An evolutionary medicine perspective on Neandertal extinction

Alexis P. Sullivan[1], Marc de Manuel[3], Tomas Marques-Bonet[3,4,5], & George H. Perry[1,2]

Departments of [1]Biology and [2]Anthropology, Pennsylvania State University, University Park, PA 16802, USA

[3]Institut de Biologia Evolutiva (CSIC/UPF), Parque de Investigación Biomédica de Barcelona (PRBB), Barcelona, Catalonia 08003, Spain

[4]CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Baldiri i Reixac 4, 08028 Barcelona, Spain

[5]Catalan Institution of Research and Advanced Studies (ICREA), Passeig de Lluís Companys, 23, 08010, Barcelona, Spain

Corresponding Author: George H. Perry
E-mail: ghp3@psu.edu

Abstract
The Eurasian sympatry of Neandertals and anatomically modern humans – beginning at least 45,000 years ago and lasting for more than 5,000 years – has long sparked anthropological interest into the factors that potentially contributed to Neandertal extinction. Among many different hypotheses, the "differential pathogen resistance" extinction model posits that Neandertals were disproportionately affected by exposure to novel infectious diseases that were transmitted during the period of spatiotemporal sympatry with modern humans. Comparisons of new archaic hominin paleogenome sequences with modern human genomes have confirmed a history of genetic admixture – and thus direct contact – between humans and Neandertals. Analyses of these data have also shown that Neandertal nuclear genome genetic diversity was likely considerably lower than that of the Eurasian anatomically modern humans with whom they came into contact, perhaps leaving Neandertal innate immune systems relatively more susceptible to novel pathogens. In this study, we compared levels of genetic diversity in genes for which genetic variation is hypothesized to benefit pathogen defense among Neandertals and African, European, and Asian modern humans, using available exome sequencing data (six chromosomes per population). We observed that Neandertals had only 31-39% as many nonsynonymous (amino acid changing) polymorphisms across 73 innate immune system genes compared to modern human populations. We also found that Neandertal genetic diversity was relatively low in an unbiased set of balancing selection candidate genes for primates – genes with the highest 1% genetic diversity genome-wide in non-human apes. In contrast, Neandertals had similar to higher levels of genetic diversity than humans in 13 major histocompatibility complex (*MHC*) genes. Thus, while Neandertals may have been relatively more susceptible to some novel pathogens and differential pathogen resistance could be considered as one potential contributing factor in their extinction, this model does have limitations.

Introduction

Following their divergence from the modern human lineage ~550 kya[1–3], Neandertals (*Homo neanderthalensis*) inhabited Eurasia from at least 300 kya[4–6] to ~41-39 kya[7]. In contrast, anatomically modern humans very likely evolved in sub-Saharan Africa, where the earliest recognized skeletal remains date to 155 kya[8,9]. Modern humans then expanded out of Africa and spread from the Middle East into Europe by at least 45 kya[10–12] and Asia by 80 kya[13,14]. Within 6 kya after modern human appearance in Europe, Neandertals no longer appear in the fossil record[15].

There are numerous hypotheses regarding Neandertal extinction. For example, some scholars have suggested that climate fluctuations between 40-30 kya played a key role in this process[16–18]. However, this hypothesis has been discounted on the basis of the simultaneous Eurasian presence of anatomically modern humans for most of the 40-30 kya period combined with the *in situ* established environmental resilience of Neandertals[19]. That is, if anything, modern humans would more likely have been adversely affected by this environment than Neandertals, given the survival and likely adaptation of the latter to the harsh Eurasian climate for hundreds of thousands of years[20].

Most other extinction hypotheses focus on Neandertal-modern human competition[21–25]. For example, potentially shorter inter-birth intervals for modern humans could have allowed more rapid population growth compared to Neandertals, facilitating eventual replacement[26,27]. Alternatively, anthropologists have speculated that the intelligence and language capabilities of anatomically modern humans were greater than those of Neandertals[28–31], perhaps facilitating competitive hunting and other advantages such as through the development of more efficient tool technologies[32]. A recent proposal is that modern humans benefitted from the early domestication of dogs, who may have aided large animal hunts to increase caloric yields for the modern humans and fuel their rapid population growth and ultimately larger sizes[33].

Thinking of the important role of disease in population dynamics, Wolff and Greenwood[34] suggested that viral disease transmission from modern humans to Neandertals could have contributed to the ultimate disappearance of the latter. Houldcroft and Underdown[35] recently echoed and expanded on these notions. This “differential pathogen resistance” model would require an anatomically modern human pathogen (or pathogens) of limited virulence for out-of-Africa migrating human populations, but that would have strongly affected immunologically naïve Neandertal populations upon contact and transmission. Moreover, the viability of this scenario also requires i) relatively fewer Neandertal pathogens/ strains at the time of contact, ii) some mechanism by which modern humans might not have been as negatively affected as Neandertals upon novel pathogen exposure, or iii) both of these factors to be present. In this paper we assess the plausibility of the differential pathogen resistance model by comparing levels of genetic diversity in genes for which genetic variation is hypothesized to benefit pathogen defense between Neandertal and modern human populations.

Recent advances in genomic sequencing technologies and ancient DNA methods have facilitated the generation of a high-quality Altai Neandertal nuclear genome sequence from Siberia (dated to ~50,000 ya)[36]. When analyzed in combination with modern human genomic data, this genome has provided convincing evidence that anatomically modern humans and Neandertals interbred[3,36–39]. The requisite intercourse demonstrates at least some level of direct contact between these populations, and thus opportunities for the transfer of infectious diseases.

Moreover, analyses of both the high-coverage diploid nuclear genome sequence from the Altai Neandertal and mitochondrial DNA sequence data that are available for multiple Neandertal individuals suggest that Neandertal genetic diversity was substantially lower than that observed within modern human populations[3,36,40,41]. Recently, Castellano et al.[42] used a DNA capture method to sequence the nuclear genome's protein-coding regions – the 'exome' – from each of two additional Neandertals: individuals who lived ~49,000 ya[43] and ~44,000 ya in Spain and Croatia, respectively. Observed levels of heterozygosity for these two Neandertals are also relatively low, suggesting that low nuclear genome genetic diversity was a general Neandertal characteristic rather than restricted to an Altai Neandertal population isolate[42]. Specifically, considering only sites with sequence coverage sufficient for single nucleotide polymorphism (SNP) identification for each of the three Neandertals, Castellano et al.[42] observed only 30.3%, 44.9%, and 45.3% as many synonymous (i.e., not amino acid-changing) SNPs in the Neandertals compared to equal-sized population samples of modern human Africans, Europeans, and Asians, respectively.

Within genes directly related to immune function, greater functional genetic diversity increases the potential responsiveness of the immune system to foreign pathogens[34,44]. Balancing selection is thought to maintain advantageous functional diversity (i.e., nonsynonymous SNPs) within these genes[45,46], and individuals with more genetic diversity across the genome tend to have higher fitness[44]. Based on population genetic theory, genetic drift is a relatively stronger force, while natural selection is relatively less effective, in smaller populations[47,48]. Thus, compared to a larger population, a population with a historically small effective population size may have lower genetic diversity in general across the genome *and* different patterns of diversity at loci affecting individual health and fitness. Indeed, along with relatively reduced overall genetic diversity, Castellano et al.[42] observed a higher proportion of predicted damaging than benign nonsynonymous (amino acid-changing) SNPs in Neandertals compared to modern humans, consistent with the reduced effectiveness of purifying selection to remove or reduce the frequencies of strongly deleterious variants in Neandertals[49–53].

In addition to purifying selection, other types of natural selection, including balancing selection, are likewise expected to be less effective in smaller populations. Thus, the generally low genetic diversity of Neandertals relative to humans may even be exacerbated at functional sites in genes related to immune function that would otherwise be preserved via balancing selection. Theoretically, such a difference could have facilitated the differential morbidity following contact and infectious disease transfer between Neandertals and modern humans potentially required under the Pleistocene epidemiological scenarios (the differential pathogen resistance model) of Wolff and Greenwood[34] and Houldcroft and Underdown[35].

In this study we specifically compared the levels and patterns of genetic variation between Neandertal and modern human populations at i) 73 genes associated with innate immune functions, ii) 164 virus-interacting protein genes, iii) 13 major histocompatibility complex (*MHC*) genes, and iv) the 1% of genes across the genome with the consistently highest levels of genetic diversity among four non-human ape species. Relatively lower genetic variation at these loci in Neandertal populations would be consistent with the differential pathogen resistance hypothesis that a greater susceptibility to novel pathogens relative to the modern humans with which they interacted could have been one contributing factor in the extinction of Neandertals. In contrast, similar levels of Neandertal and modern human genetic diversity would raise major questions about the plausibility of this epidemiological extinction hypothesis.

Results

Castellano et al.[42] produced a database of Neandertal and African, European, and Asian human exonic SNP genotypes based on the paleogenomic data for the three Neandertal individuals described above and high-coverage genome sequencing data for three human individuals from each of three continental regions. Ancestral and derived alleles were determined via comparisons to the gorilla and orangutan genome sequences[42]. Genotypes were estimated for only those positions covered by a minimum of six independent sequencing reads; the authors also tested higher minimum coverage cutoffs (up to 20x) that did not significantly affect the observed patterns of Neandertal versus modern human genetic diversity, so the lower coverage cutoff was used to include a larger number of sites in the analysis[42]. For our analysis, we considered only the autosomal sites with sufficient coverage for SNP genotyping among all individuals, so that the numbers of identified SNPs per population and their allele frequency distributions could be compared directly among the Neandertal and modern human African, European, and Asian samples (n = 6 chromosomes per population) as measures of genetic diversity.

Patterns of genetic diversity and the effectiveness of selection

Following Castellano et al.[42], we used the program PolyPhen-2[54] to infer whether each nucleotide variant was nonsynonymous (amino acid changing; potentially functional) or synonymous (resulting in no change to the amino acid sequence; typically neutral with respect to fitness), and for the nonsynonymous mutations a "damaging" or "benign" prediction of their effects on protein structure and function. On the overall dataset, we observed patterns of genetic diversity that were similar to the primary findings of Castellano et al.[42]. Specifically, Neandertals had relatively fewer total SNPs than any of the modern human populations; this was true for both nonsynonymous and synonymous SNPs (**Figure 1; Supplemental Table 1**). In addition, Neandertals also had a higher proportion of nonsynonymous to total (nonsynonymous + synonymous) SNPs compared to the human populations (Neandertals = 0.500; human populations = 0.440-0.452; Fisher's Exact Test for the Neandertal vs. European comparison; $P = 2.29\text{x}10^{-10}$; see **Supplemental Table 1** for all comparisons) and a considerably higher proportion of predicted damaging to total nonsynonymous SNPs (Neandertal = 0.435; human = 0.291-0.297; $P < 2.2\text{x}10^{-16}$ for the Neandertal-European comparison; **Supplemental Table 2**). Together, these results are consistent with the notion that purifying natural selection against potentially damaging nonsynonymous SNPs was relatively less effective in smaller Neandertal populations.

Innate immune system, virus-interacting protein, and MHC gene diversity comparisons

To assess the plausibility of the differential pathogen resistance model, we next performed analyses focused on patterns of genetic diversity within two subsets of genes for which genetic diversity itself is thought to play an important role in pathogen defense-related immune functions. The first set was comprised of 73 innate immune receptor, signaling adaptor molecule, and complement pathway genes ("Innate Immune System Genes"; **Supplemental Table 3**), for example the toll-like receptor[55–57] and mannose-associated serine protease genes[58]. Many of the proteins encoded by these genes are involved in the innate immune system's first line of defense against diverse external microorganisms. The second set was comprised of 164 genes encoding for virus-interacting proteins that also have known antiviral activity or broader roles in the immune system[59] (**Supplemental Table 4**). The second set was comprised of 13 *MHC* genes (**Supplemental Table 5**), critical immune system loci with among the strongest evidence for long-term balancing selection and long-term maintenance of allelic diversity in vertebrate genomes[60–63]. The patterns of Neandertal and modern human genetic diversity for these two gene sets were compared to those for the remaining 11,086 genes in our genome-wide dataset.

Neandertal and human nonsynonymous genetic diversity was similar between the genome-wide and innate immune system gene sets, with fewer nonsynonymous SNPs in Neandertals compared to any human population in both cases (**Figure 2A-D**). In fact, the relative number of Neandertal vs. human nonsynonymous SNPs is even slightly higher for the genome-wide set than for the innate immune system genes; e.g., with 54.1% as many Neandertal as European human nonsyonymous SNPs in the genome-wide set compared to only 39.0% for the innate immune genes. The direction of this result is consistent with expectations under a model of relatively reduced balancing selection effectiveness in Neandertals (i.e., if innate immune gene functional genetic diversity confers a fitness advantage). However, the observed Neandertal-human difference between the two gene sets is not significant based on Fisher's Exact Tests (P = 0.33 for Neandertals-Europeans; see **Supplemental Table 6** for all comparisons). Based on the results of a permutation analysis with 10,000 sets of 73 random genes from the genome-wide set, the proportion of the number of Neandertal-human nonsynonymous SNPs is also not significantly lower than expected by chance (P = 0.1944 for Neandertals-Europeans; see **Supplemental Figure 1** for all comparisons). Regardless, our results do not provide any support for the possibility that strong balancing selection has maintained similar levels of nonsynonymous diversity in Neandertal and human innate immune system genes despite the lower effective population size and genome-wide diversity of Neandertals.

The relative number of Neandertal vs. human nonsynonymous SNPs observed among the virus-interacting protein genes was also low (**Figure 2E-F**), in this case at similar proportions to the genome-wide gene set (**Supplemental Table 7**; **Supplemental Figure 2**).

In contrast, there was a greater number of total nonsynonymous SNPs observed across the 13 *MHC* genes for Neandertals (17) than for any of the three human populations (9, 13, and 9 for the African, European, and Asian population samples, respectively; **Figure 2G**). Compared to both the genome-wide set of genes and the innate immune system genes, the number of Neandertal relative to human *MHC* gene nonsynonymous SNPs is significantly greater than expected by chance when assessed with Fisher's Exact Tests (e.g., Neandertals-Europeans for *MHC* vs. genome-wide set; P = 0.02; see **Supplemental Table 8** for all comparisons) and, for some but not all population comparisons, with permutation analyses (P = 0.1556 for Neandertals-Europeans; P = 0.0759 for Neandertals-Asians; P = 0.0222 for Neandertals-Africans; **Supplemental Figure 3**).

Additionally, we observed 9/17 (52.9%) of the Neandertal *MHC* nonsynonymous SNPs at intermediate frequency (i.e., with the minor allele observed on 3 out of the 6 chromosomes in the population; **Figure 2H**), a significantly higher proportion than observed for Neandertal nonsynonymous SNPs in the genome-wide, non-immune gene set (260/2527; 10.3%; Fisher's Exact Test; $P = 1.56 \times 10^{-5}$; **Supplemental Table 9**). The proportion of Neandertal *MHC* intermediate frequency nonsynonymous SNPs is also relatively higher for Neandertals than for any of the human populations, although with the small sample sizes not all comparisons were statistically significant (Fisher's Exact Tests; P = 0.43 for Neandertals-Africans; P = 0.02 for Neandertals-Europeans; P = 0.09 for Neandertals-Asians; **Supplemental Table 10**). The Neandertal-human *MHC* genetic diversity patterns remained similar after we excluded one random individual from each human population and the Altai Neandertal individual, whose genomic *MHC* region may harbor some DNA segments introgressed from humans[64] (**Supplemental Figure 4**).

In combination, the relatively large number of Neandertal *MHC* nonsynonymous SNPs and the high proportion of those variants observed at intermediate frequencies suggest that at least for

these critical immune loci, the heterozygous fitness advantage for Neandertals was sufficiently strong to offset the lower effective size of this population, leading to similar or even higher putatively functional diversity compared to modern human populations.

Balancing selection candidate gene comparisons

Finally, we sought to compare patterns of Neandertal-human genetic diversity across genes for which evidence of balancing selection has been identified without respect to gene function. However, were we to analyze candidate genes detected using human population genomic data, we would introduce bias into our Neandertal vs. human comparison; thus we cannot use gene lists from the majority of genome-wide analyses of balancing selection in mammals that have been published to date[46,60,63,65–68].

Instead, we analyzed genome-wide sequence data from 55 total individuals from population samples of four non-human ape species, chimpanzee (*Pan troglodytes ellioti*), bonobo (*Pan paniscus*), gorilla (*Gorilla gorilla gorilla*), and orangutan (*Pongo abelii*)[69], to identify an unbiased (with respect to our analysis) set of candidate balancing selection genes. Specifically, we estimated per-species nucleotide sequence diversity as the average proportion of pairwise differences (π) for the coding regions of each gene. There were 7,259 genome-wide genes with ≥ 500 "callable" sites across all four species (see *Methods*) for which there was also at least one variable site in our Neandertal-modern human genetic diversity database. Within each species, we computed π value percentiles for each gene, and then summed the per-gene percentile values across species. Using this approach, we identified the 73 (top 1%) genes with the consistently highest genetic diversity among these non-human ape species ("top 1% ape diversity genes"; **Supplemental Table 11**).

Relative to all 7,259 genes and following correction for multiple tests, the top 1% ape diversity set was significantly enriched for genes with immune system-related Gene Ontology functional categories including "MHC protein complex" (observed = 3 genes; expected = 0.09 genes; adjusted P = 0.0015), "positive regulation of leukocyte activation" (observed = 6 genes; expected = 0.62 genes; adjusted P = 0.0061), and "immune system process" (observed = 14 genes; expected = 4.92 genes; adjusted P = 0.0098). Full results from the Gene Ontology enrichment analysis are provided in **Supplemental Table 12** (see also **Supplemental Figure 5**). In addition to the *MHC* loci, the top 1% ape diversity set contains *OAS1*, another gene for which high allelic diversity appears likely to have been maintained across multiple species by long-term balancing selection[70]. Together, these features suggest that our top 1% ape diversity genes are likely at least enriched for those affected by balancing selection in hominoid primates. Thus, it is appropriate to compare Neandertal and modern human genetic diversity at these loci as an additional component of our assessment of the underlying mechanics of the differential pathogen resistance model for Neandertal extinction.

Across the top 1% ape diversity genes, we observed only 27%-58% as many nonsynonymous SNPs for Neandertals compared to the modern human populations (**Figure 3C**). The magnitude of this difference was similar to that for the remaining genome-wide genes (**Figure 3A**; Fisher's Exact Test; P = 0.69 for Neandertals-Europeans; see **Supplemental Table 13** for all comparisons). Thus, the pattern of Neandertal vs. modern human nonsynonymous genetic diversity for the top 1% ape diversity gene set was more similar to that observed for the innate immune system genes than the *MHC* genes.

Discussion

The differential pathogen resistance model of Neandertal extinction posits that after contact with anatomically modern humans who migrated Out of Africa, Neandertals may have been

disproportionately affected by transferred infectious diseases. Our goal in this study was to assess the underlying plausibility of this model by testing whether Neandertal genetic diversity at critical immune system loci was significantly lower than that for the modern human populations with whom they came into contact – a possibility, given lower levels of general, genome-wide Neandertal genetic diversity[42,71,72]. The model would *not* require Neandertals to be depauperate of distinct and evolutionarily advantageous genetic variants at immune system loci. Indeed, evidence of the adaptive introgression of Neandertal toll-like receptor and *MHC* gene variants into modern human populations strongly suggests otherwise[73–75]. Rather, the mechanics of the differential pathogen resistance model could still be relevant if there were simply *fewer* such variants and substantially *lower* functional genetic diversity overall at immune system loci in Neandertals compared to sympatric modern human populations.

Are our results consistent with the differential pathogen resistance model for Neandertal extinction? Not in full. While Neandertal nonsyonymous genetic diversity at innate immune system, virus-interacting protein, and non-human ape high-diversity gene loci was indeed low compared to that observed for humans, Neandertal *MHC* diversity was similar or even higher than that for humans, in multiple respects. Thus, future models that incorporate epidemiological mechanisms as contributing factors to Neandertal extinction should proceed with caution, as there are specific genes for which balancing selection in Neandertals appears to have overcome the lower effective size and lower levels of genome-wide genetic diversity in this population. That said, there still may have been an aggregate differential pathogen resistance effect from lower functional diversity across innate immune system and other critical pathogen defense genes. Looking forward, this hypothesis could be explored further with the combination of expanded Neandertal paleogenomic population data and broad experimental/ functional comparisons of Neandertal vs. modern human innate immune gene diversity.

Materials and Methods

We downloaded the Neandertal exome DNA capture data published by Castellano et al.[42] (http://cdna.eva.mpg.de/neandertal/exomes/VCF). Specifically, we considered the SNP genotype data for autosomal chromosomes from the "combined" VCF files from this dataset, in which SNP genotypes for each of the 13 individuals in the dataset (three modern humans of African descent, three modern humans of European descent, three modern humans of Asian descent, three Neandertal individuals, and one individual from the archaic hominin Denisovan population) were provided for only the individuals with a minimum of six independent sequencing reads at that position. There were a total of 69,230 autosomal SNPs in this dataset. We removed n = 32,416 SNPs for which genotypes were not estimated for all of the Neandertal and modern human individuals and n = 17 SNPs with more than 2 identified alleles (e.g., A/T/C variably present at one position).

The remaining 36,797 autosomal SNPs were submitted to PolyPhen-2's online HumDiv server via batch query for identification of nonsynonymous and synonymous SNPs and estimates of predicted functional consequences for the nonsynonymous SNPs[42,54]. PolyPhen-2 classified each SNP as "missense" (nonsynonymous), "coding-synon" (synonymous), "nonsense" (stop codon), "utr-5" (5' untranslated region), "utr-3" (3' untranslated region), "intron" (non-coding variants). Our subsequent analysis focused on the nonsynonymous (n = 16,139) and synonymous (n = 18,095) SNPs. For each nonsynonymous SNP, PolyPhen-2 also provided "benign" (n = 10,558), "possibly damaging" (n = 2,222), and "probably damaging" (n = 3,317) predictions[54]. PolyPhen-2 failed to make functional predictions for a small number of nonsynonymous SNPs (n = 42), which were removed from our subsequent analyses. SNPs in the predicted possibly and probably damaging categories were combined into one "predicted damaging" category for our analyses. PolyPhen-2 assembled gene names from the UCSC

knownGene transcripts/database (hg19/GRCh37), and these gene names were used for later identification. The annotated database of nonsynonymous and synonymous SNP genotypes and population frequencies that we analyzed for this study are available in **Supplemental Database 1** (https://scholarsphere.psu.edu/files/v405s943f). All analyses of the SNP genotype data were performed using the R statistical environment[76].

In addition, to compare patterns of synonymous and nonsynonymous SNP derived allele frequencies among populations (e.g., see **Figure 1**) we determined the derived and ancestral states for the alleles of each SNP by comparison to the orthologous chimpanzee, gorilla, and orangutan nucleotides based on alignments to reference genomes for those species as provided by Castellano et al.[42]. Ancestral alleles were classified as those that matched the orthologous nucleotides for all three non-human species. SNPs with variability, missing data, or nucleotides different than either of the two Neandertal/human alleles at the orthologous positions among the three non-human species (n = 538 nonsynonymous and n = 738 synonymous SNPs) were not considered in our analyses of derived allele frequencies. Note that our comparisons of the patterns of genetic diversity among gene categories (e.g., see **Figures 2 and 3**) did not require derived allele frequency information; thus, the SNPs with variability, missing data, or nucleotides different than the two Neandertal/human alleles were still included in that analysis to avoid bias against highly variable loci (e.g., those with cross-species polymorphisms maintained by long-term balancing selection).

To identify the set of genes with consistently high levels of genetic diversity among non-human apes, we analyzed whole-genome genetic diversity data that were previously published by Prado-Martinez et al.[69] for 87 great ape individuals. Briefly, in that study DNA from each individual was sequenced on an Illumina platform to ~25x sequence coverage. Reads were mapped to the human reference genome (hg18). Genotypes were estimated only for sites that met criteria for sequence coverage, base quality, and mapping quality ("callable sites"), as described[69]. For our genetic diversity analysis, we focused on a subset of these data (n = 55 individuals), considering the one population from each ape species or species group with the largest sample size: chimpanzee (*Pan troglodytes ellioti*; n = 10), bonobo (*Pan paniscus*; n = 13), gorilla (*Gorilla gorilla gorilla*; n = 27), and orangutan (*Pongo abelii*; n = 5). We obtained hg18 gene coding region coordinates from the knownCanonical gene database using the table browser at the UCSC Genome Bioinformatics Site (genome.ucsc.edu/index.html). For each gene also in the Neandertal-human genetic diversity database we estimated nucleotide diversity as the average proportion of pairwise differences (π) for the callable sites of the coding regions[77]. The great ape gene diversity data analyzed in this study are available in **Supplemental Database 2** (https://scholarsphere.psu.edu/files/5d86p0267). Gene Ontology enrichment analyses were performed using the WEB-based Gene SeT AnaLysis Toolkit (WebGestalt). Functional category enrichments were statistically evaluated with hypergeometric tests and p-values were adjusted for multiple tests using the method of Benjamini & Hochberg[78].

Acknowledgments
We thank Sergi Castellano and Luis Barreiro for their discussions on this project, and David Enard and Martin Kuhlwilm for helpful analytical suggestions. The computational resource instrumentation used in this study was funded by the National Science Foundation (OCI–0821527).

References

1. Krings, M. *et al.* Neandertal DNA sequences and the origin of modern humans. *Cell* **90,** 19–30 (1997).
2. Green, R. E. *et al.* A complete Neanderthal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* **134,** 416–426 (2008).
3. Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505,** 43–9 (2014).
4. Endicott, P., Ho, S. Y. W. & Stringer, C. Using genetic evidence to evaluate four palaeoanthropological hypotheses for the timing of Neanderthal and modern human origins. *J. Hum. Evol.* **59,** 87–95 (2010).
5. Bischoff, J. L. *et al.* High-resolution U-series dates from the Sima de los Huesos hominids yields 600- 66+ ∞ kyrs: implications for the evolution of the early Neanderthal lineage. *J. Archaeol. Sci.* **34,** 763–770 (2007).
6. Bischoff, J. L. *et al.* The Sima de los Huesos Hominids Date to Beyond U/Th Equilibrium (>350 kyr) and Perhaps to 400-500 kyr: New Radiometric Dates. *J. Archaeol. Sci.* **30,** 275–280 (2003).
7. Higham, T. *et al.* The timing and spatiotemporal patterning of Neanderthal disappearance. *Nature* **512,** 306–309 (2014).
8. Clark, J. D. *et al.* Stratigraphic, chronological and behavioural contexts of Pleistocene Homo sapiens from Middle Awash, Ethiopia. *Nature* **423,** 747–752 (2003).
9. White, T. D. *et al.* Pleistocene Homo sapiens from Middle Awash, Ethiopia. *Nature* **423,** 742–747 (2003).
10. Prat, S. *et al.* The oldest anatomically modern humans from far southeast europe: Direct dating, culture and behavior. *PLoS One* **6,** (2011).
11. Demeter, F. *et al.* Anatomically modern human in Southeast Asia (Laos) by 46 ka. *Proc. Natl. Acad. Sci.* **109,** 14375–14380 (2012).
12. Higham, T. *et al.* The earliest evidence for anatomically modern humans in northwestern Europe. *Nature* **479,** 521–524 (2011).
13. Reyes-Centeno, H. Out of Africa and into Asia: Fossil and genetic evidence on modern human origins and dispersals. *Quat. Int.* (2016). doi:10.1016/j.quaint.2015.11.063
14. Reyes-Centeno, H., Hubbe, M., Hanihara, T., Stringer, C. & Harvati, K. Testing modern human out-of-Africa dispersal models and implications for modern human origins. *J. Hum. Evol.* **87,** 95–106 (2015).
15. Adler, D. S. *et al.* Dating the demise: Neandertal extinction and the establishment of modern humans in the southern Caucasus. *J. Hum. Evol.* **55,** 817–833 (2008).
16. Tzedakis, P. C., Hughen, K. a, Cacho, I. & Harvati, K. Placing late Neanderthals in a climatic context. *Nature* **449,** 206–208 (2007).
17. Valet, J. P. & Valladas, H. The Laschamp-Mono lake geomagnetic events and the extinction of Neanderthal: A causal link or a coincidence? *Quat. Sci. Rev.* **29,** 3887–3893 (2010).
18. Golovanova, L. V. *et al.* Significance of Ecological Factors in the Middle to Upper Paleolithic Transition. *Curr. Anthropol.* **51,** 655–691 (2010).
19. Lowe, J. *et al.* Volcanic ash layers illustrate the resilience of Neanderthals and early modern humans to natural hazards. **109,** 1–6 (2012).
20. Holliday, T. W. Postcranial evidence of cold adaptation in European Neandertals. *Am. J. Phys. Anthropol.* **104,** 245–258 (1997).
21. Raichlen, D. a., Armstrong, H. & Lieberman, D. E. Calcaneus length determines running economy: Implications for endurance running performance in modern humans and Neandertals. *J. Hum. Evol.* **60,** 299–308 (2011).
22. Sandgathe, D. M., Dibble, H. L., Goldberg, P., Mcpherron, S. P. & Hodgkins, J. On the Role of Fire in Neandertal Adaptations in Western Europe: Evidence from Pech de l'Azé

IV and Roc de Marsal, France. *PaleoAnthropology* 216–242 (2011). doi:10.4207/PA.2011.ART54
23. Banks, W. E. *et al.* Neanderthal extinction by competitive exclusion. *PLoS One* **3,** 1–8 (2008).
24. Horan, R. D., Bulte, E. & Shogren, J. F. How trade saved humanity from biological exclusion: An economic theory of Neanderthal extinction. *J. Econ. Behav. Organ.* **58,** 1–29 (2005).
25. Gilpin, W., Feldman, M. W. & Aoki, K. Ecocultural model predicts Neanderthal extinction through competition with modern humans. *Proc. Natl. Acad. Sci. USA,* (2015). doi:10.1073/pnas.1524861113
26. Trinkaus, E. Neanderthal Pubic Morphology and Gestation Length. *Curr. Anthropol.* **25,** 509–514 (1984).
27. Ponce de Leon, M. S. *et al.* Neanderthal brain size at birth provides insights into the evolution of human life history. *Proc. Natl. Acad. Sci.* **105,** 13764–13768 (2008).
28. Marwick, B. Pleistocene Exchange Networks as Evidence for the Evolution of Language. *Cambridge Archaeol. J.* **13,** 67–81 (2003).
29. Chase, P. G. & Dibble, H. L. Middle paleolithic symbolism: A review of current evidence and interpretations. *J. Anthropol. Archaeol.* **6,** 263–296 (1987).
30. Davidson, I. *et al.* The Archaeoloy of Perception: Traces of Depiction and Language. **30,** 125–155 (2013).
31. Maricic, T. *et al.* A recent evolutionary change affects a regulatory element in the human FOXP2 gene. *Mol. Biol. Evol.* **30,** 844–852 (2013).
32. Benazzi, S. *et al.* The makers of the Protoaurignacian and implications for Neandertal extinction. *Science (80-.).* **348,** 793 (2011).
33. Shipman, P. How do you kill 86 mammoths? Taphonomic investigations of mammoth megasites. *Quat. Int.* **359-360,** 1–9 (2014).
34. Wolff, H. & Greenwood, A. D. Did viral disease of humans wipe out the Neandertals? *Med. Hypotheses* **75,** 99–105 (2010).
35. Houldcroft, C. J. & Underdown, S. J. *Neanderthal Genomics Suggests a Pleistocene Time Frame for the First Epidemiologic Transition*. *bioRxiv* **480,** (2015).
36. Green, R. E. *et al.* A Draft Sequence of the Neandertal Genome. *Science (80-.).* **328,** 710–722 (2010).
37. Sankararaman, S., Patterson, N., Li, H., Pääbo, S. & Reich, D. The Date of Interbreeding between Neandertals and Modern Humans. *PLoS Genet.* **8,** (2012).
38. Vernot, B. & Akey, J. M. Complex History of Admixture between Modern Humans and Neandertals. *Am. J. Hum. Genet.* **96,** 448–453 (2015).
39. Simonti CN, Vernot B, Bastarache L, Bottinger E, Carrell DS, Chisholm RL, Crosslin DR, Hebbring SJ, Jarvik GP, Kullo IJ, Li R, Pathak J, Ritchie MD, Roden DM, Verma SS, Tromp G, Prato JD, Bush WS, Akey JM, Denny JC, and C. J. The phenotypic legacy of admixture between modern humans and Neanderthals. *Science (80-.).* **in press.,** 737–742 (2016).
40. Briggs, A. W. *et al.* Targeted Retrieval and Analysis of Five Neandertal mtDNA Genomes. *Science (80-.).* **325,** 318–321 (2009).
41. Dalén, L. *et al.* Partial genetic turnover in neandertals: Continuity in the east and population replacement in the West. *Mol. Biol. Evol.* **29,** 1893–1897 (2012).
42. Castellano, S. *et al.* Patterns of coding variation in the complete exomes of three Neandertals. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 6666–71 (2014).
43. Wood, R. E. *et al.* A new date for the neanderthals from el sidrón cave (asturias, northern spain). *Archaeometry* **55,** 148–158 (2013).
44. Markert, J. a *et al.* Neutral locus heterozygosity, inbreeding, and survival in Darwin's ground finches (Geospiza fortis and G. scandens). *Heredity (Edinb).* **92,** 306–315 (2004).

45. Qutob, N. *et al.* Signatures of historical demography and pathogen richness on MHC class i genes. *Immunogenetics* **64,** 165–175 (2012).
46. Andres, A. M. *et al.* Targets of balancing selection in the human genome. *Mol. Biol. Evol.* **26,** 2755–2764 (2009).
47. Gravel, S. When is selection effective? *bioRxiv* 010934 (2014). doi:10.1101/010934
48. Henn, A. B. M. *et al.* Distance from Sub-Saharan Africa Predicts Mutational Load in Diverse Human Genomes. *bioRxiv* (2015).
49. Juric, I., Aeschbacher, S. & Coop, G. The Strength of Selection Against Neanderthal Introgression. *bioRxiv* 030148 (2015). doi:10.1101/030148
50. Harris, K. & Nielsen, R. The Genetic Cost of Neanderthal Introgression. *bioRxiv* 030387 (2015). doi:10.1101/030387
51. Do, R., Balick, D., Li, H. & Adzhubei, I. No evidence that natural selection has been less effective at removing deleterious mutations in Europeans than in West Africans. *arXiv Prepr. arXiv …* **47,** (2015).
52. Zhao, Z., Fu, Y. X., Hewett-Emmett, D. & Boerwinkle, E. Investigating single nucleotide polymorphism (SNP) density in the human genome and its implications for molecular evolution. *Gene* **312,** 207–213 (2003).
53. Hughes, A. L. *et al.* Widespread purifying selection at polymorphic sites in human protein-coding loci. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 15754–15757 (2003).
54. Adzhubei, I. a *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7,** 248–249 (2010).
55. Akira, S., Takeda, K. & Kaisho, T. Toll-like receptors: critical proteins linking innate and acquired immunity. *Nat Immunol* **2,** 675–680 (2001).
56. Netea, M. G. & Joosten, L. A. B. TLRs of Our Fathers. *Immunity* **44,** 218–220 (2016).
57. Ferrer-Admetlla, A. *et al.* Balancing Selection Is the Main Force Shaping the Evolution of Innate Immunity Genes. *J. Immunol.* **181,** 1315–1322 (2008).
58. Fujita, T. Evolution of the lectin–complement pathway and its role in innate immunity. *Nat. Rev. Immunol.* **2,** 346–353 (2002).
59. Enard, D., Cai, L., Gwenapp, C. & Petrov, D. A. Viruses are a dominant driver of protein adaptation in mammals. *bioRxiv* (2015). doi:10.1101/029397
60. Azevedo, L., Serrano, C., Amorim, A. & Cooper, D. N. Trans-species polymorphism in humans and the great apes is generally maintained by balancing selection that modulates the host immune response. *Hum. Genomics* 4–9 (2015). doi:10.1186/s40246-015-0043-1
61. Lenz, T. L., Eizaguirre, C., Kalbe, M. & Milinski, M. Evaluating patterns of convergent evolution and trans-species polymorphism at mhc immunogenes in two sympatric stickleback species. *Evolution (N. Y).* **67,** 2400–2412 (2013).
62. Loisel, D. a, Rockman, M. V, Wray, G. a, Altmann, J. & Alberts, S. C. Ancient polymorphism and functional variation in the primate MHC-DQA1 5' cis-regulatory region. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 16331–16336 (2006).
63. Leffler, E. M. *et al.* Multiple Instances of Ancient Balancing Selection Shared Between Humans and Chimpanzees. *Science (80-.).* **339,** 1–8 (2013).
64. Kuhlwilm, M. *et al.* Ancient gene flow from early modern humans into Eastern Neanderthals. *Nature* **530,** 429–433 (2016).
65. DeGiorgio, M., Lohmueller, K. E. & Nielsen, R. A Model-Based Approach for Identifying Signatures of Ancient Balancing Selection in Genetic Data. *PLoS Genet.* **10,** e1004561 (2014).
66. Key, F. M., Teixeira, J. C., de Filippo, C. & Andrés, A. M. Advantageous diversity maintained by balancing selection in humans. *Curr. Opin. Genet. Dev.* **29,** 45–51 (2014).
67. Gao, Z., Przeworski, M. & Sella, G. Footprints of ancient-balanced polymorphisms in genetic variation data from closely related species. *Evolution (N. Y).* **69,** 431–446 (2015).

68. Rasmussen, M. D., Hubisz, M. J., Gronau, I. & Siepel, A. Genome-Wide Inference of Ancestral Recombination Graphs. *PLoS Genet.* **10,** (2014).
69. Prado-Martinez, J. *et al.* Great ape genetic diversity and population history. *Nature* **499,** 471–475 (2013).
70. Ferguson, W., Dvora, S., Fikes, R. W., Stone, A. C. & Boissinot, S. Long-term balancing selection at the antiviral gene oas1 in central African chimpanzees. *Mol. Biol. Evol.* **29,** 1093–1103 (2012).
71. Burbano, H. a *et al.* Targeted Investigation of the Neandertal Genome by Array-Based Sequence Capture. *Science (80-.).* **723,** 723–726 (2010).
72. Parks, M. *et al.* Ancient population genomics and the study of evolution. *Phil Trans R Soc B* (2015).
73. Deschamps, M. *et al.* Genomic Signatures of Selective Pressures and Introgression from Archaic Hominins at Human Innate Immunity Genes. *Am. J. Hum. Genet.* **98,** 5–21 (2016).
74. Abi-Rached, L. *et al.* The shaping of modern human immune systems by multiregional admixture with archaic humans. *Science* **334,** 89–94 (2011).
75. Dannemann, M., Andrés, A. M. & Kelso, J. Introgression of Neandertal- and Denisovan-like Haplotypes Contributes to Adaptive Variation in Human Toll-like Receptors. *Am. J. Hum. Genet.* **98,** 22–33 (2016).
76. R Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria.* (2015).
77. Nei, M. & Li, W. H. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. U. S. A.* **76,** 5269–5273 (1979).
78. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **57,** 289–300 (1995).

Figure Legends

Figure 1. Genome-wide SNP derived allele frequency distributions for Neandertals and three modern human populations. The number of SNPs observed for each population at exome sites with sufficient sequence coverage for SNP identification in all individuals in the study, binned by derived allele frequency (of n = 6 chromosomes per population). (A) Synonymous SNPs. (B) Nonsynonymous SNPs, with "predicted benign" and "predicted damaging" SNPs separately indicated, based on PolyPhen-2[54] estimates.

Figure 2. Patterns of Neandertal and modern human nonsynonymous SNP diversity genome wide, in innate immune genes, in virus-interacting protein genes, and in *MHC* genes. The number of nonsynonymous SNPs per population for sites with sufficient sequence coverage for SNP identification in all individuals in the study, with PolyPhen-2 "predicted benign" and "predicted damaging" SNPs indicated separately, and the proportion of the nonsynonymous SNPs with minor allele frequencies = 1, 2, and 3 (of n = 6 total chromosomes per population). (A-B) Nonsynonymous SNPs in the genome-wide set of n = 11,086 genes that excludes innate immune receptor, signaling adaptor molecule, complement pathway, virus-interacting protein, and *MHC* genes. (C-D) Nonsynonymous SNPs in the set of n = 73 innate immune receptor, signaling adaptor molecule, and complement pathway genes. (E-F) Nonsynonymous SNPs in the set of n = 164 virus-interacting protein genes with known antiviral or immune activities. (G-H) Nonsynonymous SNPs in the set of n = 13 *MHC* genes.

Figure 3. Patterns of Neandertal and modern human nonsynonymous SNP diversity genome wide and for genes with high genetic diversity in non-human apes. The number of nonsynonymous SNPs per population for sites with sufficient sequence coverage for SNP identification in all individuals in the study, with PolyPhen-2 "predicted benign" and "predicted damaging" SNPs indicated separately, and the proportion of the nonsynonymous SNPs with minor allele frequencies = 1, 2, and 3 (of n = 6 total chromosomes per population). (A-B) Nonsynonymous SNPs in the genome-wide set of n = 7,186 genes with at least 500 "callable" sites with sufficient coverage and mapping quality for SNP identification in a great ape population genomics panel (see *Methods*), excluding the top 1% highest diversity genes among four non-human ape species. (C-D) Nonsynonymous SNPs in the set of n = 73 genes with the consistently highest levels of genetic diversity among the four non-human ape species.

Figure 1: Genome-wide SNP derived allele frequency distributions for Neandertals and three modern human populations

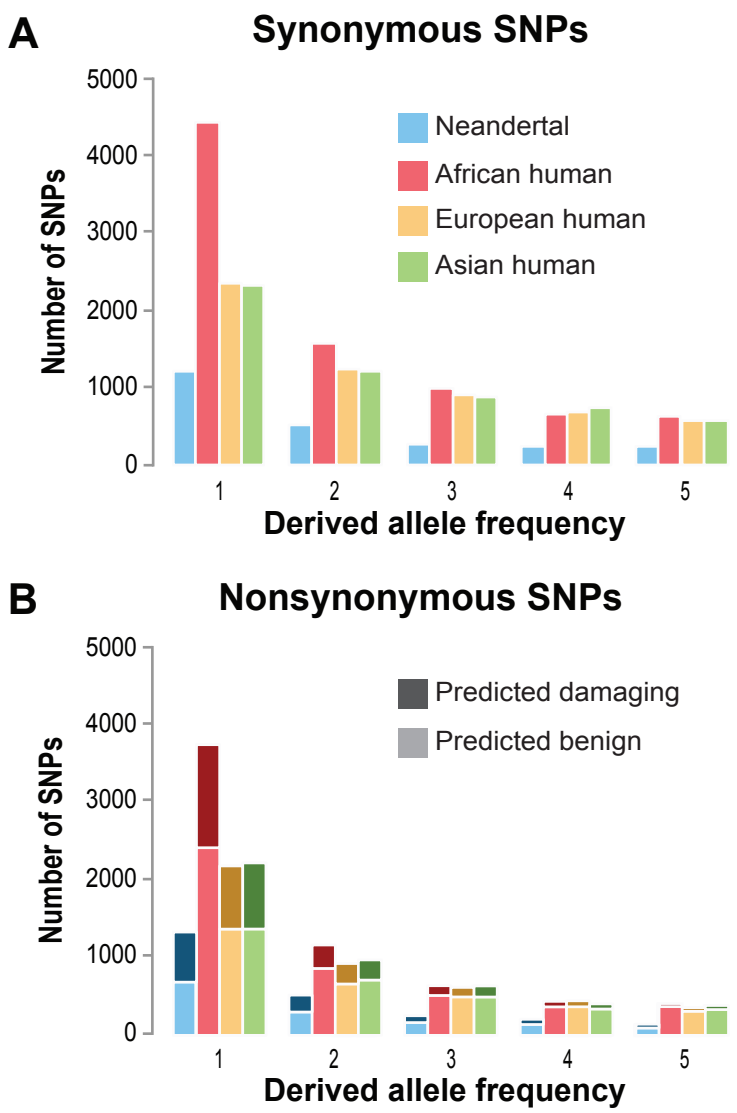


Figure 2: Patterns of Neandertal and modern human nonsynonymous SNP diversity genome wide, in virus-interacting protein genes, in innate immune genes, and in *MHC* genes

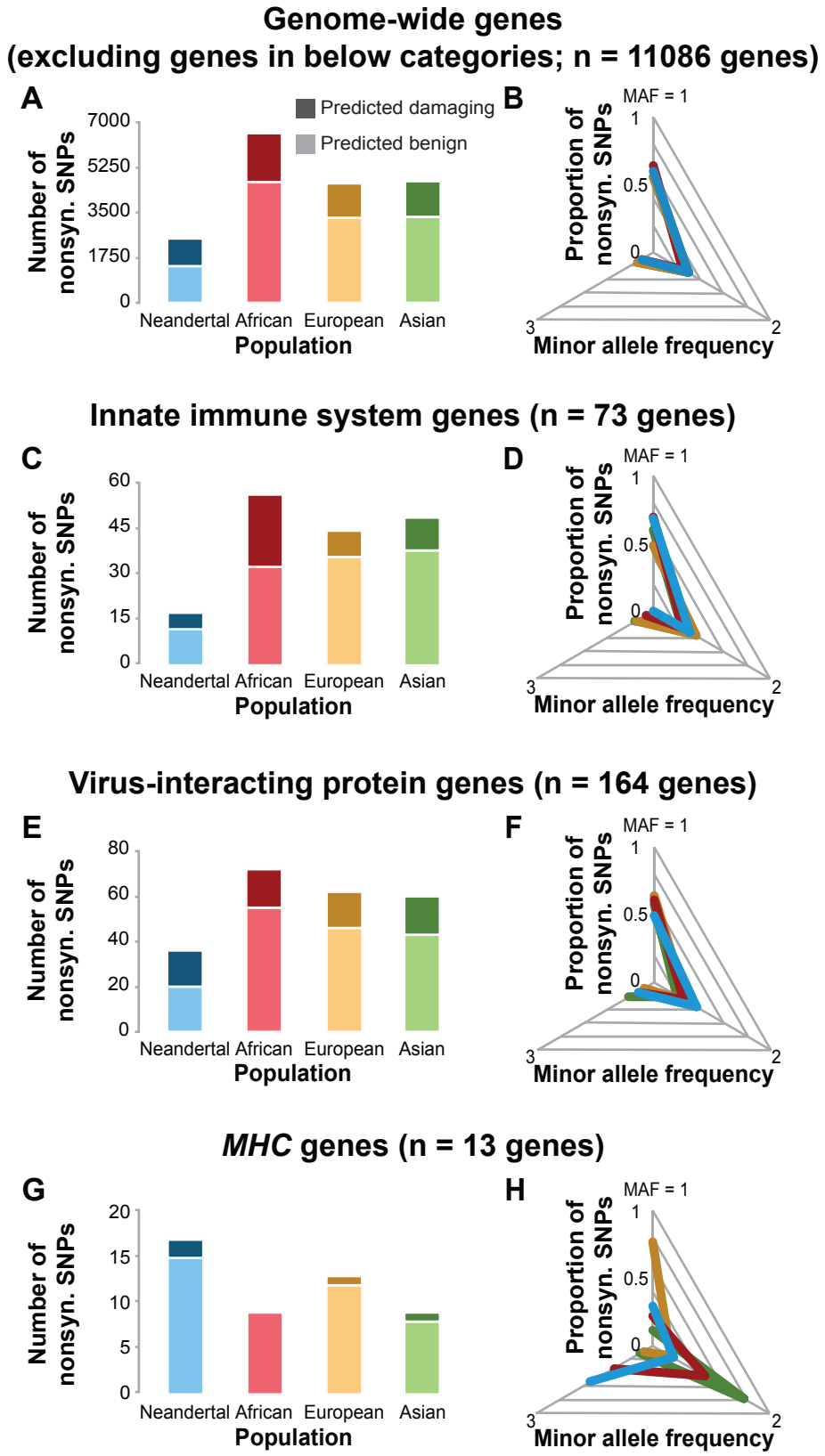
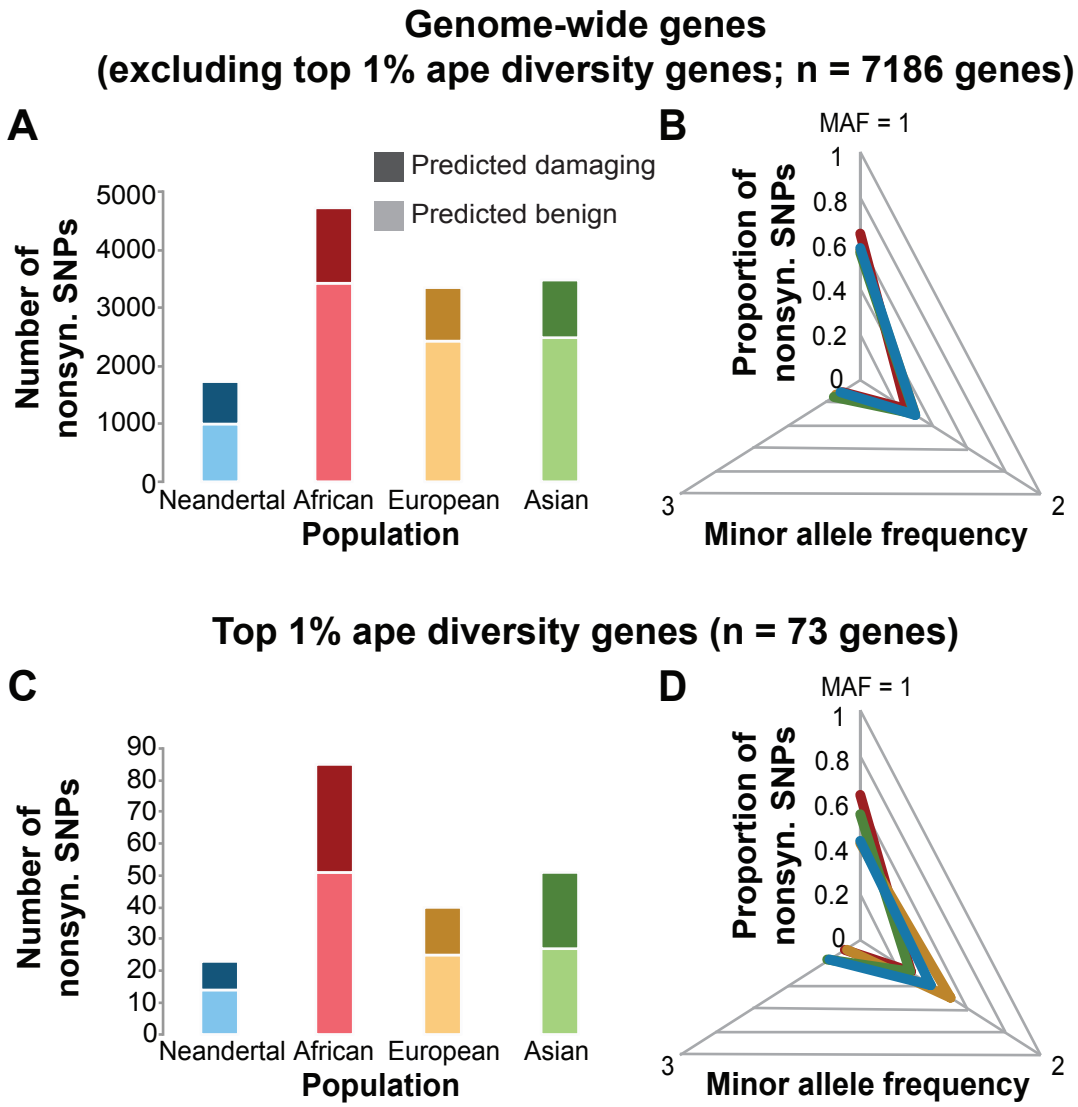


Figure 3: Patterns of Neandertal and modern human nonsynonymous SNP diversity genome wide and for genes with high genetic diversity in non-human apes



Supplementary Materials

Supplemental Database 1: Database of nonsynonymous and synonymous SNP genotypes and population frequencies (https://scholarsphere.psu.edu/files/v405s943f)

Supplemental Database 2: Database of great ape gene diversity data (https://scholarsphere.psu.edu/files/5d86p0267)

Supplemental Figure 1: Innate immune system gene permutation analyses – 10,000 sets of 73 randomly selected genes containing nonsynonymous SNPs

Supplemental Figure 2: Virus-interacting protein gene permutation analyses – 10,000 sets of 164 randomly selected genes containing nonsynonymous SNPs

Supplemental Figure 3: *MHC* gene permutation analyses – 10,000 sets of 13 randomly selected genes containing nonsynonymous SNPs

Supplemental Figure 4: Patterns of Neandertal and modern human nonsynonymous SNP diversity in *MHC* genes (n = 13) excluding the Altai Neandertal and one random modern human per population

Supplemental Figure 5: Significantly enriched gene ontology categories (red) among top 1% ape diversity genes

Supplemental Table 1: A comparison of genome-wide nonsynonymous SNPs versus total (nonsynonymous + synonymous) SNPs between Neandertal and modern human populations

Supplemental Table 2: PolyPhen-2 predictions for genome-wide nonsynonymous SNPs – damaging versus not damaging – for Neandertal and modern human populations

Supplemental Table 3: List of Innate immune system genes

Supplemental Table 4: List of virus-interacting protein genes with known antiviral or broader immune activities

Supplemental Table 5: List of *MHC* genes

Supplemental Table 6: A comparison of nonsynonymous SNPs (benign + damaging) in innate immune system genes versus genome-wide genes (not including innate immune genes) between Neandertal and modern human populations

Supplemental Table 7: A comparison of nonsynonymous SNPs (benign + damaging) in virus-interacting protein genes versus genome-wide genes (not including virus-interacting protein genes) between Neandertal and modern human populations

Supplemental Table 8: A comparison of nonsynonymous SNPs (benign + damaging) in *MHC* genes versus genome-wide genes (not including *MHC* genes) between Neandertal and modern human populations

Supplemental Table 9: Neandertal Minor Allele Frequency (MAF) comparison of nonsynonymous SNPs in MHC genes

Supplemental Table 10: A comparison of nonsynonymous SNP (benign + damaging) Minor Allele Frequencies (MAFs) in *MHC* genes between Neandertal and modern human populations

Supplemental Table 11: List of top 1% ape diversity genes

Supplemental Table 12: Results of the top 1% ape diversity genes Gene Ontology enrichment analysis

Supplemental Table 13: A comparison of nonsynonymous SNPs (benign + damaging) in top 1% ape diversity genes between Neandertal and modern human populations