bioRxiv preprint doi: https://doi.org/10.1101/051466; this version posted May 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Title: Adaptively introgressed Neandertal haplotype at the OAS locus functionally impacts innate immune responses in humans.

Authors

Aaron J. Sams[1,*†], Anne Dumaine[2†], Yohann Nédélec[2,3†], Vania Yotova[2], Caroline Alfieri[2,4], Jerome E. Tanner[2], Philipp W. Messer[1*††], and Luis B. Barreiro[2,5,*††]

Affiliations

[1]Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY

[2]Department of Genetics, Sainte-Justine Hospital Research Centre, University of Montreal, Montreal, Qc, Canada

[3]Department of Biochemistry, University of Montreal, Montreal, Qc, Canada

[4]Department of Microbiology and Immunology, University of Montreal, Montreal, Qc, Canada

[5]Department of Pediatrics, University of Montreal, Montreal, Qc, Canada

[†]These authors contributed equally to this work

[††]These authors jointly supervised this work

[*]Correspondence to: (A.J.S.) as2847@cornell.edu, (P.W.M.) messer@cornell.edu, (L.B.B.) luis.barreiro@umontreal.ca

bioRxiv preprint doi: https://doi.org/10.1101/051466; this version posted May 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

The 2'-5' oligoadenylate synthetase (OAS) locus encodes for three OAS enzymes (OAS1-3) involved in innate immune response. This region harbors high amounts of Neandertal ancestry in non-African populations; yet, strong evidence of positive selection in the OAS region is still lacking. Here we used a combination of neutral coalescent simulations and neutrality tests to firmly demonstrate a signal of adaptive introgression at the OAS locus. Furthermore, we characterized the functional consequences of the Neandertal haplotype in the transcriptional regulation of OAS genes at baseline and infected conditions. We found that cells from people with the Neandertal-like haplotype express lower levels of *OAS3* upon infection, as well as distinct isoforms of *OAS1* and *OAS2*. Notably, the Neandertal-introgressed haplotype reintroduced an ancestral splice variant of *OAS1* encoding a more active protein, suggesting that adaptive introgression occurred as a means to resurrect adaptive variation that had been lost outside Africa.

Introduction

Whole genome sequencing of several archaic human genomes(Green et al., 2010; Reich et al., 2010; Meyer et al., 2012; Prüfer et al., 2014; Sawyer et al., 2015) representing Neandertals and an as yet geographically and paleontologically unknown population referred to as Denisovans has revealed gene flow between these populations and the ancestors of present-day humans. Neandertal ancestry makes up approximately 0.5-2 percent of the ancestry of most living humans, with

bioRxiv preprint doi: https://doi.org/10.1101/051466; this version posted May 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

higher amounts of Neandertal ancestry found outside of Africa(Sankararaman et al., 2014; Vernot and Akey, 2014; Llorente et al., 2015). While it seems that there may have been widespread purifying selection against Neandertal ancestry on the ancestral African genomic background(Sankararaman et al., 2014; Vernot and Akey, 2014), some positive selection on Neandertal genes (adaptive introgression) has also been observed(Racimo et al., 2015; 2016). Neandertals and other archaic populations inhabited Eurasia for several hundred thousand years(Hublin, 2009), therefore some adaptive introgression is expected, particularly across phenotypes that are strongly influenced by direct interactions with the surrounding environment (Racimo et al., 2015), such as our immune response to infectious agents (Ségurel and Quintana-Murci, 2014).

The OAS locus on chromosome 12, which harbors three genes (*OAS1, OAS2, OAS3*) encoding the 2'-5' oligoadenylate synthetase enzymes has received considerable attention due to its clear signs of multiple archaic haplotypes in populations outside of Africa(Mendez et al., 2012; 2013), and the critical role of OAS genes in the innate immune response to viruses (Player and Torrence, 1998). The major Neandertal haplotype at the OAS locus spans a ~190 kilobase region between two surrounding recombination hotspots. This haplotype is absent from sub-Saharan African samples in the 1000 Genomes Project data, yet found at relatively high frequencies outside of Africa, reaching highest frequencies in European population samples (up to 41%, Figure 1). Indeed, the Neandertal haplotypes in the OAS region are among the most common Neandertal haplotypes among Europeans(Sankararaman et al., 2014).

bioRxiv preprint doi: https://doi.org/10.1101/051466; this version posted May 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The elevated frequency of the introgressed haplotype along with the key role OAS genes play in protective immunity against viral infections raises the possibility that introgressed Neandertal haplotypes at OAS may have been adaptive in modern humans. While some studies provide suggestive evidence of adaptive introgression at the OAS locus (Sankararaman et al., 2014; Racimo et al., 2016), strong evidence of positive selection in the OAS region is still lacking. Indeed, most studies failed to reject a model of neutral evolution for the Neandertal haplotype when using standard neutrality tests(Mendez et al., 2013; Deschamps et al., 2016). We hypothesize that the overall lack of signals of selection in the OAS region stems from the low power of standard neutrality tests to detect adaptive introgression(Sankararaman et al., 2016), and the particular genomic architecture of the region, which is characterized by the presence of two strong recombination hotspots (Figure 2A). The presence of these recombination hotspots is likely to significantly reduce power to detect signals of selection, particularly for tests based on haplotype-lengths or levels of linkage disequilibrium associated with the selected allele. Here we circumvent these issues by testing the hypothesis of adaptive introgression using extensive neutral coalescent simulations specifically tailored to match the genomic features of the OAS region in combination with ancient DNA data from Eurasia. We firmly demonstrate a population genetic signal of adaptive introgression at the OAS locus and characterize the functional consequences of the Neandertal haplotype in the transcriptional regulation of OAS genes in macrophages and peripheral blood mononuclear cells (PBMCs) at baseline and infected conditions.

bioRxiv preprint doi: https://doi.org/10.1101/051466; this version posted May 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Results

To investigate the hypothesis of non-neutral evolution at the OAS locus we simulated neutral sequence data using a demographic model based on several previously inferred parameters of human demographic history(Gravel et al., 2011; Tennessen et al., 2012; Vernot and Akey, 2015)(Figure S1) and explicitly incorporated estimated recombination rates in the OAS region (Methods). Our model included a single pulse of Neandertal introgression occurring over a span of 500-1000 years (sampled from a uniform distribution) into the ancestral Eurasian population after their population split from Africa. We investigated patterns of haplotype homozygosity surrounding Neandertal-like sites (NLS), derived alleles that are shared between Neandertals and a non-African population sample (CEU), but absent in a sub-Saharan African sample (YRI). This class of sites is frequently used as a conservative indicator of Neandertal introgression(Yang et al., 2012; Fu et al., 2014; Sankararaman et al., 2014). In each simulation we chose a NLS at random to measure two summary statistics, NLS frequency, and $H_{D/A}$, a statistic which measures the ratio of average pairwise homozygosity tract lengths among haplotypes carrying a derived versus ancestral allele at each NLS (Methods). This statistic is based on the *H* statistic that provides a general measure for haplotype lengths and has been shown to be robust in detecting both hard and soft selective sweeps(Schlamp et al., 2016). If Neandertal haplotypes have experienced a greater rate of positive selection than ancestral haplotypes, this should lead to a relative increase of *H* among the subgroup of haplotypes that carry Neandertal alleles versus those carrying only ancestral alleles, leading to higher values of $H_{D/A}$

bioRxiv preprint doi: https://doi.org/10.1101/051466; this version posted May 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

compared to a neutrally evolving locus with similar recombination rates (Figure S2). Additionally, we expected the frequency of NLS at the OAS locus to be elevated compared to the NLS frequency distribution observed across neutral simulations.

In practice, we observed both signals. We first examined the probability in our model that neutrally evolving Neandertal like sites would be segregating at their current frequencies. We found that the highest frequency NLS (DAF = 38.4%) fall in the extreme 1% of all simulations (lowest $P = 0.0055$) (Figure 2A). Next, we compared $H_{D/A}$ at each NLS in the OAS region to distributions of simulated NLS drawn to match local recombination rate and frequency (Methods), thereby controlling for associations between frequency and recombination, and haplotype lengths. We found several NLS in the OAS region that have $H_{D/A}$ values greater than 99% of comparable simulated NLS (lowest $P = 0.0023$) (Figure 2B). Further, we measured mean $H_{D/A}$ across all NLS in the OAS region and compared this to randomly sampled distributions of simulated NLS (again matching frequency and recombination rate). We found that the true mean $H_{D/A}$ in OAS is greater than four standard errors from the neutral expectation ($P < 10^{-5}$) (Figure 2D). Lastly, we considered the joint probability of observing a neutral NLS in our simulations with frequencies and $H_{D/A}$ values as high as those observed in the OAS region. Only 413 simulated NLS (out of 1,000,079) had both a higher frequency and $H_{D/A}$ value than the highest frequency/ $H_{D/A}$ pair in our NLS sample ($P < 5 \times 10^{-4}$) (Figure 2E). Combined, we identified the strongest evidence to date in support of adaptive introgression in the OAS region.

bioRxiv preprint doi: https://doi.org/10.1101/051466; this version posted May 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We further supported and contextualized our simulation results by examining the OAS region in a dataset of 230 ancient Eurasians(Mathieson et al., 2015). Assuming neutrality, the expected frequency of an allele in contemporary European populations can be predicted as a linear combination of allele frequencies sampled from representative ancient populations that have contributed ancestry to present-day European populations in different proportions(Lazaridis et al., 2014; Mathieson et al., 2015) (Methods). Therefore, using ancient allele frequencies estimated by Mathieson and colleagues, we calculated the expected allele frequency in four present day European samples from the 1,000 Genomes Project(Consortium, 2015) at 11 NLS falling within the bounds of the three OAS genes. To set up our null expectations we performed a similar analysis on a dataset of approximately one million SNPs scattered around the genome, generated by Mathieson and colleagues (by merging 213 ancient samples dated between 6500 and 300 BCE with sequencing data from four European samples from the 1,000 Genomes Project). We found that OAS NLS SNPs are outliers in the genome with respect to deviations from ancient frequencies. More specifically, we found that the allele frequency of 6 out of the 11 OAS SNPs tested in the *OAS1-OAS3* region have increased above the frequency predicted by ancient Eurasian samples by more than 20%, significantly more than what we observed for other SNPs genome-wide with comparable present-day frequencies (lowest $P = 0.00476$, Figure 2C, Table S1). Our findings at this single locus are consistent with results from the genome-wide selection scan performed by Mathieson and colleagues(2015) where the OAS region was also identified as an outlier.

bioRxiv preprint doi: https://doi.org/10.1101/051466; this version posted May 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Our population genetic results provide evidence that Neandertal alleles at the OAS locus have likely experienced positive selection at some point after their introduction into the human population, suggesting a possible functional role of these alleles in human innate immune responses. To study this possibility, we analyzed RNA-sequencing data collected on primary macrophages from 99 European-descent individuals, before and after *in-vitro* infection with *Salmonella typhimurium.* After 2 hours of infection, we found that all OAS genes were strongly up-regulated (up to 19-fold, $P < 1 \times 10^{-10}$, Figure S3), confirming the ability of *Salmonella* to activate the interferon (IFN) production pathway(Nauciel and Espinasse-Maes, 1992; Gulig et al., 1997; LaRock et al., 2015). Using genotype data available for the same individuals (673 SNPs spanning the OAS region, see methods) we tested if NLS were associated with variation in the expression levels of *OAS1*, *OAS2* or *OAS3,* in either infected or non-infected macrophages. We found that SNPs linked with the Neandertal haplotype impacted the expression levels of *OAS3* (i.e., they were expression quantitative trait loci, or *cis* eQTL for *OAS3*) (Figure 3A, false discovery rate (FDR) < 5%). Interestingly, these *cis* eQTL showed a much stronger effect in infected macrophages (best $P_{salmonella} = 3.5 \times 10^{-3}$ *vs* best $P_{\text{non-infected}} = 0.027$), supporting an interaction between the Neandertal haplotype and the *OAS3* response to *Salmonella* infection.

In addition to overall changes in expression, we took advantage of the power of RNA-sequencing data to test if NLS in the OAS regions influenced the ratio of alternative isoforms used for each of the OAS genes (i.e., alternative splicing QTL: asQTL). We found that SNPs associated with the Neandertal haplotype are

bioRxiv preprint doi: https://doi.org/10.1101/051466; this version posted May 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

significant asQTL for *OAS1* and *OAS2* in both infected and non-infected macrophages (FDR<< 1%; Figure 3B-C). The effect of the splice site variant rs10774671 at determining what isoform is primarily encoded by *OAS1* was particularly strong ($P \leq 2x10^{-32}$). The ancestral G allele at this SNP (AG at acceptor site) retains the splice site whereas the derived allele, A, (AA at acceptor site) disrupts the splice site leading to the usage of a distinct isoform (Figure S4). The Neandertal haplotype harbors the ancestral allele (encoding the p46 isoform), which is associated with high enzyme activity(Bonnevie-Nielsen et al., 2005), whereas the derived allele – present in virtually all non-Neandertal haplotypes found outside Africa (98.4% in CEU) – is associated with reduced enzymatic activity.

Because OAS genes are primarily involved in the control of viral infections we decided to validate our functional findings on peripheral blood mononuclear cells (PBMCs) from 30 individuals stimulated/infected with viral-ligands (polyI:C and gardiquimod), and live viruses (Influenza, Herpes simplex virus (HSV) 1 and HSV2). The individuals were chosen based on their genotype for the NLS rs1557866, a SNP that is a strong proxy for the presence or absence of the Neandertal haplotype in the OAS region (9 were homozygous for the Neandertal haplotype, 9 were heterozygous, and 12 homozygous for the modern human sequence).

As expected, we found that all viral-associated immune triggers led to a marked increase in *OAS1-3* gene expression levels, as measured by real-time PCR (up to

bioRxiv preprint doi: https://doi.org/10.1101/051466; this version posted May 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

30-fold, $P \leq 5.4x10^{-6}$, Figure 4A), concomitantly with the up-regulation of type-I and type-II interferon genes (Figure S5). Confirming the QTL results obtained in macrophages, we found that rs10774671 was a strong asQTL for *OAS1* in both non-infected and infected PBMCs ($P \leq 5.9x10^{-5}$, Figure 4B). Likewise, we found that the presence of the Neandertal haplotype was associated with increased expression levels of *OAS3,* particularly in PBMCs infected with influenza (P = 0.01) and the synthetic ligand gardiquimod ($P = 4.0x10^{-4}$), which mimics a single strand RNA infection (Figure 4B). Interestingly, the Neandertal haplotype harbors additional regulatory variants that only impact expression levels in a cell-type and immune stimuli specific fashion. For example, we found that the Neandertal haplotype is associated with increased expression levels of *OAS1* and *OAS2* in non-infected ($P<0.01$) and gardiquimod-stimulated PBMCs ($P<0.0004$), but not in macrophages nor in PBMCs treated with other viral agents. Collectively, our functional data shows a pervasive impact of the Neandertal haplotype on the regulation of *OAS* genes. Moreover, they show that the regulatory impact of the Neandertal haplotype varies depending on the cell type and the immune stimuli the cells are responding to.

Discussion

The Neandertal lineage was present in Eurasia for at least 400,000 years(Meyer et al., 2016), providing ample time for Neandertals to adapt to local disease environments. The admixture process, which likely fostered the transmission of pathogens between Neandertals and humans migrating out of Africa, could have

bioRxiv preprint doi: https://doi.org/10.1101/051466; this version posted May 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

also led to the exchange of genes useful in responding to local pathogens. Here, we have demonstrated that a previously reported case of Neandertal introgression at the OAS locus(Mendez et al., 2013) displays signatures of positive selection in the European population. Additionally, we have strengthened the case for adaptive introgression by providing direct functional evidence of a role for the Neandertal OAS haplotype to the regulatory responses in innate immune cells to infectious agents.

Under a model of constant directional selection with codominant fitness effects for Neandertal alleles we can use present-day frequencies and a plausible establishment frequency at the time of introgression to estimate the selection coefficient associated with the Neandertal haplotype at the OAS locus (Methods). Assuming an establishment frequency of 0.02, and introgression approximately 2000-2400 generations ago the selection coefficient in heterozygotes would be $s \sim$ 0.0014 - 0.0017. However, the observed allele frequency shift of 0.26 in only 200-340 generations (maximum shift in CEU from ancient samples; see Figure 2C**)** predicts that the selection coefficient associated with the Neandertal haplotype was 2.6-5.4 times larger during recent human evolution ($s \sim 0.0044$ - 0.0075). These results suggest that the Neandertal OAS haplotype has not experienced constant directional selection in humans.

Our results show that the Neandertal haplotype at OAS is associated with several regulatory variants that reduce expression of *OAS3* in response to infection, as well as encode alternate isoforms of *OAS1* and *OAS2*. These dramatic functional

bioRxiv preprint doi: https://doi.org/10.1101/051466; this version posted May 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

implications of the Neandertal OAS haplotype support our case for adaptive introgression at OAS. Yet, because distinct functional polymorphisms segregate together in the same haplotype, inferring the exact variant(s) targeted by positive selection remains a daunting task. We speculate, however, that the direct target of selection is likely to have been the splice variant identified in *OAS1*.

The Neandertal haplotype carries the ancestral allele (G) of the OAS1 splice variant (rs10774671), which is common both inside and outside of Africa. However, outside of Africa, the only haplotypes carrying this ancestral splice site are closely related to the Neandertal haplotype, with a few exceptions being rare recombined haplotypes (~2% of all haplotypes with the ancestral allele). This pattern reflects the possibility that Neandertal introgression, in effect, served as a means to resurrect the ancestral splice site from local extinction outside of Africa, probably following the out-of-Africa exodus. The Neandertal-introgressed allele encodes a protein variant (p46) that is associated with higher enzymatic activity(Bonnevie-Nielsen et al., 2005). The adaptive potential of this variant is supported by the observation that this variant (or other variants in strong LD with it) was shown to be associated with: *(i)* reduced infection and replication rates of West Nile virus ((Lim et al., 2009), but see (Bigham et al., 2011)), *(ii)* improved resistance to hepatitis C virus (HCV) infection(Awady et al., 2011; Kwon et al., 2012), and *(iii)* variable symptomology of Tick-Borne Encephalitis (TBE) Virus-Induced Disease (homozygous individuals for the Neandertal haplotype show the most severe symptoms of TBE). Strikingly, West Nile, hepatitis C and TBE are

bioRxiv preprint doi: https://doi.org/10.1101/051466; this version posted May 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

all members of the *Flaviviridae* family, suggesting that Flaviviruses might have been the main drivers of selection in *OAS1*.

The differential responses of homozygous carriers of the Neandertal OAS haplotype to the different viruses described above suggest that the Neandertal haplotype is not uniformly beneficial in humans. Thus, it is plausible that alleles at OAS, and particularly the *OAS1* splice variant, might be evolving under balancing selection. This hypothesis is supported by the observation that the *OAS1* splice variant (rs10774671) is found at high frequency worldwide (0.11-0.7), and that *OAS1* is among the most diverse genes in both humans and non-human primates. Indeed, a recent analysis of genome-wide sequence data from a total of 55 individuals from four non-human ape species, chimpanzee (*Pan troglodytes ellioti*), bonobo (*Pan paniscus*), gorilla (*Gorilla gorilla gorilla*), and orangutan (*Pongo abelii*), identified *OAS1* as in the top 1% of genes showing the largest levels of nucleotide diversity among ape species, consistent with a scenario of long-term balancing selection (*OAS2* and *OAS3* are ranked in the 60th and 36th percentile of the genome-wide distribution, respectively). Further supporting the idea of balancing selection on the introgressed haplotypes, our functional data suggest that the Neandertal haplotype contributes a range of gene expression responses in a cell-type and stimulus-specific manner.

In conclusion, our study demonstrates that the frequency and haplotype distribution of Neandertal-like sites can be used in a neutral simulation framework that accounts for local genomic context to investigate the history of selection at a

bioRxiv preprint doi: https://doi.org/10.1101/051466; this version posted May 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

candidate locus for which genome-wide tests of selection provide ambiguous results. When combined with functional data, our results provide the strongest evidence to date in support of adaptive introgression in the OAS region. More generally, our study raises the possibility that adaptive introgression might not necessarily occur to select newly introduced variants but rather as a means to resurrect adaptive variation into modern humans who had lost it due to demographic events.

Materials and Methods

1. Genome alignments and identification of Neandertal-like sites

Human/chimpanzee ancestral states were computed by parsimony using alignments from the UCSC Genome Browser for the human reference (hg19) and three outgroups chimpanzee (panTro2), orangutan (ponAbe2), and rhesus macaque (rheMac2)(Kent et al., 2002). Ancestral state was assumed to be the chimpanzee allele (if available) if its state was confirmed by matching either orangutan or macaque. All sites with no inferred ancestral state were removed from our analysis.

We filtered the Altai Neandertal genome(Prüfer et al., 2014) using the map35_50 set of minimum filters provided at (https://bioinf.eva.mpg.de/altai_minimal_filters/). We combined this filtered dataset with the CEU (Utah residents (CEPH) with Northern and Western European Ancestry) and YRI (Yoruba in Ibadan, Nigeria) samples from the 1000

bioRxiv preprint doi: https://doi.org/10.1101/051466; this version posted May 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Genomes Project Phase 3(Consortium, 2015), which we downloaded from (https://mathgen.stats.ox.ac.uk/impute/1000GP%20Phase%203%20haplotypes%206%20October%202014.html).

We first extracted all variants that were biallelic in our human (CEU+YRI), Neandertal, and chimpanzee alignments. We considered as Neandertal-like sites (NLS) only those variants where the African sample (YRI) had a derived allele frequency of zero and both CEU and Neandertal carried the derived allele. Finally, we required the derived allele to be present in CEU in at least two copies in order to calculate our haplotype-based test statistic ($H_{D/A}$).

2. Demographic model and neutral coalescent simulations

We performed coalescent simulations of the demographic history of the European, African, and Neandertal populations applied by Vernot and Akey(2015) based on previously inferred demographic models (Gravel et al., 2011; Tennessen et al., 2012) with some minor adjustments, including removing the East Asian component of the model and allowing for Neandertal admixture to occur over a time span from 500 to 1000 years (Figure S1). Simulations were performed with macs(Chen et al., 2009) in order to explicitly simulate the genetic map (downloaded with the 1000 Genomes samples at the link above) of the 2Mb region centered on OAS (chr12:112400000-114400000). All simulations were performed with the parameters specified in Figure S1, assuming 25 years per generation and a mutation rate of 2.5×10^{-8} per bp per generation. A sample macs command is given at the end of this section.

bioRxiv preprint doi: https://doi.org/10.1101/051466; this version posted May 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Data were thinned in a manner similar to Sankararaman et al. 2014 (2014) to account for imperfect SNP ascertainment in the 1000 Genomes dataset, such that SNPs with minor allele count of 1, 2, 3, 4, 5, 6, 7, 8, 9, and >=10 were accepted with probabilities 0.25, 0.5, 0.75, 0.8, 0.9, 0.95, 0.96, 0.97, 0.98, and 0.99, respectively. Additionally, we only kept SNPs that were polymorphic in the simulated CEU sample. Finally, we performed an additional thinning of SNPs with uniform probability of 0.05 of removal to account for slightly elevated SNP density in the simulated data. The resulting simulated datasets had an average SNP density of 4.9 SNPs per kb compared to 3.6 in the real data. This is a slightly larger than ideal difference in SNP density, but we note that neither derived allele frequency, nor our primary haplotype-based test statistic (described below) should be particularly sensitive to SNP density**.** In fact, Figure S6 illustrates that our statistic is conservative with respect to SNP density.

To estimate the probabilities of adaptive introgression conditional on the simulated locus containing a NLS (SNPs in which the non-African and Neandertal sample contain a derived allele that is absent from the African sample) we randomly selected a single NLS from the central fifth of the chromosome (positions 0.4-0.6 in the unit scaled locus) in which we calculated our test statistics. If an NLS was found, its position on the chromosome was recorded and test statistics were calculated. If no NLS was found, we repeated the simulation step. This approach allowed us to estimate under our demographic model, the proportion of independent loci that will have no obvious signal of introgression under neutrality as the proportion of successful simulation runs to the total number of attempts. We

bioRxiv preprint doi: https://doi.org/10.1101/051466; this version posted May 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

obtained an approximate 21 percent chance of observing Neandertal introgression at the OAS locus under a neutral model. Additionally, we repeated this simulation approach, setting Neandertal introgression to zero, in a set of 10,000 simulations in order to estimate the likelihood of observing a NLS in the case of incomplete lineage sorting (ILS), to estimate the contribution of ILS to our null distribution. We observed a roughly 0.6% chance that a NLS is produced by ILS at the OAS locus, suggesting that our null distribution is not significantly biased by ILS.

Sample macs command:

```
macs 416 2000000 -t 0.000731 -R oas_recrates.txt -I 3 216 2 198 0 -n 1
58.0027359781 -n 2 0.205198358413 -n 3 70.0410396717 -eg 0 1 482.67144247 -
eg 1e-08 3 570.175408787 -em 2e-08 1 3 0.409323665059 -em 3e-08 3 1
0.409323665059 -eg 0.00699726402189 3 16.9958723831 -en 0.00699727402189
1 1.98002735978 -en 0.03146374829 3 0.759313629055 -en 0.0561902522706 3
1.15461721321 -em 0.0561902622706 1 3 4.386 -em 0.0561902722706 3 1 4.386
-em 0.056088019083 3 2 17.3006865925 -em 0.0569023606299 3 2 0.0 -ej
0.08129305892 3 1 -en 0.08129306892 1 1.98002735978 -en 0.202462380301 1
1.0 -ej 0.89014331235 2 1 -h 1e3 1e3
```

3. Frequency and haplotype-based tests of neutrality

We examined the consistency of genetic variation with our neutral model using several approaches. First, we examined the likelihood of observing (Neandertal) allele frequencies as high as the OAS locus. Under neutrality, allele frequency is

bioRxiv preprint doi: https://doi.org/10.1101/051466; this version posted May 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

not dependent upon recombination rate, therefore, we can estimate the likelihood of our observed NLS frequency in the OAS region. For example, mean allele frequency of all NLS observed between chr12:113344000-113450000 is 0.35 (Figure 2A). Frequencies this great or greater were observed in our simulations less than one percent of the time (0.0085), suggesting that this haplotype frequency is unlikely under neutrality.

Additionally, we wanted to examine if the haplotypes carrying NLS at the OAS locus are longer than expected under neutrality when conditioning on the observed frequencies and the underlying genetic map, which would provide an additional signature of selection on introgressed haplotypes. For this purpose, we modified a simple haplotype statistic *H*(Schlamp et al., 2016), which measures the average length of pairwise homozygosity tracts in base pairs – a quantity that is very straightforward to interpret. As selective sweeps are expected to create long haplotypes around the selected site, the *H* statistic should be higher in samples containing positively selected haplotypes compared to samples containing neutrally evolving haplotypes, when frequency and recombination are properly controlled, similar to other statistics based on haplotype lengths, such as *EHH, iHS,* and *nSL* (Sabeti, P.C. et al., 2002; Voight et al., 2006; Ferrer-Admetlla et al., 2014). However, in contrast to these other statistics, *H* does not require specification of analysis parameters such as minimum haplotype homozygosity levels below which haplotypes are no longer extended.

bioRxiv preprint doi: https://doi.org/10.1101/051466; this version posted May 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Under adaptive introgression, we specifically expect the introgressed haplotypes to be longer than the ancestral haplotypes. We therefore defined our test statistic, $H_{D/A}$, as:

$$H_{D/A} = \ln(H_D/H_A);$$

where H_D and H_A are H calculated across haplotypes carrying derived NLS allele versus the ancestral allele.

We calculated this statistic for every simulated NLS and for every NLS in our true sample. We compared our simulation results to the observed data in two ways. First, for each observed SNP we gathered all simulated NLS within 0.025 frequency and 20 kb. We then calculate the empirical likelihood of observing each observed NLS under our neutral demographic model (Figure 2A). All NLS within the bounds of the three OAS genes are tightly linked, therefore, the peak signal at the center of this locus suggests that the probability of observing these $H_{D/A}$ values are less than one percent.

However, to examine a single hypothesis test of the null hypothesis, that the true Neandertal haplotype is not significantly longer than expected under neutrality, we performed a randomization test. First, we calculated mean $H_{D/A}$ across this central locus (chr12:113344000-113450000). Next, we randomly chose one control SNP per true NLS (using the same matching conditions above) and calculated the mean of the control set. We repeated this procedure 100,000 times to create a normally

bioRxiv preprint doi: https://doi.org/10.1101/051466; this version posted May 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

distributed null distribution, which we then used to calculate a Z-score (4.5) and corresponding P (3.35×10^{-6}) for our one-sided hypothesis test (Figure 2B).

Finally, we asked how unusual our observation of NLS frequency and $H_{D/A}$ scores were across all simulations. Across approximately one million simulations (1,000,079), we only observed 413 simulations that produced an NLS with a frequency and $H_{D/A}$ score as great or greater than the highest value observed among the true NLS (Figure 2C). In other words, the likelihood of our true sample under the neutral demographic scenario is less than 5×10^{-4}.

4. Analysis of ancient Eurasian data

We utilized supplementary data table 3 from Mathieson et al. (2015). This table includes maximum likelihood allele frequency estimates for three ancient population samples (HG- Hunter-gatherer, EF- Early farmer, SA- Steppe ancestry) and four present day European samples from the 1,000 Genomes Project (see "Genome-wide scan for selection" section of methods in(Mathieson et al., 2015)). We intersect this table with allele frequencies for 1,000 Genomes Yorubans (YRI) and the Altai Neandertal genotypes and only analyze sites for which we have data for all samples (1,004,612 SNPs).

To calculate the expected allele frequency in modern samples under drift, we used the estimated proportions (m) of (HG, EF, SA) in each of the four present-day samples estimated by Mathieson et al. (2015): CEU = (0.196, 0.257, 0.547), GBR

bioRxiv preprint doi: https://doi.org/10.1101/051466; this version posted May 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

437 = (0.362, 0.229, 0.409), IBS = (0, 0.686, 0.314) and TSI = (0, 0.645, 0.355). We
438 calculated the expected frequency E[p] of site as:

$$E[p] = \sum_{i}^{\{HG,EF,SA\}} (p_i \, x \, m_i)$$

439

440 Next, we calculated the absolute difference between observed and expected allele
441 frequency in all four present-day European samples at all available sites. To test
442 the null hypothesis that OAS NLS have not changed in frequency more than
443 expected under neutrality at 11 SNPs in the central OAS region
444 (chr12:113200000-113600000), we first calculated the fraction of all autosomal
445 SNPs in the dataset at similar present-day frequency (within 1 percent in the
446 folded frequency spectrum) of each OAS NLS. Finally, we calculated the fraction
447 of autosomal SNPs with an absolute observed minus expected frequency
448 difference greater than or equal to the OAS NLS. These results are given in Table
449 1 and illustrated in Figure 2A.

450 This test does not explicitly incorporate variance in estimated ancient allele
451 frequency. However, any bias in ancient allele frequency estimation should be
452 distributed randomly across the genome. Therefore, our comparison to a genome-
453 wide distribution of SNPs at similar present-day frequency should incorporate
454 most of this error. Nonetheless, the selection test performed by Mathieson et
455 al.(2015) does incorporate such error, so we can also look to the Ps from that test

bioRxiv preprint doi: https://doi.org/10.1101/051466; this version posted May 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

to ensure consistency with our results. In fact, when considering OAS as a single locus, these results are highly significant and consistent with ours (see Table S1).

5. Estimation of selection coefficients

To estimate the selection coefficient s under constant positive selection for a given starting frequency (x_0), final frequency (x_1), and number of generations between these estimates (Δt), we assumed a model of standard logistic growth of a codominant allele:

$$x_1 = x_0/[x_0+(1- x_0)\exp(-s\Delta t)].$$

This equation can be easily solved to obtain s, given x_0 , x_1, and Δt.

6. Sample collection

Buffy coats from 99 healthy European-descent donors were obtained from Indiana Blood Center (Indianapolis, IN, USA). Only individuals self-reported as currently healthy and not under medication were included in the study. The project was approved by the ethics committee at the CHU Sainte-Justine (protocol #4022). The individuals recruited in this study were males aged 18 to 55 years old.

7. DNA Extraction and genotyping

DNA from each of the blood donors was extracted using the Gentra Pure Gene blood kit (Qiagen). Genotyping of each individual was then performed by Illumina's HumanOmni5Exome bead Chip array and complemented with imputed

bioRxiv preprint doi: https://doi.org/10.1101/051466; this version posted May 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

475 data from the 1000 Genomes data using Impute2(Howie et al., 2009). Here, we
476 only focused on genetic diversity surrounding the OAS region –
477 chr12:113229549-113574044 (~ 344Kb) spanning from the beginning of *RPH3A*
478 to the end of *RASAL1* – for a total of 673 SNPs with a MAF above 10%.

479 *8. Isolation of monocytes and differentiation of macrophages*

480 Blood mononuclear cells were isolated by Ficoll-Paque centrifugation. Monocytes
481 were purified from peripheral blood mononuclear cells (PBMCs) by positive
482 selection with magnetic CD14 MicroBeads (Miltenyi Biotech) using the
483 autoMACS Pro Separator. All samples had purity levels above 90%, as measured
484 by flow cytometry using an antibody against CD14 (BD Biosciences). Monocytes
485 were then cultured for 7 days in RPMI-1640 (Fisher) supplemented with 10%
486 heat-inactivated FBS (FBS premium, US origin, Wisent), L-glutamine (Fisher)
487 and M-CSF (20ng/mL; R&D systems). Cell cultures were fed every 2 days with
488 complete medium supplemented with the cytokines previously mentioned. Before
489 infection, we systematically verified that the differentiated macrophages presented
490 the expected phenotype for non-activated macrophages (CD1a+, CD14+, CD83−,
491 and HLA-DRlow (BD Biosciences)).

492 *9. Bacterial preparation and infection of macrophages*

493 The day prior to infection, aliquots of *Salmonella typhimurium* (Keller strain) were
494 thawed and bacteria were grown overnight in Tryptic Soy Broth (TSB) medium.
495 Bacterial culture was diluted to mid-log phase prior to infection and supernatant

bioRxiv preprint doi: https://doi.org/10.1101/051466; this version posted May 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

density was checked at OD600. Monocyte-derived macrophages were then infected with *Salmonella typhimurium* at a multiplicity of infection (MOI) of 10:1. A control group of non-infected macrophages was treated the same way but using medium without bacteria. After 2 hours in contact with the bacteria, macrophages were washed and cultured for another hour in the presence of 50 mg/ml of gentamycin in order to kill all extracellular bacteria present in the medium. The cells were then washed a second time and cultured in complete medium with 3 mg/ml gentamycin for an additional 2 hours, the time point to which we refer in the main text.

10. Infection/stimulation of PBMC

PBMCs from a subset of 30 individuals used to derive macrophages were cultured in RPMI-1640 (Fisher) supplemented with 10% heat-inactivated FBS (FBS premium, US origin, Wisent) and 1% L-glutamine (Fisher). The 30 individuals were chosen based on their genotype for kgp4570197, a SNP which derived allele is of Neandertal origin and that we used as a proxy to identify individuals harbouring the Neandertal haplotype in the OAS region. From the 30 individuals, and based on this SNP, 9 individuals were homozygous for the Neandertal haplotype, 9 were heterozygous, and 12 homozygous for the modern human sequence.

For each of the tested individuals, PBMCs (1 million per condition) were stimulated/infected with one of the following viral-associated immune challenges: polyI:C (10 µg/ml, TLR3 agonist), gardiquimod (0.5µg/ml, TLR7 and TLR8

bioRxiv preprint doi: https://doi.org/10.1101/051466; this version posted May 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

agonist), Influenza PR8 WT (multiplicity of infection (MOI) of 0.05:1), Herpes simplex virus (HSV) 1 ($1.55x10^2$ CPE), and HSV2 ($19.5x10^4$ CPE). PBMCs were stimulated/infected for 4 hours with TLR ligands and Influenza, and 6h with HSV1 and HSV2. A control group of non-infected PBMC was treated the same way but with only medium.

11. RNA extraction, RNA-seq library preparation, and sequencing

Total RNA was extracted from the non-infected and infected/stimulated cells using the miRNeasy kit (Qiagen). RNA quantity was evaluated spectrophotometrically, and the quality was assessed with the Agilent 2100 Bioanalyzer (Agilent Technologies). Only samples with no evidence of RNA degradation (RNA integrity number > 8) were kept for further experiments. RNA-sequencing libraries were prepared using the Illumina TruSeq protocol. Once prepared, indexed cDNA libraries were pooled (6 libraries per pool) in equimolar amounts and sequenced with single-end 100bp reads on an Illumina HiSeq2500. Results based on the entire dataset are described elsewhere (Nédélec *et al*., under revision). Here, we only studied transcript-level and gene-level expression estimates for *OAS1*, *OAS2* and *OAS3*.

12. Quantifying gene expression values from RNA-seq data

Adaptor sequences and low quality score bases (Phred score < 20) were first trimmed using Trim Galore (version 0.2.7). The resulting reads were then mapped to the human genome reference sequence (Ensembl GRCh37 release 65) using

bioRxiv preprint doi: https://doi.org/10.1101/051466; this version posted May 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

539 TopHat (version 2.0.6) and using a hg19 transcript annotation database
540 downloaded from UCSC. Gene-level expression estimates were calculated using
541 featureCounts (version 1.4.6-p3) and transcript-level expression values were
542 obtained using RSEM under default parameters.

543 *13. Quantitative real time PCR*

544 For the PBMC samples we measured the expression levels of OAS and interferon
545 genes using real time PCR. 100ng of high-quality RNA was reverse-transcribed
546 into cDNA using the qScript cDNA SuperMix (Quanta Biosciences). Quantitative
547 real time PCR was performed using 96.96 Dynamic Array™ IFCs and the
548 BioMark™ HD System from Fluidigm. For the TaqMan gene assays, we used the
549 following TaqMan Gene Expression Assay (Applied BioSystems) to quantify the
550 expression levels of interferon genes: *IFNA1* (Hs03044218), *IFNA6*
551 (Hs00819627), and *IFNG* (Hs00989291). To quantify the overall expression levels
552 of OAS genes, we used probes that capture all common isoforms of *OAS1*
553 (Hs00973635), *OAS2* (Hs00942643), and OAS3 (Hs00196324). Custom-made
554 probes were designed to specifically target the short-isoform of *OAS2* (Forward
555 Primer Sequence CTGCAGGAACCCGAACAGTT; Reverse Primer Sequence
556 ACTCATGGCCTAGAGGTTGCA; Reporter Sequence
557 AGAGAAAAGCCAAAGAA). As housekeeping genes we used: *GAPDH*
558 (Hs02758991), *GUSB* (Hs99999908), *HPRT1* (Hs99999909), and *POLR2A*
559 (Hs00172187). The results reported in the manuscript used *POLR2A* as a reference
560 but all conclusions remain unchanged when using any of the other housekeeping
561 genes.

bioRxiv preprint doi: https://doi.org/10.1101/051466; this version posted May 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We start by doing a preamplification of the cDNA using the PreAmp Master Mix (Fluidigm). Preamplified cDNA was then diluted 2X on a solution of 10 mM Tris–HCl (pH 8.0) and 0.1 mM EDTA. In order to prepare samples for loading into the integrated fluid circuit (IFC), a mix was prepared consisting of 360 µL TaqMan Fast Advanced Master Mix (Applied BioSystems) and 36 µL 20× GE Sample Loading Reagent (Fluidigm). 2.75 µL of this mix was dispensed to each well of a 96-well assay plate and mixed with 2.25 µL of preamplified cDNA. Following priming of the IFC in the IFC Controller HX, 5 µL of the mixture of cDNA and loading reagent were dispensed in each of the sample inlet of the 96.96 IFC. For the TaqMan gene assays, 5 µL of mixes consisting of 2.5 µL 20× TaqMan Gene Expression Assay (Applied BioSystems) and 2.5 µL 2X Assay Loading Reagent (Fluidigm) were dispensed to each detector inlet of the 96.96 IFC. After loading the assays and samples into the IFC in the IFC Controller HX, the IFC was transferred to the BioMark HD and PCR was performed using the thermal protocol GE 96 × 96 Fast v1.pcl. This protocol consists of a Thermal Mix of 70 °C, 30 min; 25 °C, 10 min, Hot Start at 95 °C, 1 min, PCR Cycle of 35 cycles of (96 °C, 5 s; 60 °C, 20 s). Data was analysed using Fluidigm Real-Time PCR Analysis software using the Linear (Derivative) Baseline Correction Method and the Auto (Detectors) Ct Threshold Method.

To quantify the expression levels of the *OAS1* isoform associated with the derived allele at the splicing variant rs10774671 we used SybrGreen and the following forward (GCTGAGGCCTGGCTGAATTA), and reverse (CCACTTGTTAGCTGATGTCCTTGA) primers. PCR was performed using the

bioRxiv preprint doi: https://doi.org/10.1101/051466; this version posted May 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

thermal protocol 50 °C, 2 min; 95 °C, 10 min, PCR Cycle of 40 cycles of (95 °C, 15 s; 60 °C, 1 min). A melting curve was also performed to check for non-specific amplification.

14. Genotype–Phenotype Association Analysis

eQTL, asQTL were performed against *OAS1*, *OAS2* and *OAS3*. We examined associations between SNP genotypes and the phenotype of interest using a linear regression model, in which phenotype was regressed against genotype. In particular, expression levels were considered as the phenotype when searching for eQTL and the percentage usage of each isoform in each gene when mapping asQTL. To avoid low power caused by rare variants, only SNPs in the OAS region with a minor allele frequency of 10% across all individuals were tested (i.e., 673 SNPs within the region chr12:113229549-113574044). In all cases, we assumed that alleles affected the phenotype in an additive manner. For the eQTL and asQTL analyses on macrophages we mapped Salmonella-infected, and non-infected samples separately. For the non-infected and infected/stimulated PBMCs we only tested expression levels against the SNPs identified as eQTL or asQTL in the macrophage data (specifically, the SNPs for which boxplots are shown in Figure 3).

Author Contributions: Conception and design: AJS, PWM, LBB; Acquisition of data: AJS, AD, YN, VY, LBB; Analysis and interpretation of data: AJS, AD, YN, PWM, LBB; Contributed unpublished, essential data, or reagents: CA, JET; Drafting or revising the article: AJS, PWM, LBB.

bioRxiv preprint doi: https://doi.org/10.1101/051466; this version posted May 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Acknowledgements: We thank all members from the Barreiro and Messer lab for helpful discussions and comments on the manuscript. We thank Dr Silvia Vidal for the gift of the Influenza PR8 WT used in this study, Dr Hugo Soudeyns and Doris Ransy for advice with the viral infections, and Marc Montero and George (PJ) Perry for sharing their analysis on nucleotide diversity levels of OAS genes in non-human primate species.

Funding Statement: This study was funded by grants from the Canadian Institutes of Health Research (301538 and 232519), the Human Frontiers Science Program (CDA-00025/2012) and the Canada Research Chairs Program (950-228993) (to L.B.B.). Y.N. was supported by a fellowship from the *Réseau de Médecine Génétique Appliquée* (RMGA).

References:

Awady, El, M.K., Anany, M.A., Esmat, G., Zayed, N., Tabll, A.A., Helmy, A., Zayady, El, A.R., Abdalla, M.S., Sharada, H.M., Raziky, El, M., Akel, El, W., Abdalla, S., Bader El Din, N.G., 2011. Single nucleotide polymorphism at exon 7 splice acceptor site of OAS1 gene determines response of hepatitis C virus patients to interferon therapy. Journal of Gastroenterology and Hepatology. 26, 843–850.

Bigham, A.W., Buckingham, K.J., Husain, S., Emond, M.J., Bofferding, K.M., Gildersleeve, H., Rutherford, A., Astakhova, N.M., Perelygin, A.A., Busch, M.P., Murray, K.O., Sejvar, J.J., Green, S., Kriesel, J., Brinton, M.A., Bamshad, M., 2011. Host Genetic Risk Factors for West Nile Virus Infection and Disease Progression. PLoS ONE. 6, e24745.

Bonnevie-Nielsen, V., Field, L.L., Lu, S., Zheng, D.-J., Li, M., Martensen, P.M., Nielsen, T.B., Beck-Nielsen, H., Lau, Y.L., Pociot, F., 2005. Variation in antiviral 2',5'-oligoadenylate synthetase (2“5”AS) enzyme activity is controlled by a single-nucleotide polymorphism at a splice-acceptor site in the OAS1 gene. The American Journal of Human Genetics. 76, 623–633.

Chen, G.K., Marjoram, P., Wall, J.D., 2009. Fast and flexible simulation of DNA

bioRxiv preprint doi: https://doi.org/10.1101/051466; this version posted May 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

638 sequence data. Genome Research. 19, 136–142.
639 Consortium, T.1.G.P., 2015. A global reference for human genetic variation.
640 Nature. 526, 68–74.
641 Deschamps, M., Laval, G., Fagny, M., Itan, Y., Abel, L., Casanova, J.-L., Patin,
642 E., Quintana-Murci, L., 2016. Genomic Signatures of Selective Pressures and
643 Introgression from Archaic Hominins at Human Innate Immunity Genes.
644 American Journal of Human Genetics. 98, 5–21.
645 Ferrer-Admetlla, A., Liang, M., Korneliussen, T., Nielsen, R., 2014. On detecting
646 incomplete soft or hard selective sweeps using haplotype structure. Molecular
647 Biology and Evolution. 31, 1275–1291.
648 Fu, Q., Li, H., Moorjani, P., Jay, F., Slepchenko, S.M., Bondarev, A.A., Johnson,
649 P.L.F., Aximu-Petri, A., Prüfer, K., de Filippo, C., Meyer, M., Zwyns, N.,
650 Salazar-García, D.C., Kuzmin, Y.V., Keates, S.G., Kosintsev, P.A., Razhev,
651 D.I., Richards, M.P., Peristov, N.V., Lachmann, M., Douka, K., Higham,
652 T.F.G., Slatkin, M., Hublin, J.-J., Reich, D., Kelso, J., Viola, T.B., Pääbo, S.,
653 2014. Genome sequence of a 45,000-year-old modern human from western
654 Siberia. Nature. 514, 445–449.
655 Gravel, S., Henn, B.M., Gutenkunst, R.N., Indap, A.R., Marth, G.T., Clark, A.G.,
656 Yu, F., Gibbs, R.A., Bustamante, C.D., Project, 1.G., 2011. Demographic
657 history and rare allele sharing among human populations. Proceedings of the
658 National Academy of Sciences of the United States of America. 108, 11983–
659 11988.
660 Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M.,
661 Patterson, N., Li, H., Zhai, W., Fritz, M.H.-Y., 2010. A draft sequence of the
662 Neandertal genome. Science. 328, 710–722.
663 Gulig, P.A., Doyle, T.J., Clare-Salzler, M.J., Maiese, R.L., Matsui, H., 1997.
664 Systemic infection of mice by wild-type but not Spv- Salmonella typhimurium
665 is enhanced by neutralization of gamma interferon and tumor necrosis factor
666 alpha. Infection and Immunity. 65, 5191–5197.
667 Howie, B.N., Donnelly, P., Marchini, J., 2009. A Flexible and Accurate Genotype
668 Imputation Method for the Next Generation of Genome-Wide Association
669 Studies. PLoS Genetics. 5, e1000529.
670 Hublin, J.J., 2009. The origin of Neandertals. Proceedings of the National
671 Academy of Sciences. 106, 16022–16027.
672 Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M.,
673 Haussler, D., 2002. The human genome browser at UCSC. Genome Research.
674 12, 996–1006.
675 Kwon, Y.-C., Kang, J.-I., Hwang, S.B., Ahn, B.-Y., 2012. The ribonuclease l-
676 dependent antiviral roles of human 2′,5′-oligoadenylate synthetase family
677 members against hepatitis C virus. FEBS letters. 587, 156–164.
678 LaRock, D.L., Chaudhary, A., Miller, S.I., 2015. Salmonellae interactions with
679 host processes. Nature Reviews Microbiology. 13, 191–205.
680 Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K.,
681 Sudmant, P.H., Schraiber, J.G., Castellano, S., Lipson, M., Berger, B.,
682 Economou, C., Bollongino, R., Fu, Q., Bos, K.I., Nordenfelt, S., Li, H., de
683 Filippo, C., Prüfer, K., Sawyer, S., Posth, C., Haak, W., Hallgren, F.,

bioRxiv preprint doi: https://doi.org/10.1101/051466; this version posted May 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

684 Fornander, E., Rohland, N., Delsate, D., Francken, M., Guinet, J.-M., Wahl, J.,
685 Ayodo, G., Babiker, H.A., Bailliet, G., Balanovska, E., Balanovsky, O.,
686 Barrantes, R., Bedoya, G., Ben-Ami, H., Bene, J., Berrada, F., Bravi, C.M.,
687 Brisighelli, F., Busby, G.B.J., Cali, F., Churnosov, M., Cole, D.E.C., Corach,
688 D., Damba, L., van Driem, G., Dryomov, S., Dugoujon, J.-M., Fedorova, S.A.,
689 Romero, I.G., Gubina, M., Hammer, M., Henn, B.M., Hervig, T., Hodoglugil,
690 U., Jha, A.R., Karachanak-Yankova, S., Khusainova, R., Khusnutdinova, E.,
691 Kittles, R., Kivisild, T., Klitz, W., Kucinskas, V., Kushniarevich, A., Laredj,
692 L., Litvinov, S., Loukidis, T., Mahley, R.W., Melegh, B., Metspalu, E.,
693 Molina, J., Mountain, J., Näkkäläjärvi, K., Nesheva, D., Nyambo, T., Osipova,
694 L., Parik, J., Platonov, F., Posukh, O., Romano, V., Rothhammer, F., Rudan,
695 I., Ruizbakiev, R., Sahakyan, H., Sajantila, A., Salas, A., Starikovskaya, E.B.,
696 Tarekegn, A., Toncheva, D., Turdikulova, S., Uktveryte, I., Utevska, O.,
697 Vasquez, R., Villena, M., Voevoda, M., Winkler, C.A., Yepiskoposyan, L.,
698 Zalloua, P., Zemunik, T., Cooper, A., Capelli, C., Thomas, M.G., Ruiz-
699 Linares, A., Tishkoff, S.A., Singh, L., Thangaraj, K., Villems, R., Comas, D.,
700 Sukernik, R., Metspalu, M., Meyer, M., Eichler, E.E., Burger, J., Slatkin, M.,
701 Pääbo, S., Kelso, J., Reich, D., Krause, J., 2014. Ancient human genomes
702 suggest three ancestral populations for present-day Europeans. Nature. 513,
703 409–413.
704 Lim, J.K., Lisco, A., McDermott, D.H., Huynh, L., Ward, J.M., Johnson, B.,
705 Johnson, H., Pape, J., Foster, G.A., Krysztof, D., Follmann, D., Stramer, S.L.,
706 Margolis, L.B., Murphy, P.M., 2009. Genetic Variation in OAS1 Is a Risk
707 Factor for Initial Infection with West Nile Virus in Man. PLoS Pathogens. 5,
708 e1000321.
709 Llorente, M.G., Jones, E.R., Eriksson, A., Siska, V., Arthur, K.W., Arthur, J.W.,
710 Curtis, M.C., Stock, J.T., Coltorti, M., Pieruccini, P., Stretton, S., Brock, F.,
711 Higham, T., Park, Y., Hofreiter, M., G, B.D., Bhak, J., Pinhasi, R., Manica,
712 A., 2015. Ancient Ethiopian genome reveals extensive Eurasian admixture
713 throughout the African continent. Science.
714 Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg,
715 S.A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., Sirak, K.,
716 Gamba, C., Jones, E.R., Llamas, B., Dryomov, S., Pickrell, J., Arsuaga, J.L.,
717 de Castro, J.M.B., Carbonell, E., Gerritsen, F., Khokhlov, A., Kuznetsov, P.,
718 Lozano, M., Meller, H., Mochalov, O., Moiseyev, V., Guerra, M.A.R.,
719 Roodenberg, J., Vergès, J.M., Krause, J., Cooper, A., Alt, K.W., Brown, D.,
720 Anthony, D., Lalueza-Fox, C., Haak, W., Pinhasi, R., Reich, D., 2015.
721 Genome-wide patterns of selection in 230 ancient Eurasians. Nature.
722 Mendez, F.L., Watkins, J.C., Hammer, M.F., 2012. A Haplotype at STAT2
723 Introgressed from Neanderthals and Serves as a Candidate of Positive
724 Selection in Papua New Guinea. The American Journal of Human Genetics.
725 91, 265–274.
726 Mendez, F.L., Watkins, J.C., Hammer, M.F., 2013. Neandertal origin of genetic
727 variation at the cluster of OAS immunity genes. Molecular Biology and
728 Evolution. 30, 798–801.
729 Meyer, M., Arsuaga, J.L., de Filippo, C., Nagel, S., Aximu-Petri, A., Nickel, B.,

bioRxiv preprint doi: https://doi.org/10.1101/051466; this version posted May 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

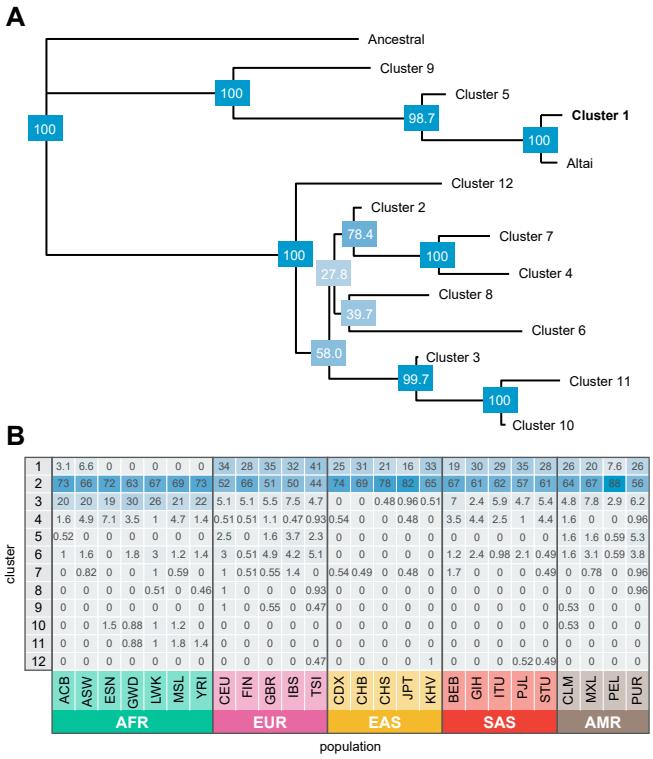
730 Martínez, I., Gracia, A., de Castro, J.M.B., Carbonell, E., Viola, B., Kelso, J.,
731 Prüfer, K., Pääbo, S., 2016. Nuclear DNA sequences from the Middle
732 Pleistocene Sima de los Huesos hominins. Nature.
733 Meyer, M., Kircher, M., Gansauge, M.T., Li, H., Racimo, F., Mallick, S.,
734 Schraiber, J.G., Jay, F., Prufer, K., de Filippo, C., 2012. A high-coverage
735 genome sequence from an archaic denisovan individual. Science. 338, 222–
736 226.
737 Nauciel, C., Espinasse-Maes, F., 1992. Role of Gamma Interferon and Tumor-
738 Necrosis-Factor-Alpha in Resistance to Salmonella-Typhimurium Infection.
739 Infection and Immunity. 60, 450–454.
740 Player, M.R., Torrence, P.F., 1998. The 2–5 A system: Modulation of viral and
741 cellular processes through acceleration of RNA degradation. Pharmacology &
742 Therapeutics. 78, 55–113.
743 Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S.,
744 Heinze, A., Renaud, G., Sudmant, P.H., de Filippo, C., Li, H., Mallick, S.,
745 Dannemann, M., Fu, Q., Kircher, M., Kuhlwilm, M., Lachmann, M., Meyer,
746 M., Ongyerth, M., Siebauer, M., Theunert, C., Tandon, A., Moorjani, P.,
747 Pickrell, J., Mullikin, J.C., Vohr, S.H., Green, R.E., Hellmann, I., Johnson,
748 P.L.F., Blanche, H., Cann, H., Kitzman, J.O., Shendure, J., Eichler, E.E., Lein,
749 E.S., Bakken, T.E., Golovanova, L.V., Doronichev, V.B., Shunkov, M.V.,
750 Derevianko, A.P., Viola, B., Slatkin, M., Reich, D., Kelso, J., Pääbo, S., 2014.
751 The complete genome sequence of a Neanderthal from the Altai Mountains.
752 Nature. 505, 43–49.
753 Racimo, F., Marnetto, D., Huerta-Sanchez, E., 2016. The landscape of uniquely
754 shared archaic alleles in present-day human populations. bioRxiv. 045237.
755 Racimo, F., Sankararaman, S., Nielsen, R., Huerta-Sanchez, E., 2015. Evidence
756 for archaic adaptive introgression in humans. Nature Reviews Genetics.
757 Reich, D., Green, R.E., Kircher, M., Krause, J., Patterson, N., Durand, E.Y., Viola,
758 B., Briggs, A.W., Stenzel, U., Johnson, P.L.F., 2010. Genetic history of an
759 archaic hominin group from Denisova Cave in Siberia. Nature. 468, 1053–
760 1060.
761 Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z.P., Richter, D.J., Schaffner,
762 S.F., Gabriel, S.B., Platko, J.V., Patterson, N.J., McDonald, G.J., 2002.
763 Detecting recent positive selection in the human genome from haplotype
764 structure. Nature. 419, 832–837.
765 Sankararaman, S., Mallick, S., Dannemann, M., Prüfer, K., Kelso, J., Pääbo, S.,
766 Patterson, N., Reich, D., 2014. The genomic landscape of Neanderthal
767 ancestry in present-day humans. Nature. 507, 354–357.
768 Sankararaman, S., Mallick, S., Patterson, N., Reich, D., 2016. The Combined
769 Landscape of Denisovan and Neanderthal Ancestry in Present-Day Humans.
770 Current Biology.
771 Sawyer, S., Renaud, G., Viola, B., Hublin, J.-J., Gansauge, M.-T., Shunkov, M.V.,
772 Derevianko, A.P., Prüfer, K., Kelso, J., Pääbo, S., 2015. Nuclear and
773 mitochondrial DNA sequences from two Denisovan individuals. Proceedings
774 of the National Academy of Sciences. 201519905.
775 Schlamp, F., Made, J., Stambler, R., Chesebrough, L., Boyko, A.R., Messer, P.W.,

bioRxiv preprint doi: https://doi.org/10.1101/051466; this version posted May 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

776 2016. Evaluating the performance of selection scans to detect selective sweeps
777 in domestic dogs. Molecular Ecology. 25, 342–356.
778 Ségurel, L., Quintana-Murci, L., 2014. Preserving immune diversity through
779 ancient inheritance and admixture. Current opinion in immunology. 30C, 79–
780 84.
781 Tennessen, J.A., Bigham, A.W., O'Connor, T.D., Fu, W., Kenny, E.E., Gravel, S.,
782 McGee, S., Do, R., Liu, X., Jun, G., Kang, H.M., Jordan, D., Leal, S.M.,
783 Gabriel, S., Rieder, M.J., Abecasis, G., Altshuler, D., Nickerson, D.A.,
784 Boerwinkle, E., Sunyaev, S., Bustamante, C.D., Bamshad, M.J., Akey, J.M.,
785 Broad GO, Seattle GO, on behalf of the NHLBI Exome Sequencing Project,
786 2012. Evolution and Functional Impact of Rare Coding Variation from Deep
787 Sequencing of Human Exomes. Science. 337, 64–69.
788 Vernot, B., Akey, J.M., 2014. Resurrecting surviving Neandertal lineages from
789 modern human genomes. Science. 343, 1021.
790 Vernot, B., Akey, J.M., 2015. Complex History of Admixture between Modern
791 Humans and Neandertals. The American Journal of Human Genetics.
792 Voight, B.F., Kudaravalli, S., Wen, X., Pritchard, J.K., 2006. A map of recent
793 positive selection in the human genome. PLoS Biology. 4, e72.
794 Yang, M.A., Malaspinas, A.S., Durand, E.Y., Slatkin, M., 2012. Ancient structure
795 in Africa unlikely to explain Neanderthal and non-African genetic similarity.
796 Molecular Biology and Evolution. 29, 2987–2995.
797
798
799
800
801
802
803

bioRxiv preprint doi: https://doi.org/10.1101/051466; this version posted May 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

804



805
806

807 Figure 1: Neandertal introgresssed haplotypes in the OAS region. (A) Neighbor-
808 joining tree of 5008 phased haplotypes spanning chr12: 113344739-113449528
809 (hg19) from phase 3 of the 1000 genomes project. Haplotypes condensed into 12
810 core haplotypes based on majority allele in clusters with pairwise differences of 85
811 or less. The figure illustrates that the Altai haplotype is very similar to “cluster 1”
812 haplotypes found in several human populations. Bootstrap values (1000 replicates)
813 are provided in blue boxes at each node. (B) Frequencies of the 12 core haplotypes
814 within each 1000 genomes project population sample. The most common
815 Neandertal-like haplotype, cluster 1, is found only outside of sub-Saharan African
816 samples, with the exception of recently admixed populations. Population codes
817 can be found at http://www.1000genomes.org/category/population/.
818
819
820
821

bioRxiv preprint doi: https://doi.org/10.1101/051466; this version posted May 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

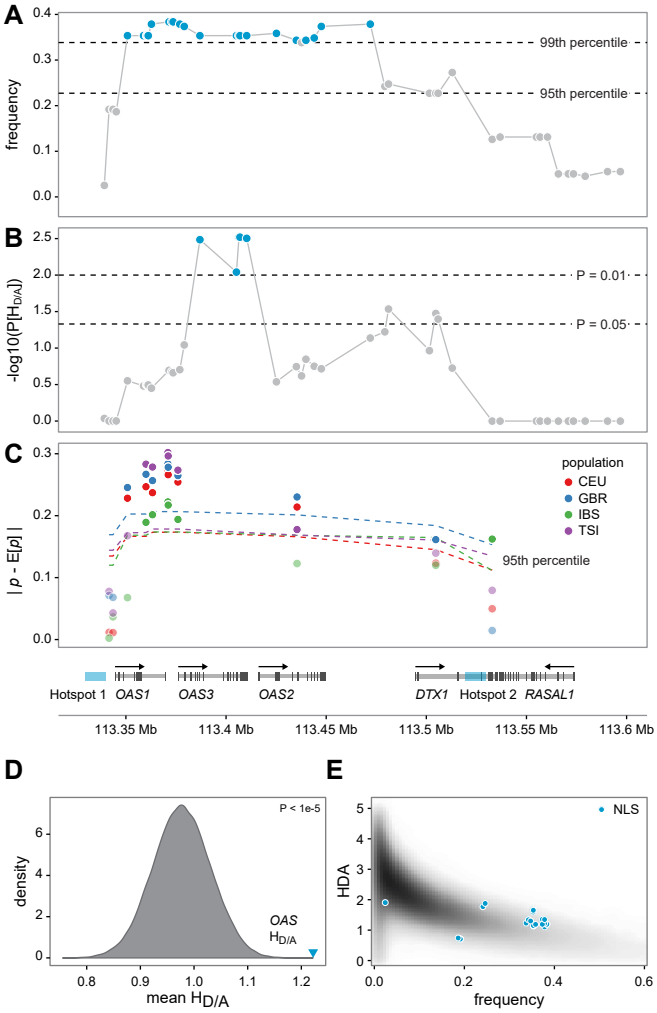


Figure 2: OAS-introgressed haplotypes show multiple signatures of positive selection. (A) Comparison of frequency (y-axis) of NLS in the OAS locus in CEU sample with respect to neutral expectations (dashed lines) based on coalescent simulations. (B) The $H_{D/A}$ values associated with several NLS are significantly larger (-log10 Ps in the y-axis) than expected under a neutral model of evolution. $H_{D/A}$ at each NLS in the OAS region were compared against the distributions of simulated NLS drawn to match local recombination rate and allele frequency. (C) Absolute difference between observed and expected allele frequency in all four present-day European samples (y-axis) based on ancient DNA data. Dashed lines represent the 95th percentile of the expected distribution based on similar deviations calculated on a dataset of approximately one million SNPs scattered around the genome and with comparable present-day frequencies to those found for NLS in the OAS region (D) Comparison of mean $H_{D/A}$ across OAS locus to randomized frequency and genomic position-matched SNPs from simulations. OAS locus is a strong outlier ($p < 10^{-5}$) compared to simulations. (E) Joint density distribution of $H_{D/A}$ and Neandertal-like site derived frequency. Cluster of NLS in OAS are significant outliers to neutral distribution ($p < 5 \times 10^{-4}$).

bioRxiv preprint doi: https://doi.org/10.1101/051466; this version posted May 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

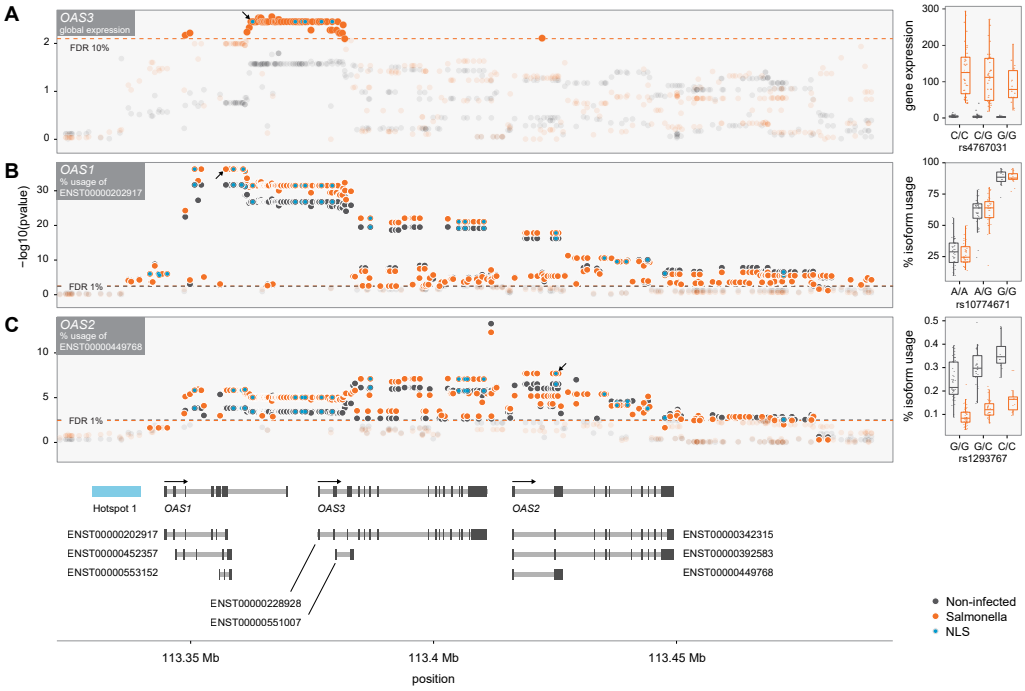


Figure 3: Pervasive impact of the Neandertal haplotype on the regulation of *OAS* genes in primary macrophages. (A) –log 10 Ps (y-axis) for the association between genotypes for SNPs with a MAF > 10% in the OAS region and expression levels of *OAS3* in non-infected (black) and Salmonella-infected macrophages (orange). The dashed line shows the P cutoff corresponding to an FDR of 10%. The right panel shows a boxplot for the association between genotypes at the NLS rs4767031 (x-axis) and the expression levels of *OAS3* (y-axis)(B) –log 10 Ps (y-axis) for the association between genotypes in the OAS regions and the percentage usage of isoform ENST00000202917 (i.e., p46 in the text) in non-infected (black) and Salmonella-infected macrophages (orange). The dashed line shows the P cutoff corresponding to an FDR of 1%. The right panel shows a boxplot for the association between genotypes at the splicing variant rs10774671 (x-axis) and the percentage usage of isoform p46 (y-axis). (C) Similar to (B) but for the percentage usage of isoform ENST00000449768 of OAS2. In all the panels NLS are highlighted by blue dots. The arrows on panel A-C highlight the location of the SNPs for which the boxplots are shown on the right.

bioRxiv preprint doi: https://doi.org/10.1101/051466; this version posted May 3, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

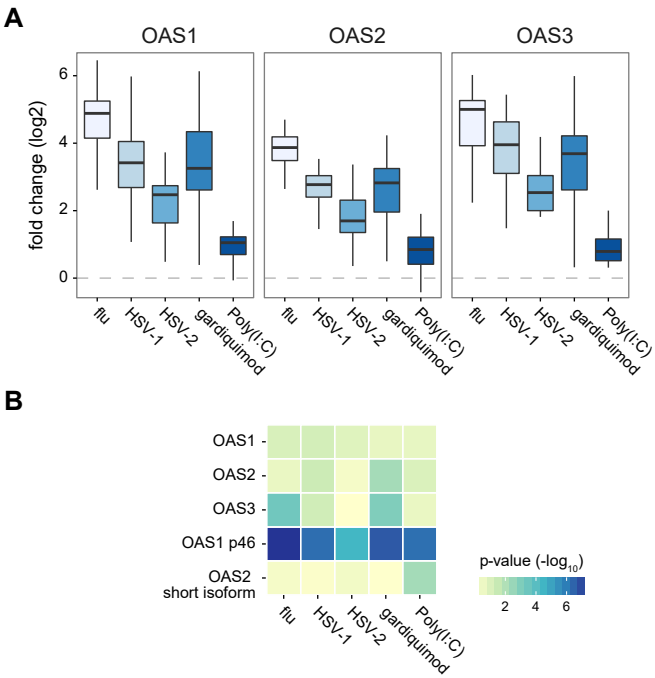
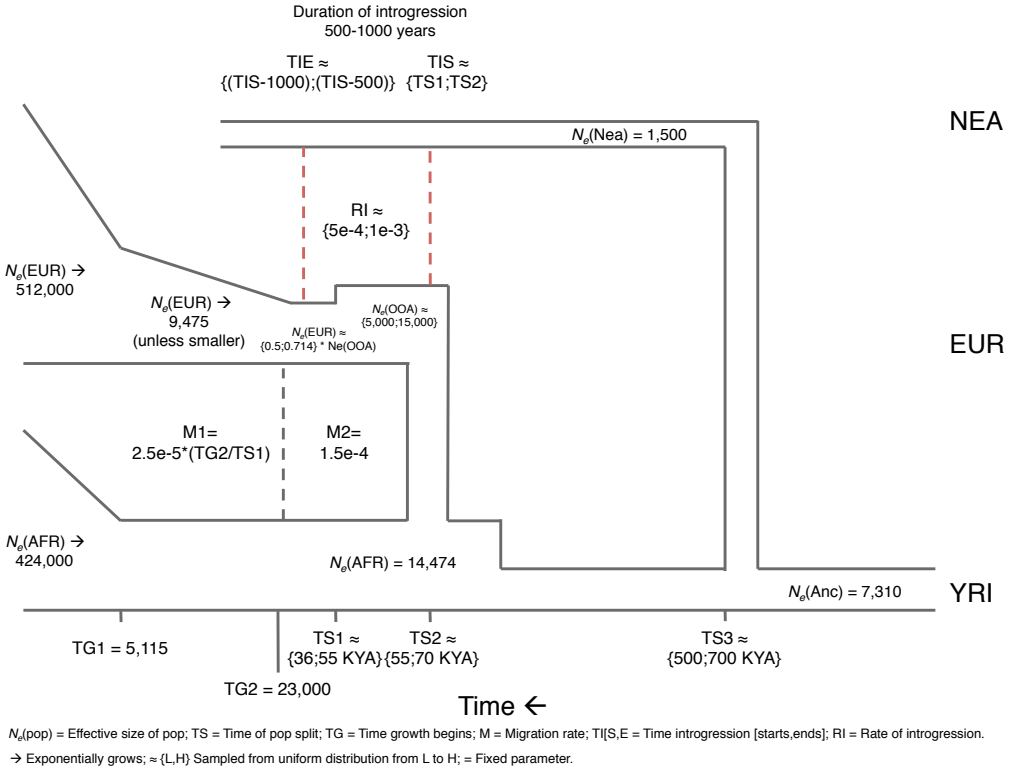
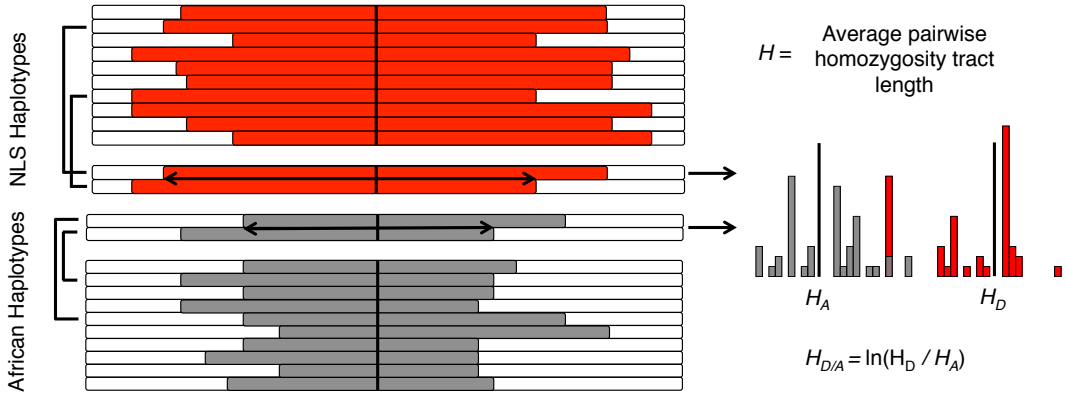


Figure 4: The Neandertal haplotype in the OAS regions has a different impact on the regulation of OAS genes depending on the viral agents PBMCs are exposed to. (A) Log 2 fold induction (y-axis) of *OAS1*, *OAS2* and *OAS3* in response to different viral agents or viral-associated immune stimuli (B) –log10 P for the association between genotype status for the Neandertal haplotype and overall expression levels of OAS genes and the expression of specific isoforms of *OAS1* and *OAS2* (those identified in Figure 3 as associated with NLS).

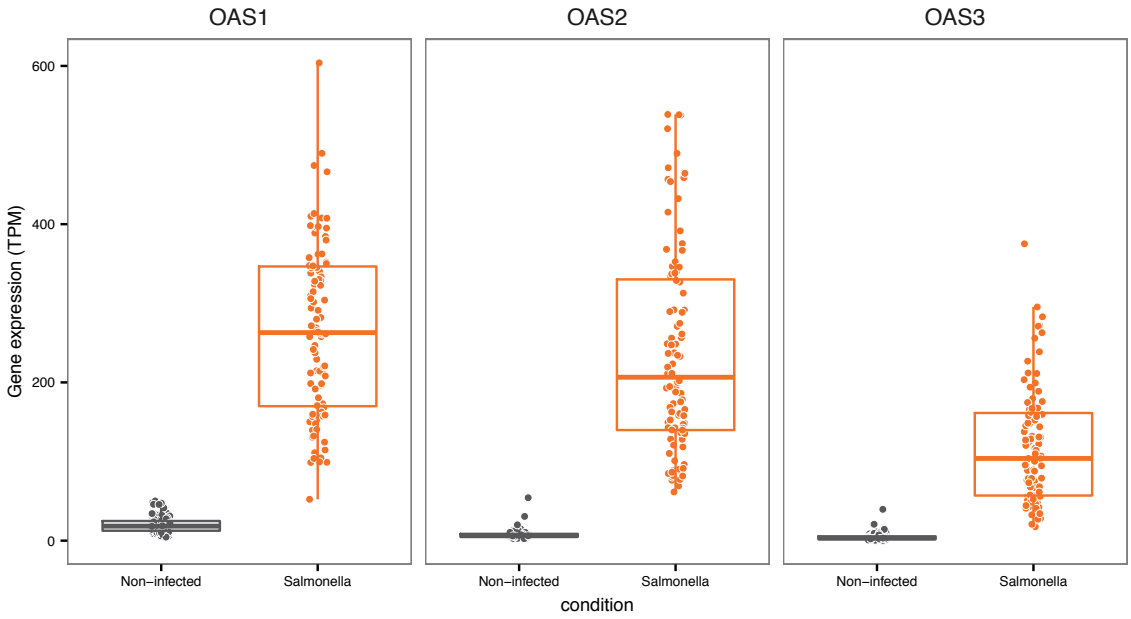
Supplementary Figures:



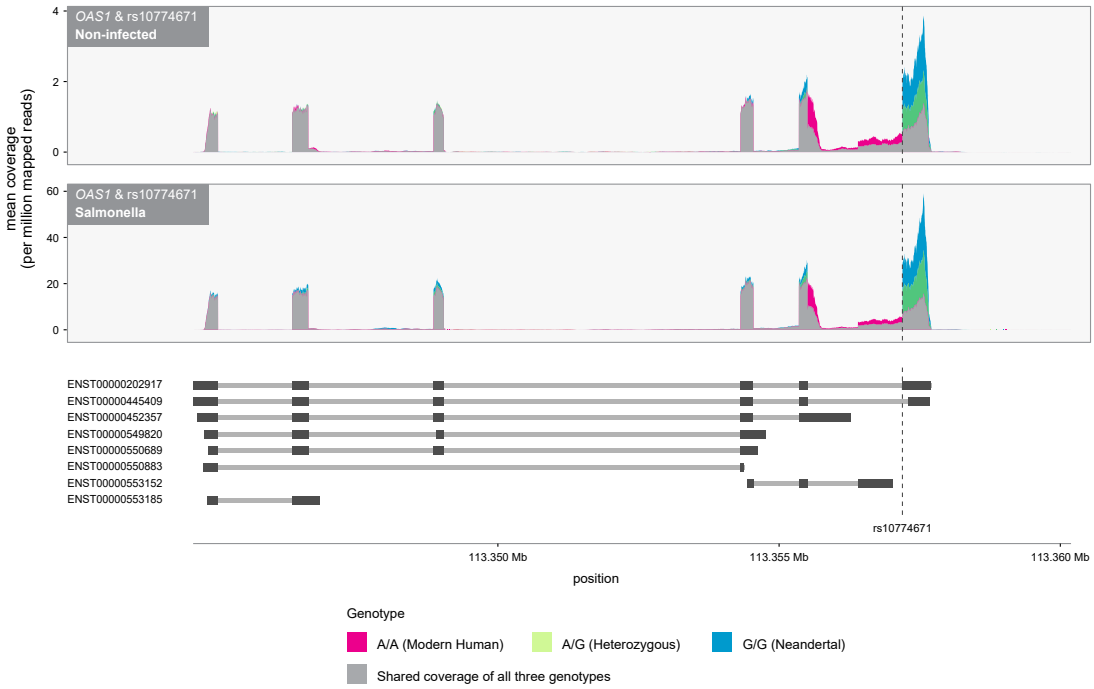
Supplementary Figure 1: Demographic model for neutral coalescent simulations.



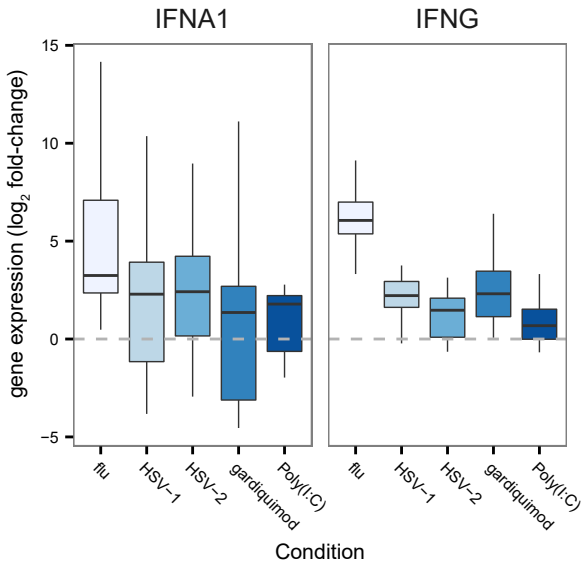
Supplementary Figure 2: Illustration of $H_{D/A}$ statistic.



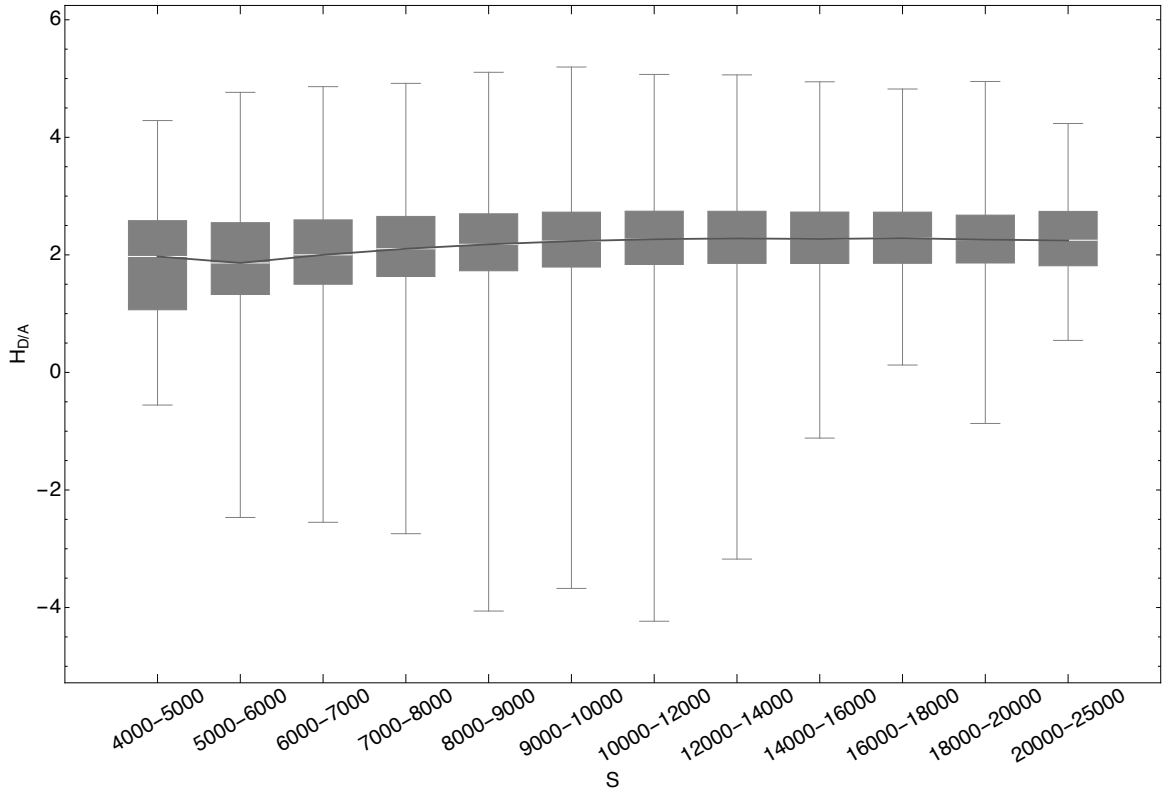
Supplementary Figure 3: Expression levels of OAS genes in primary macrophages before and after infection with *Salmonella*.



Supplementary Figure 4: The splicing variant rs10774671 is a strong asQTL for *OAS1*. Plotted is the normalized average coverage at which each base was sequenced along the genomic regions encoding the gene *OAS1*. Individuals were stratified according to their genotype at rs10774671. Below the figure are gene models from Ensembl database. Individuals carrying the G allele at rs10444671 (i.e., the Neandertal allele) primarily express the transcript ENST00000202917 (referred to as p46 in the text) whereas individuals carrying the A derived allele loose the splice site, which leads to the usage of a distinct isoform.



Supplementary Figure 5: Log 2 fold induction (y-axis) of *IFNA1* (type I interferon) and *IFNG* (type II interferon) in PBMCs upon infection/stimulation with several viral agents.



Supplementary Figure 6: $H_{D/A}$ does not vary substantially with SNP density. $H_{D/A}$ statistic plotted against number of segregating sites (S) per locus across all neutral coalescent simulations. The slight increase in $H_{D/A}$ with elevated S is conservative with respect to our one-tailed test for elevated $H_{D/A}$.

Supplementary Tables:

	CEU	GBR	IBS	TSI	Mathieson et al. 2015 P
113341598	0.856	0.407	0.967	0.256	0.233
113343563	0.862	0.428	0.529	0.524	0.335
113350796	0.0123	0.0194	0.395	0.0557	0.0249
113360025	0.00850	0.0114	0.0317	0.00633	0.000277
113363284	0.0126	0.0170	0.0288	0.00732	0.0000288
113370966	0.00654	0.00902	0.0190	0.00476	2.29E-08
113371114	0.00735	0.0102	0.0209	0.00518	0.000000315
113375983	0.00947	0.0140	0.0333	0.00787	0.00459
113435450	0.0167	0.0284	0.132	0.0409	0.0000375
113504725	0.0852	0.0797	0.134	0.0797	0.157
113532885	0.354	0.853	0.0112	0.218	0.0236
p <= 0.01					
0.01<p<=0.05					

Supplementary Table 1: Proportion of similar frequency genome-wide SNPs with deviations from expected frequency as great or greater than 11 Neandertal-like sites. Fraction of all non-OAS SNPs for which absolute difference in expected (based on aDNA samples) and observed allele frequency that are as great or greater than NLS derived allele frequency difference. Final column reports P from Mathieson et al. 2015 selection test based on same aDNA samples.