that rs460879-T in East Asians and Europeans is of independent origin and is of Neanderthal origin in East Asians (see next subsection). The eQTL (rs164749-T) also has an independent origin (similar to rs460879-T, data not shown).

In addition, we explored the possibility of selective pressure. As mentioned earlier, rs460879-T elevates Hcy level (Paré et al. 2009; Lange et al. 2010). However, an elevated Hcy level is a risk factor for congenital birth defects (Zhao et al. 2014) and therefore appears to be disadvantageous. This could be a negative side effect of some unknown advantageous phenotype. One possibility of the advantageous phenotype is pigmentation, because the aforementioned rs460879-T is also associated with three pigmentation traits (freckles, red hair, and burning) in Europeans (Sulem et al. 2008; Cerqueira et al. 2012). However, further evidence for the association between rs460879-T and pigmentation traits in East Asians is needed to convincingly conclude that the pigmentation traits are the underlying advantageous phenotypes.

**Independent Origins of rs460879-T in East Asians and Non-East Asians**

An allele in strong LD with the Neanderthal introgression (rs460879-T) was found to be associated with Hcy level and pigmentation in four previous studies. However, this allele was also found at moderate frequency in South Asians, Europeans, Americans, and Africans (“non-East Asians”) (fig. 2B) in whom this Neanderthal introgression is rare (fig. 2A). We propose that rs460879-T has independent origins, i.e., from Neanderthal introgression and descending within the AMH. We examined this scenario by reconstructing a phylogenetic tree for haplotypes carrying rs460879-T (fig. 1B).

In the phylogenetic tree, almost all East Asian haplotypes coalesced with Neanderthal first (96.9%, 500/516), indicating a Neanderthal origin. In contrast, all African haplotypes (100%, 275/275), and almost all European (99.4%, 487/490), American (99.4%, 336/338), and South Asian (93.8%, 196/209) haplotypes did not, thus indicating a non-Neanderthal origin. This observation suggests that the earlier scenario is highly likely to hold true.

By using a previously described method (Ding, Hu, Xu, Wang, Jin 2014), we showed that the nonintrogressive rs460879-T-carrying haplotypes in East Asians derived from recent gene flows from Europeans, while the introgressive rs460879-T-carrying haplotypes in non-East Asians were from recent gene flows from East Asians ($P < 2.2 \times 10^{-16}$, single-tailed Welch’s two sample t-test, Materials and Methods, supplementary table S3, Supplementary Material online).

Based on the earlier analyses, we conclude that rs460879-T in East Asians originated from Neanderthals, while rs460879-T in non-East Asians was from the AMH lineage. It is likely that rs460879-T existed before the AMH–Neanderthal divergence and was lost in East Asians during the out-of-Africa migration. It was reintroduced subsequently into East Asians by Neanderthal introgression.

In conclusion, the Neanderthal introgression at $DPEP1$ was under positive selection and reached high frequency in East Asians. An allele (rs460879-T) in strong LD with the introgression was found to be associated with elevated Hcy level and pigmentation traits. This allele existed before the AMH–Neanderthal divergence, became lost in East Asians, and was reintroduced by Neanderthal introgression. Our findings shed new light on the contribution of Neanderthal introgression to the adaptation of non-Africans—instead of contributing a de novo important allele—the Neanderthal introgression reintroduced a previously existing important allele into populations in which the allele had been lost.

**Supplementary Material**

Supplementary text, figure S1, and tables S1–S5 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).
Acknowledgments

This research was supported by the National Science Foundation of China (Grants 31271338 and 31330038) and the Chinese National Basic Research Program (Grant 2012CB944600). We thank Dr. Hua Zhong for technical assistance. We thank editors from the Oxford Language Editing for copyediting the letter. The authors declare that there is no conflict of interest.

References


