Neandertal Origin of Genetic Variation at the Cluster of *OAS* Immunity Genes

Fernando L. Mendez,[1,2] Joseph C. Watkins,[3] and Michael F. Hammer*[,1,2]

[1]Department of Ecology and Evolutionary Biology, University of Arizona

[2]ARL Division of Biotechnology, University of Arizona

[3]Mathematics Department, University of Arizona

***Corresponding author:** E-mail: mfh@u.arizona.edu; mfh@email.arizona.edu.

Associate editor: Yoko Satta

Abstract

Analyses of ancient DNA from extinct humans reveal signals of at least two independent hybridization events in the history of non-African populations. To date, there are very few examples of specific genetic variants that have been rigorously identified as introgressive. Here, we survey DNA sequence variation in the *OAS* gene cluster on chromosome 12 and provide strong evidence that a haplotype extending for ~185 kb introgressed from Neandertals. This haplotype is nearly restricted to Eurasians and is estimated to have diverged from the Neandertal sequence ~125 kya. Despite the potential for novel functional variation, the observed frequency of this haplotype is consistent with neutral introgression. This is the second locus in the human genome, after *STAT2*, carrying distinct haplotypes that appear to have introgressed separately from both Neandertals and Denisova.

Key words: **human origins, hybridization, Neandertal, TMRCA, linkage disequilibrium.**

Letter

Recent genome-level analyses of ancient DNA suggest that two independent processes of interbreeding and introgression occurred between archaic humans and non-African populations (Green et al. 2010; Reich et al. 2010). The initial evidence for admixture was mostly based on 1) the excess of derived sites shared between non-Africans and Neandertals relative to those shared between Neandertals and sub-Saharan Africans, 2) the fact that Neandertals share haplotype-tagging single-nucleotide polymorphisms (SNPs) with non-Africans that differentiate Africans and non-Africans, and 3) the observation that regions in non-African genomes that most closely match Neandertals are also deeply diverged from another non-African genomes (Green et al. 2010). Analyses based on simulated frequency spectra or global patterns of linkage disequilibrium (LD) of present-day Europeans support the hypothesis that recent admixture with Neandertals accounts for their greater similarity with non-Africans (Sankararaman et al. 2012; Yang et al. 2012). Models with ancient population structure may produce some (but probably not all) of the same patterns (Green et al. 2010; Durand et al. 2011; Eriksson and Manica 2012).

A complementary approach to study introgression focuses on genealogical patterns at individual loci and the expectation that introgressive haplotypes from divergent populations will simultaneously produce an increase in the time to the most recent common ancestor (TMRCA) and in the extent of LD (Wall 2000; Nordborg 2001). The presence of these genetic patterns and even the sharing of polymorphism with ancient DNA sequences may not be sufficient to rule out ancestral shared polymorphism as a possible explanation (Eriksson and Manica 2012). This is especially true when, as in the case of Neandertals and anatomically modern humans (AMH), the split time is recent relative to the expected coalescence time. We are aware of only two cases in which a model-testing framework was used to formally reject the hypothesis that sequence similarity between modern human and archaic haplotypes was due to ancestral polymorphism. The first involved a divergent haplotype at the *OAS1* locus that introgressed from Denisova into the ancestors of Melanesians (Mendez et al. 2012b), and the second involved a haplotype at the *STAT2* locus that introgressed from Neandertals (Mendez et al. 2012a). In both cases, levels of LD associated with the introgressive haplotype reject a model of evolution in a single population. In addition, in the case of *STAT2*, the upper 99% confidence interval for the divergence between the introgressive haplotype and the Neandertal sequence postdates the origin of AMH.

In a previous study, we noted that the human reference sequence for *OAS1* is very similar to that of Neandertals (Mendez et al. 2012b). To support or reject the hypothesis of Neandertal introgression at this locus we 1) infer the length of the physical extent of the sequence similarity between the reference and Neandertal, 2) estimate the TMRCA between Neandertal and reference haplotypes, 3) estimate the time of diversification of the reference haplotype within modern humans, and 4) assess the geographic distribution of the reference haplotype. We started by phasing resequence data covering the 3′ end of the *OAS1* gene in three African (Biaka, Mandenka, and San) and three non-African (Papuan, Han, and Basque) populations (Mendez et al. 2012b), only including sites present in at least three chromosomes in our sample. This analysis reveals the existence of eight haplotypes

Table 1. *OAS1* 3′ Haplotypes.

Haplotypes[a]	Base Position[b]																	Populations[c]					
	6900	6922	6935	6980	7193	7202	7209	7216	7237	7267	7409	7442	7494	7565	7663	7688	7708	Africans			Non-Africans		
Ancestral	A	C	C	C	G	C	G	A	G	AT	C	G	C	C	T	G	A						
Neandertal	.	T	.	T	.	.	.	.	.	.	.	.	.	.	A	N	N	B	M	S	P	H	F
Denisova	.	.	.	.	.	.	A	.	C	.	.	A	.	N	A	N	.						
R	.	T	.	T	.	.	.	.	.	.	.	.	.	.	A	A	.	0	0	0	0	8	6
K	G	.	T	.	.	.	.	.	.	.	.	.	G	.	.	.	G	0	1	2	0	0	0
D	.	.	.	.	.	.	A	.	C	–[d]	.	A	.	.	A	A	.	0	0	0	7	0	0
P	.	.	.	.	.	.	A	.	C	.	.	A	.	.	A	A	.	0	0	0	11	0	0
S	.	T	.	T	.	.	A	G	C	.	T	A	.	T	A	A	.	1	0	3	0	0	0
F	.	T	.	T	.	.	A	.	C	.	.	A	.	.	A	A	.	17	20	10	0	0	0
A	.	T	.	T	A	.	A	.	C	.	.	A	.	.	A	A	.	12	11	5	12	21	26
H	.	T	.	T	A	T	A	.	C	.	.	A	.	.	A	A	.	0	0	0	0	3	0

[a]Sites observed less than three times were removed.
[b]Positions according to hg19 abbreviated by removing the leading "11335" digits.
[c]Biaka (B), Mandenka (M), San (S), Papuan (P), Han (H), French Basque (F).
[d]The "–" indicates deletion of the dinucleotide.

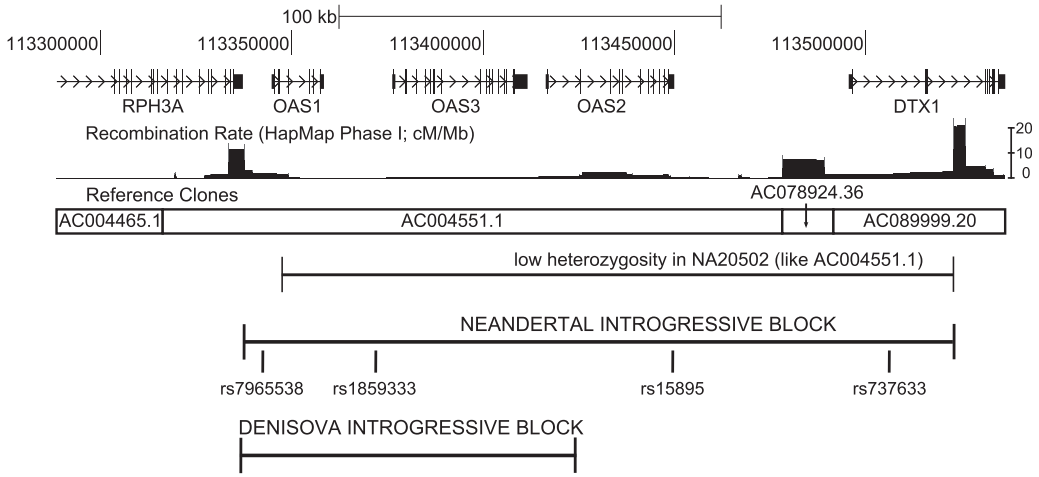

Fig. 1. Schematic representation of the region under study. From the top: chromosomal positions (hg19), genes in this region, estimated local recombination rate (based on HapMap Phase I data), clones used to generate the reference sequence, low heterozygosity region for NA20502, span of the Neandertal introgressive block, diagnostic SNPs, and approximate extent of the Denisova introgressive haplotype in Mendez et al. (2012b).

(R, K, D, P, S, F, A, and H) within this region (table 1). Of the 15 sites in table 1 with Neandertal sequence coverage, haplotype R ("reference" sequence) completely agrees with the Neandertal sequence. We then searched for sequences that extend the range of haplotype R beyond the clone used to generate reference sequence for *OAS1*, *OAS2*, and *OAS3* (i.e., AC004551.1 in fig. 1). After examining high coverage publicly available complete genome sequence data, we identified three individuals that carry two copies of haplotype R (supplementary table S1, Supplementary Material online). The sequence of NA20502 serves as good proxy for the extended R haplotype for the following two reasons: 1) it shares extremely high sequence similarity with the reference clone sequence (i.e., with only four variable sites in 131037 bases between 113347000 and the end of the AC004551.1 clone) and 2) it shows low levels of heterozygosity beyond the end of the AC004551.1 clone (fig. 1). By combining information from both the reference sequence and NA20502, we inferred that the Neandertal-like haplotype extends for the ~185 kb between the two hotspots in figure 1 (i.e., 113337000 to 113522000). Outside this region, SNPs associated with haplotype R are no longer shared with the Neandertal sequence but are shared with other human sequences. A brief analysis of 1000 genomes data for this region (1000 Genomes Project Consortium et al. 2012) confirms that the introgressive region is bounded by the two hotspots (supplementary material, Supplementary Material online). Figure 1 shows that the regions that have introgressed from Neandertal and Denisova overlap extensively.

To estimate the time of divergence between the Neandertal sequence and haplotype R, we computed the number of mutations with sequence coverage in chimpanzee, gorilla, and Neandertal and that have accumulated in the combined reference/NA20502 sequence since the most

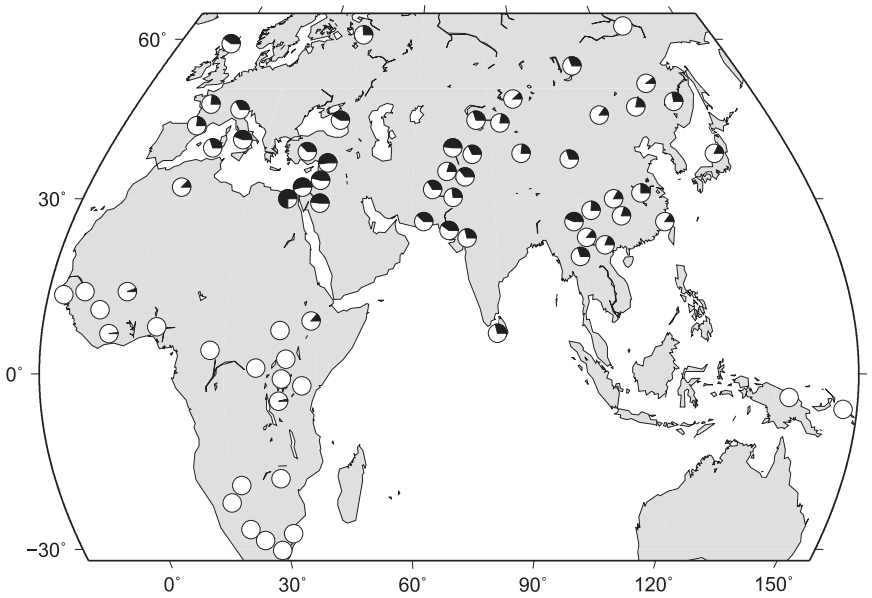

FIG. 2. Geographic distribution of the R haplotype in Old World populations. The frequency of the R haplotype is represented by the filled portion of pie chart. The presence of the R haplotype was inferred using the SNPs rs1859333, rs10850094, or rs1051042, depending on the data set used (supplementary material, Supplementary Material online).

recent common ancestor (MRCA) with chimpanzee within the ~185 kb. We inferred that 931 mutations with outgroup coverage have accumulated in the modern human lineage, but only 675 of those have Neandertal sequence coverage. Of these 675 mutations, 14 are ancestral in Neandertals, indicating that they have accumulated in the R haplotype since the MRCA with Neandertals. Assuming a human-chimpanzee sequence divergence time of 6 Mya, we estimate that the TMRCA between the Neandertal and R haplotype sequences is 124 kya, with 95% and 99% confidence upper bounds of 193 and 224 kya, respectively (supplementary material, Supplementary Material online).

To estimate the TMRCA of European chromosomes carrying the introgressive haplotype R, we use methods that incorporate either LD or intra-allelic sequence variation. The LD-based method considers the fraction of haplotypes that have persisted since the MRCA (supplementary material, Supplementary Material online). First, we chose four SNPs (rs7965538, rs1859333, rs15895, and rs737633) (fig. 1), at which the combination of allelic states distinguishes Neandertal from sub-Saharan Africans in the Human Genome Diversity Project (HGDP) panel (Li et al. 2008). By using a genetic map for Icelandic individuals (Kong et al. 2010), we estimated the genetic distance between rs7965538 and rs737633 as 0.1783 cM. Restricting our analysis to European HapMap phase III phased data (hapmap.org), we estimated the frequency of the (inferred) original (Neandertal-like) haplotype to be ~4.6% (19/410 chromosomes) and the average frequency of the Neandertal state for each of the four SNPs (but possibly on different haplotypes) as ~33%. With these values, we estimated a TMRCA in Europeans of 43 kya with a 95% upper bound at 86 kya.

We also estimated the TMRCA of the R haplotype using the number of derived sites in the reference sequence that are not shared by all R haplotype chromosomes in homozygous individuals (supplementary table S1, Supplementary Material online). To obtain a mutation rate for this region, we measured the human-chimpanzee sequence divergence as ~1% for this region and assumed that human and chimpanzee sequences diverged 6 Mya. The three observed reference mutations (present in the human reference sequence but not in all R haplotypes) in ~131 kb result in an estimate for the TMRCA of 21 kya and 95% confidence upper bound of 71 kya (supplementary material, Supplementary Material online). This value may be an underestimate, since some more basal R haplotype sequences might not be represented in the data set that we used.

The R haplotype carries six polymorphic sites that were introduced into Eurasians from Neandertals and that are predicted to affect protein sequences in the *OAS* genes (four in *OAS1* and two in *OAS2*). Interestingly, one of these (rs15895) has been associated with variable symptomology in Tick-Borne Encephalitis Virus-Induced Disease in northeastern Europeans (Barkhash et al. 2010). The R haplotype encodes the ancestral version of the sequence, which encodes an OAS2 protein that is eight amino acids shorter than that encoded by all other haplotype backgrounds due to an earlier stop codon. The functional consequences of Neandertal introgression of the other five polymorphisms (rs1293767, rs10774671, rs1131476, rs1051042, and rs11352835) are unclear (supplementary material, Supplementary Material online).

We investigated the geographic distribution of the Neandertal-like haplotype by first identifying a set of SNPs that serve as tag-SNPs for haplotype R and then using public data to assess the frequency of these variants (supplementary material, Supplementary Material online). Haplotype R is mostly restricted to Eurasia, but it is also observed in North and East Africa and in a few instances in West Africa (fig. 2). We tested whether the frequency of haplotype R in

the European HGDP populations was unusually high. Considering that it is not observed in the sub-Saharan farmers of the HGDP, the average frequency of diagnostic SNPs of the R haplotype is marginally significant ($P \sim 0.05$). However, when we allow SNPs with an average allele frequency of 1–4% in sub-Saharan farmers, mimicking the estimated percentage of Neandertal contribution (Green et al. 2010), the frequency of the R haplotype is not unusually high ($P \sim 0.18$), suggesting that we do not have to invoke positive directional selection if we assume that the R haplotype is introgressive in Eurasians only (supplementary fig. S1, Supplementary Material online). Nevertheless, the striking geographic differentiation of specific haplotypes into three regions (Africa, Eurasia, and Melanesia) (table 1) may hint at the possibility of balancing selection acting at the *OAS1* locus.

In summary, several lines of evidence support the hypothesis that a region of DNA containing the *OAS* genes introgressed from Neandertals into AMH. The haplotype R, which has a recent common ancestor in Europeans, has a high degree of similarity with the Neandertal sequence for up to ~185 kb. The estimated divergence time for haplotype R and the Neandertal sequence of only ~125 kya is considerably more recent than the estimated divergence time for Neandertal and AMH populations of ~300 kya (Green et al. 2010, Reich et al. 2010). SNPs diagnostic of haplotype R are broadly distributed in Eurasians and North Africans and are nearly absent in sub-Saharan Africans. Taken together, these observations strongly favor the hypothesis of recent introgression over the alternative hypothesis of ancestral shared polymorphism, providing further evidence that Eurasian ancestors originated in distinct African and Neandertal populations. Similar analyses of additional introgressive candidates will help to assess whether ancient population structure has made a contribution of any significance to the greater sharing of derived SNPs between non-Africans and Neandertals (Eriksson and Manica 2012).

It is interesting to note that two introgression events, one from Neandertals and one from Denisova, have affected both the *OAS1* (fig. 1) and *STAT2* loci (Mendez et al. 2012a, 2012b). Genomic scans should help to determine whether co-occurrence of introgression from these two taxa is a rare or a common phenomenon and whether some loci (e.g., those involved in immunity) are particularly susceptible to the effects of archaic admixture (Abi-Rached et al. 2011).

Supplementary Material

Supplementary material, supplementary tables S1 and S2, and supplementary figures S1–S3 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank Jennifer Reid for technical assistance and August Woerner, Krishna Veeramah, and Ryan Gutenkunst for comments on the manuscript. This work was funded by a National Science Foundation HOMINID grant (BCS0423670) to M.F.H.

References

1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* 491:56–65.

Abi-Rached L, Jobin MJ, Kulkarni S, et al. (23 co-authors). 2011. The shaping of modern human immune systems by multiregional admixture with archaic humans. *Science* 334:89–94.

Barkhash AV, Perelygin AA, Babenko VN, Myasnikova NG, Pilipenko PI, Romaschenko AG, Voevoda MI, Brinton MA. 2010. Variability in the 2′-5′-oligoadenylate synthetase gene cluster is associated with human predisposition to tick-borne encephalitis virus-induced disease. *J Infect Dis.* 202:1813–1818.

Durand EY, Patterson N, Reich D, Slatkin M. 2011. Testing for ancient admixture between closely related populations. *Mol Biol Evol.* 28: 2239–2252.

Eriksson A, Manica A. 2012. Effect of ancient population structure on the degree of polymorphism shared between modern human populations and ancient hominins. *Proc Natl Acad Sci U S A.* 109: 13956–13960.

Green RE, Krause J, Briggs AW, et al. (56 co-authors). 2010. A draft sequence of the Neandertal genome. *Science* 328:710–722.

Kong A, Thorleifsson G, Gudbjartsson DF, et al. (15 co-authors). 2010. Fine-scale recombination rate differences between sexes, populations and individuals. *Nature* 467:1099–1103.

Li JZ, Absher DM, Tang H, et al. (11 co-authors). 2008. Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 319:1100–1104.

Mendez FL, Watkins JC, Hammer MF. 2012a. A haplotype at *STAT2* introgressed from Neandertals 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.

Nordborg M. 2001. On detecting ancient admixture. In: Donnelly P, Foley RA, editors. Genes, fossils and behaviour: an integrated approach to human evolution. NATO Science Series: Life Sciences, Vol. 310. Amsterdam: IOS Press.

Reich D, Green RE, Kircher M, et al. (28 co-authors). 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468:1053–1060.

Sankararaman S, Patterson N, Li H, Paabo S, Reich D. 2012. The date of interbreeding between Neandertals and modern humans. *PLoS Genet.* 8:e1002947.

Wall JD. 2000. Detecting ancient admixture in humans using sequence polymorphism data. *Genetics* 154:1271–9.

Yang MA, Malaspinas AS, Durand EY, Slatkin M. 2012. Ancient structure in Africa unlikely to explain Neanderthal and non-African genetic similarity. *Mol Biol Evol.* 28:2239–2252.