



# Denisovan admixture facilitated environmental adaptation in Papua New Guinean populations

Danat Yermakovich<sup>a</sup>, Mathilde André<sup>a</sup> , Nicolas Brucato<sup>b</sup>, Jason Kariwiga<sup>c,d</sup>, Matthew Leavesley<sup>c,e</sup> , Vasili Pankratov<sup>a</sup> , Mayukh Mondal<sup>a,f</sup>, François-Xavier Ricaut<sup>b,1,2</sup> , and Michael Dannemann<sup>a,1,2</sup>

Affiliations are included on p. 9.

Edited by Marcus Feldman, Stanford University, Stanford, CA; received March 21, 2024; accepted May 16, 2024

Neandertals and Denisovans, having inhabited distinct regions in Eurasia and possibly Oceania for over 200,000 y, experienced ample time to adapt to diverse environmental challenges these regions presented. Among present-day human populations, Papua New Guineans (PNG) stand out as one of the few carrying substantial amounts of both Neandertal and Denisovan DNA, a result of past admixture events with these archaic human groups. This study investigates the distribution of introgressed Denisovan and Neandertal DNA within two distinct PNG populations, residing in the highlands of Mt Wilhelm and the lowlands of Daru Island. These locations exhibit unique environmental features, some of which may parallel the challenges that archaic humans once confronted and adapted to. Our results show that PNG highlanders carry higher levels of Denisovan DNA compared to PNG lowlanders. Among the Denisovan-like haplotypes with higher frequencies in highlander populations, those exhibiting the greatest frequency difference compared to lowlander populations also demonstrate more pronounced differences in population frequencies than frequency-matched nonarchaic variants. Two of the five most highly differentiated of those haplotypes reside in genomic areas linked to brain development genes. Conversely, Denisovan-like haplotypes more frequent in lowlanders overlap with genes associated with immune response processes. Our findings suggest that Denisovan DNA has provided genetic variation associated with brain biology and immune response to PNG genomes, some of which might have facilitated adaptive processes to environmental challenges.

human evolution | archaic admixture | local adaptation

The initial human settlement of New Guinea is estimated to have occurred by at least 50,000 B.P. (1, 2). Today, the distribution of PNG (Papua New Guinean) populations across the region is uneven, often occurring in areas characterized by significant environmental disparities (3, 4). These environmental challenges, such as exposure to high altitudes or region-specific pathogens, have been demonstrated to correlate with phenotypic variations among PNG populations inhabiting distinct environments (5). These challenges have also been established as factors contributing to the emergence of local genetic adaptation signatures (6, 7). Moreover, PNG, like their counterparts in near and remote Oceania, bear a significant share of approximately 3 to 4% Denisovan DNA, ranking among the highest proportions globally (8, 9). This is in addition to the ~2% of Neandertal ancestry that are found in PNG and all present-day non-Africans (10, 11). Our understanding of particularly the functional aspects of Denisovan DNA in its carriers and its potential contribution to adaptive processes in populations, such as the PNG, remains limited. This restricted knowledge is attributable to a variety of challenges. First, the scarcity of fossil fragments has hindered the ability to comprehensively reconstruct Denisovan physiology and their historical geographic distribution (12). These aspects make it challenging to postulate hypotheses regarding their functions in present-day individuals. Some insights have been derived from the Denisovan genome sequence (8), such as the presence of only one copy of the amylase gene—a feature shared with Neandertals, hinting at potential differences in starch digestion compared to modern humans (13). Additionally, attempts to predict Denisovan phenotypes from genomic data have provided insights into their skeletal physiology (14). Due to the limited fossil record originating from only a handful of locations, predicting the precise geographical range of Denisovans presents another significant challenge. However, both the distribution of these fossils and the distribution of populations carrying Denisovan DNA collectively suggest a relatively broad habitat (12, 15). It is plausible that Denisovans may have inhabited various regions across the Asian continent, spanning from Siberia to Southeast Asia. The existing uncertainty about their geographical distribution further complicates the formulation of hypotheses regarding potential environmental

## Significance

Understanding Denisovan biology remains challenging due to the scarcity of fossil evidence and gaps in our knowledge about their natural habitat. To enhance our understanding, this study explores Denisovan DNA inherited through past admixtures between modern humans and this ancient group. Focusing on two distinct Papuan populations residing at different sea levels, our analysis reveals significant frequency differences in Denisovan DNA, impacting genomic regions linked to immunity and brain development. These findings provide valuable insights into the functional roles of Denisovan DNA, shedding light on its adaptive signatures in the environments associated with the two Papuan populations. This suggests that the adaptive processes observed may parallel those of Denisovans facing similar environmental challenges.

Author contributions: D.Y., N.B., F.-X.R., and M.D. designed research; D.Y., M.A., N.B., V.P., M.M., F.-X.R., and M.D. performed research; D.Y., J.K., M.L., V.P., M.M., and M.D. contributed new reagents/analytic tools; D.Y., M.A., and M.D. analyzed data; and D.Y., N.B., F.-X.R., and M.D. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2024 the Author(s). Published by PNAS. This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

<sup>1</sup>F.-X.R. and M.D. contributed equally to this work.

<sup>2</sup>To whom correspondence may be addressed. Email: francois-xavier.ricaut@univ-tlse3.fr or michael.dannemann@ut.ee.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2405889121/-/DCSupplemental>.

Published June 18, 2024.

adaptations of Denisovans. The interbreeding between modern humans and Denisovans has been suggested to occur in multiple waves involving distinct Denisovan populations (7, 16–18). These findings may imply the existence of diverse Denisovan groups with potentially differing genetic compositions, adding another layer of complexity to the interpretation of their phenotypic characteristics. Another challenge arises when attempting to annotate the phenotypic effects of Denisovan DNA in contemporary individuals. In the case of introgressed Neandertal DNA, the utilization of phenotypic association data played a pivotal role in studying its influence on modern humans and drawing potential insights into Neandertal biology (19–23). However, unlike Neandertal DNA, which is present in populations with accessible GWAS data, Denisovan DNA lacks readily available association information from phenotype and expression cohorts, rendering such annotation approaches less practicable. Nonetheless, several studies have provided valuable insights into how Denisovan DNA has influenced phenotypic variation and facilitated adaptation in modern humans (24–30). Notably, among them are well-documented instances where Denisovan DNA has been associated with adaptations related to high-altitude environments, metabolic processes, and immune responses. The environmental challenges encountered by PNG as they dispersed across various regions could potentially mirror the challenges that Denisovans themselves had to contend with. Analyzing the fate of Denisovan DNA within PNG populations residing in diverse environmental conditions may offer a pathway to assess the functional capacity of Denisovan DNA and unveil some of its adaptive potential.

## Results

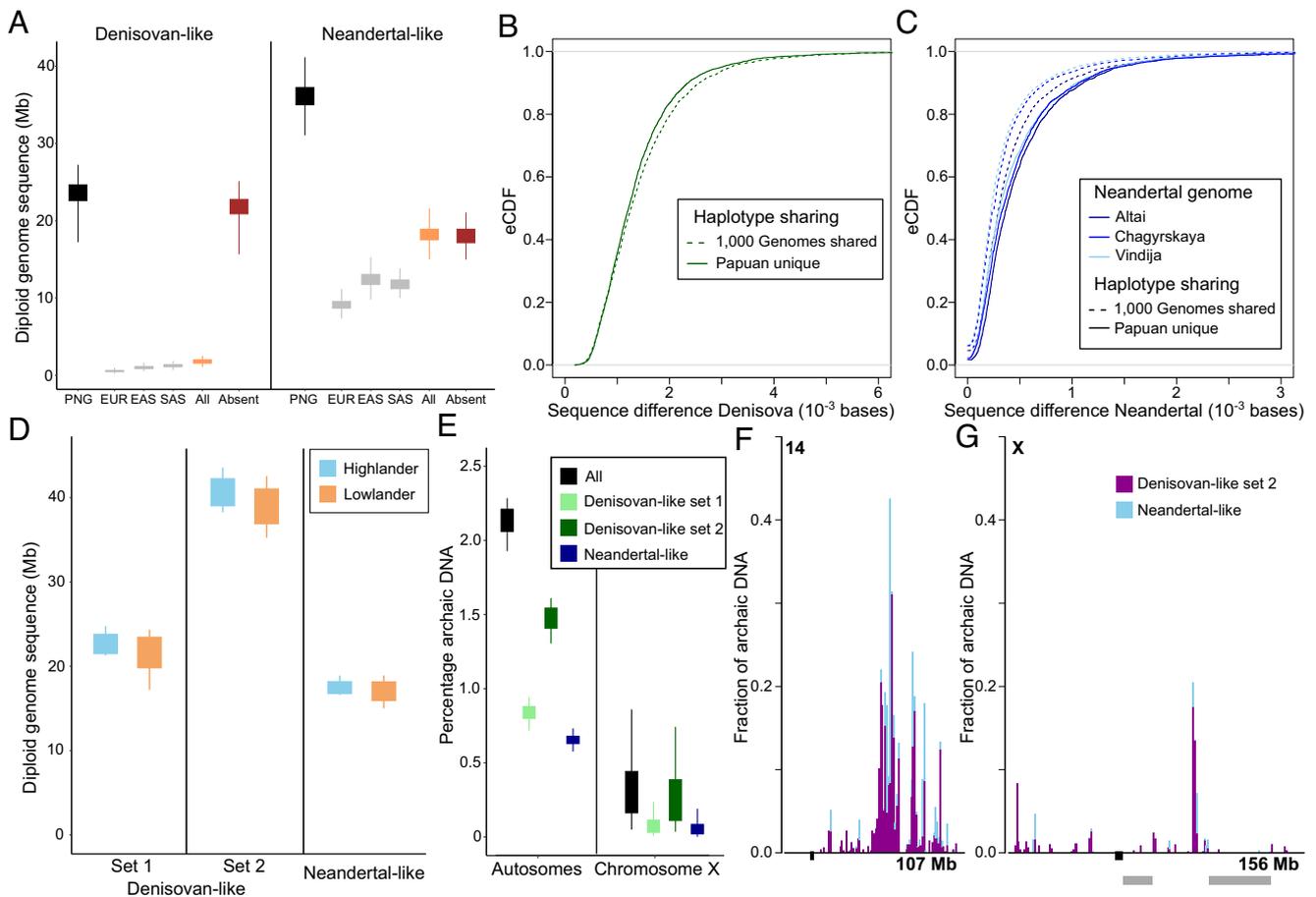
**Identification of Archaic DNA in PNG Highlanders and Lowlanders.** In this study, our objective was to evaluate the degree to which DNA inherited from past interbreeding with archaic humans has contributed to shaping local adaptation of two geographically distinct PNG populations, each confronting unique environmental challenges. To address this task, we characterized the genomic landscape of Neandertal and Denisovan DNA in the genomes of 74 PNG individuals residing in the lowlands of Daru Island and 54 individuals inhabiting the highlands of Mount Wilhelm (*Dataset S1* and *Methods*) (6). The habitats of both populations are marked by notable environmental differences, primarily in terms of altitude. Individuals from the PNG highlander group, included in this study, reside at elevations ranging from 2,300 to 2,700 m above sea level. In contrast, the PNG lowlanders residing on Daru Island, located off the southern coast of Papua New Guinea, inhabit an environment approximately 100 m above sea level. These altitude differentials are associated with various other environmental factors, including variations in food availability and pathogen exposure. We reconstructed introgressed archaic haplotypes utilizing a previously established method to detect archaic DNA within contemporary populations (22). This approach identifies archaic SNPs (aSNPs) within the genomes of present-day individuals. These aSNPs function as markers of introgressed variants, with distinctive attributes such as allele sharing signatures with Neandertals and Denisovans. Moreover, these aSNPs reside on haplotypes of a length that exceeds those of segments that result from incomplete lineage sorting (ILS) between modern and archaic humans. ILS segments can produce comparable allele-sharing patterns to those seen in introgressed haplotypes but are on average considerably older and shorter (*Methods*). When employing this methodology in the analysis of PNG genomes within our study, we identified 168,152 aSNPs (*Dataset S2*). These aSNPs were associated with

10,431 unique core haplotypes spanning across 50.0 to 65.5 megabases (Mb) of diploid archaic DNA within the autosome and chromosome X of the analyzed PNG individuals.

**Replication of Introgression Map Results with an Alternative Method.** We were able to replicate a large fraction of our identified archaic haplotypes using another alternative method (HMMIX) designed for reconstructing archaic DNA in contemporary populations (31). This alternative approach employs a hidden Markov model to identify genomic regions characterized by a high density of SNPs that are absent in an unadmixed outgroup population. A significant proportion of 92.0% of the archaic haplotypes we identified were also captured by the alternative approach (with a posterior probability exceeding 0.8; 96.0% with posterior probability > 0.5; *Methods*). Generally, the alternative method detected a notably higher count of haplotypes (total haplotypes in PNG with posterior probability > 0.8, HMMIX: 483,376; our method: 172,121, *SI Appendix, Fig. S1A*). However, many HMMIX-specific haplotypes with a posterior probability larger than 0.8 were characterized by an absence of aSNPs (81.2% HMMIX-specific haplotypes carried no aSNPs), possibly due to the underlying method of HMMIX which does not involve archaic genome information. These haplotypes also exhibited shorter lengths compared to the ones we observed and omitted because their lengths aligned with ILS (*SI Appendix, Fig. S1 B and C*).

**Most Likely Archaic Source of Introgressed Haplotypes.** Next, our goal was to assign the introgressed haplotypes we identified to their most probable archaic origins. To achieve this, we assessed their sequence similarity in comparison to the genomic sequences of three high-coverage Neandertals from the Altai Mountains (10), Vindija Cave (11), and Chagyrskaya Cave (32), as well as the Denisovan individual (8).

We observed that on average, ~61% of the haplotypes per individual displayed a closer sequence similarity to Neandertals than to the Denisovan individual (Fig. 1A). This result stands in stark contrast to the anticipated genome-wide estimates of Neandertal ancestry (~2%) and Denisovan ancestry (~4%) within these populations (8, 11, 18). Yet, this figure aligns with prior studies that have attributed this disparity to the substantial differences between the sequenced Denisovan and the introgressing Denisovan population (27, 28). To examine potential factors contributing to the variance between our haplotype-based ancestry estimates and those derived from genome-wide analysis, our study investigated the geographical distribution of archaic haplotypes in PNG (*SI Appendix*). First, assuming that PNG share one major pulse of Neandertal admixture with other non-African populations and possibly a second minor pulse with present-day mainland Asians (33), we scanned for the presence of Neandertal-like haplotypes from PNG individuals in present-day Eurasian individuals that match the aSNP content and genomic location (*Methods, Fig. 1A*). Our findings revealed that roughly half of these Neandertal-like haplotypes were exclusively found in PNG. One possible explanation for the presence of these PNG-specific Neandertal haplotypes is that they originated from shared Neandertal admixture pulses but subsequently disappeared from Eurasian populations. However, in our analysis of the proportion of unique Neandertal variants in both 1,000 Genomes populations and superpopulations, we estimate that such variants can account for no more than 20% of the total PNG-specific Neandertal variants (*SI Appendix*). To further explore alternative sources for the PNG-specific fraction of Neandertal-like haplotypes, we assessed their sequence relationship with Neandertals and the Denisovan and compared those measures to the sequence similarity estimates of Neandertal-like haplotypes in PNG that are also found in 1,000



**Fig. 1.** Archaic DNA in PNG genomes. (A) Amount of diploid genome sequences reconstructed in PNG individuals based on archaic haplotypes exhibiting higher sequence similarity with the Denisovan individual (*Left*) and Neandertals (*Right*). Boxplots are employed to visualize the distributions, with the outer whiskers indicating the minimum and maximum values. The recovered sequence amounts are presented in Mb for various categories, including total sequences (black), sequences found in 1,000 Genomes Europeans (EUR, gray), East Asians (EAS, gray), South Asians (SAS, gray), and the combined sequences from all three Eurasian populations (All, orange). Additionally, the amount of genome sequences recovered using haplotypes that were not detected in any of the Eurasian populations (Absent) is displayed in red. (B and C) Empirical cumulative density distributions (eCDF, y-axis) are displayed for the sequence similarity measures (in number of differences per  $10^{-3}$  bases, x-axis) of archaic haplotypes in PNG, specifically those exhibiting a greater sequence similarity with Neandertals than with the Denisovan individual. Panels (B) and (C) show these distributions in comparison to the Denisovan and three Neandertals, respectively. Haplotypes found in 1,000 Genomes Eurasians are represented by dashed lines, while PNG-specific haplotypes are delineated with solid lines. (D) Boxplots illustrating the amount of Denisovan-like (sets 1 and 2) and Neandertal-like diploid genome sequence in PNG highlanders (blue) and lowlanders (orange). Outer whiskers represent the minimum and maximum values. (E) Boxplots displaying the percentage of archaic (black/gray), Denisovan-like (set 1: green; set 2: dark green), and Neandertal-like (blue) ancestry on the autosomes and X chromosome in the PNG cohort. The outer whiskers indicate the 95% CI borders. (F and G) The proportion of Denisovan-like (set 2, purple) and Neandertal-like (blue) DNA within one-megabase windows is depicted for chromosome 14 (F), which harbors the highest levels of archaic DNA, and chromosome X (G), characterized by the least amount of archaic DNA. Gray areas below the x-axis denote regions that were previously reported to be devoid of archaic ancestry in present-day populations. Black areas indicate centromere regions.

Genomes Eurasians. We found that PNG-specific Neandertal-like haplotypes exhibited a closer sequence relationship to the Denisovan individual (Fig. 1B) and displayed greater sequence divergence from Neandertals (Fig. 1C) compared to the PNG Neandertal-like haplotypes identified in Eurasians. These results suggest that some of the PNG-specific Neandertal-like haplotypes may have originated from an archaic population that had a closer sequence similarity to the Denisovan and a larger divergence from Neandertal genomes compared to Neandertal-like haplotype found in PNG and 1,000 Genomes Eurasians. Several scenarios are consistent with this observation. For instance, a scenario involving admixture between PNG and an archaic population carrying both considerable Neandertal and Denisovan ancestry would support these findings. However, to date, no such scenarios have been reported. Moreover, there are factors that might lead to the misidentification of the ancestral origins of archaic haplotypes. For example, previous studies have reported multiple pulses of admixture between Oceanian populations and distinct Denisovan groups (7, 16, 18). Those introgressing Denisovan groups showed substantially different sequence affinities

with the sequenced Denisovan. Furthermore, the large sequence divergence of the introgressing Denisovan population compared to the sequenced Denisovan is in contrast with the comparably closer sequence affinity of high-coverage Neandertal genomes to the introgressing Neandertal population. A previous study has shown that this contrast and other factors such as ILS pose an additional challenge in accurately annotating archaic ancestries to introgressed haplotypes, particularly those of true Denisovan ancestry (27). Finally, the higher number of available high-coverage Neandertal genomes might potentially further amplify the issue of misclassification due to ILS. It is also worth noting that admixture between archaic human groups has previously been reported (10, 34). For example, the Altai Neandertal has been demonstrated to harbor Denisovan ancestry in its genome, introducing an additional challenge and further increasing the likelihood of misclassifying genuine Denisovan introgressed haplotypes.

Considering all these observations collectively, we hypothesize that the PNG-specific set of haplotypes with a closer sequence similarity to Neandertals is composed of a mix of true Neandertal

and misclassified Denisovan haplotypes. Based on that hypothesis, in our study, we opted to categorize different sets of archaic haplotypes. First, archaic haplotypes displaying a closer sequence similarity to the Denisovan than to Neandertals were defined as “Denisovan-like set 1”—a set of archaic haplotypes with a high likelihood of Denisovan ancestry, but likely lacking a substantial proportion of misclassified Denisovan haplotypes. Second, alongside Denisovan-like set 1, we introduced “Denisovan-like set 2”, comprising PNG-specific haplotypes closer in sequence to Neandertals and all haplotypes in Denisovan-like set 1—a more comprehensive set of genuine Denisovan haplotypes compared to set 1, but also comprising a fraction of true PNG-specific Neandertal haplotypes. Third, we defined “Neandertal-like” haplotypes in PNG as those exhibiting a closer sequence similarity to at least one Neandertal genome compared to the Denisovan genome and were also present in 1,000 Genomes Eurasian populations—a set of archaic haplotypes with a high likelihood of Neandertal ancestry. The use of each of these three sets in our analyses possesses distinct advantages and limitations, which we thoroughly discuss in *SI Appendix*. We would like to note that Denisovan-like set 2 accounted for 69.6% of all archaic haplotypes in PNG—a number substantially closer to the figure derived from other genome-wide ancestry estimates (*SI Appendix*). Our analysis revealed that PNG highlanders and lowlanders carry similar amounts of Neandertal-like DNA ( $P = 0.98$ , Mann–Whitney  $U$  test). However, notably, PNG highlanders exhibited ~1.5% (Denisovan-like set 1,  $P = 0.10$ ) and ~3% (Denisovan-like set 2,  $P = 0.003$ , Fig. 1*D* and *Dataset S1*) more Denisovan-like DNA relative to the number of their lowland counterparts.

**Genomic Distribution of Archaic Haplotypes.** Introgressed segments combined from all studied PNG individuals covered approximately 21.15% of the human genome. Among those segments, we identified 21 archaic haplotypes characterized by archaic allele frequencies surpassing 70% (99.8 percentile) in the combined dataset of the two PNG populations. Interestingly, only eight of these haplotypes exhibit overlaps with protein-coding genes, posing a challenge in establishing a functional link of those haplotypes on the biology of their carriers (*Dataset S3*). Among the top five haplotypes featuring the highest archaic allele frequencies, only the Neandertal-like haplotype with the highest frequency of 89% (chr9:112,041,006–112,068,271) shares an overlap with a protein-coding gene, namely *SUSD1*. Notably, *SUSD1* remains relatively understudied, although a recent investigation has suggested its involvement in neurodevelopmental disorders (35).

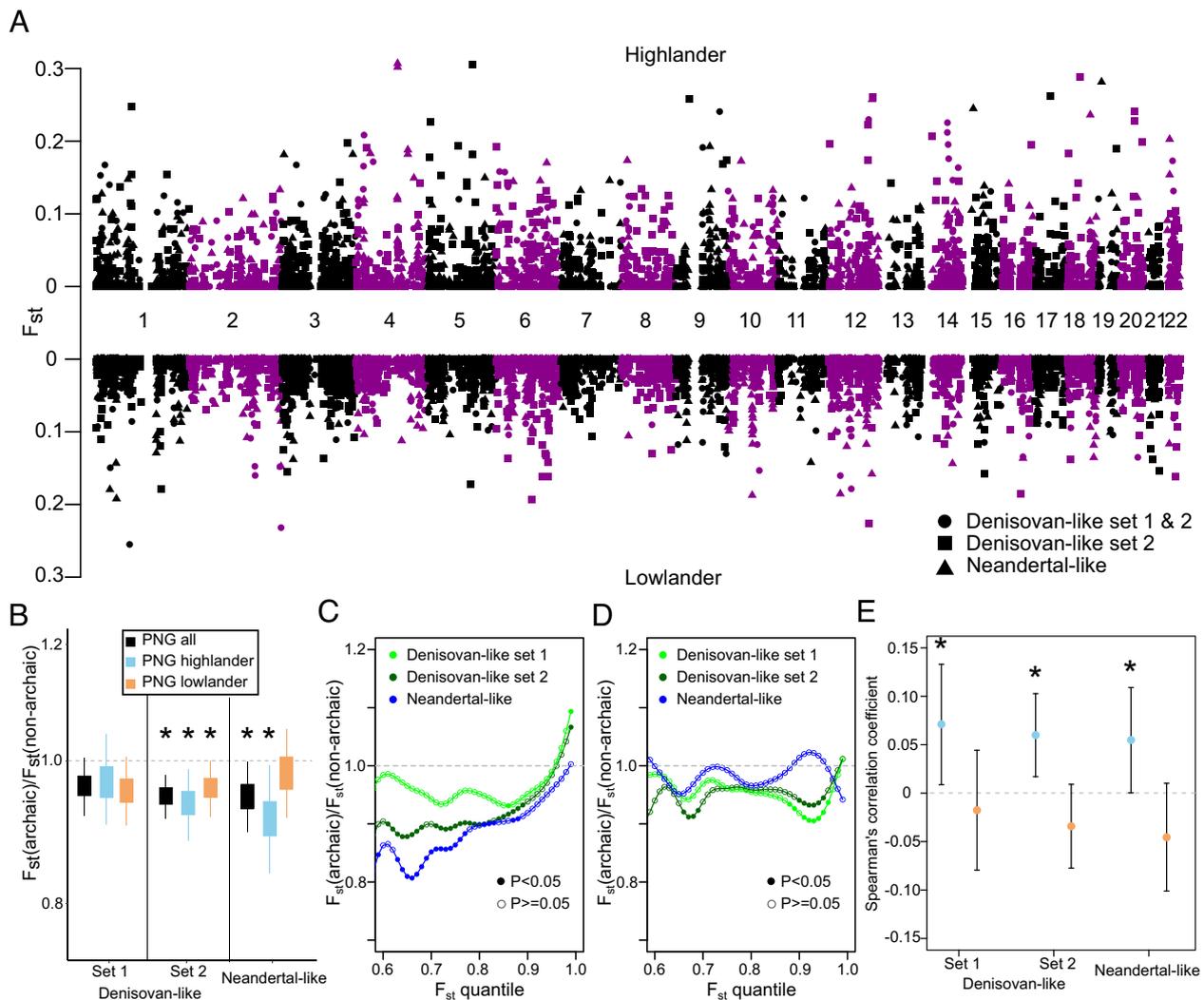
Additionally, we studied the archaic DNA content within regions previously identified as potential sites for negative selection acting on introgressed DNA (27, 28). Our analysis confirmed that within all five autosomal genomic regions previously reported to be devoid of archaic DNA, there was either an absence or only negligible traces of archaic DNA present (*SI Appendix*, Fig. S2). Consistent with these previous studies, we also noted substantially lower levels of archaic DNA on the X chromosome (*SI Appendix*, Fig. S2). PNG individuals exhibited an average 7.8-fold decrease in archaic DNA content on chromosome X in comparison to autosomes. This reduction was more prominent for Neandertal-like DNA, with a 12.6-fold decrease, compared to Denisovan-like DNA, which showed an 11.8-fold and 6.9-fold reduction for sets 1 and 2, respectively (Fig. 1*E–G*).

**Signatures of Local Adaptation in PNG Lowlanders and Highlanders.** Next, we aimed to investigate the extent of population differentiation between the PNG highlanders and lowlanders. This analysis was designed to provide us with insights into the

potential impact of selection on archaic haplotypes within each of these two examined populations. To achieve this, we conducted a comparative analysis of the allele frequency difference for each distinct archaic haplotype within both the PNG highlander and lowlander populations by calculating their respective  $F_{st}$  values (*Methods*, Fig. 2*A*, and *Dataset S3*). We then determined whether archaic haplotypes exhibited altered levels of population differentiation between these two examined groups by comparing their  $F_{st}$  distribution to those of a genomic background of 1,000 randomly generated sets of nonarchaic variants, chosen to match the allele frequency distribution of the archaic haplotypes (*Methods* and Fig. 2*B*). We also required that the alleles within the random sets of nonarchaic variants matching the archaic allele frequency must also exhibit a higher frequency within the respective PNG population. Overall, our analysis revealed that the mean  $F_{st}$  value for archaic haplotypes (0.0178) was significantly lower than the mean values observed in the random nonarchaic sets (ranging from 0.018 to 0.019,  $P < 0.001$ ). This trend was consistent for both Neandertal-like ( $P = 0.04$ ) and Denisovan-like (set 1:  $P = 0.002$ ; set 2:  $P = 0.08$ ) haplotypes. When partitioning the  $F_{st}$  distribution according to whether the higher archaic allele frequency was detected in highlanders or lowlanders, we found consistently lower mean values for all sets of archaic haplotypes ( $P$  values ranging from 0.004 to 0.63). When we stratified our analysis according to 1% bin  $F_{st}$  quantiles, we observed that the  $F_{st}$  distribution of Denisovan-like haplotypes with a higher frequency in highlanders started to trend to higher  $F_{st}$  values compared to randomly selected nonarchaic background sets and was significantly higher for the haplotypes within the top 1% of  $F_{st}$  values for both sets of Denisovan-like haplotypes set 1:  $P = 0.01$ ; set 2:  $P = 0.04$ , Fig. 2*C* and *D*). The observation that the most highly differentiated Denisovan-like haplotypes with a higher frequency in highlanders show larger levels of population differentiation than random frequency-matched nonarchaic variants would be consistent with elevated levels of local selective pressures that exceed those of random nonarchaic variants.

Next, to further investigate signatures of selection on archaic haplotypes in PNG, we employed a computational approach to reconstruct the allele frequency trajectories of all archaic haplotypes over the course of the past 1,000 generations in both highlanders and lowlanders, utilizing an approximate full-likelihood method (*Methods* and *Dataset S4*). We found that log-likelihood ratios (logLRs) that signal deviations from neutral allele frequency patterns showed a significant correlation with  $F_{st}$  values for both sets of Denisovan-like (Spearman correlation, set 1:  $P = 0.05$ ; set 2:  $P = 0.01$ ) and also for Neandertal-like ( $P = 0.03$ ) haplotypes among highlanders, but this correlation was absent among lowlanders ( $P > 0.05$ , Fig. 2*E*). Our results indicate that the differences in frequencies among highlanders for some of the most distinct archaic haplotypes have been driven by substantial increases in allele frequencies over the past 1,000 generations, matching the time of between ~25,000 and 20,000 B.P. when highlands were settled permanently (4, 36).

**Gene Content of Genomic Regions Overlapping Highly Differentiated Archaic Haplotypes.** A key emphasis of this study was to explore the potential phenotypic impact of archaic haplotypes exhibiting signs of local adaptation in highlanders or lowlanders. Previous studies on the health of Papua New Guinea (PNG) populations, including some focusing on the two PNG groups analyzed in this study, have indicated distinctive physiological traits among highlander groups. These traits include increased ventilatory lung function and hemoglobin concentration, a deeper chest, shorter stature, and smaller waist (5, 37–39). Furthermore, genetic variants exhibiting signatures of selection in the PNG highlanders



**Fig. 2.** Frequency differences of archaic haplotypes between PNG highlanders and lowlanders. (A) Manhattan plot visualizes  $F_{st}$  values, representing frequency differences between PNG highlanders and lowlanders for Denisovan-like haplotypes (sets 1 and 2: circles; set 2 only: squares) and Neandertal-like haplotypes (triangles). The Manhattan plot is symmetrically divided across the  $x$ -axis, with data points displayed on both the upper and lower sides conditioned on whether a haplotype is found at a higher frequency in highlanders or lowlanders, respectively. (B) Distributions of the ratio of the mean  $F_{st}$  value for archaic haplotypes compared to each mean  $F_{st}$  value obtained from 1,000 nonarchaic background sets (*Methods*). These distributions are presented for distinct categories of archaic haplotypes: The full sets of Denisovan-like (set 1: *Left*; set 2: *Middle*) and Neandertal-like (*Right*) haplotypes are shown in black. Additionally, they include subsets of haplotypes having higher frequency in highlanders (blue) or lowlanders (orange) for both sets of Denisovan-like and Neandertal-like haplotypes. Whiskers in the boxplots represent the 95% CI for these distributions. Distributions that exhibit a significant deviation from a ratio of 1 ( $P < 0.05$ ) are denoted by an asterisk. (C and D) Distribution of the ratio between  $F_{st}$  quantiles for sets of Denisovan-like and Neandertal-like haplotypes displaying a higher frequency in PNG highlanders (C) and PNG lowlanders (D) relative to the quantile  $F_{st}$  values derived from 1,000 matching nonarchaic background sets (*Methods*). Distributions are smoothed using a cubic spline smoothing algorithm. Quantiles for which the ratio significantly deviates from one are shown with filled circles. A gray line is included to represent the neutral expectation of one. (E) Spearman's correlation coefficients together with their 95% CI ( $y$ -axis) calculated between  $F_{st}$  values and log-likelihood ratios derived from reconstructed allele frequencies for sets of Denisovan-like and Neandertal-like haplotypes are shown. Haplotypes are categorized into two sets: those exhibiting a larger archaic allele frequency in PNG highlanders (blue) and in lowlanders (orange). Correlation coefficients that exhibit a significant deviation from a value of 0 ( $P < 0.05$ ) are denoted by an asterisk.

under investigation have been associated with variations in red blood cell composition and cardiovascular phenotypes. This observation aligns with findings in other high-altitude populations (6). To investigate whether similar phenotypes are influenced by archaic introgression, we examined the gene content of highly differentiated archaic haplotypes and the phenotypic associations of aSNPs linked to them. First, we evaluated the gene content within genomic regions containing different sets of archaic haplotypes showing signatures of population differentiation. Subsequently, to explore potential phenotypic consequences of genes within these regions, we conducted tests for functional enrichment within the gene ontology (GO) (40). In total, we carried out six enrichment analyses for genes that overlapped with archaic haplotypes falling within the top 1% of the  $F_{st}$  distributions for each possible combination of

higher archaic allele frequency in highlanders or lowlanders with Neandertal-like or Denisovan-like haplotypes sets (*SI Appendix, Methods, and Dataset S5*).

No enriched GO category was observed for both sets of high- $F_{st}$  Denisovan-like haplotypes with higher frequencies in highlanders [sets 1 and 2; FWER (Family-wise error rate)  $> 0.05$ , *Dataset S5*]. Nevertheless, two of the top five most differentiated Denisovan-like haplotypes (associated with both sets of Denisovan-like haplotypes) overlapped a total of 16 protein-coding genes. Ten of those genes had a direct link to the brain and included pivotal developmental genes like *NEUROD2* and *PAX5* (*SI Appendix*). This finding aligns with the finding that among the five GO categories exhibiting the lowest  $P$  values (Denisovan-like sets 1 and 2), three were associated with fear response. We also found that aSNPs associated with the

two highly differentiated haplotypes modified regulatory regions, like 3' and 5'UTRs (untranslated region) for several of those genes (e.g., *IKZF3*, *PPP1R1B*, *SI Appendix* and *Dataset S6*). Although a formal GO enrichment analysis did not yield significant results, the notable concentration of brain-related genes within genomic regions that overlap two of the most differentiated Denisovan-like set 1 and 2 haplotypes exhibiting substantial frequency increases in PNG highlanders suggests a potential role for Denisovan DNA in influencing the biology of the brain in this population. The associated aSNPs demonstrate potential regulatory effects on some of these genes.

Genes located in genomic regions overlapping high-Fst Denisovan-like set 2 haplotypes displaying a higher frequency in lowlanders exhibited a GO enrichment for the category “cellular response to organic substance” (FWER = 0.04, *Dataset S5*). In addition, we found three categories showing borderline significance linked to cytokine and protozoan responses (FWER: 0.06 to 0.07). These three categories encompassed between 3 and 10 haplotype-associated genes that were a subset from the broader pool of 19 haplotype-associated genes identified within the cellular response to organic substance category. This observation implies a potential connection between the specific response to organic substances and pathogenic elements. These findings were specific to Denisovan-like set 2, as several of the haplotypes overlapping with the 17 genes were PNG-specific Neandertal-like. Notably, among the Denisovan-like set 2 haplotypes associated with genes in these categories was a haplotype encompassing four members of the Guanylate-binding proteins family [*GBP1*, *GBP2*, *GBP4*, and *GBP7* (Guanylate-Binding Protein 7)]—proteins crucial in immune response mechanisms (41). This specific haplotype had been previously identified in Melanesians as a candidate for positive selection (6, 27) and carries an archaic variant introducing a missense variant to *GBP7* (Fig. 3B and *SI Appendix*). Additionally, this region shows two distinct haplotypes that are both present in homozygous and heterozygous states in archaic humans, suggesting that this diversity might have been a result of admixture among these groups as well (Fig. 3C). Apart from *GBP7*, three other genes (*DRC7*, *ENTHD1*, and *RTP5*) harbored missense aSNPs associated with high-Fst Denisovan-like set 2 haplotypes with a higher frequency in lowlanders. However, none of these genes were associated with the discussed GO categories. Nevertheless, we identified multiple aSNPs altering regulatory regions of genes linked to those categories, including the 3'UTRs of *GBP4* and *GBP7* (refer to *SI Appendix* and *Dataset S6*), suggesting a potential regulatory effect of those variants on *GBP4* and *GBP7*.

Among the two groups of high-Fst Neandertal-like haplotypes, only the set exhibiting higher frequencies in highlanders displayed enriched GO categories. These categories comprised 22 associations related to regulatory and metabolic functions, largely influenced by members of the zinc finger family located within highlander-specific and high-Fst Neandertal-like haplotypes (*Dataset S5* and *SI Appendix*).

**Phenotypic Inferences of aSNPs Associated with Highly Differentiated Archaic Haplotypes.** To better understand the potential impact of archaic DNA within the genomes of PNG, we conducted an analysis using data from UK Biobank (42) and Biobank Japan (43). This analysis aimed to identify associations involving aSNPs linked to archaic haplotypes reconstructed within the two PNG populations. It is important to note that this approach has limitations due to notably lower levels of Denisovan DNA in UK and Japanese populations compared to PNG, resulting in a significantly reduced number of PNG aSNPs tested in these cohorts (Fig. 3A). Moreover, this limitation was compounded by

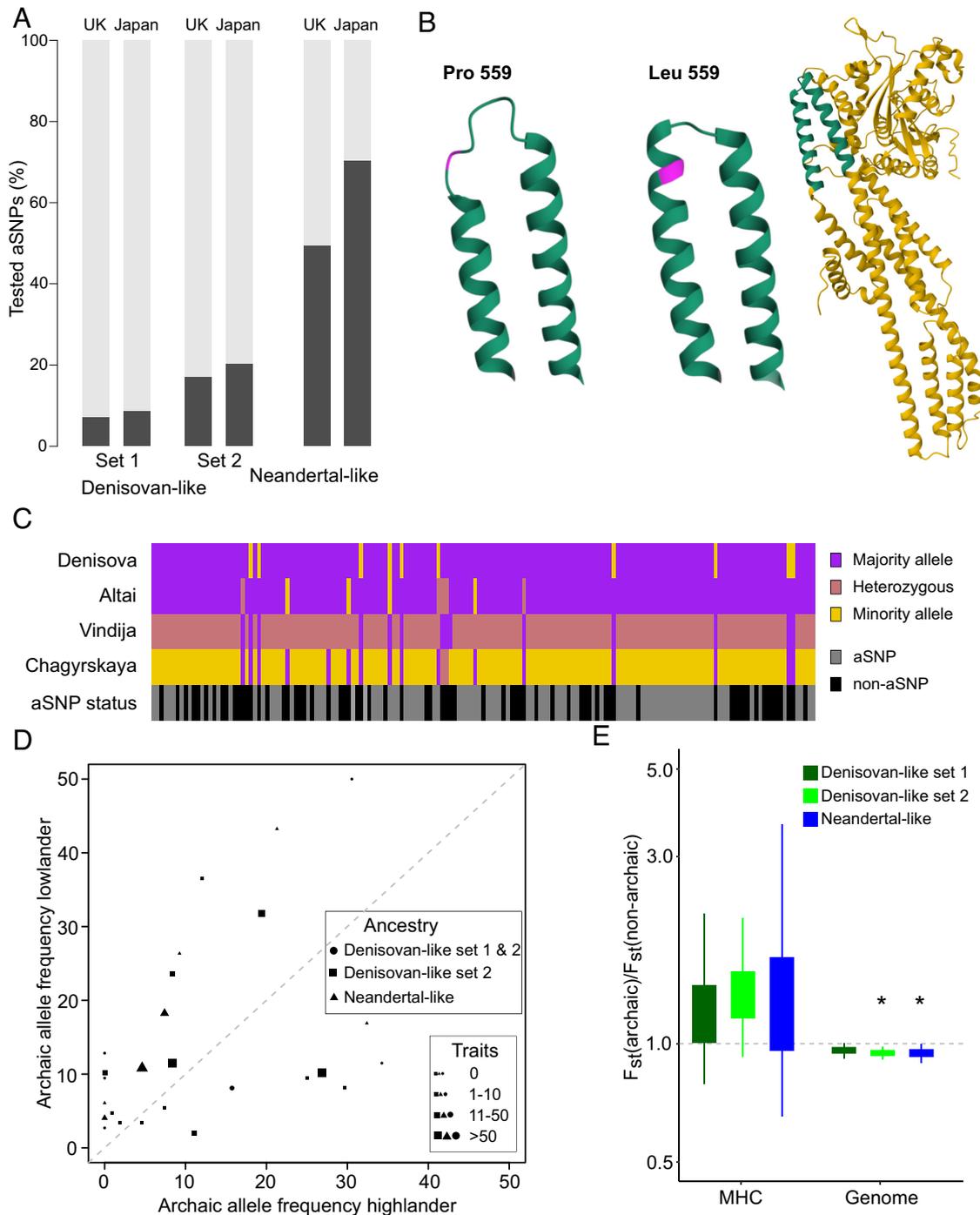
the possibility that aSNPs might have been tested on a different archaic haplotype in the biobanks compared to those in PNG.

In total, 379 archaic haplotypes harbored aSNPs exhibiting significant phenotype associations in at least one of the two biobank cohorts (*Dataset S7*). Among these associations were 48 archaic haplotypes carrying aSNPs that alter the protein sequence of a gene. In general, we observed that a higher percentage of all Denisovan-like haplotypes in set 2 carried missense aSNPs (5.9%) compared to Neandertal-like haplotypes (4.7%,  $P = 0.02$ , Fisher's exact test, *Dataset S8*). A total of 125 of all 568 missense-carrying haplotypes contained multiple missense aSNPs, notably a Denisovan-like (set 1 and 2) haplotype carrying 10 missense aSNPs. This specific haplotype had been previously noted in a region involving *MUC19*, which has shown signs of introgression and selection among both archaic and modern humans (44) (*SI Appendix*). Traits commonly associated with missense-carrying haplotypes included blood biomarkers, measurements related to bone density and body fat, as well as occurrences of diabetes, some of which correlate with the phenotypic difference previously reported for PNG populations.

When we explored aSNP trait associations for all archaic PNG haplotypes, we identified four Denisovan-like haplotypes that exhibited substantial pleiotropy, displaying between 77 and 178 associations with various medical and nondisease phenotypes (*Dataset S7* and Fig. 3D). Intriguingly, all of these haplotypes were situated within the major histocompatibility complex (MHC), a crucial immune-related region in the human genome. Notably, two of these haplotypes ranked among the top 10% in the Fst distribution. In total, we found 19 Denisovan-like and 7 Neandertal-like haplotypes located within the MHC. We found that Denisovan-like (both sets) and Neandertal-like haplotypes showed comparable but nominally higher mean Fst values relative to other sets of frequency-matched nonarchaic sets of SNPs within the MHC (Fig. 3E). The relative mean Fst values for Neandertal-like and Denisovan-like haplotypes compared to random background sets were substantially larger than their respective genome-wide counterparts, which tentatively showed lower mean Fst values compared to nonarchaic random background sets (all  $P < 0.001$ ). While these results indicate a difference in population differentiation between PNG highlanders and lowlanders for archaic haplotypes in the MHC compared to other genomic regions, it is important to interpret these findings within the context of the high evolutionary dynamics characterizing the MHC.

## Discussion

In this study, we investigated the genomic and phenotypic impact of Denisovan and Neandertal DNA within two PNG populations living in distinct environmental regions—the mountainous terrain surrounding Mount Wilhelms and Daru Island. We found that Denisovan-like haplotypes exhibiting the most significant frequency differentiation between both populations and that had risen to higher frequencies among highlanders exceeded the level of population differentiation seen in nonarchaic variants with comparable frequencies. These findings imply that Denisovan DNA played a significant role in adaptive processes for these populations. Highly differentiated haplotypes that exhibited higher frequencies in highlanders overlapped several genes associated with early brain development, a result also observed for some of the most highly differentiated nonarchaic haplotypes in these populations (6). This result might reflect adaptive patterns to highlander-specific environmental factors. For instance, high-altitude-induced hypoxia has been linked to adaptive changes in neurons (45). Similarly, dietary



**Fig. 3.** Phenotypic implications of archaic admixture in PNG populations. (A) The percentage of aSNPs tested in two biobank cohorts, namely the UK Biobank and Biobank Japan, is depicted. The relative distribution of aSNPs is categorized according to the ancestral origin of the haplotype linked to them in PNG. Only aSNPs with a minor allele frequency of at least 1% in both biobank cohorts are considered. (B) A three-dimensional representation of the human GBP7 is depicted on the right. A distinct segment of the protein is emphasized in green. Two renditions of this protein segment are displayed: One reflects the modern human reference protein sequence (*Middle*), while the other showcases an amino acid substitution introduced by the archaic allele chr1:89,132,390:A>G, resulting in a Leucine to Proline alteration at position 559 (*Left*). (C) The genetic composition across 163 variable sites in the Denisovan and three Neandertal individuals within the region of the introgressed Denisovan-like set 2 haplotype at chr1:89,054,418-89,200,767 is depicted. Alleles are color-coded based on their prevalence in the four archaic genomes (major allele in purple, minor allele in yellow, and heterozygous sites in orange). The analysis only covers genetic positions with genotype information available for all four archaic individuals. The *Bottom* panel highlights positions that represent aSNP locations in PNG (gray). (D) Scatterplot depicting the frequencies for sets of archaic haplotypes spanning the Major Histocompatibility Complex (MHC) in PNG highlanders (x-axis) and lowlanders (y-axis). Denisovan-like haplotypes are represented by circles (sets 1 and 2) and squares (set 2), while Neandertal haplotypes are denoted by triangles. The symbol sizes correspond to the number of trait associations associated with aSNPs segregating on these haplotypes (*Methods*). A gray diagonal line has been included to represent the region for haplotypes with identical frequencies in both populations. (E) Frequency distributions illustrate the ratio of both the mean  $F_{st}$  value for genome-wide sets of archaic haplotypes (*Right*) and sets of haplotypes overlapping the MHC (*Left*) compared to the mean  $F_{st}$  values obtained from 1,000 nonarchaic background sets genome-wide and within the MHC, respectively (*Methods*). Whiskers in the boxplots represent the 95% CI for these distributions. Distributions with a significant deviation from one are highlighted by an asterisk ( $P < 0.05$ ).

variations resulting from differences in food availability have been demonstrated to profoundly affect brain development (46). Notably, none of the PNG individuals examined in our study carried a

previously described Denisovan-like haplotype associated with high-altitude adaptation (25). Therefore, at this juncture, it remains challenging to predict the extent and manner in which introgressed

Denisovan DNA impacts these genes and their associated phenotypes. In this context, it is worth noting that Neandertal DNA has been demonstrated to significantly impact several behavioral and neurological phenotypes in its carriers today (19, 21, 47–49). Our findings imply that comparable patterns might also hold true for Denisovan DNA. One set of highly differentiated Denisovan-like haplotypes, present at high frequencies among lowlanders, exhibited a significant overlap with genes associated with pathogen response. One plausible explanation for this result could be the exposure to tropical diseases, such as malaria. A total of 94% of all malaria deaths in the Western Pacific region in 2021 were accounted for in Papua New Guinea (50, 51). Although malaria is highly prevalent in the geographical region, it is nearly absent among PNG highlanders (52). Our findings imply that Denisovan DNA may have played a role in the adaptation to defense of malaria and/or other tropical diseases. The influence of Denisovan-like haplotypes stands out when compared to introgressed Neandertal-like haplotypes. We do not observe a similar degree of population differentiation or functional significance for Neandertal DNA for the most highly differentiated haplotypes in PNG populations. However, we did find that highly differentiated Neandertal-like haplotypes with greater prevalence among highlanders were notably enriched for genes involved in transcriptional processes. These findings align with earlier reports highlighting the importance of Neandertal DNA in the regulation of gene expression (53–57).

Our findings offer potential phenotype-associated candidates that can contribute to a deeper understanding of the role of Denisovan DNA in contemporary adaptive processes. Furthermore, these candidates can serve as a basis for reconstructing the phenotypic characteristics of Denisovans and shedding light on the adaptive mechanisms within this archaic human group. These identified candidates present a valuable resource for functional testing through experimental assays (58, 59) and can collaborate with prediction algorithms to further explore their phenotypic significance (60). Ultimately, a significant expansion of available association data (61) will be a crucial component in advancing our understanding of the phenotypic potential of Denisovan DNA and its contribution to the adaptation of modern humans. It will also bring us another step closer to learning more about our extinct relatives and unique aspects of their biology. In our study, we examined two sets of Denisovan-like introgressed haplotypes. This decision arises from the ongoing challenge of confidently annotating Denisovan DNA within present-day populations. Both sets of Denisovan-like haplotypes are potentially prone to either omitting genuine Denisovan sequences or incorporating misclassified Neandertal haplotypes. The low number of archaic genomes, especially Denisovans, as well as our likely incomplete knowledge about the complex landscape of archaic admixture in Oceania are key factors limiting the assessment of the accuracy of the exact number of true Denisovan haplotypes in both sets. Future research incorporating additional genomic data from modern and archaic humans and enhanced genomic and evolutionary methodologies will be key to improve our understanding of archaic admixture. These advancements will also facilitate a more precise interpretation of our findings, including an evaluation of the accuracy levels of the Denisovan haplotype sets we have generated.

## Methods

**Genomic Datasets.** This study included whole-genome sequencing data (8–43× coverage, [Dataset S1](#); human genome version hg38) derived from a cohort of 128 unrelated adult PNG individuals, consisting of 74 individuals from the lowlands of Daru Island and 54 individuals from the highland region around Mount Wilhelm (6, 62–64). Andre et al. utilized the Broad Institute's

GATK v4.2.0.0 “Germline short variant discovery” pipeline (65) to generate genotype data from a diverse cohort comprising populations from Papua New Guinea, Oceania, and additional individuals of African, European, and Asian descent sourced from the 1000 Genomes Project dataset. To ensure data quality, a coverage-based genomic mask was established to filter out sites with limited accessibility, particularly those derived from next-generation short-read sequencing methods. Subsequently, the genotype dataset was refined to include only biallelic SNPs with a call rate of at least 95%. Phasing was conducted without a reference panel using shapeit v4.2.2, which incorporates an automatic imputation step prior to phasing. Furthermore, we analyzed high-coverage whole-genome sequencing data from the 1,000 Genomes cohort (66) and four archaic humans: the Altai, Chagyrskaya, and Vindija Neandertals (10, 11, 32) and the Denisovan (8). Genotype information for all four archaic genomes was only available for the human genome version hg19. To harmonize our genotype datasets, genotypes were converted to hg38 coordinates using UCSC genome browser's liftover tool (67). In addition, genotypes were filtered using the provided genomic masks from each dataset. Genotype data from the 1,000 Genomes cohort were limited to variable single nucleotide variant positions within the dataset. When combining Yoruba (YRI) population and PNG populations, YRI positions without genotype information in the 1,000 Genomes cohort were considered to be homozygous for the human reference allele. Genotype information for the PNG and 1,000 Genomes individuals was available in phased format with the exception of chromosome X in PNG. We therefore processed the unphased genotype data for chromosome X using the same pipelines as previously used for the autosomes (6). Briefly, we kept biallelic variants that had genotype information for more than 95% of individuals and passed the same quality filters as applied in the autosomal dataset. Next, we filtered out the pseudoautosomal parts PAR1 and PAR2 and phased the remaining variants using shapeit v4.2.2 with default parameters (68).

**Archaic Introgression Map.** We employed a previously established methodology (22) to characterize introgressed archaic haplotypes in individuals of both PNG ( $N = 128$ ) and three 1,000 Genomes (66) Eurasian superpopulations (Europe,  $N = 633$ ; South Asia,  $N = 601$ , East Asia,  $N = 585$ ). This approach identifies these haplotypes by leveraging distinctive characteristics, including shared allele signatures, haplotype structure, and haplotype length, which are indicative of ancestral interbreeding between modern humans and Neandertals as well as Denisovans. Following the approach, we first identified aSNPs within the genomes of four distinct populations under analysis: PNG, East Asians, South Asians, and Europeans. These aSNPs were defined as those containing an allele that met the following criteria: i) it was absent in the 1,000 Genomes Yoruba population, ii) it was present in at least one of the three high-coverage Neandertal genomes [Vindija (11), Chagyrskaya (32), Altai (10)], or the Denisovan genome (8), and iii) it was present in at least one individual within the four populations we were examining. Next, within each of the four populations, we computed pairwise measures of linkage disequilibrium (LD) represented as  $r^2$  between all identified aSNPs within that particular population. We collapsed sets of aSNPs that showed  $r^2 > 0.8$  and defined them as an archaic haplotype. All aSNPs displaying no LD of  $r^2 > 0.8$  with any other aSNP were removed. Subsequently, we calculated the nucleotide distance for all remaining haplotypes by measuring the span between the two furthest aSNPs within each haplotype. We then assessed the compatibility of each haplotype's length with the genomic phenomenon ILS. ILS refers to the retention of ancestral genetic variation shared by some modern and archaic human populations, which predates the divergence of these human groups. As a result, ILS can lead to similar allele-sharing patterns as aSNPs. However, because ILS segments are considerably older, the average length of ILS haplotypes is much shorter compared to Neandertal or Denisovan haplotypes. These archaic haplotypes typically span tens of kilobases in size, reflecting the relatively recent admixture between modern and archaic humans approximately 55,000 B.P. (69). Building upon the methodology introduced by Huerta-Sanchez et al. (25) and incorporating more recent estimates for divergence and mutation rates (70), we computed the likelihood of each inferred haplotype's length being compatible with ILS. This calculation was conducted using recombination rate estimates obtained from two separate cohorts (71, 72). We calculated false discovery rates (FDR) by adjusting the acquired  $P$  values for multiple testing using the Benjamini-Hochberg procedure (73). Haplotypes that exhibited a length compatible with ILS for both recombination maps (FDR > 0.05) or had fewer

than 10 aSNPs were excluded from further analysis. Finally, separately for PNG and each 1,000 Genomes Eurasian superpopulation and using the remaining haplotypes and their corresponding aSNPs, we reconstructed haplotypes within individuals based on each individual's alignment with the archaic allele at the aSNP sites.

Next, we assigned archaic haplotypes to their most probable archaic source by evaluating their sequence similarity with the genomes of three high-coverage Neandertals and the Denisovan. For haplotype alleles that were present in a heterozygous state in the unphased archaic human genomes, we defined a distance of 0.5. Haplotypes were defined as Denisovan-like when they exhibited a stronger sequence affinity with the Denisovan genome than with the other three Neandertals. All remaining haplotypes were categorized as Neandertal-like. Moreover, we further categorized Neandertal-like and Denisovan-like haplotypes within the PNG population based on their presence in any of the three Eurasian populations we examined. Archaic haplotypes in PNG were regarded as shared with Eurasian populations if they exhibited an overlap of at least 80% and were composed of the same set of aSNPs in the overlapping region as the matching archaic haplotypes in Eurasians. Finally, we generated three sets of archaic haplotypes: i) Denisovan-like set 1 which is composed of all haplotypes with the closest sequence similarity to the Denisovan; ii) Denisovan-like set 2, composed of all haplotypes in (i) and all Neandertal-like haplotypes that are found in PNG, but not Eurasia; and iii) Neandertal-like set with all Neandertal-like haplotype shared between PNG and 1,000 Genomes Eurasians.

**Evaluation of Introgression Map.** To assess the performance of our approach, we leveraged HMMIX, an independent method utilizing a Hidden Markov model designed to infer putative introgressed segments (31). We executed the method using default parameters for all 128 PNG samples. To closely mirror the comparison with our approach, we selected the Yoruba population from the 1,000 Genomes cohort as the outgroup. In addition, we used the four archaic genomes to annotate HMMIX's results with shared archaic variants. In comparing our results to the output of HMMIX, we examined the overlap between fragments identified as "Archaic" by HMMIX and our inferred archaic haplotypes, considering overlaps of any length.

**Measures of Population Differentiation.** We assessed the extent of population differentiation between PNG highlanders and lowlanders by calculating *F<sub>st</sub>* values for all identified archaic haplotypes in these two groups. For each haplotype, we randomly selected one candidate aSNP for the analysis, additionally conditioning the selected aSNP to be within the same 1% allele frequency bin as the aSNP with the median allele frequency value of a given archaic haplotype. We used VCFtools (0.1.14) software and computed the Weir and Cockerham *F<sub>st</sub>* estimate (74). Furthermore, we created 1,000 control sets of randomly selected nonarchaic SNPs that matched the number of candidate aSNPs and their frequency distribution within the combined PNG population dataset. Moreover, for aSNPs exhibiting differences in archaic allele frequencies among PNG populations, we selected nonarchaic SNPs matched in frequency, where the allele corresponding to the archaic allele also demonstrates a higher frequency in the PNG population, akin to the archaic allele.

**Computational Reconstruction of Allele Frequency Trajectories for aSNPs.** We employed computational methods to reconstruct allele frequency trajectories, utilizing genomic data from both PNG highlanders and lowlanders, applying a modification of the pipeline detailed in André et al. (6). In brief, our approach involved selecting three random representative aSNPs from each archaic haplotype. Subsequently, we extracted the local genealogical tree for each aSNP, utilizing the RELATE software (75) (v1.1.8). These generated trees served as input for CLUES (76), an approximate full-likelihood method for testing signatures of selection (v1). CLUES facilitated the reconstruction of allele frequencies and assigned log-likelihood ratios (log(LR)) to indicate support for non-neutrality. We did not assess aSNPs with a minor allele frequency below 5%. Furthermore, we excluded data points for which the CLUES algorithm did not yield a log(LR) value.

**Functional Enrichment Analysis.** We performed functional enrichment analysis in the GO (40) using the R package GOfuncR (77). Specifically, we examined six sets of archaic candidate haplotypes that ranked within the top 1 percentile of their respective *F<sub>st</sub>* distributions. These six sets were created by splitting all

archaic haplotypes by i) ancestry (Denisovan-like sets 1 and 2 or Neandertal-like) and ii) frequency (higher in highlanders or lowlanders). For each of the six sets, we defined background haplotypes, which included all remaining haplotypes that did not fall within the top 1 percentile of their respective *F<sub>st</sub>* distributions. To account for the genomic clustering of functionally related gene groups, we ran the GO enrichment software with the parameter `circ_chrom=TRUE`. This setting allowed us to create 10,000 test sets that randomly shifted the coordinates of both candidate and background haplotypes on the circularized version of the same chromosome. FWER were determined by comparing empirical enrichment *P* values for each GO category to the minimum enrichment *P* values observed in the entire GO across all test sets.

**Phenotypic and Regulatory Annotation of aSNPs.** We conducted a screening of phenotypic association data from two biobank cohorts, the UK Biobank (42) (4,280 phenotypes, <https://www.nealelab.is/uk-biobank>) and Biobank Japan (43) (220 phenotypes), to identify associations with aSNPs linked to the archaic haplotypes observed in PNG highlanders and lowlanders. We included in our analysis all associations involving aSNPs with allele frequencies exceeding 1% and meeting the stringent genome-wide significance threshold of  $P < 5 \times 10^{-8}$ . In addition, we leveraged ENSEMBL's variant effect predictor (78) (ensembl-vep version: 109.3) to annotate the aSNPs identified in PNG with their predicted molecular consequences.

**Protein Visualization.** We employed AlphaFold2 (79) to predict the 3D protein folding structures for two variants of GBP7. The first variant was constructed based on the reference protein sequence, while the second variant incorporated a Proline substitution in place of Leucine at amino acid position 559. This substitution was introduced by chr1:89,132,390:A>G aSNP, with Guanine being the archaic allele at this site. This aSNP resides on the Denisovan-like set 2 haplotype (chr1:89,054,418-89,200,767), which encompasses this gene. To visualize the resulting protein structures for both variants, we utilized the Mol\* viewer (80) (Fig. 3B).

**Data, Materials, and Software Availability.** Code availability: [https://github.com/SillySabertooth/papuan\\_archaic\\_admixture/](https://github.com/SillySabertooth/papuan_archaic_admixture/) (81). Previously published data were used for this work (Genotype data Papuan dataset: European Genome-Phenome data repository access: EGAD00001010142, EGAD00001010143, EGAD50000000050 (62-64); 1,000 Genomes dataset: <https://www.internationalgenome.org/>; Archaic humans: <http://cdna.eva.mpg.de/neandertal/>. Phenotype data: UK Biobank GWAS summary statistics: <https://www.nealelab.is/uk-biobank/>; Biobank Japan GWAS summary statistics: <https://phweb.jp/>; ENSEMBL's variant effect predictor: <https://www.ensembl.org/info/docs/tools/vep/index.html>) (82). All other data are included in the article and/or supporting information.

**ACKNOWLEDGMENTS.** This work was supported by the Estonian Research Council grant TK(TK214). D.Y., M.A., V.P., and M.D. were supported by the European Union through Horizon 2020 Research and Innovation Program under Grant No. 810645 and the European Union through the European Regional Development Fund Project No. MOBEC008. N.B. and F.-X.R. were supported by the French Ministry of Research grant Agence Nationale de la Recherche (ANR PAPUADEVOL 20-CE12-0003-01), the French Ministry of Foreign and European Affairs (French Prehistoric Mission in Papua New Guinea), and the Labex TULIP, France, and the Leakey foundation. We would like to thank Hernán A. Burbano and Kay Prüfer for commenting on this manuscript. Data analyses were carried out in part in the High-Performance Computing Center of the University of Tartu.

Author affiliations: <sup>a</sup>Center of Genomics, Evolution and Medicine, Institute of Genomics, University of Tartu, Tartu 51010, Estonia; <sup>b</sup>Centre de Recherche sur la Biodiversité et l'Environnement, Université de Toulouse, Centre National de la Recherche Scientifique, Institut de Recherche pour le Développement, Toulouse Institut National Polytechnique, Université Toulouse 3-Paul Sabatier, cedex 9, Toulouse 31062, France; <sup>c</sup>Strand of Anthropology, Sociology and Archaeology, School of Humanities and Social Sciences, University of Papua New Guinea, PO Box 320, University 134, National Capital District, Papua New Guinea; <sup>d</sup>School of Social Science, University of Queensland, St. Lucia, QLD 4072, Australia; <sup>e</sup>The Australian Research Council Centre of Excellence for Australian Biodiversity and Heritage & College of Arts, Society and Education, James Cook University, Cairns, QLD 4870, Australia; and <sup>f</sup>Institute of Clinical Molecular Biology, Christian-Albrechts-Universität zu Kiel, Kiel 24118, Germany

1. C. Clarkson *et al.*, Human occupation of northern Australia by 65,000 years ago. *Nature* **547**, 306–310 (2017).
2. J. F. O'Connell *et al.*, When did Homo sapiens first reach Southeast Asia and Sahul? *Proc. Natl. Acad. Sci. U.S.A.* **115**, 8482–8490 (2018).
3. N. Brucato *et al.*, Papua New Guinean genomes reveal the complex settlement of North Sahul. *Mol. Biol. Evol.* **38**, 5107–5121 (2021).
4. G. R. Summerhayes, J. H. Field, B. Shaw, D. Gaffney, The archaeology of forest exploitation and change in the tropics during the Pleistocene: The case of Northern Sahul (Pleistocene New Guinea). *Quat. Int.* **448**, 14–30 (2017).
5. M. André *et al.*, Phenotypic differences between highlanders and lowlanders in Papua New Guinea. *PLoS One* **16**, e0253921 (2021).
6. M. André *et al.*, Positive selection in the genomes of two Papua New Guinean populations at distinct altitude levels. *Nat. Commun.* **15**, 3352 (2024).
7. J. Choin *et al.*, Genomic insights into population history and biological adaptation in Oceania. *Nature* **592**, 583–589 (2021).
8. M. Meyer *et al.*, A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222–226 (2012).
9. M. Larena *et al.*, Philippine Ayta possess the highest level of Denisovan ancestry in the world. *Curr. Biol.* **31**, 4219–4230.e10 (2021).
10. K. Prüfer *et al.*, The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014).
11. K. Prüfer *et al.*, A high-coverage Neanderthal genome from Vindija Cave in Croatia. *Science* **358**, 655–658 (2017).
12. S. Peyrégne, V. Slon, J. Kelso, More than a decade of genetic research on the Denisovans. *Nat. Rev. Genet.* **25**, 83–103 (2023), 10.1038/s41576-023-00643-4.
13. G. H. Perry, L. Kistler, M. A. Kelaita, A. J. Sams, Insights into hominin phenotypic and dietary evolution from ancient DNA sequence data. *J. Hum. Evol.* **79**, 55–63 (2015).
14. D. Gokhman *et al.*, Reconstructing Denisovan anatomy using DNA methylation maps. *Cell* **179**, 180–192.e10 (2019).
15. J. C. Teixeira, A. Cooper, Using hominin introgression to trace modern human dispersals. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 15327–15332 (2019).
16. S. R. Browning, B. L. Browning, Y. Zhou, S. Tucci, J. M. Akey, Analysis of human sequence data reveals two pulses of archaic Denisovan admixture. *Cell* **173**, 53–61.e9 (2018).
17. M. Mondal, J. Bertranpetit, O. Lao, Approximate Bayesian computation with deep learning supports a third archaic introgression in Asia and Oceania. *Nat. Commun.* **10**, 246 (2019).
18. G. S. Jacobs *et al.*, Multiple deeply divergent Denisovan ancestries in Papuans. *Cell* **177**, 1010–1021.e32 (2019).
19. E. McArthur, D. C. Rinker, J. A. Capra, Quantifying the contribution of Neanderthal introgression to the heritability of complex traits. *Nat. Commun.* **12**, 4481 (2021).
20. M. Dannemann, J. Kelso, The contribution of Neanderthals to phenotypic variation in modern humans. *Am. J. Hum. Genet.* **101**, 578–589 (2017).
21. C. N. Simonti *et al.*, The phenotypic legacy of admixture between modern humans and Neanderthals. *Science* **351**, 737–741 (2016).
22. M. Dannemann, The population-specific impact of neanderthal introgression on human disease. *Genome Biol. Evol.* **13**, evaa250 (2021).
23. L. Skov *et al.*, The nature of Neanderthal introgression revealed by 27,566 Icelandic genomes. *Nature* **582**, 78–83 (2020).
24. D. Koller *et al.*, Denisovan and Neanderthal archaic introgression differentially impacted the genetics of complex traits in modern populations. *BMC Biol.* **20**, 249 (2022).
25. E. Huerta-Sánchez *et al.*, Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* **512**, 194–197 (2014).
26. F. Racimo *et al.*, Archaic adaptive introgression in TBX15/WARS2. *Mol. Biol. Evol.* **34**, 509–524 (2017).
27. B. Vernot *et al.*, Excavating Neanderthal and Denisovan DNA from the genomes of Melanesian individuals. *Science* **352**, 235–239 (2016).
28. S. Sankararaman, S. Mallick, N. Patterson, D. Reich, The combined landscape of Denisovan and Neanderthal ancestry in present-day humans. *Curr. Biol.* **26**, 1241–1247 (2016).
29. D. M. Vespasiani *et al.*, Denisovan introgression has shaped the immune system of present-day Papuans. *PLoS Genet.* **18**, e1010470 (2022).
30. N. Brucato *et al.*, Chronology of natural selection in Oceanian genomes. *iScience* **25**, 104583 (2022).
31. L. Skov *et al.*, Detecting archaic introgression using an unadmixed outgroup. *PLoS Genet.* **14**, e1007641 (2018).
32. F. Mafessoni *et al.*, A high-coverage Neanderthal genome from Chagyrskaya cave. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 15132–15136 (2020).
33. B. Vernot, J. M. Akey, Complex history of admixture between modern humans and Neanderthals. *Am. J. Hum. Genet.* **96**, 448–453 (2015).
34. V. Slon *et al.*, The genome of the offspring of a Neanderthal mother and a Denisovan father. *Nature* **561**, 113–116 (2018).
35. D. Glavan *et al.*, Identification of transcriptome alterations in the prefrontal cortex, hippocampus, amygdala and hippocampus of suicide victims. *Sci. Rep.* **11**, 18853 (2021).
36. D. Gaffney, A. Ford, G. Summerhayes, Crossing the Pleistocene-Holocene transition in the New Guinea Highlands: Evidence from the lithic assemblage of Kiowa rockshelter. *J. Anthropol. Archaeol.* **39**, 223–246 (2015).
37. J. E. Cotes, H. R. Anderson, J. M. Patrick, Lung function and the response to exercise in the New Guineans: Role of genetic and environmental factors. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **268**, 349–361 (1974).
38. N. Senn *et al.*, Population hemoglobin mean and anemia prevalence in Papua New Guinea: New metrics for defining malaria endemicity? *PLoS One* **5**, e9375 (2010).
39. A. J. Woolcock, M. H. Colman, C. R. Blackburn, Factors affecting normal values for ventilatory lung function. *Am. Rev. Respir. Dis.* **106**, 692–709 (1972).
40. M. Ashburner *et al.*, Gene Ontology: Tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
41. K. Tretina, E.-S. Park, A. Maminska, J. D. MacMicking, Interferon-induced guanylate-binding proteins: Guardians of host defense in health and disease. *J. Exp. Med.* **216**, 482–500 (2019).
42. C. Bycroft *et al.*, The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
43. A. Nagai *et al.*, Overview of the BioBank Japan Project: Study design and profile. *J. Epidemiol.* **27**, S2–S8 (2017).
44. F. A. Villanea *et al.*, The MUC19 gene in Denisovans, Neanderthals, and modern humans: An evolutionary history of recurrent introgression and natural selection. *bioRxiv* [Preprint] (2023). 10.1101/2023.09.25.559202 (Accessed 2 May 2024).
45. C. A. Eichstaedt *et al.*, Genetic and phenotypic differentiation of an Andean intermediate altitude population. *Physiol. Rep.* **3**, e12376 (2015).
46. S. E. Cusick, M. K. Georgieff, The role of nutrition in brain development: The golden opportunity of the "First 1000 Days". *J. Pediatr.* **175**, 16–21 (2016).
47. M. Dannemann *et al.*, Neanderthal introgression partitions the genetic landscape of neuropsychiatric disorders and associated behavioral phenotypes. *Transl. Psychiatry* **12**, 433 (2022).
48. Z. Chen *et al.*, The contribution of Neanderthal introgression and natural selection to neurodegenerative diseases. *Neurobiol. Dis.* **180**, 106082 (2023).
49. C. A. Trujillo *et al.*, Reintroduction of the archaic variant of *NOVA1* in cortical organoids alters neurodevelopment. *Science* **371**, eaax2537 (2021).
50. W. H. Organization, World malaria report 2022 (World Health Organization, 2022). <https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2022>. Accessed 2 May 2024.
51. U. Kitur, T. Adair, I. Riley, A. D. Lopez, Estimating the pattern of causes of death in Papua New Guinea. *BMC Public Health* **19**, 1322 (2019).
52. I. Müller, M. Bockarie, M. Alpers, T. Smith, The epidemiology of malaria in Papua New Guinea. *Trends Parasitol.* **19**, 253–259 (2003).
53. D. Yermakovich *et al.*, Long-range regulatory effects of Neanderthal DNA in modern humans. *Genetics* **223**, iyac188 (2022), 10.1093/genetics/iyac188.
54. M. Dannemann, K. Prüfer, J. Kelso, Functional implications of Neanderthal introgression in modern humans. *Genome Biol.* **18**, 61 (2017).
55. H. Quach *et al.*, Genetic adaptation and neanderthal admixture shaped the immune system of human populations. *Cell* **167**, 643–656.e17 (2016).
56. M. Silver, L. Quintana-Murci, M. Rotival, Impact and evolutionary determinants of neanderthal introgression on transcriptional and post-transcriptional regulation. *Am. J. Hum. Genet.* **104**, 1241–1250 (2019).
57. R. C. McCoy, J. Wakefield, J. M. Akey, Impacts of neanderthal-introgressed sequences on the landscape of human gene expression. *Cell* **168**, 916–927.e12 (2017).
58. I. Gallego Romero, A. J. Lea, Leveraging massively parallel reporter assays for evolutionary questions. *Genome Biol.* **24**, 26 (2023).
59. A. A. Pollen, U. Kilik, C. B. Lowe, J. G. Camp, Human-specific genetics: New tools to explore the molecular and cellular basis of human evolution. *Nat. Rev. Genet.* **24**, 687–711 (2023).
60. C. M. Brand, L. L. Colbran, J. A. Capra, Predicting archaic hominin phenotypes from genomic data. *Annu. Rev. Genomics Hum. Genet.* **23**, 591–612 (2022).
61. H. M. Natri *et al.*, Genetic architecture of gene regulation in Indonesian populations identifies QTLs associated with global and local ancestries. *Am. J. Hum. Genet.* **109**, 50–65 (2022).
62. M. Leavesley *et al.*, Papua New Guinean Genome Altitude Project Dataset 2. European Genome-phenome Archive. <https://ega-archive.org/datasets/EGAD00001010142>. Accessed 2 May 2024.
63. M. Leavesley *et al.*, Papua New Guinean Genome Altitude Project Dataset 1. European Genome-phenome Archive. <https://ega-archive.org/datasets/EGAD00001010143>. Accessed 2 May 2024.
64. M. Leavesley *et al.*, Papua New Guinean Lowlanders dataset. European Genome-phenome Archive. <https://ega-archive.org/datasets/EGAD500000000050>. Accessed 2 May 2024.
65. R. Poplin *et al.*, Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv* [Preprint] (2017). 10.1101/201178 (Accessed 2 May 2024).
66. M. Byrska-Bishop *et al.*, High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. *Cell* **185**, 3426–3440.e19 (2022).
67. W. J. Kent *et al.*, The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
68. O. Delaneau, J.-F. Zagury, M. R. Robinson, J. L. Marchini, E. T. Dermitzakis, Accurate, scalable and integrative haplotype estimation. *Nat. Commun.* **10**, 5436 (2019).
69. S. Sankararaman, N. Patterson, H. Li, S. Pääbo, D. Reich, The date of interbreeding between Neanderthals and modern humans. *PLoS Genet.* **8**, e1002947 (2012).
70. M. Dannemann, A. M. Andrés, J. Kelso, Introgression of neanderthal- and Denisovan-like haplotypes contributes to adaptive variation in human toll-like receptors. *Am. J. Hum. Genet.* **98**, 22–33 (2016).
71. A. Kong *et al.*, A high-resolution recombination map of the human genome. *Nat. Genet.* **31**, 241–247 (2002).
72. International HapMap 3 Consortium *et al.*, Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).
73. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc.* **57**, 289–300 (1995).
74. P. Danecek *et al.*, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
75. L. Speidel, M. Forest, S. Shi, S. R. Myers, A method for genome-wide genealogy estimation for thousands of samples. *Nat. Genet.* **51**, 1321–1329 (2019).
76. A. J. Stern, P. R. Wilton, R. Nielsen, An approximate full-likelihood method for inferring selection and allele frequency trajectories from DNA sequence data. *PLoS Genet.* **15**, e1008384 (2019).
77. S. Grote, GOfuncR: Gene ontology enrichment using FUNC (R package version 1.24.0, 2018). <https://www.bioconductor.org/packages/3.8/bioc/html/GOfuncR.html>. Accessed 2 May 2024.
78. W. McLaren *et al.*, The ensembl variant effect predictor. *Genome Biol.* **17**, 122 (2016).
79. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
80. D. Sehnal *et al.*, Mol\* Viewer: Modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Res.* **49**, W431–W437 (2021).
81. D. Yermakovich, Software for project "Denisovan admixture facilitated environmental adaptation in Papua New Guinean populations". GitHub. [https://github.com/SillySabertooth/papuan\\_archaic\\_admixture](https://github.com/SillySabertooth/papuan_archaic_admixture). Accessed 2 May 2024.
82. W. McLaren *et al.*, Variant Effect Predictor. ENSEMBL. <https://www.ensembl.org/info/docs/tools/vep/index.html>. Accessed 2 May 2024.