Title: Archaic introgression and variation in pharmacogenes and implications for local adaptation

Authors: Tadeusz H. Wroblewski [*1], Kelsey E. Witt [*2], Seung-been Lee [3], Ripan S. Malhi [4], Emilia Huerta-Sanchez [2], Fernando Villanea [&5], Katrina G. Claw [&1]

Affiliations:

[1] Division of Biomedical Informatics and Personalized Medicine, Colorado Center for Personalized Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

[2] Ecology and Evolutionary Biology and Center for Computational and Molecular Biology, Brown University, Providence, RI, USA

[3] Precision Medicine Institute, Macrogen Inc., Seoul, Republic of Korea

[4] Department of Anthropology, Carl R. Woese Institute for Genomic Biology, University of Illinois Urbana-Champaign, USA

[5] Department of Anthropology, University of Colorado Boulder, Boulder CO, USA

*co-first authors

[&]co-senior authors/corresponding authors

Abstract

Modern humans carry Neanderthal and Denisovan (archaic) genome elements which may have been adaptive. These effects may be particularly evident in pharmacogenes – genes responsible for the processing of exogenous substances such as food, pollutants, and medications. However, the health implications and contribution of archaic ancestry in pharmacogenes of modern humans remains understudied. We characterize eleven key cytochrome P450 (*CYP450*) genes involved in drug metabolizing reactions in three Neanderthal and one Denisovan individuals and examine archaic introgression in modern human populations. We infer the metabolizing efficiency of these eleven genes in archaic individuals and show important genetic differences relative to modern human variants. We identify archaic-specific SNVs in each *CYP450* gene, including some that are potentially damaging, which may result in altered metabolism in modern human people carrying these variants. We highlight four genes which display interesting patterns of archaic variation: *CYP2B6* – we find a large number of unique variants in the Vindija Neanderthal, some of which are shared with a small subset of African modern humans; *CYP2C9* – containing multiple variants that are shared between Europeans and Neanderthals; *CYP2A6*12* – a variant defined by a hybridization event that was found in humans and Neanderthals, suggesting the recombination event predates both species; and *CYP2J2* – in which we hypothesize a Neanderthal variant was re-introduced in non-African populations by archaic admixture. The genetic variation identified in archaic individuals suggests that selective pressures from the environment may be involved in the evolution of *CYP450* genes.

Introduction

The cytochrome P450 (*CYP450*) genes encode oxidase enzymes that function in metabolism of endogenous small molecules and in detoxification of exogenous (or xenobiotic) compounds. This gene family is present in all mammals, with 57 active genes and 58 pseudogenes coding for CYP450 enzymes in humans (Thomas 2007). Xenobiotic-substrate enzymes have been studied extensively because of their roles in the absorption, distribution, metabolism, and excretion (ADME) of pharmaceuticals and drug development. The evolution of xenobiotic CYP450 enzymes may be driven by an organism's need to metabolically detoxify foreign compounds, often toxic chemicals produced by plants, fungi, and bacteria, in the local environment. The *CYP450* genes show evidence of positive selection and high allele frequency variation in humans (Thomas 2007; Fuselli et al. 2010). It has been suggested that the shift from hunting and gathering to food production in humans may have profoundly changed the selective effect of some CYP450 enzymes (Fuselli et al. 2010; Fuselli 2019). Although these genes have been studied extensively in humans and other mammals, the extent of genetic variation in archaic individuals and its relation to modern human variation has not been addressed. Genetic variation in ADME genes varies extensively across modern human populations, and identifying alleles inherited through archaic introgression may be informative about the origin of specific pharmacogenetic (PGx) variants and their resulting phenotypes (e.g., metabolizer status).

Throughout human evolutionary history, our species has adapted to numerous distinct and varied challenging environments, which present the need to metabolize new xenobiotic substances, such as food, pollutants, and medications (Fan et al. 2016). Modern humans encountered new environmental challenges as they dispersed first throughout and then outside of the African continent, but they also encountered other hominin species already adapted for life in those regions of the world: Neanderthals and Denisovans (Durvasula and Sankararaman 2020; Bergström et al. 2021). The direct sequencing of multiple Neanderthal and Denisovan genomes has revealed a complex history of admixture between these archaic humans and the ancestors of modern humans (Browning et al. 2018; Villanea and Schraiber 2018). Most modern humans carry a small but significant portion of archaic ancestry, which has been targeted by natural selection (Sankararaman et al. 2016). Purifying selection — or negative natural selection — has removed most archaic variants in functional genomic regions (Petr et al. 2019; Zhang et al. 2020; Schaefer et al. 2021), while some archaic variants may have been lost through genetic bottlenecks or drift in modern human populations. However, there are functional regions for which archaic variants are found at very high frequency in living humans through the effects of positive natural selection (Mendez et al. 2012; Huerta-Sánchez et al. 2014; Racimo et al. 2015; Dannemann et al. 2016). One of the most well-known examples of adaptation through archaic admixture is high-altitude adaptation in Tibetans, a population in which the Denisovan variant of the gene *EPAS1* is found at extremely high frequency, and has conferred them with resistance to hypoxic stress in high-altitude environments (Huerta-Sánchez et al. 2014).

The *CYP450* genes are a prime target for adaptation through archaic introgression as humans encountered novel exogenous substances as they expanded their range. Neanderthals and Denisovans may have possessed *CYP450* variants fine-tuned to metabolizing substances found in their native habitats of Eurasia and Siberia. As modern humans expanded outside of Africa, they likely faced novel environmental factors which may have influenced selective pressures. Advantageous archaic *CYP450* variants introduced to modern humans through admixture may have been retained in the modern human gene pool by natural selection.

Here, we investigate the evolution and predicted phenotypes of *CYP450* genes in archaic individuals to increase our understanding of introgressed genetic variants and the different selective regimes and environments acting on these enzymes. We examine genetic variation and predict metabolizer phenotypes in eleven *CYP450* genes in three Neanderthal and one Denisovan individuals using publicly available high-resolution whole genome sequencing data. In addition, we examine the effect of archaic introgression in modern human populations for these eleven genes. These 11 genes encode enzymes that when combined are responsible for up to 75% of the metabolism of commonly prescribed drugs (Evans and Relling 1999), which could have implications for modern human health and disease as well as drug safety and efficacy.

Results

Archaic and shared human CYP450 variation

In this study, we investigated genetic variation in eleven *CYP450* genes: *CYP1A2, CYP2A6, CYP2B6, CYP2C8, CYP2C9, CYP2C19, CYP2D6, CYP2E1, CYP2J2, CYP3A4,* and *CYP3A5*. We identified a total of 1,623 single nucleotide variants (SNVs) in the four archaic individuals (one Denisovan from Denisova Cave and three Neanderthals from the Vindija, Denisova Cave/Altai, and Chagyrskaya sites) for the eleven *CYP450* genes investigated (Figure 1, Supplemental Table 1). Of the variants identified in the archaic individuals, 81.2% (n=1318) were intronic, 6.8% (n=111) were in the promoter region, 4.7% (n=77) were exonic, 5.1% (n=83) were in the untranslated regions (3'/5'-UTR), 1.9% (n=31) were non-coding RNA (ncRNA), and 0.2% (n=3) were splicing variants (Figure 1). Across all genes investigated, the Vindija Neanderthal presented the highest number of total variants (n=943), followed by the Chagyrskaya Neanderthal, Altai Neanderthal, and the Denisovan individual (n=667, n=584, and n=574, respectively).

Structural variation (SV) was additionally investigated in each of the eleven *CYP450* genes in archaic individuals using sequencing read data and previously validated methods (Lee, Wheeler, Patterson, et al. 2019; Lee, Wheeler, Thummel, et al. 2019). SV was only observed in *CYP2A6* and *CYP2D6* in Neanderthal individuals (Figure 2, Supplemental Figure 1). The Altai and Vindija Neanderthal individuals exhibited a greater structural variation burden, presenting the homozygous partial deletion hybrid variant *CYP2A6*12/*12* (Figure 2a-b) and a gene multiplication variation event: *CYP2D6*2/*2x4* for Altai (i.e., 4 copies of *CYP2D6*) and *CYP2D6*2/*2x3* for

Vindija (Supplemental Figure 1a-b). The Chagyrskaya Neanderthal was heterozygous with a *CYP2A6*1/*12* diplotype (Figure 2c) and there were no SV identified in the diplotype *CYP2D6*2/*41* (Supplemental Figure 1c). The Denisovan individual did not show any copy number variation in any of the eleven *CYP450* genes investigated (Figure 2d, Supplemental Figure 1d).

To further assess the genomic landscape of the archaic individuals in the context of modern humans, we identified variation in modern human individuals from the 1000 Genomes Project (Supplemental Figure 2, Supplemental Table 2a, 1000 Genomes Project Consortium et al. 2015), as well as Papuans sequenced as part of the Simons Genome Diversity Project (Mallick et al. 2016) and compared the archaic and modern human *CYP450* haplotypes. To determine genetic variation that could be a result of potential introgression, we identified archaic variants that were present in non-African modern human populations. As the majority of archaic admixture occurred in Eurasia, shared genetic variants between archaic humans and Africans are expected to be shared ancestrally, but variants exclusively shared between non-Africans and archaic humans are expected to be the result of archaic admixture (see *U* statistic in Racimo et al., 2016). A total of 155 archaic SNVs were shared between archaic and modern humans (Table 1, Supplemental Table 3); 140 of these SNVs were shared only between Neanderthals and non-African modern humans, 11 were shared between only the Denisovan individual and non-African modern humans, and 4 of the shared SNVs were identified in the Neanderthals, the Denisovan, and non-African modern humans (Supplemental Table 3). *CYP2C19* contained the greatest number of archaic variants shared with non-African modern humans (*n*=64). Although most shared variants were present at a frequency of less than 2% in any non-African modern human population, some *CYP450* genes contain archaic variants that are found in human populations with a frequency of 10% and greater (*n*=41). These variants include SNVs in *CYP2C8* and *CYP2J2* that were at frequencies greater than 10% in Europeans and admixed Americans, SNVs in *CYP2C9* that were at elevated frequencies in Europeans, admixed Americans, and South Asians (including *rs1799853*, the causative SNV for the *CYP2C9*2* haplotype, that was identified at frequencies of 8-15% in European populations and admixed American populations), and a SNV in *CYP1A2* (*rs2470890*) found at elevated frequencies in all non-African populations (Supplemental Table 3). Of the 41 SNVs found at elevated frequencies, only four were exonic (*rs2470890* in *CYP1A2, rs10509681* in *CYP2C8, rs11572080* in *CYP2C8*, and *rs1799853* in *CYP2C9*) and the remainder were intronic. Of all the non-African modern human populations, Papuans had the fewest shared archaic SNVs, but for these sites, archaic variants were observed at higher allele frequencies than any other human population, and all but one of these archaic variants were exclusive to Papuans (Supplemental Table 3). For example, we identified the exonic variant *rs3915951* in *CYP2D6* that had an allele frequency of 12.5% (*n*=1) in Papuans and was not found in any other modern human population (Supplemental Table 3). This observation of high frequency of archaic alleles exclusive to Papuans is consistent with either founder effects or a higher degree of archaic admixture into their ancestral populations.

Super-divergent CYP2B6 haplotypes shared between the Vindija Neanderthal and African individuals

The surprisingly large number of heterozygous SNVs in the Vindija Neanderthal *CYP2B6* gene was an outlier relative to other Neanderthal diplotypes and warranted further scrutiny. We identified 334 variants sites in the Vindija Neanderthal, while the other three archaic genomes had fewer than 100 variants each (Denisovan: n=75; Altai: n=72; and Chagyrskaya: n=76, Figure 1). The Vindija Neanderthal had far more variant sites than the other Neanderthal or Denisovan individuals, with 91.9% (307/334) of these variant sites being heterozygous (Figure 3). To determine how the divergent Vindija *CYP2B6* haplotype compared to other regions of the genome, we calculated the pairwise distance between the Chagyrskaya Neanderthal and Vindija Neanderthal with 29.1 kb windows (*CYP2B6* gene length) across the genome and identified the Vindija *CYP2B6* region as within the top 1% of windows based on pairwise distance between Neanderthals (pairwise distance = 0.00175, Supplemental Figure 3).

Given that *CYP2B6* shares significant homology with the pseudogene, *CYP2B7*, which is located nearby (40.6 kb), we wanted to ensure that the elevated variation identified in *CYP2B6* in the Vindija Neanderthal was not an effect of read mis-mapping with the paralog gene (Zanger and Klein 2013). To confirm a low likelihood of paralog gene mis-mapping error, we assessed the read depth at the *CYP2B6* locus in the Vindija Neanderthal genome and identified no read depth elevation nor structural variation present (Supplemental Figure 4).

Furthermore, we identified 11 individuals from 1000 Genomes Project samples (1000 Genomes Project Consortium et al. 2015) who carry a related haplotype, all also showing a non-elevated read depth and a similar increase in variation to the Vindija Neanderthal (Figure 3, Supplemental Figure 4, Supplemental Table 2b). Human individuals carrying this haplotype were found exclusively in all African populations (ASW, ESN, GWD, LWK, MSL, YRI) at low frequencies (Supplemental Table 2b). The two divergent *CYP2B6* haplotypes (in the Vindija Neanderthal and in the 11 Africans) uniquely share some SNVs (n=206), however they also present 92 African and 58 Vindija SNVs that are unique to each individual haplotype (Supplemental Figure 5). Comparatively, the other archaic individuals harbor far fewer unique SNVs in their *CYP2B6* haplotypes (26 SNVs across the three individuals).

Phased Diplotypes and Predicted Metabolizer Phenotypes

Pharmacogene haplotypes are typically identified as star alleles - haplotype patterns composed of SNVs, indels, and structural variants (SVs) in pharmacogenes that are usually associated with enzyme activity levels. We identified the star allele composition of the eleven *CYP450* genes in archaic individuals using variant and sequencing read data using previously validated methods (Lee, Wheeler, Patterson, et al. 2019; Lee, Wheeler, Thummel, et al. 2019). We identify the phased diplotype and predicted phenotype for each individual with the two primary star haplotypes called by Stargazer (Lee, Wheeler, Patterson, et al. 2019; Lee, Wheeler, Thummel, et al. 2019, Figure 4, Supplemental Table 4).

All three Neanderthal individuals displayed the reference diplotype (**1/*1*) for *CYP2J2* and *CYP3A5* and

presented the following non-reference homozygous diplotypes: *CYP1A2*1F/*1F*, *CYP2B6*22/*22*, *CYP2C8*3/*3*, *CYP2C19*8/*8*, and *CYP2E1*7/*7* (Figure 4). The Altai and Vindija Neanderthals presented the homozygous SV diplotype *CYP2A6*12/*12*, while the Chagyrskaya Neanderthal was heterozygous for **12* (*CYP2A6*1/*12*, Figure 4). For *CYP2C9*, the Chagyrskaya Neanderthal presented a **2/*2* diplotype, the Altai Neanderthal presented a **1/*2* diplotype, and we were unable to accurately determine the diplotype for the Vindija Neanderthal because of a lack of read coverage at specific star variant locations (see Methods, Figure 4). The Altai and Vindija Neanderthals presented gene copy number variation for *CYP2D6,* presenting **2/*2x4* and **2/*2x3* diplotypes, respectively; the Chagyrskaya displayed a *CYP2D6*2/*41* diplotype. For *CYP3A4*, the Chagyrskaya Neanderthal presented a reference diplotype (**1/*1*), the Altai Neanderthal presented a **1/*1B* diplotype, and the Vindija Neanderthal presented a homozygous **1B/*1B* diplotype.

We additionally predicted metabolizer status phenotypes based on the diplotypes for each individual using Stargazer, as designated by the Clinical Pharmacogenetics Implementation Consortium (CPIC, https://cpicpgx.org, Figure 4). While Stargazer is designed for human phenotype prediction, we can use it to analyze the landscape for overall functionality in archaic individuals. All three Neanderthal individuals were predicted to have the same phenotype for the following enzymes: a normal metabolizer phenotype for CYP2J2 and CYP3A5; an intermediate metabolizer phenotype for CYP2C8 and CYP2E1; a slow metabolizer phenotype for CYP2A6; a poor metabolizer phenotype for CYP2C19; an ultra-rapid metabolizer phenotype for CYP2B6; and the diplotype for CYP1A2 currently has an unknown functional impact based on PharmVar (https://www.pharmvar.org/). For the CYP2C9 enzyme, the phenotype for the Chagyrskaya Neanderthal was predicted as a normal metabolizer, the Altai Neanderthal phenotype was predicted as an intermediate metabolizer, and the phenotype for the Vindija Neanderthal could not be accurately determined. For CYP3A4, the Chagyrskaya Neanderthal had a predicted normal metabolizer phenotype and the diplotypes for the Altai and Vindija Neanderthal individuals currently have an unknown functional impact. The gene copy number increase of *CYP2D6* in the Altai and Vindija Neanderthal individuals predict an ultra-rapid metabolizer phenotype while the Chagyrskaya was predicted to have a normal metabolizer phenotype for this enzyme.

The Denisovan individual had notably different variation from that observed in the Neanderthal individuals (Figure 4). The Denisovan individual displayed the reference diplotype (**1/*1*) for *CYP2C8* and *CYP2C19* and presented the following non-reference diplotypes: *CYP1A2*1/*1F*, *CYP2A6*1/*9, CYP2B6*6/*6*, *CYP2C9*1/*8*, *CYP2D6*1/*4, CYP2E1*1/*7*, *CYP2J2*1/*7*, *CYP3A4*1B/*1B*, and *CYP3A5*1/*3* (Figure 4). The predicted phenotypes for the Denisovan individual showed a normal metabolizer phenotype for the CYP2C8, CYP2C9, CYP2C19, and CYP2E1 enzymes, an intermediate metabolizer phenotype for the CYP2B6, CYP2D6, and CYP3A5 enzymes, and a slow metabolizer phenotype for CYP2A6 enzyme. The diplotypes identified for the CYP1A2, CYP2J2, and CYP3A4 enzymes currently have an unknown functional impact.

Potentially-Function Altering SNVs

While Stargazer can make predictions from known star alleles, a number of novel missense variants were identified in the dataset. We identified potentially damaging or function altering SNVs in each gene using various programs that can predict the functional impact of genetic variation: Combined Annotation Dependent Depletion (CADD) integrates multiple weighted metrics to identify deleterious variants (Rentzsch et al. 2019; Rentzsch et al. 2021), Sorting Intolerant From Tolerant (SIFT) predicts the functional impact of amino acid substitution caused by genetic variation (Vaser et al. 2016), and Polymorphism Phenotyping (PolyPhen) identifies the predicted structural and functional outcome of amino acid substitution (Adzhubei et al. 2010). Variants considered potentially damaging or function-altering were defined as SNVs that are in the top 1% of deleterious variants by CADD score (PHRED-normalized CADD score ≥ 20), deleterious by SIFT prediction (SIFT score ≤ 0.05) or damaging by Polyphen2 prediction (Polyphen2 score between 0.15-1.0 as possibly damaging and 0.85-1.0 as probably damaging). We identified 23 potentially damaging variants in the eleven *CYP450* genes investigated, of which 20 were considered deleterious by CADD (PHRED-normalized CADD score was ≥ 20), 16 were predicted by SIFT as deleterious, and 16 had a Polyphen prediction of damaging: 3 as "possibly damaging" and 13 as "probably damaging" (Figure 5, Supplemental Table 5). Of the 23 variants identified, 7 SNVs were considered novel and not observed in any variant annotation databases utilized (see Methods). *CYP2D6* had the highest number of deleterious variants found in exonic and splicing sites (4 SNVs total) and the other genes each had 1-3 potentially deleterious variants. In the *CYP2J2* gene, we have identified numerous deleterious SNVs in the Chagyrskaya and Vindija Neanderthal individuals in heterozygous form but did not find any of these Neanderthal deleterious SNVs in modern human populations. No potentially deleterious SNVs were identified in *CYP2A6*. Four SNVs were diagnostic for star allele haplotypes as discussed above (Figure 4), including the variant accounting for the *CYP2C9*2* allele (*rs1799853*), the variant defining the *CYP2C19*8* allele (*rs41291556*), and two variants (*rs3892097* and *rs1065852*) that compose the *CYP2D6*4* haplotype (Figure 5).

Most archaic variants present in non-African modern humans were either intronic (n=140) or exonic synonymous (n=4) mutations, making their impact on metabolism unclear; however, we identify three exonic variants with a CADD score ≥ 20, suggesting that they are likely to impact function (*rs3915951*, *rs41291556*, and *rs1799853*, **bolded** in Supplemental Table 3). The first variant, *rs3915951,* is an exonic missense variant in *CYP2D6* that was identified as heterozygous in a single Papuan. The second variant is the causative exonic missense SNV for *CYP2C19*8* (*rs41291556*) and had a frequency of 1.5% in Central Europeans and 0.5% in the Telugu population in the 1000 Genomes Project dataset (Supplemental Table 3). The third variant is the causative missense SNV for *CYP2C9*2* (*rs1799853*) and was found globally at less than 5% frequency, with the exception of in European and admixed American populations, where the allele frequency ranges from 8-15% (Supplemental Table 3).

Introgression of archaic CYP450 alleles into modern humans

Each *CYP450* gene had a range of archaic SNVs shared with modern human populations, from 2 shared SNVs (*CYP1A2*) to 64 shared SNVs (*CYP2C19*, Table 1). The frequency of shared SNVs also varied extensively across global human populations, likely reflecting patterns of past gene flow with diverse archaic populations and genetic drift (Sankararaman et al. 2014; Sankararaman et al. 2016), but overall the shared archaic SNVs were generally at low frequency across modern human populations (Supplemental Table 3). To confirm if the archaic *CYP450* SNVs identified in modern human populations corresponded to an archaic haplotype inherited from Neanderthals, we calculated sequence divergence between each *CYP450* gene haplotypes and the Vindija Neanderthal using Haplostrips (Supplemental Figure 6a-k). Interestingly, *CYP2J2* showed evidence of haplotype sharing between the archaic individuals and modern humans because there were 8 archaic SNVs shared between Neanderthals and all non-African modern human populations at frequencies higher than 2%, making it a possible example of archaic introgression of a functional gene (Supplemental Table 3). We visualized these distances grouped by geographic region (Figure 6, Supplemental Figure 6i), and found that all non-African super-populations carry Neanderthal-like *CYP2J2* alleles at low frequencies: 0.4% in Southeast Asia, 4.8% in Europe, 3.9% in East Asia, and 7.9% in admixed Americans. The Neanderthal-like *CYP2J2* haplotype was not found in African populations.

*Introgression of the CYP2A6*12 hybrid allele*

We identified the *CYP2A6*12* allele in all three Neanderthal individuals (Figure 3), and this may suggest inheritance of the hybrid allele through human-Neanderthal introgression. The CYP2A6 enzyme metabolizes coumarin, nicotine, and other plant secondary metabolites. The *CYP2A6* gene is located adjacent to the inactive *CYP2A7* gene and several allelic variants of *CYP2A6* have been created by unequal crossover and gene conversion events between these genes (Oscarson et al. 2002). The *CYP2A6*12* is a hybrid allele where exon 1 and 2 originate from *CYP2A7* and exons 3–9 originate from *CYP2A6* (Figure 4), and causes a 50% reduction in CYP2A6 protein levels and a 40% decrease in CYP2A6 coumarin 7-hydroxylation activity (Oscarson et al. 2002), leading to slow metabolism of various substrates of the CYP2A6 enzyme. The *CYP2A6*12* allele is found at low frequencies in global populations, including African American (0.4%), Canadian First Nation (0.5%), and Japanese (0.8%) individuals and is absent in African populations (Oscarson et al. 2002).

However, it is possible that the unequal crossover event that created *CYP2A6*12* predates the split time of the modern humans and Neanderthal-Denisovans, and is present in humans and Neanderthals because of Incomplete Lineage Sorting (ILS, suggested in Lin et al. 2015), but this does not explain why the allele is not present in Africans. To assess if the *CYP2A6*12* haplotype could have survived the disruption from recombination in this region from the time since the split of modern humans and the archaic individuals, we calculated the probability of a haplotype carrying *CYP2A6*12* to be maintained in both human and Neanderthal lineages through ILS. To calculate this probability, we used a previously published ILS equation (see Methods, Huerta-Sánchez et al. 2014). Using the

31.9 kb *CYP2A6*12* hybrid segment (Oscarson et al. 2002) and a regional recombination rate of 0.77 cM/Mb (Myers et al. 2005), we estimate that it is unlikely for segments exceeding 17.1 kb to be maintained since the human-Neanderthal branch split (at α=0.05) and it is therefore improbable that the 31.9 kb *CYP2A6*12* haplotype would have been maintained by ILS in both lineages (p=0.0014).

Discussion

We investigated a panel of *CYP450* genes which are responsible for 75% of all drug metabolizing reactions and represent the bulk of drug metabolizing enzymes (DME) for which therapeutic recommendations exist (Evans and Relling 1999). DMEs are subject to evolutionary processes (e.g., mutation, drift, natural selection) because of their role as detoxifiers of xenobiotic substances. Here, we analyze data from modern and archaic humans to gain insight into the evolutionary history of these loci and to predict metabolizing phenotypes of archaic individuals.

Archaic CYP450 metabolizing phenotypes

We find that the majority of the archaic *CYP450* genes in our panel conferred a normal/intermediate or unknown metabolizer phenotype (*CYP1A2*, *CYP2C8*, *CYP2C9*, *CYP2E1*, *CYP2J2*, *CYP3A4* and *CYP3A5*), consistent with previous studies showing strong purifying selection (Nelson et al. 2012) at these loci. The genes that showed either a poor/slow (*CYP2A6*, *CYP2C19*) or ultra-rapid (*CYP2B6*, *CYP2D6*) metabolizer status in the archaic individuals have also been known to have high variation in human populations (Zhou et al. 2017) and may indicate increased natural population variation or adaptation to the environment. However, it is important to note that the predicted phenotypes are based on modern human phenotypes, and it is unclear how the archaic-specific variation (in particular from the novel function-altering variants) and the archaic genomic landscape would affect metabolism in archaic individuals. These variants of uncertain significance would benefit from further functional tests, individually as well as in the haplotypes found in archaic individuals, measuring enzyme activity and protein abundance.

We highlight three *CYP450* genes with high archaic genetic diversity: *CYP2C9, CYP2A6,* and *CYP2J2.* First, *CYP2C9* contains a number of SNVs that are shared between Neanderthals and Europeans and are rare or absent from Africans, suggesting the possibility of archaic introgression for these variants. The CYP2C9 enzyme is responsible for metabolizing multiple classes of medications and is an important DMEs in humans (Miners and Birkett 1998; Henderson et al. 2018). Substrates for CYP2C9 include non-steroidal anti-inflammatories (Tracy et al. 1995; Miners et al. 1996; Hamman et al. 1997; Yamazaki et al. 1998), angiotensin II blockers (e.g. losartan) (Stearns et al. 1995), S-warfarin (Rettie et al. 1994; Yamazaki et al. 1998), phenytoin (Giancarlo et al. 2001), and tolbutamide (Miners and Birkett 1996). In the *CYP2C9* gene, more than 60 functional haplotypes have been identified through the PharmVar consortium (Gaedigk et al. 2018; Sangkuhl et al. 2021). The shared *CYP2C9*

variants do not show the usual patterns between genetic and geographic distance as human populations expanded out of Africa and into Eurasia, Oceania, and the Americas. If the *CYP2C9*2* variant was indeed introgressed from Neanderthals, the elevated frequency in European populations suggests that it may have been adaptive for processing certain xenobiotics that human populations were exposed to in Western Eurasia. Changes in *CYP450* genes as a result of adaptation to new diets has been previously identified, such as selection for slower-metabolizing *CYP2D6* variants in agricultural populations (Fuselli et al. 2010). However, we cannot rule out that the elevated frequency of the Neanderthal allele in Europeans was the result of strong founder's effect. Further exploration of this haplotype would be needed to determine whether this haplotype was introgressed and then targeted by positive selection.

Second, *CYP2A6* – describes a variant (**12*) defined by a recombination event and found in modern humans, Neanderthals, and Denisovans. The *CYP2A6* gene is expressed mainly in the liver and represents between 1% and 10% of the total liver CYP450 protein (Haberl et al. 2005). The CYP2A6 enzyme metabolizes drugs and pro-carcinogenic compounds including tegafur (Komatsu et al. 2000), valproic acid (Kiang et al. 2006; Tan et al. 2010), and coumarin (Miles et al. 1990), and is the primary enzyme involved in nicotine metabolism (Hukkanen et al. 2005). At present, more than 10 different allelic variants are known that cause absent or reduced enzyme activity (Oscarson et al. 2002; Benowitz et al. 2006), suggesting that eliminating or reducing *CYP2A6* function is a recurring evolutionary strategy. Accordingly, there is marked diversity in *CYP2A6* enzyme function in global populations, where only approximately 1% of European and Middle Eastern populations are poor metabolizers, but up to 20% of East Asian populations show severely reduced enzyme activity (Haberl et al. 2005). An alternative to explain the prevalence of *CYP2A6*12* is that this hybrid allele evolved multiple times, independently in modern non-Africans and Neanderthals, in addition to the low frequency of *CYP2A6*12* introgressed from Neanderthals. It has been proposed that the slow metabolism of some plant secondary metabolites may be adaptive as high levels of these toxins in human tissues may act as a deterrent to parasites (Sullivan et al. 2008; Hagen et al. 2009; Hagen et al. 2013), perhaps explaining why this hybrid allele co-occurs in both Neanderthals and some human populations. Future directions include examining *CYP2A6* variation in more ancestral species, such as primates, to predict the timing of the CYP2A6*12 allele more accurately.

Finally, *CYP2J2* – for which Neanderthal variants may have been introduced to non-African populations by archaic admixture and is now found in some modern populations. The CYP2J2 enzyme accounts for roughly 1-2% of hepatic CYP protein expression, but also displays high expression in the lung, kidney, heart, placenta, salivary gland, and skeletal muscle (Bièche et al. 2007; Murray 2016). CYP2J2 is involved in the oxidation pathways of polyunsaturated fatty acids (PUFAs) (Murray 2016) and mediates the oxidation of drugs including ebastine (Hashizume et al. 2002), astemizole (Matsumoto and Yamazoe 2001; Lafite et al. 2007), terfenadine (Matsumoto and Yamazoe 2001), and ebastine (Lee et al. 2012). The majority of established *CYP2J2* polymorphisms occur at low frequencies, with the most common variant, *CYP2J2*7,* occurring at frequencies between 1-20% across global

populations (King et al. 2002; Dreisbach et al. 2005; King et al. 2005; Wang et al. 2006; Polonikov et al. 2007; Polonikov et al. 2008; Murray 2016). Introgressed variants in *CYP2J2* may have been adaptive and selected for in modern human populations, possibly due to their role in the oxidation of PUFAs. Oxidation of PUFAs brings adverse effects to food by producing off-flavors and may have been instrumental in modern human expansion to new environments. Further work to determine the functional effects of the shared *CYP2J2* variants is needed and to test associations to PUFA levels in modern human populations.

Super-divergent CYP2B6 in the Vindija Neanderthal

The highly divergent haplotypes in the Vindija Neanderthal and a small number of African individuals (n=11) for *CYP2B6* was an unexpected finding. CYP2B6 is responsible for metabolizing drugs such as the prodrug cyclophosphamide (Huang et al. 2000; Zanger and Klein 2013); efavirenz, a non-nucleoside reverse transcriptase inhibitor (Ward et al. 2003; Desta et al. 2007; Zanger and Klein 2013); the antidepressant bupropion (Faucette et al. 2000; Hesse et al. 2000; Zanger and Klein 2013); and ketamine (Desta et al. 2012). Genetic polymorphisms in *CYP2B6* have been identified to alter enzyme activity (Ariyoshi et al. 2001; Lang et al. 2001; Kirchheiner et al. 2003; Ward et al. 2003; Hofmann et al. 2008) and there have been over 37 *CYP2B6* haplotypes identified (Gaedigk et al. 2018; Desta et al. 2021).

The Vindija Neanderthal had far more variant sites in *CYP2B6* compared to the other Neanderthal and Denisovan individuals as well as in other modern human populations. The excess of heterozygous sites did not seem to be due to a mapping error and the shared variation with 11 African individuals from the 1000 Genomes Project strongly suggesting that these sites may be part of the same haplotype, with some divergent haplotypes between the African modern human and the Vindija haplotype. This pattern seems consistent with early human-Neanderthal admixture allowing for private variation to accumulate on the African and Vindija Neanderthal haplotypes. Similarly, introgression from an archaic population with higher levels of heterozygosity could have resulted in two divergent haplotypes being passed to the human-Neanderthal shared lineage, each assorting into African and Neanderthals.

In fact, a study that aimed to identify archaic gene flow in Africans that pre-dates the split time of Neanderthals, Denisovans, and modern humans identified *CYP2B6* as a candidate region in the African Mende population, three of whom have the divergent *CYP2B6* haplotype (Durvasula and Sankararaman 2020). These haplotypes were likely introgressed from a population that lived long before humans diverged from Neanderthals and Denisovans and show much more divergence from one another than the other modern and archaic human haplotypes do. While it is difficult to determine the exact relationship between the two divergent haplotypes because the Vindija genome is un-phased, making it difficult to infer haplotypes, both Neanderthals and modern Africans harbor ancient haplotypes of *CYP2B6.* Given the number of differences between the two *CYP2B6* divergent haplotypes, they likely both existed long before the modern human-Neanderthal split at ~500,000 ybp.

CYP450 genetic introgression

The sharing of archaic variants in the eleven *CYP450* genes we studied in modern humans is largely consistent with our general understanding of archaic introgression in human populations. For example, the number of Neanderthal SNVs is significantly higher than the number of Denisovan SNVs for all populations except Papuans (Sankararaman et al. 2016), which is consistent with the majority of shared variants in humans being Neanderthal in origin. Papuans and other Melanesian populations have much more Denisovan ancestry than most other populations (Meyer et al. 2012) and may have had gene flow from a different Denisovan population than the one encountered by the ancestors of Eurasians (Browning et al. 2018), and thus it is not surprising that they have an exclusive set of archaic alleles for the *CYP450* genes, many of which are Denisovan in origin.

Interestingly, of the 23 potentially damaging variants we found in the eleven *CYP450* genes, 7 SNVs were found uniquely in the archaic individuals. Neanderthal and Denisovan genome elements in modern humans are the targets of natural selection; in particular, most of the depletion of archaic ancestry has been predicted to be the result of purifying selection against weakly deleterious alleles (Harris and Nielsen 2016; Juric et al. 2016). After archaic and modern human admixture, increased pressure from purifying selection would remove archaic deleterious alleles, as they would result in marked health and fitness detriments for modern humans. The strength of purifying selection against these deleterious alleles has been estimated to be extremely high, and likely removed a significant portion of Neanderthal ancestry from the modern human gene pool in just a few generations (Harris and Nielsen 2016; Juric et al. 2016; Petr et al. 2019).

Limitations

A potential limitation of this study is the nature of ancient DNA; degraded DNA and chemical damage — in particular for transition point mutations (cytosine (C) → thymine (T)) — result in low read depth and sequencing errors for ancient genomes. In order to distinguish true polymorphism at heterozygous sites from sequencing errors and archaic DNA (aDNA) damage, we excluded base pairs that are represented in only a single read for each position in the BAM read alignment. Because DNA damage and sequencing error is distributed randomly, we expect false polymorphism to only appear once at each position, whereas any polymorphism represented in multiple reads is statistically likely to represent true heterozygosity. To further account for false positive variant calls, we filtered calls for allelic bias, removing any heterozygous calls that had an allelic read ratio less than 0.2 or greater than 0.8. The nature of ancient genomes resulted in low read depth for some star alleles that were called (Figure 3), but given this limitation, we also manually checked the sequencing reads for all diplotypes reported (Supplemental Table 4).

As a result of our novel filtering method, the archaic variant call files we generated contain a larger number of heterozygous sites compared to the variant call files that are usually used to represent the archaic genomes (Meyer et al. 2012; Prüfer et al. 2014; Prüfer et al. 2017; Mafessoni et al. 2020). The method used to make the previously

generated variant call files likely used a more stringent method of filtering, resulting in more sites being called as homozygous. To ensure that our results were not biased by our variant calling method, we repeated all analyses (including calling star alleles, identifying archaic SNVs in modern humans, and comparing human and archaic haplotypes) using the original variant call files, and found that while the number of identified SNVs is smaller, the patterns identified (such as the divergent *CYP2B6* haplotype in the Vindija and allele sharing between archaic and modern humans) remain.

Other limitations were the required phasing of the archaic *CYP450* sequences with the program Beagle, which uses a human reference panel based on the 1000 Genomes panel for haplotype assessment, and the calling of the final diplotype with current known human star alleles. While this is the best reference panel available, it is uncertain if phasing based on modern humans reflects the true arrangement of alleles along chromosomes in archaic individuals. This limitation is somewhat ameliorated by finding homozygous allele calls for most *CYP450* genes in the archaic individuals (Supplemental Table 3), which is consistent with the much lower heterozygosity of Neanderthals and Denisovans along the entire genome (Meyer et al. 2012; Prüfer et al. 2014; Prüfer et al. 2017; Mafessoni et al. 2020), and identifying that many of the heterozygous star allele calls were composed of only a single nucleotide variant, effectively deeming the phasing non-impactful. The final diplotype calls also did not consider any of the novel potentially deleterious SNVs but will require further functional validation of these variants.

Lastly, a further limitation to determining if archaic *CYP450* alleles were inherited by modern humans through introgression is incomplete lineage sorting (ILS), in which modern humans, Neanderthals and Denisovans all share *CYP450* alleles ancestrally, as could be the case for *CYP2A6*12*. However, we expect haplotypes shared through ILS to accumulate neutral mutations over time, slowly driving divergence between haplotypes. We expect true archaically introgressed alleles to be less divergent, as is the case in *CYP2J2*.

Conclusions

Understanding the impact of archaic variation on modern human health is still in its early stages. While a few genes have been identified for which the archaic alleles play an important role in human adaptation to their environment, such as *EPAS1*, the vast contribution of archaic ancestry to human health remains unknown. Our results suggest that interactions between modern and archaic humans may have resulted in the introduction of novel *CYP450* variants into modern human populations, helping them adapt to novel environments as they expanded out of Africa. Important insights will continue to emerge from the careful inspection of pharmacologically relevant and highly studied genes such as the *CYP450* genes investigated here.

Materials and Methods

Samples

We investigate population-specific *CYP450* gene variation by combining data from the 1000 Genomes Project (1000 Genomes Project Consortium et al. 2015), the Neanderthal genome project (Prüfer et al. 2014), the Chagyrskaya Neanderthal genome project (Mafessoni et al. 2020), and the Denisovan Genome Project (Meyer et al. 2012). We extracted coding variation from four archaic human genomes; pertaining to a single Denisovan individual from the Denisova cave in the Altai Mountains (~21X coverage), a Neanderthal individual from Croatia (~42X coverage), and two Neanderthal individuals from the Altai Mountains: one from the Denisova cave (~30X coverage), and one from the Chagyrskaya cave (~28x coverage). These individuals are estimated to be at least 50,000 years old.

Variant Calling

Sequencing data for the three Neanderthal individuals and one Denisovan individual are publicly available, and processing of the BAM files has been outlined in previous literature (Meyer et al. 2012; Prüfer et al. 2014; Prüfer et al. 2017; Mafessoni et al. 2020). Sample BAM files were separated by chromosome for easier downstream processing using samtools (version 1.10 with hts lib 1.10) 'view' function (Li et al. 2009). Variant calling was performed for each *CYP450* gene of interest, using pypgx (version 0.1.37) 'bam2vcf' which implements the Genome Analysis Toolkit (GATK, version 4.1.9.0) 'HaplotypeCaller' function (options -emit-ref-confidence GVCF; -minimum-mapping-quality 10) followed by GATK 'GenotypeGVCF' to merge individual samples (Poplin et al. 2018). Variants were called against the HumanG1Kv37 hg19 reference assembly. GATK 'VariantFiltration' was additionally utilized to annotate variants with a quality score 50. Variants from modern human individuals from the 1000 Genomes Project (1000 Genomes Project Consortium et al. 2015) included for further analysis were called using the same procedure.

For individual SNV analysis, variant call format (VCF) files were further filtered by the PHRED-scaled quality score (QUAL) 40 and gene regions were determined as the RefSeq start and end coordinates of the gene with 2000 base-pairs upstream to account for the promoter region using bcftools (version 1.10.2 with hts lib 1.10.2, (Li et al. 2009; O'Leary et al. 2016). Read depth for each variant was reported from the GDF files generated from pypgx bam2gdf (described below). Variants were additionally filtered for allele balance (AB) bias and polymorphic sites that had a BAM read depth ≤ 1 for one allele were removed to adjust for the random effects of aDNA damage being called as true variation. AB was calculated as the ratio of minor allelic read depth to major allelic read depth, and variant calls were filtered with an AB > 0.2.

Variant annotation was conducted through ANNOVAR (-protocol refGeneWithVer, knownGene, avsnp150, ljb26_pp2hvar) which utilizes dbSNP150, RefSeq, and the UCSC genome browser to identify known variants and the gene locations that variants occur at; the occurrence of novel SNVs were additionally confirmed with Gnomad (v2.1.1, (Sherry et al. 2001; Kent et al. 2002; Wang et al. 2010; O'Leary et al. 2016; Karczewski et

al. 2021). Variant locations were identified with ANNOVAR, which annotates gene location based on RefSeq and UCSC genome browser genomic databases. Variants are defined as exonic, splicing, ncRNA, 3' and 5'UTR, intronic, and in the promoter region. If a location varied between the RefSeq and UCSC gene annotation, the called location is determined by an ordering precedence: exonic/splicing > ncRNA > 3'/5'UTR > intronic > promoter region.

Potentially pathogenic variants were identified using three algorithms to rank variant deleteriousness: Combined Annotation Dependent Depletion (CADD v1.6), Sorting Intolerant From Tolerant (SIFT 4G v2.0.0), and Polymorphism Phenotyping (PolyPhen v2, (Adzhubei et al. 2010; Vaser et al. 2016; Rentzsch et al. 2019). Variants were considered as deleterious if the PHRED-normalized CADD score was ≥ 20, a variant was identified to be deleterious by the SIFT prediction, or a variant was identified as probably/possibly damaging by PolyPhen2 prediction.

Comparisons of pairwise distance between Neanderthal individuals was conducted by calculating the pairwise difference across windows of two Neanderthal genomes. In our case, we used the Chagyrskaya and Vindija VCFs provided by their original publications (Prüfer et al. 2017; Mafessoni et al. 2020) and chose a window size corresponding to the *CYP2B6* gene length (29.1-kb). As the archaic genomes are unphased, we could not separate the two haplotypes for each individual, so we compared genotypes rather than haplotypes. To calculate pairwise distance, we added +1 to our total number of sites if the two genotypes were homozygous for the opposite allele (i.e., 0/0 vs. 1/1), and added 1/2 if one of the genotypes was heterozygous, then divided the total number of different sites by the length of the window.

Star allele calling and metabolizer phenotype prediction

CYP450 diplotypes and metabolizer phenotypes were determined using Stargazer, a bioinformatics tool for identifying star alleles of pharmacogenes by detecting SNVs, indels, and SVs from next-generation sequencing data (Lee, Wheeler, Patterson, et al. 2019; Lee, Wheeler, Thummel, et al. 2019). Briefly, small nucleotide variants (SNVs and indels) observed from each *CYP* gene were input into Stargazer, which utilizes the statistical phasing software BEAGLE (Browning and Browning 2007) to phase the variants into haplotypes. Phased variants were then matched to star alleles for each of the two haplotypes from an individual. Of note, diplotype calls were considered "non-determinant" if any of the variants used to define a star allele in Stargazer had a read depth ≤ 1. In addition to small nucleotide variants, SV calls were made by Stargazer using per-base depth of coverage data computed with the pypgx 'bam2gdf' command (https://github.com/sbslee/pypgx). Three control genes were used to normalize read depth and estimate accurate gene copy number: vitamin D receptor (*VDR)*, epidermal growth factor receptor (*EGFR)*, and ryanodine receptor 1 (*RYR1)*. Final calls for star alleles were determined as the call that was consistent in the highest number of control genes. SV calls were additionally confirmed with visual inspection of read depth profiles for the three control genes. Finally, phenotype prediction was conducted in Stargazer by converting the

called diplotype into an activity score, which is then used to predict the gene phenotype.

Identifying archaic alleles in modern populations

To identify the presence of archaic *CYP450* variants in modern human populations, we calculated the allele frequency of the archaic variants described above in modern humans. We used modern human genomes from all populations sequenced for the 1000 Genomes Project (1000 Genomes Project Consortium et al. 2015), as well as the Papuans sequenced as part of the Simons Genome Diversity Project (Mallick et al. 2016). We examined all SNVs that were shared between archaic and modern humans, but also focused on a subset of SNVs that were more likely to be introgressed, as they were found outside of Africa but not in African populations. To exclude SNVs that were ancestrally shared between archaic humans and modern humans, archaic SNVs had to have a frequency in African populations of less than 1 percent and be present in at least one non-African population with an allele frequency greater than 1 percent. For all results described in the text, we used the archaic VCF generated using the methods in this paper for comparison with modern humans. For comparison, we repeated these analyses with the published VCFs that are associated with the original archaic genome sequencing studies (Meyer et al. 2012; Prüfer et al. 2014; Prüfer et al. 2017; Mafessoni et al. 2020).

To explore SNV sharing between the *CYP2B6* haplotypes, we divided the archaic and modern humans into four groups: the Vindija genome (containing the divergent haplotype), the other three archaic genomes, the eleven Africans with the divergent haplotypes, and the rest of the modern Africans without the divergent haplotypes. Each group was scored for the presence or absence of a given SNV, and the sharing of these SNVs was summarized in a Venn diagram (Supplemental Figure 4).

We used Haplostrips (Marnetto and Huerta-Sánchez 2017) to calculate the distance between haplotypes of *all CYP450 genes* in modern humans in the 1000 Genomes Panel relative to a reference haplotype. Distances in Haplostrips are Manhattan distances, simply the number of SNVs with different alleles between the two sequences. Haplotypes are re-ordered by decreasing similarity with the archaic reference haplotype. Here, the Haplostrip is polarized to a Vindija Neanderthal reference haplotype (a consensus of the two archaic chromosomes) or a Denisovan reference haplotype, and each subsequent haplotype is ordered by genetic similarity, from most related to least related. For this analysis, we looked at haplotypes composed of each *CYP450* gene at the gene coordinates, plus 5000 base pairs downstream and upstream to capture more linked neutral variation, found in the VCF files from the 1000 Genomes Project and the various archaic genome projects. We used the genetic distances calculated by Haplostrips to rank the proximity of modern human haplotypes to archaic haplotypes.

Incomplete Lineage Sorting Calculation

Incomplete Lineage Sorting (ILS) was assessed using a previously published equation (Huerta-Sánchez et al. 2014). Regional recombination rates were determined using HapMap (Myers et al. 2005). For these calculations,

we used a generation time of 29 years (Langergraber et al. 2012; Zeberg and Pääbo 2021) and a branch length of 550,000 years with a 50,000 year timeframe of Neanderthal-Human interbreeding (Prüfer et al. 2014; Zeberg and Pääbo 2021).

Data Visualization

All data visualization was conducted in R Studio Suite (R version 4.0.2, https://www.r-project.org/) with the use of the CRAN library *ggplot2* for graphical visualization of variant frequency, star allele diplotype calls, deleterious variant occurrence, and haplotype distance visualization (R Core Team 2008; Wickham 2009). The R package *venneuler* was used to create the Venn diagram in Supplemental Figure 4 (https://CRAN.R-project.org/package=venneuler).

Acknowledgements: We would like to thank George Perry and Omer Gokcumen for comments on early versions of our manuscript, and anonymous reviewers for their feedback.

Funding: THW and KGC are supported by the NHGRI grant to KGC R35HG011319. KEW is supported by the NIH grant to EHS 1R35GM128946-01. EHS was also supported by an Alfred P. Sloan Fellowship.

Data Availability: The scripts created to analyze and generate this data can be found at https://github.com/kelsey-witt/archaic_pgx and https://github.com/the-claw-lab/aDNA_PGx_2021.

Table 1: A summary of archaic variants identified in modern human populations for *CYP450* genes. For each population, the number of SNVs found in each population is listed, along with the median allele frequencies for all SNVs identified in the populations in the region. "East Asians", "Europeans", "admixed Americans", and "South Asians" refer to the populations from the 1000 Genomes dataset, while "Papuans" refers to the Simons Genome Diversity Project Papuan population. Variants of interest have been highlighted with a section sign (§).

	Number of Archaic Variants		Number of SNVs Identified in Each Population				
***CYP450* Gene**	**Total Identified**	**Identified in Modern Human Populations**	**East Asians**	**Europeans**	**Admixed Americans**	**South Asians**	**Papuans**
CYP1A2	46	2	2§	2§	1§	1§	0
CYP2A6	69	9	0	6	6	6	3§
CYP2B6	426	20	4	14	13	14	5§
CYP2C8	165	15	15	15§	15§	15	0
CYP2C9	162	24	24	24§	24§	24§	1§
CYP2C19	332	64	52	59	3	58	5§
CYP2D6	44	3	0	0	0	0	3§
CYP2E1	84	1	0	0	0	0	1§
CYP2J2	94	11	9	8§	9§	9	1§
CYP3A4	93	6	2	1	0	6	6§
CYP3A5	108	0	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*

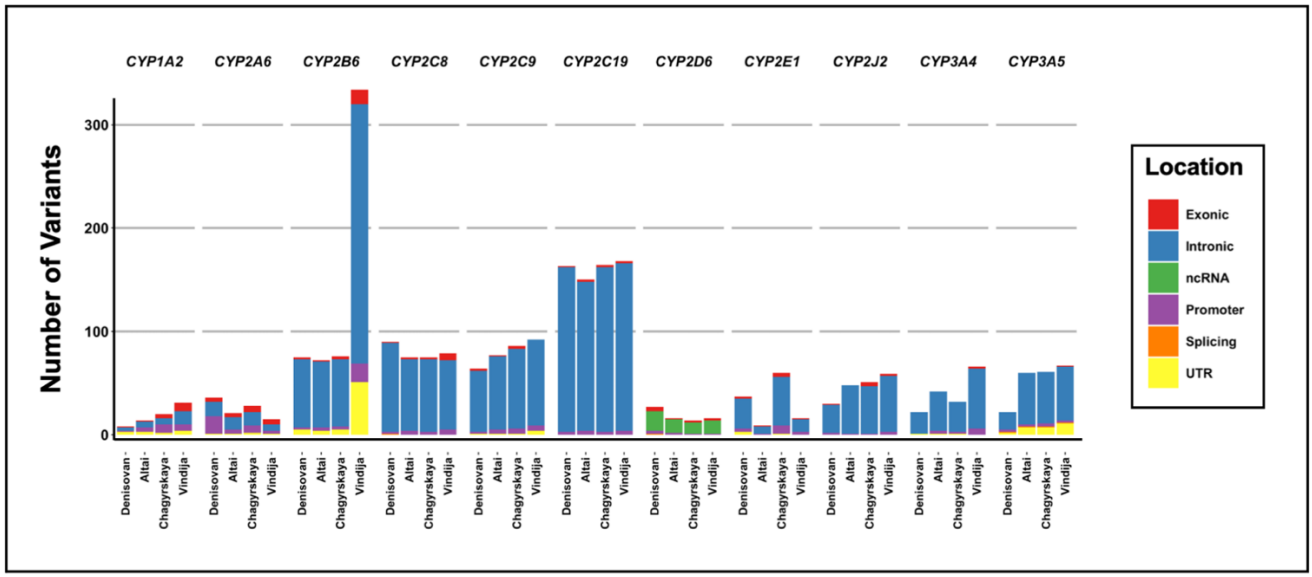

Figure 1: *CYP450* genetic variation and locations in archaic individuals. The total number and relative gene location of SNVs identified in each archaic individual displayed as a stacked bar plot. Variants were filtered by the following parameters: PHRED-scaled quality score ≥ 40 and heterozygous variant calls were removed if there was a single BAM read for one allele or the allele balance (AB) ≤ 0.2. Gene regions were determined by the start and end coordinates of the gene with an additional 2000 base-pairs (bp) for the promoter region, and gene location was identified using ANNOVAR. 3' & 5' untranslated regions are grouped together (UTR), and the promoter region (Promoter) has been characterized as any upstream variants that are within 2000 bp of the gene.

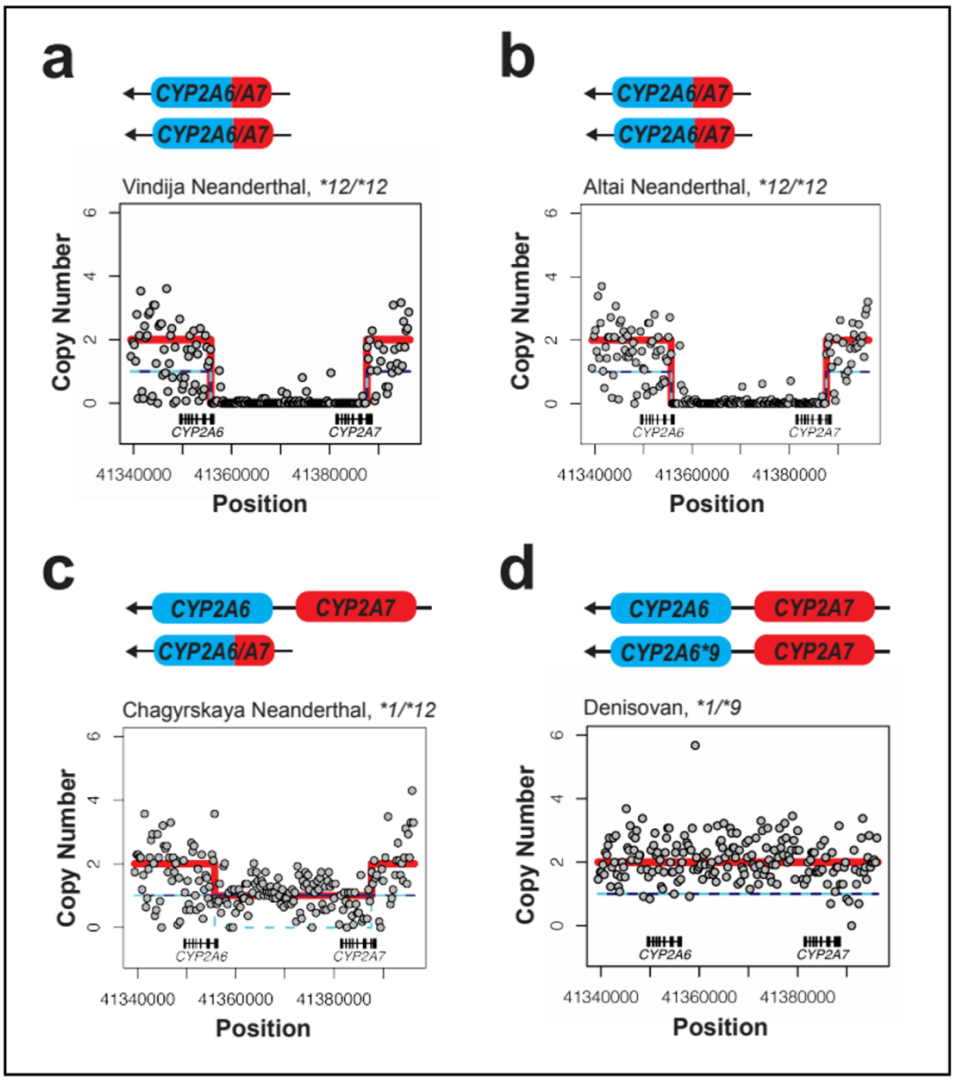

Figure 2: Structural variants detected in the *CYP2A6* and *CYP2A7* genes using Stargazer.

Copy number variation plots demonstrating the structural variation across the Neanderthal and Denisovan individuals for *CYP2A6* and *CYP2A7*. Copy number plots are displayed as Stargazer's output for each sample, with the gray dots indicating the copy number estimate based on read data (displayed as the copy number normalized to the *VDR* control gene region). The dark blue and cyan lines indicate Stargazer's copy number assessment for each haplotype, while the thicker red line indicates the overall copy number profile from each gene. Above each plot, there is a schematic representation of the predicted structure of *CYP2A6* (blue) and *CYP2A7* (red) for each individual, with an arrow indicating the direction of transcription. The hg19 genetic coordinates are presented on the *x axis* with the gene regions indicated directly above. (a) The Vindija Neanderthal and (b) Altai Neanderthal both present a *CYP2A6*12/*12* diplotype, characterized by a partial deletion hybridization event in both gene copies, (c) the Chagyrskaya Neanderthal demonstrates a **1/*12* diplotype, and (d) the Denisovan individual presents no structural variation, displaying a **1/*9* diplotype, which is characterized by an upstream SNV.

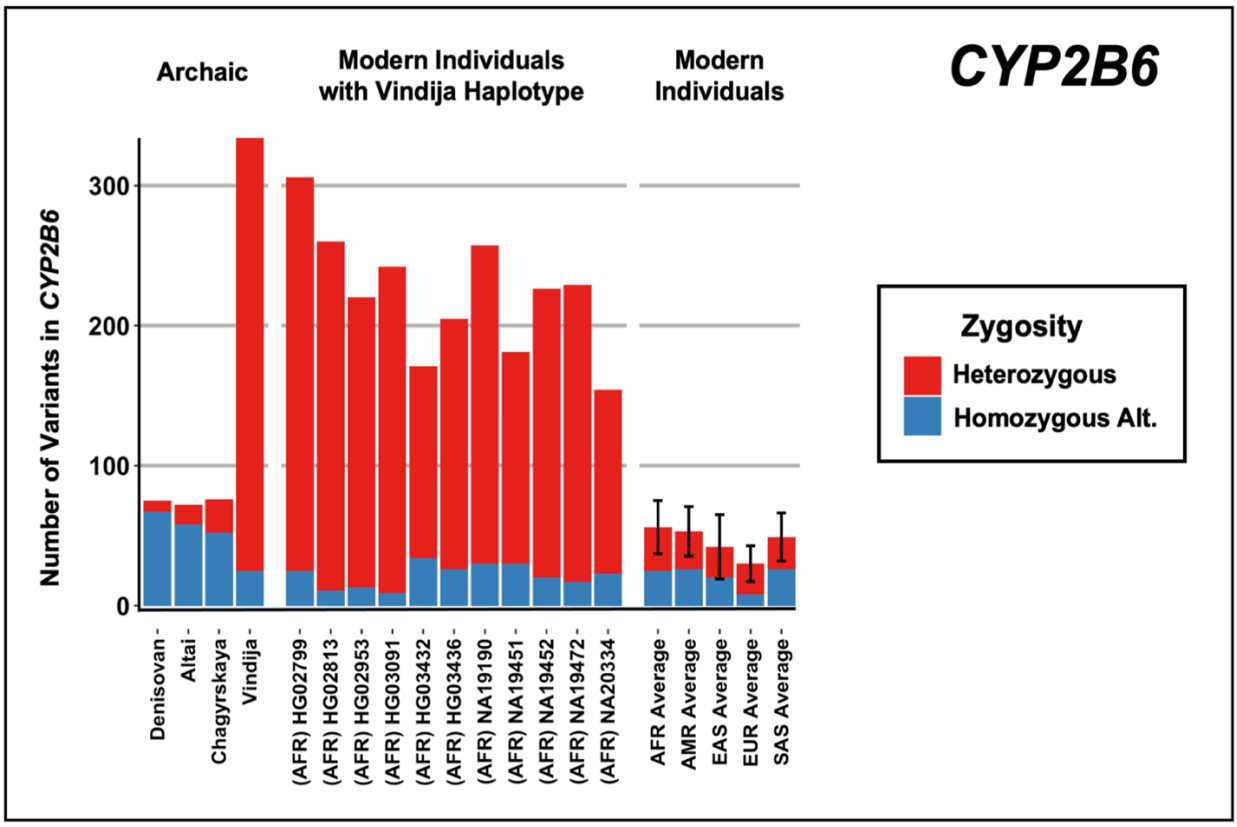

Figure 3: *CYP2B6* genetic variation in archaic individuals compared to modern humans from the 1000 Genomes Project. The total number SNVs identified in each individual displayed as a stacked bar plot. Bar plots have been faceted by the archaic individuals, the modern individuals identified to share a large number of variants with the Vindija *CYP2B6* divergent haplotype, and the average number of variants for 10 modern individuals from each super population of the 1000 Genomes Project. The height of each stacked bar plot represents the total number of variants for each individual and the number of variants for the modern individuals have been reported as the mean with error bars representing the standard deviation of all variants. Variants were filtered by the following parameters: PHRED-scaled quality score ≥ 40 and heterozygous variant calls were removed if there was a single BAM read for one allele or the allele balance (AB) ≤ 0.2. Gene regions were determined by the start and end coordinates of the gene with an additional 2000 base-pairs (bp) for the promoter region, and zygosity is defined as heterozygous if the variant site has one alternative allele and homozygous alternative (Homozygous Alt.) if the variant site has two alternative alleles at that position.

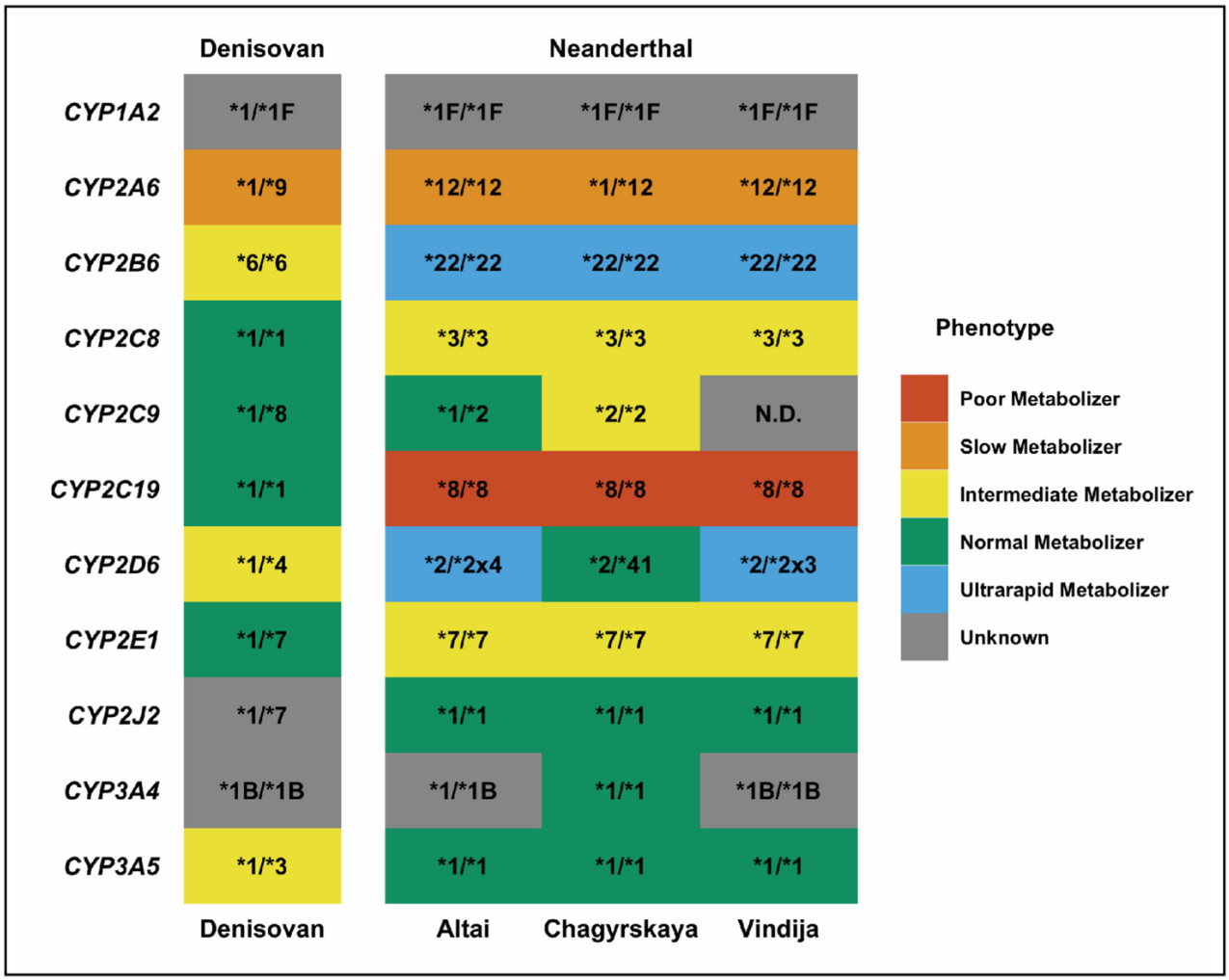

Figure 4: Phased diplotypes and predicted metabolizer phenotypes identified in archaic individuals. Heatmap showing the primary star allele diplotypes for each gene from the three Neanderthal and one Denisovan individuals. Each diplotype was determined as the two phased haplotypes from Stargazer, which are detected with the variant and read data from each gene. The fill of each tile represents the predicted phenotype determined by the combined activity score associated with the presented diplotype as identified in modern humans. Unknown phenotypes are presented in grey and indicate that the functional effect of the diplotype has not been determined. A diplotype was considered non-determinant (N.D.) if there was a lack of read coverage at specific star variant locations.

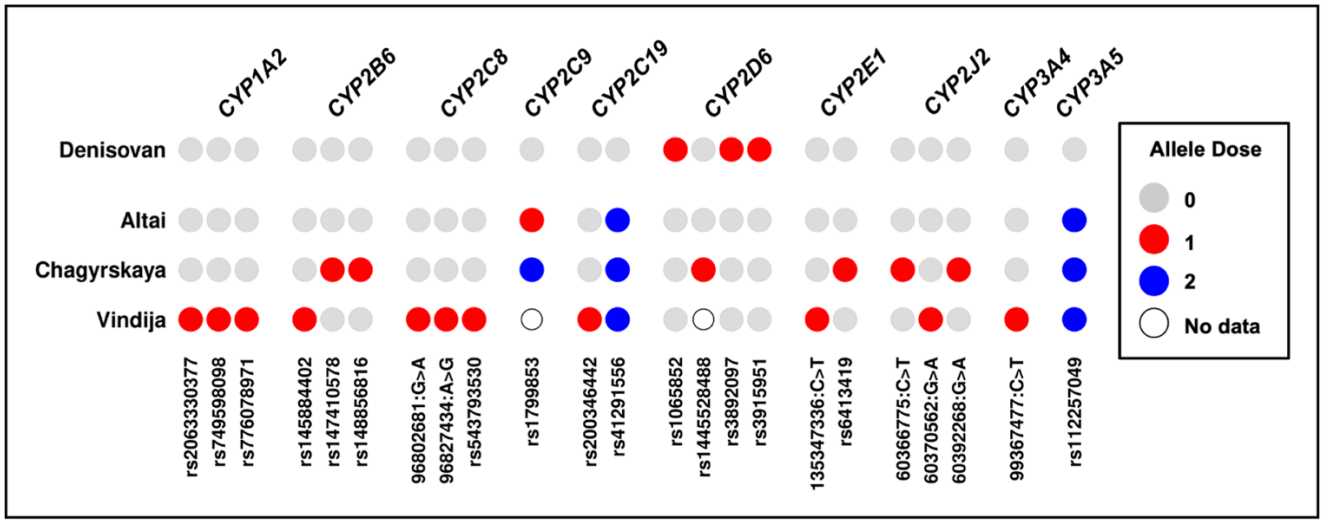

Figure 5: Potentially damaging or deleterious variants in archaic individuals. Heat map displaying the occurrence of potentially damaging or deleterious variants in each sample for the selected *CYP450* genes. Variants are identified with rsID, if no rsID number was identified, variants are considered potentially novel and the hg19 position is displayed along with mutation observed (position: reference allele > alternative allele). Variants were selected using three algorithms to assess the potential to be damaging or deleterious to the gene: CADD ≥ 20, SIFT prediction of deleterious, or PolyPhen prediction of possibly/probably damaging. Allele dosage is categorized as the number of alternate alleles at that position. Circles that are white with a black outline indicate that no call was made at the given position for that individual. An allele dose of 1 indicates an SNV is heterozygous for the reference and alternate allele, and an allele dose of 2 indicates an SNV is homozygous for the alternate allele. By these criteria, no variants were identified for *CYP2A6.*

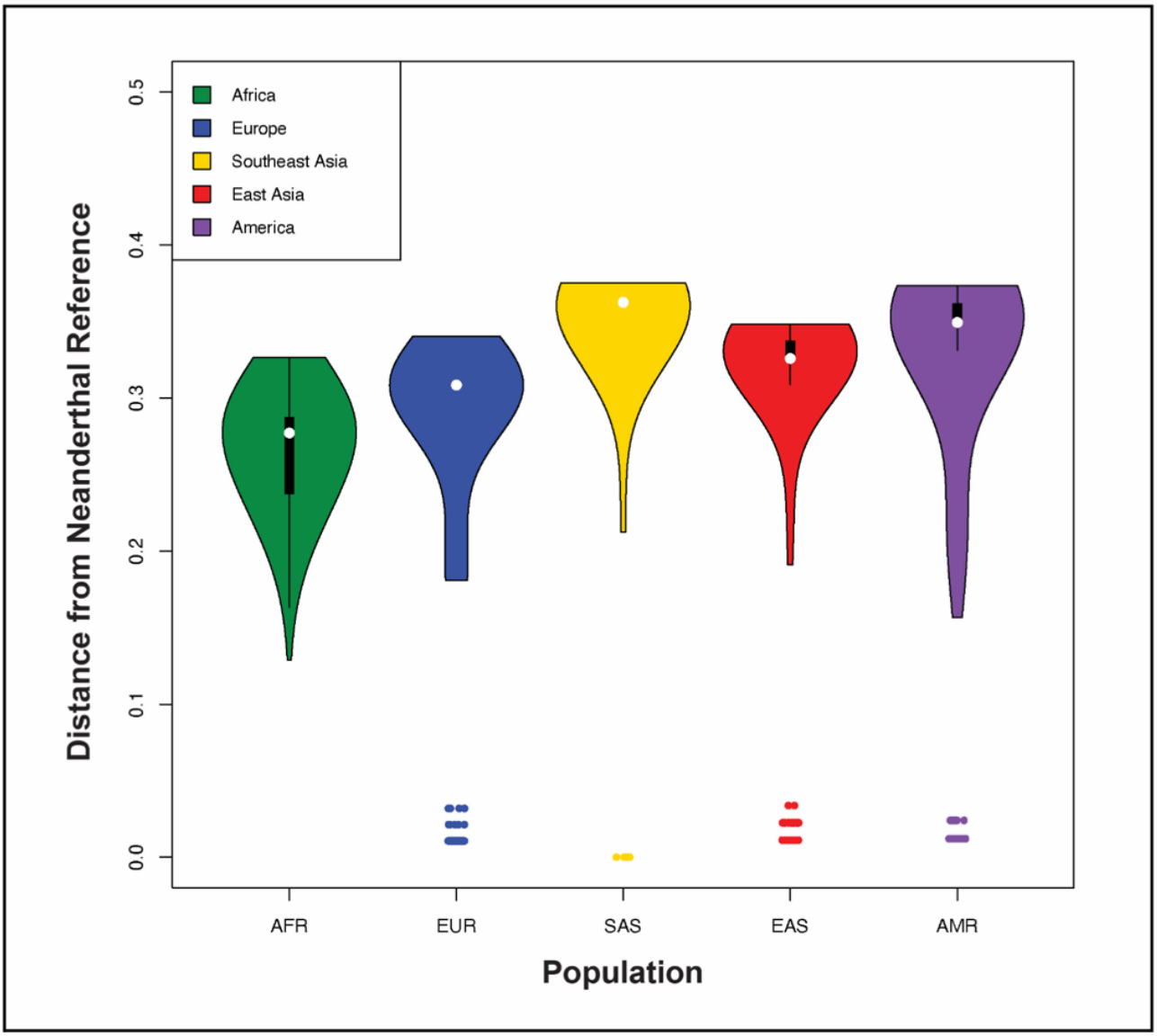

Figure 6: Haplostrips violin plot of *CYP2J2*. Haplotype distances at *CYP2J2* (hg19 1:60355979-60395470) in modern human populations relative to a Neanderthal reference haplotype (Vindija Neanderthal). The distribution of the kernel density on the violin plots represents the abundance of individual haplotypes ordered by distance to the reference. The vast majority of human-like haplotypes sit at the same distance furthest from the archaic reference, giving the violin plots a "flat top". Outlier data points near lower distances represent modern human haplotypes carrying the Neanderthal-like *CYP2J2* in all populations except African populations.

References

1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, et al. 2015. A global reference for human genetic variation. Nature. 526(7571):68–74. doi:10.1038/nature15393. http://www.ncbi.nlm.nih.gov/pubmed/26432245%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4750478.

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. Nat Methods. 7(4):248–249. doi:10.1038/nmeth0410-248. http://dx.doi.org/10.1038/nmeth0410-248.

Ariyoshi N, Miyazaki M, Toide K, Sawamura Yi, Kamataki T. 2001. A single nucleotide polymorphism of CYP2b6 found in Japanese enhances catalytic activity by autoactivation. Biochem Biophys Res Commun. 281(5):1256–1260. doi:10.1006/bbrc.2001.4524. http://dx.doi.org/10.1006/bbrc.2001.4524.

Benowitz NL, Swan GE, Jacob P 3rd, Lessov-Schlaggar CN, Tyndale RF. 2006. CYP2A6 genotype and the metabolism and disposition kinetics of nicotine. Clin Pharmacol Ther. 80(5):457–467. doi:10.1016/j.clpt.2006.08.011. http://dx.doi.org/10.1016/j.clpt.2006.08.011.

Bergström A, Stringer C, Hajdinjak M, Scerri EML, Skoglund P. 2021. Origins of modern human ancestry. Nature. 590(7845):229–237. doi:10.1038/s41586-021-03244-5. http://dx.doi.org/10.1038/s41586-021-03244-5.

Bièche I, Narjoz C, Asselah T, Vacher S, Marcellin P, Lidereau R, Beaune P, de Waziers I. 2007. Reverse transcriptase-PCR quantification of mRNA levels from cytochrome (CYP)1, CYP2 and CYP3 families in 22 different human tissues. Pharmacogenet Genomics. 17(9):731–742. doi:10.1097/FPC.0b013e32810f2e58. http://dx.doi.org/10.1097/FPC.0b013e32810f2e58.

Browning SR, Browning BL. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am J Hum Genet. 81(5):1084–1097. doi:10.1086/521987. http://dx.doi.org/10.1086/521987.

Browning SR, Browning BL, Zhou Y, Tucci S, Akey JM. 2018. Analysis of Human Sequence Data Reveals Two Pulses of Archaic Denisovan Admixture. Cell. 173(1):1–9. doi:10.1016/j.cell.2018.02.031. https://doi.org/10.1016/j.cell.2018.02.031.

Castellano S, Parra G, Sánchez-Quinto FA, Racimo F, Kuhlwilm M, Kircher M, Sawyer S, Fu Q, Heinze A, Nickel B, et al. 2014. Patterns of coding variation in the complete exomes of three Neandertals. Proc Natl Acad Sci U S A. 111(18):6666–6671. doi:10.1073/pnas.1405138111. http://dx.doi.org/10.1073/pnas.1405138111.

Chen L, Li JH, Kaur V, Muhammad A, Fernandez M, Hudson MS, Goldfine AB, Florez JC. 2020. The presence of two reduced function variants in CYP2C9 influences the acute response to glipizide. Diabet Med. 37(12):2124–2130. doi:10.1111/dme.14176. http://dx.doi.org/10.1111/dme.14176.

Chen L, Wolf AB, Fu W, Li L, Akey JM. 2020. Identifying and Interpreting Apparent Neanderthal Ancestry in African Individuals. Cell. 180(4):677–687.e16. doi:10.1016/j.cell.2020.01.012. http://dx.doi.org/10.1016/j.cell.2020.01.012.

Dannemann M, Andrés AM, Kelso J. 2016. Introgression of Neandertal- and Denisovan-like Haplotypes Contributes to Adaptive Variation in Human Toll-like Receptors. The American Journal of Human Genetics. 98(2):399. doi:10.1016/j.ajhg.2016.01.012. http://dx.doi.org/10.1016/j.ajhg.2016.01.012.

Desta Z, El-Boraie A, Gong L, Somogyi AA, Lauschke VM, Dandara C, Klein K, Miller NA, Klein TE, Tyndale RF, et al. 2021. PharmVar GeneFocus: CYP2B6. Clin Pharmacol Ther. 110(1):82–97. doi:10.1002/cpt.2166. http://dx.doi.org/10.1002/cpt.2166.

Desta Z, Moaddel R, Ogburn ET, Xu C, Ramamoorthy A, Venkata SLV, Sanghvi M, Goldberg ME, Torjman MC, Wainer IW. 2012. Stereoselective and regiospecific hydroxylation of ketamine and norketamine. Xenobiotica. 42(11):1076–1087. doi:10.3109/00498254.2012.685777. http://dx.doi.org/10.3109/00498254.2012.685777.

Desta Z, Saussele T, Ward B, Blievernicht J, Li L, Klein K, Flockhart DA, Zanger UM. 2007. Impact of CYP2B6 polymorphism on hepatic efavirenz metabolism in vitro. Pharmacogenomics. 8(6):547–558. doi:10.2217/14622416.8.6.547. http://dx.doi.org/10.2217/14622416.8.6.547.

Dreisbach AW, Japa S, Sigel A, Parenti MB, Hess AE, Srinouanprachanh SL, Rettie AE, Kim H, Farin FM, Hamm LL, et al. 2005. The Prevalence of CYP2C8, 2C9, 2J2, and soluble epoxide hydrolase polymorphisms in African Americans with hypertension. Am J Hypertens. 18(10):1276–1281. doi:10.1016/j.amjhyper.2005.04.019. http://dx.doi.org/10.1016/j.amjhyper.2005.04.019.

Durvasula A, Sankararaman S. 2020. Recovering signals of ghost archaic introgression in African populations. Science Advances. 6(7):eaax5097. doi:10.1126/sciadv.aax5097. [accessed 2021 Apr 14]. https://advances.sciencemag.org/content/6/7/eaax5097.

Evans WE, Relling MV. 1999. Pharmacogenomics: translating functional genomics into rational therapeutics. Science. 286(5439):487–491. doi:10.1126/science.286.5439.487. http://dx.doi.org/10.1126/science.286.5439.487.

Fan S, Hansen MEB, Lo Y, Tishkoff SA. 2016. Going global by adapting local: A review of recent human adaptation. Science. 354(6308):54–59. doi:10.1126/science.aaf5098. http://dx.doi.org/10.1126/science.aaf5098.

Faucette SR, Hawke RL, Lecluyse EL, Shord SS, Yan B, Laethem RM, Lindley CM. 2000. Validation of bupropion hydroxylation as a selective marker of human cytochrome P450 2B6 catalytic activity. Drug Metab Dispos. 28(10):1222–1230. https://www.ncbi.nlm.nih.gov/pubmed/10997944.

Fuselli S. 2019. Beyond drugs: the evolution of genes involved in human response to medications. Proc Biol Sci. 286(1913):20191716. doi:10.1098/rspb.2019.1716. http://dx.doi.org/10.1098/rspb.2019.1716.

Fuselli S, de Filippo C, Mona S, Sistonen J, Fariselli P, Destro-Bisol G, Barbujani G, Bertorelle G, Sajantila A. 2010. Evolution of detoxifying systems: the role of environment and population history in shaping genetic diversity at human CYP2D6 locus. Pharmacogenet Genomics. 20(8):485–499. doi:10.1097/FPC.0b013e32833bba25. http://dx.doi.org/10.1097/FPC.0b013e32833bba25.

Gaedigk A, Ingelman-Sundberg M, Miller NA, Steven Leeder J, Whirl-Carrillo M, Klein TE, the PharmVar Steering Committee. 2018. The Pharmacogene Variation (PharmVar) Consortium: Incorporation of the Human Cytochrome P450 (CYP) Allele Nomenclature Database. Clinical Pharmacology & Therapeutics. 103(3):399–401. doi:10.1002/cpt.910. http://dx.doi.org/10.1002/cpt.910.

Giancarlo GM, Venkatakrishnan K, Granda BW, von Moltke LL, Greenblatt DJ. 2001. Relative contributions of CYP2C9 and 2C19 to phenytoin 4-hydroxylation in vitro: inhibition by sulfaphenazole, omeprazole, and ticlopidine. Eur J Clin Pharmacol. 57(1):31–36. doi:10.1007/s002280100268. http://dx.doi.org/10.1007/s002280100268.

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH-Y, et al. 2010. A Draft Sequence of the Neandertal Genome. Science. 328(5979):710–722. doi:10.1126/science.1188021. http://www.sciencemag.org/content/328/5979/710.abstract.

Haberl M, Anwald B, Klein K, Weil R, Fuss C, Gepdiremen A, Zanger UM, Meyer UA, Wojnowski L. 2005. Three haplotypes associated with CYP2A6 phenotypes in Caucasians. Pharmacogenet Genomics. 15(9):609–624. doi:10.1097/01.fpc.0000171517.22258.f1. http://dx.doi.org/10.1097/01.fpc.0000171517.22258.f1.

Hagen EH, Roulette CJ, Sullivan RJ. 2013. Explaining human recreational use of “pesticides”: the neurotoxin regulation model of substance use vs. the hijack model and implications for age and sex differences in drug consumption. Front Psychiatry. 4:142. https://www.frontiersin.org/articles/10.3389/fpsyt.2013.00142/full.

Hagen EH, Sullivan RJ, Schmidt R, Morris G, Kempter R, Hammerstein P. 2009. Ecology and neurobiology of toxin avoidance and the paradox of drug reward. Neuroscience. 160(1):69–84. doi:10.1016/j.neuroscience.2009.01.077. http://dx.doi.org/10.1016/j.neuroscience.2009.01.077.

Hamman MA, Thompson GA, Hall SD. 1997. Regioselective and stereoselective metabolism of ibuprofen by human cytochrome P450 2C. Biochem Pharmacol. 54(1):33–41. doi:10.1016/s0006-2952(97)00143-3. http://dx.doi.org/10.1016/s0006-2952(97)00143-3.

Harris K, Nielsen R. 2016. The Genetic Cost of Neanderthal Introgression. Genetics. 203(2):881–891. doi:10.1534/genetics.116.186890. http://dx.doi.org/10.1534/genetics.116.186890.

Hashizume T, Imaoka S, Mise M, Terauchi Y, Fujii T, Miyazaki H, Kamataki T, Funae Y. 2002. Involvement of CYP2J2 and CYP4F12 in the metabolism of ebastine in human intestinal microsomes. J Pharmacol Exp Ther. 300(1):298–304. doi:10.1124/jpet.300.1.298. http://dx.doi.org/10.1124/jpet.300.1.298.

Henderson LM, Claw KG, Woodahl EL, Robinson RF, Boyer BB, Burke W, Thummel KE. 2018. P450 Pharmacogenetics in Indigenous North American Populations. J Pers Med. 8(1). doi:10.3390/jpm8010009. http://dx.doi.org/10.3390/jpm8010009.

Hesse LM, Venkatakrishnan K, Court MH, von Moltke LL, Duan SX, Shader RI, Greenblatt DJ. 2000. CYP2B6 mediates the in vitro hydroxylation of bupropion: potential drug interactions with other antidepressants. Drug Metab Dispos. 28(10):1176–1183. https://www.ncbi.nlm.nih.gov/pubmed/10997936.

Hofmann MH, Blievernicht JK, Klein K, Saussele T, Schaeffeler E, Schwab M, Zanger UM. 2008. Aberrant splicing caused by single nucleotide polymorphism c.516G>T [Q172H], a marker of CYP2B6*6, is responsible for decreased expression and activity of CYP2B6 in liver. J Pharmacol Exp Ther. 325(1):284–292. doi:10.1124/jpet.107.133306. http://dx.doi.org/10.1124/jpet.107.133306.

Huang Z, Roy P, Waxman DJ. 2000. Role of human liver microsomal CYP3A4 and CYP2B6 in catalyzing N-dechloroethylation of cyclophosphamide and ifosfamide. Biochem Pharmacol. 59(8):961–972. doi:10.1016/s0006-2952(99)00410-4. http://dx.doi.org/10.1016/s0006-2952(99)00410-4.

Huerta-Sánchez E, Jin X, Asan, Bianba Z, Peter BM, Vinckenbosch N, Liang Y, Yi X, He M, Somel M, et al. 2014. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature. 512(7513):194–197. doi:10.1038/nature13408. http://dx.doi.org/10.1038/nature13408.

Hukkanen J, Jacob P 3rd, Benowitz NL. 2005. Metabolism and disposition kinetics of nicotine. Pharmacol Rev. 57(1):79–115. doi:10.1124/pr.57.1.3. http://dx.doi.org/10.1124/pr.57.1.3.

Juric I, Aeschbacher S, Coop G. 2016. The Strength of Selection against Neanderthal Introgression. PLoS Genet. 12(11):e1006340. doi:10.1371/journal.pgen.1006340. http://dx.doi.org/10.1371/journal.pgen.1006340.

Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, et al. 2021. Author Correction: The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 590(7846):E53. doi:10.1038/s41586-020-03174-8. http://dx.doi.org/10.1038/s41586-020-03174-8.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. Genome Res. 12(6):996–1006. doi:10.1101/gr.229102. http://dx.doi.org/10.1101/gr.229102.

Kiang TKL, Ho PC, Anari MR, Tong V, Abbott FS, Chang TKH. 2006. Contribution of CYP2C9, CYP2A6, and CYP2B6 to valproic acid metabolism in hepatic microsomes from individuals with the CYP2C9*1/*1 genotype. Toxicol Sci. 94(2):261–271. doi:10.1093/toxsci/kfl096. http://dx.doi.org/10.1093/toxsci/kfl096.

King LM, Gainer JV, David GL, Dai D, Goldstein JA, Brown NJ, Zeldin DC. 2005. Single nucleotide polymorphisms in the CYP2J2 and CYP2C8 genes and the risk of hypertension. Pharmacogenet Genomics. 15(1):7–13. doi:10.1097/01213011-200501000-00002. http://dx.doi.org/10.1097/01213011-200501000-00002.

King LM, Ma J, Srettabunjong S, Graves J, Bradbury JA, Li L, Spiecker M, Liao JK, Mohrenweiser H, Zeldin DC. 2002. Cloning of CYP2J2 gene and identification of functional polymorphisms. Mol Pharmacol. 61(4):840–852. doi:10.1124/mol.61.4.840. http://dx.doi.org/10.1124/mol.61.4.840.

Kirchheiner J, Klein C, Meineke I, Sasse J, Zanger UM, Mürdter TE, Roots I, Brockmöller J. 2003. Bupropion and 4-OH-bupropion pharmacokinetics in relation to genetic polymorphisms in CYP2B6. Pharmacogenetics. 13(10):619–626. doi:10.1097/00008571-200310000-00005. http://dx.doi.org/10.1097/00008571-200310000-00005.

Komatsu T, Yamazaki H, Shimada N, Nakajima M, Yokoi T. 2000. Roles of cytochromes P450 1A2, 2A6, and 2C8 in 5-fluorouracil formation from tegafur, an anticancer prodrug, in human liver microsomes. Drug Metab Dispos. 28(12):1457–1463. https://www.ncbi.nlm.nih.gov/pubmed/11095583.

Lafite P, André F, Zeldin DC, Dansette PM, Mansuy D. 2007. Unusual regioselectivity and active site topology of human cytochrome P450 2J2. Biochemistry. 46(36):10237–10247. doi:10.1021/bi700876a. http://dx.doi.org/10.1021/bi700876a.

Langergraber KE, Prüfer K, Rowney C, Boesch C, Crockford C, Fawcett K, Inoue E, Inoue-Muruyama M, Mitani JC, Muller MN, et al. 2012. Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. Proc Natl Acad Sci U S A. 109(39):15716–15721. doi:10.1073/pnas.1211740109. http://dx.doi.org/10.1073/pnas.1211740109.

Lang T, Klein K, Fischer J, Nüssler AK, Neuhaus P, Hofmann U, Eichelbaum M, Schwab M, Zanger UM. 2001. Extensive genetic polymorphism in the human CYP2B6 gene with impact on expression and function in human liver. Pharmacogenetics. 11(5):399–415. doi:10.1097/00008571-200107000-00004. http://dx.doi.org/10.1097/00008571-200107000-00004.

Lee CA, Jones JP 3rd, Katayama J, Kaspera R, Jiang Y, Freiwald S, Smith E, Walker GS, Totah RA. 2012. Identifying a selective substrate and inhibitor pair for the evaluation of CYP2J2 activity. Drug Metab Dispos. 40(5):943–951. doi:10.1124/dmd.111.043505. http://dx.doi.org/10.1124/dmd.111.043505.

Lee S-B, Wheeler MM, Patterson K, McGee S, Dalton R, Woodahl EL, Gaedigk A, Thummel KE, Nickerson DA. 2019. Stargazer: a software tool for calling star alleles from next-generation sequencing data using CYP2D6

as a model. Genet Med. 21(2):361–372. doi:10.1038/s41436-018-0054-0. http://dx.doi.org/10.1038/s41436-018-0054-0.

Lee S-B, Wheeler MM, Thummel KE, Nickerson DA. 2019. Calling Star Alleles With Stargazer in 28 Pharmacogenes With Whole Genome Sequences. Clin Pharmacol Ther. 106(6):1328–1337. doi:10.1002/cpt.1552. http://dx.doi.org/10.1002/cpt.1552.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 25(16):2078–2079. doi:10.1093/bioinformatics/btp352. http://dx.doi.org/10.1093/bioinformatics/btp352.

Lin Y-L, Pavlidis P, Karakoc E, Ajay J, Gokcumen O. 2015. The evolution and functional impact of human deletion variants shared with archaic hominin genomes. Mol Biol Evol. 32(4):1008–1019. doi:10.1093/molbev/msu405. http://dx.doi.org/10.1093/molbev/msu405.

MacDonald JR, Ziman R, Yuen RKC, Feuk L, Scherer SW. 2014. The Database of Genomic Variants: a curated collection of structural variation in the human genome. Nucleic Acids Res. 42(Database issue):D986–92. doi:10.1093/nar/gkt958. http://dx.doi.org/10.1093/nar/gkt958.

Mafessoni F, Grote S, de Filippo C, Slon V, Kolobova KA, Viola B, Markin SV, Chintalapati M, Peyrégne S, Skov L, et al. 2020. A high-coverage Neandertal genome from Chagyrskaya Cave. Proc Natl Acad Sci U S A. 117(26):15132–15136. doi:10.1073/pnas.2004944117. [accessed 2021 Apr 9]. https://www.pnas.org/content/117/26/15132.

Mallick S, Li H, Lipson M, Mathieson I, Gymrek M, Racimo F, Zhao M, Chennagiri N, Nordenfelt S, Tandon A, et al. 2016. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature. 538(7624):201–206. doi:10.1038/nature18964. http://dx.doi.org/10.1038/nature18964.

Marnetto D, Huerta-Sánchez E. 2017. Haplostrips : revealing population structure through haplotype visualization. Price S, editor. Methods Ecol Evol. 8(10):1389–1392. doi:10.1111/2041-210X.12747. https://onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.12747.

Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, Daly MJ, Bustamante CD, Kenny EE. 2017. Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. Am J Hum Genet. 100(4):635–649. doi:10.1016/j.ajhg.2017.03.004. http://dx.doi.org/10.1016/j.ajhg.2017.03.004.

Matsumoto S, Yamazoe Y. 2001. Involvement of multiple human cytochromes P450 in the liver microsomal metabolism of astemizole and a comparison with terfenadine. Br J Clin Pharmacol. 51(2):133–142. doi:10.1111/j.1365-2125.2001.01292.x. http://dx.doi.org/10.1111/j.1365-2125.2001.01292.x.

Mendez FL, Watkins JC, Hammer MF. 2012. A haplotype at STAT2 Introgressed from neanderthals and serves as a candidate of positive selection in Papua New Guinea. Am J Hum Genet. 91(2):265–274. doi:10.1016/j.ajhg.2012.06.015. http://dx.doi.org/10.1016/j.ajhg.2012.06.015.

Meyer M, Kircher M, Gansauge M-T, Li H, Racimo F, Mallick S, Schraiber JG, Jay F, Prüfer K, de Filippo C, et al. 2012. A high-coverage genome sequence from an archaic Denisovan individual. Science. 338(6104):222–226. doi:10.1126/science.1224344. http://dx.doi.org/10.1126/science.1224344.

Miles JS, McLaren AW, Forrester LM, Glancey MJ, Lang MA, Wolf CR. 1990. Identification of the human liver cytochrome P-450 responsible for coumarin 7-hydroxylase activity. Biochem J. 267(2):365–371. doi:10.1042/bj2670365. http://dx.doi.org/10.1042/bj2670365.

Miners JO, Birkett DJ. 1996. Use of tolbutamide as a substrate probe for human hepatic cytochrome P450 2C9. Methods Enzymol. 272:139–145. doi:10.1016/s0076-6879(96)72017-7. http://dx.doi.org/10.1016/s0076-6879(96)72017-7.

Miners JO, Birkett DJ. 1998. Cytochrome P4502C9: an enzyme of major importance in human drug metabolism. Br J Clin Pharmacol. 45(6):525–538. doi:10.1046/j.1365-2125.1998.00721.x. http://dx.doi.org/10.1046/j.1365-2125.1998.00721.x.

Miners JO, Coulter S, Tukey RH, Veronese ME, Birkett DJ. 1996. Cytochromes P450, 1A2, and 2C9 are responsible for the human hepatic O-demethylation of R- and S-naproxen. Biochem Pharmacol. 51(8):1003–1008. doi:10.1016/0006-2952(96)85085-4. http://dx.doi.org/10.1016/0006-2952(96)85085-4.

Murray M. 2016. CYP2J2 - regulation, function and polymorphism. Drug Metab Rev. 48(3):351–368. doi:10.1080/03602532.2016.1188938. http://dx.doi.org/10.1080/03602532.2016.1188938.

Myers S, Bottolo L, Freeman C, McVean G, Donnelly P. 2005. A fine-scale map of recombination rates and hotspots across the human genome. Science. 310(5746):321–324. doi:10.1126/science.1117196. http://dx.doi.org/10.1126/science.1117196.

Nelson MR, Wegmann D, Ehm MG, Kessner D, St Jean P, Verzilli C, Shen J, Tang Z, Bacanu S-A, Fraser D, et al. 2012. An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. Science. 337(6090):100–104. doi:10.1126/science.1217876. http://dx.doi.org/10.1126/science.1217876.

Ninomiya H, Mamiya K, Matsuo S, Ieiri I, Higuchi S, Tashiro N. 2000. Genetic polymorphism of the CYP2C subfamily and excessive serum phenytoin concentration with central nervous system intoxication. Ther Drug Monit. 22(2):230–232. doi:10.1097/00007691-200004000-00016. http://dx.doi.org/10.1097/00007691-200004000-00016.

O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, et al. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion,

and functional annotation. Nucleic Acids Res. 44(D1):D733–45. doi:10.1093/nar/gkv1189. http://dx.doi.org/10.1093/nar/gkv1189.

Oscarson M, McLellan RA, Asp V, Ledesma M, Bernal Ruiz ML, Sinues B, Rautio A, Ingelman-Sundberg M. 2002. Characterization of a novel CYP2A7/CYP2A6 hybrid allele (CYP2A6*12) that causes reduced CYP2A6 activity. Hum Mutat. 20(4):275–283. doi:10.1002/humu.10126. http://dx.doi.org/10.1002/humu.10126.

Petr M, Pääbo S, Kelso J, Vernot B. 2019. Limits of long-term selection against Neandertal introgression. Proc Natl Acad Sci U S A. 116(5):1639–1644. doi:10.1073/pnas.1814338116. http://dx.doi.org/10.1073/pnas.1814338116.

Polonikov AV, Ivanov VP, Solodilova MA, Khoroshaya IV, Kozhuhov MA, Ivakin VE, Katargina LN, Kolesnikova OE. 2008. A common polymorphism G-50T in cytochrome P450 2J2 gene is associated with increased risk of essential hypertension in a Russian population. Dis Markers. 24(2):119–126. doi:10.1155/2008/626430. http://dx.doi.org/10.1155/2008/626430.

Polonikov AV, Ivanov VP, Solodilova MA, Khoroshaya IV, Kozhuhov MA, Panfilov VI. 2007. Promoter polymorphism G-50T of a human CYP2J2 epoxygenase gene is associated with common susceptibility to asthma. Chest. 132(1):120–126. doi:10.1378/chest.07-0371. http://dx.doi.org/10.1378/chest.07-0371.

Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA, Kling DE, Gauthier LD, Levy-Moonshine A, Roazen D, et al. 2018. Scaling accurate genetic variant discovery to tens of thousands of samples. bioRxiv.:201178. doi:10.1101/201178. [accessed 2021 Apr 9]. https://www.biorxiv.org/content/10.1101/201178v3.

Prüfer K, de Filippo C, Grote S, Mafessoni F, Korlević P, Hajdinjak M, Vernot B, Skov L, Hsieh P, Peyrégne S, et al. 2017. A high-coverage Neandertal genome from Vindija Cave in Croatia. Science. 358(6363):655–658. doi:10.1126/science.aao1887. [accessed 2021 Apr 9]. https://science.sciencemag.org/content/358/6363/655.

Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, Heinze A, Renaud G, Sudmant PH, de Filippo C, et al. 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature. 505(7481):43–49. doi:10.1038/nature12886. http://dx.doi.org/10.1038/nature12886.

Racimo F, Sankararaman S, Nielsen R, Huerta-Sánchez E. 2015. Evidence for archaic adaptive introgression in humans. Nat Rev Genet. 16(6):359–371. doi:10.1038/nrg3936. http://www.nature.com/doifinder/10.1038/nrg3936.

R Core Team. 2008. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. https://www.R-project.org.

Rentzsch P, Schubach M, Shendure J, Kircher M. 2021. CADD-Splice-improving genome-wide variant effect prediction using deep learning-derived splice scores. Genome Med. 13(1):31. doi:10.1186/s13073-021-00835-9. http://dx.doi.org/10.1186/s13073-021-00835-9.

Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. 2019. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res. 47(D1):D886–D894. doi:10.1093/nar/gky1016. http://dx.doi.org/10.1093/nar/gky1016.

Rettie AE, Wienkers LC, Gonzalez FJ, Trager WF, Korzekwa KR. 1994. Impaired (S)-warfarin metabolism catalysed by the R144C allelic variant of CYP2C9. Pharmacogenetics. 4(1):39–42. doi:10.1097/00008571-199402000-00005. http://dx.doi.org/10.1097/00008571-199402000-00005.

Sangkuhl K, Claudio-Campos K, Cavallari LH, Agundez JAG, Whirl-Carrillo M, Duconge J, Del Tredici AL, Wadelius M, Rodrigues Botton M, Woodahl EL, et al. 2021. PharmVar GeneFocus: CYP2C9. Clin Pharmacol Ther. 110(3):662–676. doi:10.1002/cpt.2333. http://dx.doi.org/10.1002/cpt.2333.

Sankararaman S, Mallick S, Dannemann M, Prüfer K, Kelso J, Pääbo S, Patterson N, Reich D. 2014. The genomic landscape of Neanderthal ancestry in present-day humans. Nature. 507(7492):354–357. doi:10.1038/nature12961. http://dx.doi.org/10.1038/nature12961.

Sankararaman S, Mallick S, Patterson N, Reich D. 2016. The Combined Landscape of Denisovan and Neanderthal Ancestry in Present-Day Humans. Curr Biol. 26(9):1241–1247. doi:10.1016/j.cub.2016.03.037. http://dx.doi.org/10.1016/j.cub.2016.03.037.

Sankararaman S, Patterson N, Li H, Pääbo S, Reich D. 2012. The date of interbreeding between Neandertals and modern humans. PLoS Genet. 8(10):e1002947. doi:10.1371/journal.pgen.1002947. http://dx.doi.org/10.1371/journal.pgen.1002947.

Sausville LN, Gangadhariah MH, Chiusa M, Mei S. 2018. The cytochrome P450 slow metabolizers CYP2C9* 2 and CYP2C9* 3 directly regulate tumorigenesis via reduced epoxyeicosatrienoic acid production. Cancer Res. https://cancerres.aacrjournals.org/content/78/17/4865.short.

Schaefer NK, Shapiro B, Green RE. 2021. An ancestral recombination graph of human, Neanderthal, and Denisovan genomes. Science Advances. 7(29):eabc0776. doi:10.1126/sciadv.abc0776. [accessed 2021 Jul 28]. https://advances.sciencemag.org/content/7/29/eabc0776.

Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. 2001. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 29(1):308–311. doi:10.1093/nar/29.1.308. http://dx.doi.org/10.1093/nar/29.1.308.

Sistonen J, Fuselli S, Palo JU, Chauhan N, Padh H, Sajantila A. 2009. Pharmacogenetic variation at CYP2C9, CYP2C19, and CYP2D6 at global and microgeographic scales. Pharmacogenet Genomics. 19(2):170–179. doi:10.1097/FPC.0b013e32831ebb30. http://dx.doi.org/10.1097/FPC.0b013e32831ebb30.

Stearns RA, Chakravarty PK, Chen R, Chiu SH. 1995. Biotransformation of losartan to its active carboxylic acid metabolite in human liver microsomes. Role of cytochrome P4502C and 3A subfamily members. Drug Metab Dispos. 23(2):207–215. https://www.ncbi.nlm.nih.gov/pubmed/7736913.

Sullivan RJ, Hagen EH, Hammerstein P. 2008. Revealing the paradox of drug reward in human evolution. Proc Biol Sci. 275(1640):1231–1241. doi:10.1098/rspb.2007.1673. http://dx.doi.org/10.1098/rspb.2007.1673.

Tan L, Yu J-T, Sun Y-P, Ou J-R, Song J-H, Yu Y. 2010. The influence of cytochrome oxidase CYP2A6, CYP2B6, and CYP2C9 polymorphisms on the plasma concentrations of valproic acid in epileptic patients. Clin Neurol Neurosurg. 112(4):320–323. doi:10.1016/j.clineuro.2010.01.002. http://dx.doi.org/10.1016/j.clineuro.2010.01.002.

Thomas JH. 2007. Rapid Birth–Death Evolution Specific to Xenobiotic Cytochrome P450 Genes in Vertebrates. PLoS Genet. 3(5):e67. doi:10.1371/journal.pgen.0030067. [accessed 2021 Apr 12]. https://journals.plos.org/plosgenetics/article/file?id=10.1371/journal.pgen.0030067&type=printable.

Tracy TS, Rosenbluth BW, Wrighton SA, Gonzalez FJ, Korzekwa KR. 1995. Role of cytochrome P450 2C9 and an allelic variant in the 4'-hydroxylation of (R)- and (S)-flurbiprofen. Biochem Pharmacol. 49(9):1269–1275. doi:10.1016/0006-2952(95)00048-5. http://dx.doi.org/10.1016/0006-2952(95)00048-5.

Van Booven D, Marsh S, McLeod H. 2010. Cytochrome P450 2C9-CYP2C9. Pharmacogenetics. https://www.ncbi.nlm.nih.gov/pmc/articles/pmc3201766/.

Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. 2016. SIFT missense predictions for genomes. Nat Protoc. 11(1):1–9. doi:10.1038/nprot.2015.123. http://dx.doi.org/10.1038/nprot.2015.123.

Vernot B, Akey JM. 2014. Resurrecting Surviving Neandeltal Linages from Modern Human Genomes. Science. 343(February):1017–1021. doi:10.5061/dryad.5t110.Supplementary. http://classic.sciencemag.org/content/343/6174/1017.full.pdf.

Vernot B, Tucci S, Kelso J, Schraiber JG, Wolf AB, Gittelman RM, Dannemann M, Grote S, McCoy RC, Norton H, et al. 2016. Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. Science. 352(6282):235–239. doi:10.1126/science.aad9416. http://www.sciencemag.org/cgi/doi/10.1126/science.aad9416.

Villanea FA, Schraiber JG. 2018. Multiple episodes of interbreeding between Neanderthal and modern humans. Nature Ecology & Evolution. 3(1):39–44. doi:10.1038/s41559-018-0735-8. [accessed 2021 Apr 11]. https://www.nature.com/articles/s41559-018-0735-8.

Wall JD, Yang M a., Jay F, Kim SK, Durand EY, Stevison LS, Gignoux C, Woerner A, Hammer MF, Slatkin M. 2013. Higher levels of neanderthal ancestry in East asians than in europeans. Genetics. 194(1):199–209. doi:10.1534/genetics.112.148213. http://dx.doi.org/10.1534/genetics.112.148213.

Wang H, Jiang Y, Liu Y, Lin C, Cheng G, Chen X, Hao B, Tan W, Lin D, He F. 2006. CYP2J2*7 single nucleotide polymorphism in a Chinese population. Clin Chim Acta. 365(1-2):125–128. doi:10.1016/j.cca.2005.08.007. http://dx.doi.org/10.1016/j.cca.2005.08.007.

Wang K, Li M, Hakonarson H. 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 38(16):e164. doi:10.1093/nar/gkq603. http://dx.doi.org/10.1093/nar/gkq603.

Ward BA, Gorski JC, Jones DR, Hall SD, Flockhart DA, Desta Z. 2003. The cytochrome P450 2B6 (CYP2B6) is the main catalyst of efavirenz primary and secondary metabolism: implication for HIV/AIDS therapy and utility of efavirenz as a substrate marker of CYP2B6 catalytic activity. J Pharmacol Exp Ther. 306(1):287–300. doi:10.1124/jpet.103.049601. http://dx.doi.org/10.1124/jpet.103.049601.

Wickham H. 2009. ggplot2: Elegant Graphics for Data Analysis. Springer Science & Business Media. https://play.google.com/store/books/details?id=bes-AAAAQBAJ.

Yamazaki H, Inoue K, Chiba K, Ozawa N, Kawai T, Suzuki Y, Goldstein JA, Guengerich FP, Shimada T. 1998. Comparative studies on the catalytic roles of cytochrome P450 2C9 and its Cys- and Leu-variants in the oxidation of warfarin, flurbiprofen, and diclofenac by human liver microsomes. Biochem Pharmacol. 56(2):243–251. doi:10.1016/s0006-2952(98)00133-6. http://dx.doi.org/10.1016/s0006-2952(98)00133-6.

Zanger UM, Klein K. 2013. Pharmacogenetics of cytochrome P450 2B6 (CYP2B6): advances on polymorphisms, mechanisms, and clinical relevance. Front Genet. 4:24. doi:10.3389/fgene.2013.00024. http://dx.doi.org/10.3389/fgene.2013.00024.

Zeberg H, Pääbo S. 2021. A genomic region associated with protection against severe COVID-19 is inherited from Neandertals. Proc Natl Acad Sci U S A. 118(9). doi:10.1073/pnas.2026309118. http://dx.doi.org/10.1073/pnas.2026309118.

Zhang X, Kim B, Lohmueller KE, Huerta-Sánchez E. 2020. The Impact of Recessive Deleterious Variation on Signals of Adaptive Introgression in Human Populations. Genetics. 215(3):799–812. doi:10.1534/genetics.120.303081. http://dx.doi.org/10.1534/genetics.120.303081.

Zhou Y, Ingelman-Sundberg M, Lauschke VM. 2017. Worldwide Distribution of Cytochrome P450 Alleles: A Meta-analysis of Population-scale Sequencing Projects. Clinical Pharmacology & Therapeutics. 102(4):688–700. doi:10.1002/cpt.690. http://dx.doi.org/10.1002/cpt.690.