



HAL
open science

Denisovan admixture facilitated environmental adaptation in Papua New Guinean populations

Danat Yermakovich, Mathilde André, Nicolas Brucato, Jason Kariwiga, Matthew Leavesley, Vasili Pankratov, Mayukh Mondal, Francois-Xavier Ricaut, Michael Dannemann

► **To cite this version:**

Danat Yermakovich, Mathilde André, Nicolas Brucato, Jason Kariwiga, Matthew Leavesley, et al.. Denisovan admixture facilitated environmental adaptation in Papua New Guinean populations. Proceedings of the National Academy of Sciences of the United States of America, 2024, 121 (26), pp.e2405889121. 10.1073/pnas.2405889121 . hal-04616634

HAL Id: hal-04616634

<https://hal.science/hal-04616634v1>

Submitted on 19 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 Denisovan admixture facilitated environmental adaptation in
2 Papua New Guinean populations

3

4 **Authors**

5

6 Danat Yermakovich¹, Mathilde André¹, Nicolas Brucato², Jason Kariwiga^{3,4}, Matthew
7 Leavesley^{3,5}, Vasili Pankratov¹, Mayukh Mondal^{1,6}, François-Xavier Ricaut^{2,*}, Michael
8 Dannemann^{1,*}

9

10 1 Institute of Genomics, University of Tartu, Riia 23b, 51010 Tartu, Estonia

11 2 Centre de Recherche sur la Biodiversité et l'Environnement (CRBE), Université de
12 Toulouse, CNRS, IRD, Toulouse INP, Université Toulouse 3 – Paul Sabatier (UT3),
13 Toulouse, France

14 3 Strand of Anthropology, Sociology and Archaeology, School of Humanities and Social
15 Sciences, University of Papua New Guinea, PO Box 320, University 134, National Capital
16 District, Papua New Guinea

17 4 School of Social Science, University of Queensland, St Lucia, Queensland, Australia

18 5 CABAH & College of Arts, Society and Education, James Cook University, PO Box 6811,
19 Cairns, QLD 4870, Australia

20 6 Institute of Clinical Molecular Biology, Christian-Albrechts-Universität zu Kiel 24118 Kiel,
21 Germany

22 * Authors contributed equally

23 Corresponding authors: Michael Dannemann (michael.dannemann@ut.ee) and François-
24 Xavier Ricaut (francois-xavier.ricaut@univ-tlse3.fr)

25 **Classification**

26 Biological Sciences

27 Anthropology

28 **Keywords**

29 Human evolution; Archaic admixture; Selection

30 **Abstract**

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

50 **Significance statement**

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

61 **Main**

62

63 The initial human settlement of New Guinea is estimated to have occurred by at least 50
64 thousand years ago (1, 2). Today, the distribution of PNG populations across the region is
65 uneven, often occurring in areas characterized by significant environmental disparities (3, 4).
66 These environmental challenges, such as exposure to high altitudes or region-specific
67 pathogens, have been demonstrated to correlate with phenotypic variations among PNG
68 populations inhabiting distinct environments (5). These challenges have also been
69 established as factors contributing to the emergence of local genetic adaptation signatures
70 (6, 7). Moreover, PNG, like their counterparts in near and remote Oceania, bear a significant
71 share of approximately 3-4% Denisovan DNA, ranking among the highest proportions
72 globally (8, 9). This is in addition to the ~2% of Neandertal ancestry that are found in PNG
73 and all present-day non-Africans (10, 11). Our understanding of particularly the functional
74 aspects of Denisovan DNA in its carriers and its potential contribution to adaptive processes
75 in populations, such as the PNG, remains limited. This restricted knowledge is attributable to
76 a variety of challenges. Firstly, the scarcity of fossil fragments has hindered the ability to
77 comprehensively reconstruct Denisovan physiology and their historical geographic
78 distribution (12). These aspects make it challenging to postulate hypotheses regarding their
79 functions in present-day individuals. Some insights have been derived from the Denisovan
80 genome sequence (8), such as the presence of only one copy of the amylase gene - a
81 feature shared with Neandertals, hinting at potential differences in starch digestion
82 compared to modern humans (13). Additionally, attempts to predict Denisovan phenotypes
83 from genomic data have provided new insights into their skeletal physiology (14). Due to the
84 limited fossil record originating from only a handful of locations, predicting the precise
85 geographical range of Denisovans presents another significant challenge. However, both the
86 distribution of these fossils and the distribution of populations carrying Denisovan DNA
87 collectively suggest a relatively broad habitat (12, 15). It is plausible that Denisovans may
88 have inhabited various regions across the Asian continent, spanning from Siberia to
89 Southeast Asia. The existing uncertainty about their geographical distribution further
90 complicates the formulation of hypotheses regarding potential environmental adaptations of
91 Denisovans. The interbreeding between modern humans and Denisovans has been
92 suggested to occur in multiple waves involving distinct Denisovan populations (7, 16–18).
93 These findings may imply the existence of diverse Denisovan groups with potentially
94 differing genetic compositions, adding another layer of complexity to the interpretation of
95 their phenotypic characteristics. Another challenge arises when attempting to annotate the
96 phenotypic effects of Denisovan DNA in contemporary individuals. In the case of
97 introgressed Neandertal DNA, the utilization of phenotypic association data played a pivotal

98 role in studying its influence on modern humans and drawing potential insights into
99 Neandertal biology (19–23). However, unlike Neandertal DNA, which is present in
100 populations with accessible GWAS data, Denisovan DNA lacks readily available association
101 information from phenotype and expression cohorts, rendering such annotation approaches
102 less practicable. Nonetheless, several studies have provided valuable insights into how
103 Denisovan DNA has influenced phenotypic variation and facilitated adaptation in modern
104 humans (24–30). Notably, among them are well-documented instances where Denisovan
105 DNA has been associated with adaptations related to high-altitude environments, metabolic
106 processes, and immune responses. The environmental challenges encountered by PNG as
107 they dispersed across various regions could potentially mirror the challenges that
108 Denisovans themselves had to contend with. Analyzing the fate of Denisovan DNA within
109 PNG populations residing in diverse environmental conditions may offer a pathway to assess
110 the functional capacity of Denisovan DNA and unveil some of its adaptive potential.

111 **Results**

112

113 *Identification of archaic DNA in PNG highlanders and lowlanders*

114

115 In this study, our objective was to evaluate the degree to which DNA inherited from past
116 interbreeding with archaic humans has contributed to shaping local adaptation of two
117 geographically distinct PNG populations, each confronting unique environmental challenges.
118 To address this task, we characterized the genomic landscape of Neandertal and Denisovan
119 DNA in the genomes of 74 PNG individuals residing in the lowlands of Daru Island and 54
120 individuals inhabiting the highlands of Mount Wilhelm (Supplementary Table 1, Methods)(6).
121 The habitats of both populations are marked by notable environmental differences, primarily
122 in terms of altitude. Individuals from the PNG highlander group, included in this study, reside
123 at elevations ranging from 2,300 to 2,700 meters above sea level. In contrast, the PNG
124 lowlanders residing on Daru Island, located off the southern coast of Papua New Guinea,
125 inhabit an environment approximately 100 meters above sea level. These altitude
126 differentials are associated with various other environmental factors, including variations in
127 food availability and pathogen exposure. We reconstructed introgressed archaic haplotypes
128 utilizing a previously established method to detect archaic DNA within contemporary
129 populations (22). This approach identifies archaic SNPs (aSNPs) within the genomes of
130 present-day individuals. These aSNPs function as markers of introgressed variants, with
131 distinctive attributes such as allele sharing signatures with Neandertals and Denisovans.
132 Moreover, these aSNPs reside on haplotypes of a length that exceeds those of segments
133 that result from incomplete lineage sorting (ILS) between modern and archaic humans. ILS
134 segments can produce comparable allele-sharing patterns to those seen in introgressed
135 haplotypes but are on average considerably older and shorter (Methods). When employing
136 this methodology in the analysis of PNG genomes within our study, we identified 168,152
137 aSNPs (Supplementary Table 2). These aSNPs were associated with 10,431 unique core
138 haplotypes spanning across 50.0 to 65.5 megabases (Mb) of diploid archaic DNA within the
139 autosome and chromosome X of the analyzed PNG individuals.

140

141 *Replication of introgression map results with an alternative method*

142

143 We were able to replicate a large fraction of our identified archaic haplotypes using another
144 alternative method (HMMIX) designed for reconstructing archaic DNA in contemporary
145 populations (31). This alternative approach employs a hidden Markov model to identify
146 genomic regions characterized by a high density of SNPs that are absent in an unadmixed
147 outgroup population. A significant proportion of 92.0% of the archaic haplotypes we identified

148 were also captured by the alternative approach (with a posterior probability exceeding 0.8;
149 96.0% with posterior probability > 0.5; Methods). Generally, the alternative method detected
150 a notably higher count of haplotypes (total haplotypes in PNG with posterior probability >
151 0.8, HMMIX: 483,376; our method: 172,121, Supplementary Figure 1a). However, many
152 HMMIX-specific haplotypes with a posterior probability larger than 0.8 were characterized by
153 an absence of archaic SNPs (81.2% HMMIX-specific haplotypes carried no archaic SNPs),
154 possibly due to the underlying method of HMMIX which doesn't involve archaic genome
155 information. These haplotypes also exhibited shorter lengths compared to the ones we
156 observed and omitted because their lengths aligned with incomplete lineage sorting
157 (Supplementary Figure 1b-c).

158

159 *Most likely archaic source of introgressed haplotypes*

160

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

165

166 We observed that, on average, ~61% of the haplotypes per individual displayed a closer
167 sequence similarity to Neandertals than to the Denisovan individual (Figure 1a). This result
168 stands in stark contrast to the anticipated genome-wide estimates of Neandertal ancestry
169 (~2%) and Denisovan ancestry (~4%) within these populations (8, 11, 18). Yet, this figure
170 aligns with prior studies that have attributed this disparity to the substantial differences
171 between the sequenced Denisovan and the introgressing Denisovan population (27, 28). To
172 examine potential factors contributing to the variance between our haplotype-based ancestry
173 estimates and those derived from genome-wide analysis, our study investigated the
174 geographical distribution of archaic haplotypes in PNG (Supplementary Information). First,
175 assuming that PNG share one major pulse of Neandertal admixture with other non-African
176 populations and possibly a second minor pulse with present-day mainland Asians (33), we
177 scanned for the presence of Neandertal-like haplotypes from PNG individuals in present-day
178 Eurasian individuals that match the aSNP content and genomic location (Methods, Figure
179 1a). Our findings revealed that roughly half of these Neandertal-like haplotypes were
180 exclusively found in PNG. One possible explanation for the presence of these PNG-specific
181 Neandertal haplotypes is that they originated from shared Neandertal admixture pulses but
182 subsequently disappeared from Eurasian populations. However, in our analysis of the
183 proportion of unique Neandertal variants in both 1,000 Genomes populations and super-
184 populations, we estimate that such variants can account for no more than 20% of the total

185 PNG-specific Neandertal variants (Supplementary Information). To further explore
186 alternative sources for the PNG-specific fraction of Neandertal-like haplotypes, we assessed
187 their sequence relationship with Neandertals and the Denisovan and compared those
188 measures to the sequence similarity estimates of Neandertal-like haplotypes in PNG that are
189 also found in 1,000 Genomes Eurasians. We found that PNG-specific Neandertal-like
190 haplotypes exhibited a closer sequence relationship to the Denisovan individual (Figure 1b)
191 and displayed greater sequence divergence from Neandertals (Figure 1c) compared to the
192 PNG Neandertal-like haplotypes identified in Eurasians. These results suggest that some of
193 the PNG-specific Neandertal-like haplotypes may have originated from an archaic population
194 that had a closer sequence similarity to the Denisovan and a larger divergence from
195 Neandertal genomes compared to Neandertal-like haplotype found in PNG and 1,000
196 Genomes Eurasians. Several scenarios are consistent with this observation. For instance, a
197 scenario involving admixture between PNG and an archaic population carrying both
198 considerable Neandertal and Denisovan ancestry would support these findings. However, to
199 date, no such scenarios have been reported. Moreover, there are factors that might lead to
200 the misidentification of the ancestral origins of archaic haplotypes. For example, previous
201 studies have reported multiple pulses of admixture between Oceanian populations and
202 distinct Denisovan groups (7, 16, 18). Those introgressing Denisovan groups showed
203 substantially different sequence affinities with the sequenced Denisovan. Furthermore, the
204 large sequence divergence of the introgressing Denisovan population compared to the
205 sequenced Denisovan is in contrast with the comparably closer sequence affinity of high-
206 coverage Neandertal genomes to the introgressing Neandertal population. A previous study
207 has shown that this contrast and other factors such as incomplete lineage sorting (ILS) pose
208 an additional challenge in accurately annotating archaic ancestries to introgressed
209 haplotypes, particularly those of true Denisovan ancestry (27). Finally, the higher number of
210 available high-coverage Neandertal genomes might potentially further amplify the issue of
211 misclassification due to ILS. It's also worth noting that admixture between archaic human
212 groups has previously been reported (10, 34). For example, the Altai Neandertal has been
213 demonstrated to harbor Denisovan ancestry in its genome, introducing an additional
214 challenge and further increasing the likelihood of misclassifying genuine Denisovan
215 introgressed haplotypes.

216

217 Considering all these observations collectively, we hypothesize that the PNG-specific set of
218 haplotypes with a closer sequence similarity to Neandertals is composed of a mix of true
219 Neandertal and misclassified Denisovan haplotypes. Based on that hypothesis, in our study,
220 we opted to categorize different sets of archaic haplotypes. Firstly, archaic haplotypes
221 displaying a closer sequence similarity to the Denisovan than to Neandertals were defined

222 as 'Denisovan-like set 1' - a set of archaic haplotypes with a high likelihood of Denisovan
223 ancestry, but likely lacking a substantial proportion of misclassified Denisovan haplotypes.
224 Secondly, alongside 'Denisovan-like set 1', we introduced 'Denisovan-like set 2', comprising
225 PNG-specific haplotypes closer in sequence to Neandertals and all haplotypes in
226 'Denisovan-like set 1' - a more comprehensive set of genuine Denisovan haplotypes
227 compared to set 1, but also comprising a fraction of true PNG-specific Neandertal
228 haplotypes. Thirdly, we defined 'Neandertal-like' haplotypes in PNG as those exhibiting a
229 closer sequence similarity to at least one Neandertal genome compared to the Denisovan
230 genome and were also present in 1,000 Genomes Eurasian populations - a set of archaic
231 haplotypes with a high likelihood of Neandertal ancestry. The use of each of these three sets
232 in our analyses possesses distinct advantages and limitations, which we thoroughly discuss
233 in the Supplementary Information. We would like to note that Denisovan-like set 2 accounted
234 for 69.6% of all archaic haplotypes in PNG - a number substantially closer to the figure
235 derived from other genome-wide ancestry estimates (Supplementary Information). Our
236 analysis revealed that PNG highlanders and lowlanders carry similar amounts of Neandertal-
237 like DNA ($P=0.98$, Mann-Whitney-U test). However, notably, PNG highlanders exhibited
238 ~1.5% (Denisovan-like set 1, $P=0.10$) and ~3% (Denisovan-like set 2, $P=0.003$, Figure 1d,
239 Supplementary Table 1) more Denisovan-like DNA relative to the number of their lowland
240 counterparts.

241

242 *Genomic distribution of archaic haplotypes*

243

244 Introgressed segments combined from all studied PNG individuals covered approximately
245 21.15% of the human genome. Among those segments we identified 21 archaic haplotypes
246 characterized by archaic allele frequencies surpassing 70% (99.8 percentile) in the
247 combined dataset of the two PNG populations. Interestingly, only eight of these haplotypes
248 exhibit overlaps with protein-coding genes, posing a challenge in establishing a functional
249 link of those haplotypes on the biology of their carriers (Supplementary Table 3). Among the
250 top five haplotypes featuring the highest archaic allele frequencies, only the Neandertal-like
251 haplotype with the highest frequency of 89% (chr9:112,041,006-112,068,271) shares an
252 overlap with a protein coding gene, namely *SUSD1*. Notably, *SUSD1* remains relatively
253 understudied, although a recent investigation has suggested its involvement in
254 neurodevelopmental disorders (35).

255

256 Additionally, we studied the archaic DNA content within regions previously identified as
257 potential sites for negative selection acting on introgressed DNA (27, 28). Our analysis
258 confirmed that, within all five autosomal genomic regions previously reported to be devoid of

259 archaic DNA, there was either an absence or only negligible traces of archaic DNA present
260 (Supplementary Figure 2). Consistent with these previous studies we also noted
261 substantially lower levels of archaic DNA on the X chromosome (Supplementary Figure 2).
262 PNG individuals exhibited an average 7.8-fold decrease in archaic DNA content on
263 chromosome X in comparison to autosomes. This reduction was more prominent for
264 Neandertal-like DNA, with a 12.6-fold decrease, compared to Denisovan-like DNA, which
265 showed a 11.8-fold and 6.9-fold reduction for sets 1 and 2, respectively (Figure 1e-g).

266

267 *Signatures of local adaptation in PNG lowlanders and highlanders*

268

269 Next, we aimed to investigate the extent of population differentiation between the PNG
270 highlanders and lowlanders. This analysis was designed to provide us with insights into the
271 potential impact of selection on archaic haplotypes within each of these two examined
272 populations. To achieve this, we conducted a comparative analysis of the allele frequency
273 difference for each distinct archaic haplotype within both the PNG highlander and lowlander
274 populations by calculating their respective F_{st} values (Methods, Figure 2a, Supplementary
275 Table 3). We then determined whether archaic haplotypes exhibited altered levels of
276 population differentiation between these two examined groups by comparing their F_{st}
277 distribution to those of a genomic background of 1,000 randomly generated sets of non-
278 archaic variants, chosen to match the allele frequency distribution of the archaic haplotypes
279 (Methods, Figure 2b). We also required that the alleles within the random sets of non-archaic
280 variants matching the archaic allele frequency must also exhibit a higher frequency within
281 the respective PNG population. Overall, our analysis revealed that the mean F_{st} value for
282 archaic haplotypes (0.0178) was significantly lower than the mean values observed in the
283 random non-archaic sets (ranging from 0.018 to 0.019, $P < 0.001$). This trend was consistent
284 for both Neandertal-like ($P = 0.04$) and Denisovan-like (set 1: $P = 0.002$, set 2: $P = 0.08$)
285 haplotypes. When partitioning the F_{st} distribution according to whether the higher archaic
286 allele frequency was detected in highlanders or lowlanders, we found consistently lower
287 mean values for all sets of archaic haplotypes (P values ranging from 0.004 to 0.63) When
288 we stratified our analysis according to 1% bin F_{st} quantiles we observed that the F_{st}
289 distribution of Denisovan-like haplotypes with a higher frequency in highlanders started to
290 trend to higher F_{st} values compared to randomly selected non-archaic background sets and
291 was significantly higher for the haplotypes within the top 1% of F_{st} values for both sets of
292 Denisovan-like haplotypes set 1: $P = 0.01$, set 2: $P = 0.04$, Figure 2c,d). The observation that
293 the most highly differentiated Denisovan-like haplotypes with a higher frequency in
294 highlanders show larger levels of population-differentiation than random frequency-matched

295 non-archaic variants would be consistent with elevated levels of local selective pressures
296 that exceed those of random non-archaic variants.

297

298 Next, to further investigate signatures of selection on archaic haplotypes in PNG, we
299 employed a computational approach to reconstruct the allele frequency trajectories of all
300 archaic haplotypes over the course of the past 1,000 generations in both highlanders and
301 lowlanders, utilizing an approximate full-likelihood method (Methods, Supplementary Table
302 4). We found that log-likelihood ratios (logLRs) that signal deviations from neutral allele
303 frequency patterns showed a significant correlation with F_{st} values for both sets of
304 Denisovan-like (Spearman correlation, set 1: $P=0.05$; set 2: $P=0.01$) and also for Neandertal-
305 like ($P=0.03$) haplotypes among highlanders, but this correlation was absent among
306 lowlanders ($P>0.05$, Figure 2e). Our results indicate that the differences in frequencies
307 among highlanders for some of the most distinct archaic haplotypes have been driven by
308 substantial increases in allele frequencies over the past 1,000 generations, matching the
309 time of between ~25,000 and 20,000 years ago when highlands were settled permanently
310 (4, 36).

311

312 *Gene content of genomic regions overlapping highly differentiated archaic haplotypes*

313

314 A key emphasis of this study was to explore the potential phenotypic impact of archaic
315 haplotypes exhibiting signs of local adaptation in highlanders or lowlanders. Previous studies
316 on the health of Papua New Guinea (PNG) populations, including some focusing on the two
317 PNG groups analyzed in this study, have indicated distinctive physiological traits among
318 highlander groups. These traits include increased ventilatory lung function and hemoglobin
319 concentration, a deeper chest, shorter stature, and smaller waist (5, 37–39). Furthermore,
320 genetic variants exhibiting signatures of selection in the PNG highlanders under investigation
321 have been associated with variations in red blood cell composition and cardiovascular
322 phenotypes. This observation aligns with findings in other high-altitude populations (6). To
323 investigate whether similar phenotypes are influenced by archaic introgression, we
324 examined the gene content of highly differentiated archaic haplotypes and the phenotypic
325 associations of aSNPs linked to them. First, we evaluated the gene content within genomic
326 regions containing different sets of archaic haplotypes showing signatures of population
327 differentiation. Subsequently, to explore potential phenotypic consequences of genes within
328 these regions, we conducted tests for functional enrichment within the gene ontology (GO)
329 (40). In total, we carried out six enrichment analyses for genes that overlapped with archaic
330 haplotypes falling within the top 1% of the F_{st} distributions for each possible combination of
331 higher archaic allele frequency in highlanders or lowlanders with Neandertal-like or

332 Denisovan-like haplotypes sets (Supplemental Information, Methods, Supplementary Table
333 5).

334

335 No enriched GO category was observed for both sets of high-Fst Denisovan-like haplotypes
336 with higher frequencies in highlanders (sets 1 and 2; FWER>0.05, Supplementary Table 5).
337 Nevertheless, two of the top five most differentiated Denisovan-like haplotypes (associated
338 with both sets of Denisovan-like haplotypes) overlapped a total of 16 protein coding genes.
339 Ten of those genes had a direct link to the brain and included pivotal developmental genes
340 like *NEUROD2* and *PAX5* (Supplementary Information). This finding aligns with the
341 discovery that among the five GO categories exhibiting the lowest P values (Denisovan-like
342 sets 1 and 2), three were associated with fear response. We also found that aSNPs
343 associated with the two highly differentiated haplotypes modified regulatory regions, like 3'
344 and 5'UTRs for several of those genes (e.g. *IKZF3*, *PPP1R1B*, Supplementary Information
345 and Supplementary Table S6). Although a formal Gene Ontology (GO) enrichment analysis
346 did not yield significant results, the notable concentration of brain-related genes within
347 genomic regions that overlap two of the most differentiated Denisovan-like set 1 & 2
348 haplotypes exhibiting substantial frequency increases in PNG highlanders suggests a
349 potential role for Denisovan DNA in influencing the biology of the brain in this population.
350 The associated aSNPs demonstrate potential regulatory effects on some of these genes.

351

352 Genes located in genomic regions overlapping high-Fst Denisovan-like set 2 haplotypes
353 displaying a higher frequency in lowlanders exhibited a GO enrichment for the category
354 'cellular response to organic substance' (FWER=0.04, Supplementary Table 5). In addition,
355 we found three categories showing borderline significance linked to cytokine and protozoan
356 responses (FWER: 0.06-0.07). These three categories encompassed between 3 and 10
357 haplotype-associated genes that were a subset from the broader pool of 19 haplotype-
358 associated genes identified within the 'cellular response to organic substance' category. This
359 observation implies a potential connection between the specific response to organic
360 substances and pathogenic elements. These findings were specific to Denisovan-like set 2,
361 as several of the haplotypes overlapping with the 17 genes were PNG-specific Neandertal-
362 like. Notably, among the Denisovan-like set 2 haplotypes associated with genes in these
363 categories was a haplotype encompassing four members of the Guanylate-binding proteins
364 family (*GBP1*, *GBP2*, *GBP4*, *GBP7*) - proteins crucial in immune response mechanisms (41).
365 This specific haplotype had been previously identified in Melanesians as a candidate for
366 positive selection (6, 27) and carries an archaic variant introducing a missense variant to
367 *GBP7* (Figure 3b, Supplementary Information). Additionally, this region shows two distinct
368 haplotypes that are both present in homozygous and heterozygous states in archaic

369 humans, suggesting this diversity might have been a result of admixture among these
370 groups as well (Figure 3c). Apart from *GBP7*, three other genes (*DRC7*, *ENTHD1*, *RTP5*)
371 harbored missense aSNPs associated with high Fst Denisovan-like set 2 haplotypes with a
372 higher frequency in lowlanders. However, none of these genes were associated with the
373 discussed GO categories. Nevertheless, we identified multiple aSNPs altering regulatory
374 regions of genes linked to those categories, including the 3' untranslated regions (3'UTRs) of
375 *GBP4* and *GBP7* (refer to Supplementary Information and Supplementary Table 6),
376 suggesting a potential regulatory effect of those variants on *GBP4* and *GBP7*.

377

378 Among the two groups of high-Fst Neandertal-like haplotypes, only the set exhibiting higher
379 frequencies in highlanders displayed enriched GO categories. These categories comprised
380 22 associations related to regulatory and metabolic functions, largely influenced by members
381 of the zinc finger family located within highlander-specific and high-Fst Neandertal-like
382 haplotypes (Supplementary Table 5, Supplementary Information).

383

384 *Phenotypic inferences of archaic SNPs associated with highly differentiated archaic*
385 *haplotypes*

386

387 To better understand the potential impact of archaic DNA within the genomes of PNG, we
388 conducted an analysis using data from UK Biobank (42) and Biobank Japan (43). This
389 analysis aimed to identify associations involving aSNPs linked to archaic haplotypes
390 reconstructed within the two PNG populations. It's important to note that this approach has
391 limitations due to notably lower levels of Denisovan DNA in UK and Japanese populations
392 compared to PNG, resulting in a significantly reduced number of PNG aSNPs tested in these
393 cohorts (Figure 3a). Moreover, this limitation was compounded by the possibility that aSNPs
394 might have been tested on a different archaic haplotype in the biobanks compared to those
395 in PNG.

396

397 In total 379 archaic haplotypes harbored aSNPs exhibiting significant phenotype
398 associations in at least one of the two biobank cohorts (Supplementary Table 7). Among
399 these associations were 48 archaic haplotypes carrying aSNPs that alter the protein
400 sequence of a gene. In general, we observed that a higher percentage of all Denisovan-like
401 haplotypes in set 2 carried missense aSNPs (5.9%) compared to Neandertal-like haplotypes
402 (4.7%, $P=0.02$, Fisher's exact test, Supplementary Table 8). A total of 125 of all 568
403 missens-carrying haplotypes contained multiple missense aSNPs, notably a Denisovan-like
404 (set 1 & 2) haplotype carrying 10 missense aSNPs. This specific haplotype had been
405 previously noted in a region involving *MUC19*, which has shown signs of introgression and

406 selection among both archaic and modern humans (44) (Supplementary Information). Traits
407 commonly associated with missense-carrying haplotypes included blood biomarkers,
408 measurements related to bone density and body fat, as well as occurrences of diabetes,
409 some of which correlate with the phenotypic difference previously reported for PNG
410 populations.

411

412 When we explored aSNP trait associations for all archaic PNG haplotypes, we identified four
413 Denisovan-like haplotypes that exhibited substantial pleiotropy, displaying between 77 and
414 178 associations with various medical and non-disease phenotypes (Supplementary Table
415 7, Figure 3d). Intriguingly, all of these haplotypes were situated within the Major
416 Histocompatibility Complex (MHC), a crucial immune-related region in the human genome.
417 Notably, two of these haplotypes ranked among the top 10% in the F_{st} distribution. In total,
418 we found 19 Denisovan-like and 7 Neandertal-like haplotypes located within the MHC. We
419 found that Denisovan-like (both sets) and Neandertal-like haplotypes showed comparable
420 but nominally higher mean F_{st} values relative to other sets of frequency-matched non-
421 archaic sets of SNPs within the MHC (Figure 3e). The relative mean F_{st} values for
422 Neandertal-like and Denisovan-like haplotypes compared to random background sets were
423 substantially larger than their respective genome-wide counterparts, which tentatively
424 showed lower mean F_{st} values compared to non-archaic random background sets (all
425 $P < 0.001$). While these results indicate a difference in population differentiation between
426 PNG highlanders and lowlanders for archaic haplotypes in the MHC compared to other
427 genomic regions, it's important to interpret these findings within the context of the high
428 evolutionary dynamics characterizing the MHC.

429 **Discussion**

430

431 In this study, we investigated the genomic and phenotypic impact of Denisovan and
432 Neandertal DNA within two PNG populations living in distinct environmental regions - the
433 mountainous terrain surrounding Mount Wilhelms and Daru Island. We found that
434 Denisovan-like haplotypes exhibiting the most significant frequency differentiation between
435 both populations and that had risen to higher frequencies among highlanders exceeded the
436 level of population differentiation seen in non-archaic variants with comparable frequencies.
437 These findings imply that Denisovan DNA played a significant role in adaptive processes for
438 these populations. Highly differentiated haplotypes that exhibited higher frequencies in
439 highlanders overlapped several genes associated with early brain development, a result also
440 observed for some of the most highly differentiated non-archaic haplotypes in these
441 populations (6). This result might reflect adaptive patterns to highlander-specific
442 environmental factors. For instance, high-altitude-induced hypoxia has been linked to
443 adaptive changes in neurons (45). Similarly, dietary variations resulting from differences in
444 food availability have been demonstrated to profoundly affect brain development (46).
445 Notably, none of the PNG individuals examined in our study carried a previously described
446 Denisovan-like haplotype associated with high-altitude adaptation (25). Therefore, at this
447 juncture, it remains challenging to predict the extent and manner in which introgressed
448 Denisovan DNA impacts these genes and their associated phenotypes. In this context it is
449 worth noting that Neandertal DNA has been demonstrated to significantly impact several
450 behavioral and neurological phenotypes in its carriers today (19, 21, 47–49). Our findings
451 imply that comparable patterns might also hold true for Denisovan DNA. One set of highly
452 differentiated Denisovan-like haplotypes, present at high frequencies among lowlanders,
453 exhibited a significant overlap with genes associated with pathogen response. One plausible
454 explanation for this result could be the exposure to tropical diseases, such as malaria. A total
455 of 94% of all malaria deaths in the Western Pacific region in 2021 were accounted for in
456 Papua New Guinea (50, 51). Although malaria is highly prevalent in the geographical region,
457 it is nearly absent among PNG highlanders (52). Our findings imply that Denisovan DNA
458 may have played a role in the adaptation to defense of malaria and/or other tropical
459 diseases. The influence of Denisovan-like haplotypes stands out when compared to
460 introgressed Neandertal-like haplotypes. We do not observe a similar degree of population
461 differentiation or functional significance for Neandertal DNA for the most highly differentiated
462 haplotypes in PNG populations. However, we did discover that highly differentiated
463 Neandertal-like haplotypes with greater prevalence among highlanders were notably
464 enriched for genes involved in transcriptional processes. These findings align with earlier

465 reports highlighting the importance of Neandertal DNA in the regulation of gene expression
466 (53–57).

467

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

491 **Methods**

492

493 *Genomic datasets*

494

495 This study included whole-genome sequencing data (8-43X coverage, Supplementary Table
496 1; human genome version hg38) derived from a cohort of 128 unrelated adult PNG
497 individuals, consisting of 74 individuals from the lowlands of Daru Island and 54 individuals
498 from the highland region around Mount Wilhelm (6). Andre et al. utilized the Broad Institute's
499 GATK v4.2.0.0 'Germline short variant discovery' pipeline (62) to generate genotype data
500 from a diverse cohort comprising populations from Papua New Guinea, Oceania, and
501 additional individuals of African, European, and Asian descent sourced from the 1000
502 Genomes Project dataset. To ensure data quality, a coverage-based genomic mask was
503 established to filter out sites with limited accessibility, particularly those derived from next-
504 generation short-read sequencing methods. Subsequently, the genotype dataset was refined
505 to include only biallelic SNPs with a call rate of at least 95%. Phasing was conducted without
506 a reference panel using shapeit v4.2.2, which incorporates an automatic imputation step
507 prior to phasing. Furthermore, we analyzed high coverage whole-genome sequencing data
508 from the 1,000 Genomes cohort (63) and four archaic humans: the Altai, Chagyrskaya and
509 Vindija Neandertals (10, 11, 32) and the Denisovan (8). Genotype information for all four
510 archaic genomes was only available for the human genome version hg19. To harmonize our
511 genotype datasets, genotypes were converted to hg38 coordinates using UCSC genome
512 browser's liftover tool (64). In addition, genotypes were filtered using the provided genomic
513 masks from each dataset. Genotype data from the 1,000 Genomes cohort was limited to
514 variable single nucleotide variant positions within the dataset. When combining Yoruba (YRI)
515 population and PNG populations, YRI positions without genotype information in the 1,000
516 Genomes cohort were considered to be homozygous for the human reference
517 allele. Genotype information for the PNG and 1,000 Genomes individuals was available in
518 phased format with the exception of chromosome X in PNG. We therefore processed the
519 unphased genotype data for chromosome X using the same pipelines as previously used for
520 the autosomes (6). Briefly, we kept biallelic variants that had genotype information for more
521 than 95% of individuals and passed the same quality filters as applied in the autosomal
522 dataset. Next, we filtered out the pseudo-autosomal parts PAR1 and PAR2 and phased the
523 remaining variants using shapeit v4.2.2 with default parameters (65).

524

525 *Archaic introgression map*

526

527 We employed a previously established methodology (22) to characterize introgressed
528 archaic haplotypes in individuals of both PNG (N=128) and three 1,000 Genomes (63)
529 Eurasian super populations (Europe, N=633; South Asia, N=601, East Asia, N=585). This
530 approach identifies these haplotypes by leveraging distinctive characteristics, including
531 shared allele signatures, haplotype structure, and haplotype length, which are indicative of
532 ancestral interbreeding between modern humans and Neandertals as well as Denisovans.
533 Following the approach, we first identified archaic SNPs (aSNPs) within the genomes of four
534 distinct populations under analysis: PNG, East Asians, South Asians, and Europeans. These
535 aSNPs were defined as those containing an allele that met the following criteria: (i) it was
536 absent in the 1,000 Genomes Yoruba population, (ii) it was present in at least one of the
537 three high-coverage Neandertal genomes (Vindija (11), Chagyrskaya (32), Altai (10)), or the
538 Denisovan genome (8), and (iii) it was present in at least one individual within the four
539 populations we were examining. Next, within each of the four populations, we computed
540 pairwise measures of linkage disequilibrium (LD) represented as r^2 between all identified
541 aSNPs within that particular population. We collapsed sets of aSNPs that showed $r^2 > 0.8$ and
542 defined them as an archaic haplotype. All aSNPs displaying no LD of $r^2 > 0.8$ with any other
543 aSNP were removed. Subsequently, we calculated the nucleotide distance for all remaining
544 haplotypes by measuring the span between the two furthest aSNPs within each haplotype.
545 We then assessed the compatibility of each haplotype's length with the genomic
546 phenomenon incomplete lineage sorting (ILS). ILS refers to the retention of ancestral genetic
547 variation shared by some modern and archaic human populations, which predates the
548 divergence of these human groups. As a result, ILS can lead to similar allele-sharing
549 patterns as aSNPs. However, because ILS segments are considerably older, the average
550 length of ILS haplotypes is much shorter compared to Neandertal or Denisovan haplotypes.
551 These archaic haplotypes typically span tens of kilobases in size, reflecting the relatively
552 recent admixture between modern and archaic humans approximately 55,000 years ago
553 (66). Building upon the methodology introduced by Huerta-Sanchez et al (25) and
554 incorporating more recent estimates for divergence and mutation rates (67), we computed
555 the likelihood of each inferred haplotype's length being compatible with incomplete lineage
556 sorting (ILS). This calculation was conducted using recombination rate estimates obtained
557 from two separate cohorts (68, 69). We calculated false-discovery rates (FDR) by adjusting
558 the acquired P values for multiple testing using the Benjamini-Hochberg procedure (70).
559 Haplotypes that exhibited a length compatible with ILS for both recombination maps
560 (FDR > 0.05) or had fewer than 10 aSNPs were excluded from further analysis. Lastly,
561 separately for PNG and each 1,000 Genomes Eurasian superpopulation and using the
562 remaining haplotypes and their corresponding aSNPs, we reconstructed haplotypes within
563 individuals based on each individual's alignment with the archaic allele at the aSNP sites.

564

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

581

582 *Evaluation of introgression map*

583

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

592

593 *Measures of population differentiation*

594

595 We assessed the extent of population differentiation between PNG highlanders and
596 lowlanders by calculating F_{st} values for all identified archaic haplotypes in these two groups.
597 For each haplotype, we randomly selected one candidate aSNP for the analysis, additionally
598 conditioning the selected aSNP to be within the same 1% allele frequency bin as the aSNP
599 with the median allele frequency value of a given archaic haplotype. We used VCFtools
600 (0.1.14) software and computed the Weir and Cockerham F_{st} estimate (71). Furthermore,

601 we created 1,000 control sets of randomly selected non-archaic SNPs that matched the
602 number of candidate aSNPs and their frequency distribution within the combined PNG
603 population dataset. Moreover, for aSNPs exhibiting differences in archaic allele frequencies
604 among PNG populations, we selected non-archaic SNPs matched in frequency, where the
605 allele corresponding to the archaic allele also demonstrates a higher frequency in the PNG
606 population, akin to the archaic allele.

607

608 *Computational reconstruction of allele frequency trajectories for aSNPs*

609

610 We employed computational methods to reconstruct allele frequency trajectories, utilizing
611 genomic data from both PNG highlanders and lowlanders, applying a modification of the
612 pipeline detailed in André et al (6). In brief, our approach involved selecting three random
613 representative aSNPs from each archaic haplotype. Subsequently, we extracted the local
614 genealogical tree for each aSNP, utilizing the RELATE software (72) (v1.1.8). These
615 generated trees served as input for CLUES (73), an approximate full-likelihood method for
616 testing signatures of selection (v1). CLUES facilitated the reconstruction of allele frequencies
617 and assigned log-likelihood ratios ($\log(LR)$) to indicate support for non-neutrality. We did not
618 assess aSNPs with a minor allele frequency below 5%. Furthermore, we excluded data
619 points for which the CLUES algorithm did not yield a $\log(LR)$ value.

620

621 *Functional enrichment analysis*

622

623 We performed functional enrichment analysis in the Gene Ontology (GO) (40) using the R
624 package GOfuncR (74). Specifically, we examined six sets of archaic candidate haplotypes
625 that ranked within the top 1 percentile of their respective F_{st} distributions. These six sets
626 were created by splitting all archaic haplotypes by (i) ancestry (Denisovan-like sets 1 & 2 or
627 Neandertal-like) and (ii) frequency (higher in highlanders or lowlanders). For each of the six
628 sets, we defined background haplotypes, which included all remaining haplotypes that did
629 not fall within the top 1 percentile of their respective F_{st} distributions. To account for the
630 genomic clustering of functionally related gene groups, we ran the GO enrichment software
631 with the parameter `circ_chrom=TRUE`. This setting allowed us to create 10,000 test sets that
632 randomly shifted the coordinates of both candidate and background haplotypes on the
633 circularized version of the same chromosome. Family-wise error rates (FWER) were
634 determined by comparing empirical enrichment P values for each GO category to the
635 minimum enrichment P values observed in the entire GO across all test sets.

636

637 *Phenotypic and regulatory annotation of aSNPs*

638

639 We conducted a screening of phenotypic association data from two biobank cohorts, the UK
640 Biobank (42) (4,280 phenotypes, <https://www.nealelab.is/uk-biobank>) and Biobank Japan
641 (43) (220 phenotypes), to identify associations with aSNPs linked to the archaic haplotypes
642 observed in PNG highlanders and lowlanders. We included in our analysis all associations
643 involving aSNPs with allele frequencies exceeding 1% and meeting the stringent genome-
644 wide significance threshold of $P < 5 \times 10^{-8}$. In addition, we leveraged ENSEMBL's variant effect
645 predictor (75) (ensembl-vep version: 109.3) to annotate the aSNPs identified in PNG with
646 their predicted molecular consequences.

647

648 *Protein visualization*

649

650 We employed AlphaFold2 (76) to predict the 3D protein folding structures for two variants of
651 GBP7. The first variant was constructed based on the reference protein sequence, while the
652 second variant incorporated a Proline substitution in place of Leucine at amino acid position
653 559. This substitution was introduced by chr1:89,132,390:A>G aSNP, with Guanine being
654 the archaic allele at this site. This aSNP resides on the Denisovan-like set 2 haplotype
655 (chr1:89,054,418-89,200,767), which encompasses this gene. To visualize the resulting
656 protein structures for both variants, we utilized the Mol* viewer (77) (Figure 3b).

657 **Data availability**

658

659 Genotype data

660

- 661 • Papuan dataset: European Genome-Phenome data repository access:
662 EGAD00001010142, EGAD00001010143, EGAD50000000050
- 663 • 1,000 Genomes dataset: <https://www.internationalgenome.org/>
- 664 • Archaic humans: <http://cdna.eva.mpg.de/neandertal/>

665

666 Phenotype data

667

- 668 • UK Biobank GWAS summary statistics: <https://www.nealelab.is/uk-biobank>
- 669 • Biobank Japan GWAS summary statistics: <https://pheweb.jp/>
- 670 • ENSEMBL's variant effect predictor:
671 <https://www.ensembl.org/info/docs/tools/vep/index.html>

672

673 **Code availability**

674

- 675 • https://github.com/SillySabertooth/papuan_archaic_admixture/

676 **References**

- 677 1. C. Clarkson, *et al.*, Human occupation of northern Australia by 65,000 years ago. *Nature*
678 **547**, 306–310 (2017).
- 679 2. J. F. O’Connell, *et al.*, When did Homo sapiens first reach Southeast Asia and Sahul?
680 *Proc. Natl. Acad. Sci. U. S. A.* **115**, 8482–8490 (2018).
- 681 3. N. Brucato, *et al.*, Papua New Guinean Genomes Reveal the Complex Settlement of
682 North Sahul. *Mol. Biol. Evol.* **38**, 5107–5121 (2021).
- 683 4. G. R. Summerhayes, J. H. Field, B. Shaw, D. Gaffney, The archaeology of forest
684 exploitation and change in the tropics during the Pleistocene: The case of Northern
685 Sahul (Pleistocene New Guinea). *Quat. Int.* **448**, 14–30 (2017).
- 686 5. M. André, *et al.*, Phenotypic differences between highlanders and lowlanders in Papua
687 New Guinea. *PLoS One* **16**, e0253921 (2021).
- 688 6. M. André, *et al.*, Positive selection in the genomes of two Papua New Guinean
689 populations at distinct altitude levels. *bioRxiv*, 2022.12.15.520226 (2022).
- 690 7. J. Choin, *et al.*, Genomic insights into population history and biological adaptation in
691 Oceania. *Nature* **592**, 583–589 (2021).
- 692 8. M. Meyer, *et al.*, A high-coverage genome sequence from an archaic Denisovan
693 individual. *Science* **338**, 222–226 (2012).
- 694 9. M. Larena, *et al.*, Philippine Ayta possess the highest level of Denisovan ancestry in the
695 world. *Curr. Biol.* **31**, 4219–4230.e10 (2021).
- 696 10. K. Prüfer, *et al.*, The complete genome sequence of a Neanderthal from the Altai
697 Mountains. *Nature* **505**, 43–49 (2014).
- 698 11. K. Prüfer, *et al.*, A high-coverage Neandertal genome from Vindija Cave in Croatia.
699 *Science* **358**, 655–658 (2017).
- 700 12. S. Peyrégne, V. Slon, J. Kelso, More than a decade of genetic research on the
701 Denisovans. *Nat. Rev. Genet.* (2023) <https://doi.org/10.1038/s41576-023-00643-4>.
- 702 13. G. H. Perry, L. Kistler, M. A. Kelaita, A. J. Sams, Insights into hominin phenotypic and
703 dietary evolution from ancient DNA sequence data. *J. Hum. Evol.* **79**, 55–63 (2015).
- 704 14. D. Gokhman, *et al.*, Reconstructing Denisovan Anatomy Using DNA Methylation Maps.
705 *Cell* **179**, 180–192.e10 (2019).
- 706 15. J. C. Teixeira, A. Cooper, Using hominin introgression to trace modern human
707 dispersals. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 15327–15332 (2019).
- 708 16. S. R. Browning, B. L. Browning, Y. Zhou, S. Tucci, J. M. Akey, Analysis of Human
709 Sequence Data Reveals Two Pulses of Archaic Denisovan Admixture. *Cell* **173**, 53–
710 61.e9 (2018).
- 711 17. M. Mondal, J. Bertranpetit, O. Lao, Approximate Bayesian computation with deep
712 learning supports a third archaic introgression in Asia and Oceania. *Nat. Commun.* **10**,
713 246 (2019).
- 714 18. G. S. Jacobs, *et al.*, Multiple Deeply Divergent Denisovan Ancestries in Papuans. *Cell*

- 715 **177**, 1010–1021.e32 (2019).
- 716 19. E. McArthur, D. C. Rinker, J. A. Capra, Quantifying the contribution of Neanderthal
717 introgression to the heritability of complex traits. *Nat. Commun.* **12**, 4481 (2021).
- 718 20. M. Dannemann, J. Kelso, The Contribution of Neanderthals to Phenotypic Variation in
719 Modern Humans. *Am. J. Hum. Genet.* **101**, 578–589 (2017).
- 720 21. C. N. Simonti, *et al.*, The phenotypic legacy of admixture between modern humans and
721 Neandertals. *Science* **351**, 737–741 (2016).
- 722 22. M. Dannemann, The Population-Specific Impact of Neanderthal Introgression on Human
723 Disease. *Genome Biol. Evol.* **13** (2021).
- 724 23. L. Skov, *et al.*, The nature of Neanderthal introgression revealed by 27,566 Icelandic
725 genomes. *Nature* **582**, 78–83 (2020).
- 726 24. D. Koller, *et al.*, Denisovan and Neanderthal archaic introgression differentially impacted
727 the genetics of complex traits in modern populations. *BMC Biol.* **20**, 249 (2022).
- 728 25. E. Huerta-Sánchez, *et al.*, Altitude adaptation in Tibetans caused by introgression of
729 Denisovan-like DNA. *Nature* **512**, 194–197 (2014).
- 730 26. F. Racimo, *et al.*, Archaic Adaptive Introgression in TBX15/WARS2. *Mol. Biol. Evol.* **34**,
731 509–524 (2017).
- 732 27. B. Vernot, *et al.*, Excavating Neanderthal and Denisovan DNA from the genomes of
733 Melanesian individuals. *Science* **352**, 235–239 (2016).
- 734 28. S. Sankararaman, S. Mallick, N. Patterson, D. Reich, The Combined Landscape of
735 Denisovan and Neanderthal Ancestry in Present-Day Humans. *Curr. Biol.* **26**, 1241–
736 1247 (2016).
- 737 29. D. M. Vespasiani, *et al.*, Denisovan introgression has shaped the immune system of
738 present-day Papuans. *PLoS Genet.* **18**, e1010470 (2022).
- 739 30. N. Brucato, *et al.*, Chronology of natural selection in Oceanian genomes. *iScience* **25**,
740 104583 (2022).
- 741 31. L. Skov, *et al.*, Detecting archaic introgression using an unadmixed outgroup. *PLoS*
742 *Genet.* **14**, e1007641 (2018).
- 743 32. F. Mafessoni, *et al.*, A high-coverage Neanderthal genome from Chagyrskaya Cave.
744 *Proc. Natl. Acad. Sci. U. S. A.* **117**, 15132–15136 (2020).
- 745 33. B. Vernot, J. M. Akey, Complex history of admixture between modern humans and
746 Neandertals. *Am. J. Hum. Genet.* **96**, 448–453 (2015).
- 747 34. V. Slon, *et al.*, The genome of the offspring of a Neanderthal mother and a Denisovan
748 father. *Nature* **561**, 113–116 (2018).
- 749 35. D. Glavan, *et al.*, Identification of transcriptome alterations in the prefrontal cortex,
750 hippocampus, amygdala and hippocampus of suicide victims. *Sci. Rep.* **11**, 18853
751 (2021).
- 752 36. D. Gaffney, A. Ford, G. Summerhayes, Crossing the Pleistocene–Holocene transition in
753 the New Guinea Highlands: Evidence from the lithic assemblage of Kiowa rockshelter.

- 754 *Journal of Anthropological Archaeology* **39**, 223–246 (2015).
- 755 37. J. E. Cotes, H. R. Anderson, J. M. Patrick, Lung function and the response to exercise
756 in the New Guineans: role of genetic and environmental factors. *Philos. Trans. R. Soc.*
757 *Lond. B Biol. Sci.* **268**, 349–361 (1974).
- 758 38. N. Senn, *et al.*, Population hemoglobin mean and anemia prevalence in Papua New
759 Guinea: new metrics for defining malaria endemicity? *PLoS One* **5**, e9375 (2010).
- 760 39. A. J. Woolcock, M. H. Colman, C. R. Blackburn, Factors affecting normal values for
761 ventilatory lung function. *Am. Rev. Respir. Dis.* **106**, 692–709 (1972).
- 762 40. M. Ashburner, *et al.*, Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**,
763 25–29 (2000).
- 764 41. K. Tretina, E.-S. Park, A. Maminska, J. D. MacMicking, Interferon-induced guanylate-
765 binding proteins: Guardians of host defense in health and disease. *J. Exp. Med.* **216**,
766 482–500 (2019).
- 767 42. C. Bycroft, *et al.*, The UK Biobank resource with deep phenotyping and genomic data.
768 *Nature* **562**, 203–209 (2018).
- 769 43. A. Nagai, *et al.*, Overview of the BioBank Japan Project: Study design and profile. *J.*
770 *Epidemiol.* **27**, S2–S8 (2017).
- 771 44. F. A. Villanea, *et al.*, The MUC19 gene in Denisovans, Neanderthals, and Modern
772 Humans: An Evolutionary History of Recurrent Introgression and Natural Selection.
773 *bioRxiv* (2023) <https://doi.org/10.1101/2023.09.25.559202>.
- 774 45. C. A. Eichstaedt, *et al.*, Genetic and phenotypic differentiation of an Andean
775 intermediate altitude population. *Physiol Rep* **3** (2015).
- 776 46. S. E. Cusick, M. K. Georgieff, The Role of Nutrition in Brain Development: The Golden
777 Opportunity of the “First 1000 Days.” *J. Pediatr.* **175**, 16–21 (2016).
- 778 47. M. Dannemann, *et al.*, Neandertal introgression partitions the genetic landscape of
779 neuropsychiatric disorders and associated behavioral phenotypes. *Transl. Psychiatry*
780 **12**, 433 (2022).
- 781 48. Z. Chen, *et al.*, The contribution of Neandertal introgression and natural selection to
782 neurodegenerative diseases. *Neurobiol. Dis.* **180**, 106082 (2023).
- 783 49. C. A. Trujillo, *et al.*, Reintroduction of the archaic variant of *NOVA1* in cortical organoids
784 alters neurodevelopment. *Science* **371** (2021).
- 785 50. Organization, W. H., *World malaria report 2022* (World Health Organization, 2022).
- 786 51. U. Kitur, T. Adair, I. Riley, A. D. Lopez, Estimating the pattern of causes of death in
787 Papua New Guinea. *BMC Public Health* **19**, 1322 (2019).
- 788 52. I. Müller, M. Bockarie, M. Alpers, T. Smith, The epidemiology of malaria in Papua New
789 Guinea. *Trends Parasitol.* **19**, 253–259 (2003).
- 790 53. D. Yermakovich, *et al.*, Long-range regulatory effects of Neandertal DNA in modern
791 humans. *Genetics* (2022) <https://doi.org/10.1093/genetics/iyac188>.
- 792 54. M. Dannemann, K. Prüfer, J. Kelso, Functional implications of Neandertal introgression

- 793 in modern humans. *Genome Biol.* **18**, 61 (2017).
- 794 55. H. Quach, *et al.*, Genetic Adaptation and Neanderthal Admixture Shaped the Immune
795 System of Human Populations. *Cell* **167**, 643–656.e17 (2016).
- 796 56. M. Silvert, L. Quintana-Murci, M. Rotival, Impact and Evolutionary Determinants of
797 Neanderthal Introgression on Transcriptional and Post-Transcriptional Regulation. *Am. J. Hum. Genet.* **104**, 1241–1250 (2019).
798
- 799 57. R. C. McCoy, J. Wakefield, J. M. Akey, Impacts of Neanderthal-Introgressed Sequences
800 on the Landscape of Human Gene Expression. *Cell* **168**, 916–927.e12 (2017).
- 801 58. I. Gallego Romero, A. J. Lea, Leveraging massively parallel reporter assays for
802 evolutionary questions. *Genome Biol.* **24**, 26 (2023).
- 803 59. A. A. Pollen, U. Kilik, C. B. Lowe, J. G. Camp, Human-specific genetics: new tools to
804 explore the molecular and cellular basis of human evolution. *Nat. Rev. Genet.* **24**, 687–
805 711 (2023).
- 806 60. C. M. Brand, L. L. Colbran, J. A. Capra, Predicting Archaic Hominin Phenotypes from
807 Genomic Data. *Annu. Rev. Genomics Hum. Genet.* **23**, 591–612 (2022).
- 808 61. H. M. Natri, *et al.*, Genetic architecture of gene regulation in Indonesian populations
809 identifies QTLs associated with global and local ancestries. *Am. J. Hum. Genet.* **109**,
810 50–65 (2022).
- 811 62. R. Poplin, *et al.*, Scaling accurate genetic variant discovery to tens of thousands of
812 samples. *bioRxiv* (2017) <https://doi.org/10.1101/2011178>.
- 813 63. ,High-coverage whole-genome sequencing of the expanded 1000 Genomes Project
814 cohort including 602 trios. *Cell* **185**, 3426–3440.e19 (2022).
- 815 64. W. J. Kent, *et al.*, The human genome browser at UCSC. *Genome Res.* **12**, 996–1006
816 (2002).
- 817 65. O. Delaneau, J.-F. Zagury, M. R. Robinson, J. L. Marchini, E. T. Dermitzakis, Accurate,
818 scalable and integrative haplotype estimation. *Nat. Commun.* **10**, 5436 (2019).
- 819 66. S. Sankararaman, N. Patterson, H. Li, S. Pääbo, D. Reich, The date of interbreeding
820 between Neanderthals and modern humans. *PLoS Genet.* **8**, e1002947 (2012).
- 821 67. M. Dannemann, A. M. Andrés, J. Kelso, Introgression of Neanderthal- and Denisovan-like
822 Haplotypes Contributes to Adaptive Variation in Human Toll-like Receptors. *Am. J. Hum. Genet.* **98**, 22–33 (2016).
823
- 824 68. A. Kong, *et al.*, A high-resolution recombination map of the human genome. *Nat. Genet.*
825 **31**, 241–247 (2002).
- 826 69. International HapMap 3 Consortium, *et al.*, Integrating common and rare genetic
827 variation in diverse human populations. *Nature* **467**, 52–58 (2010).
- 828 70. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful
829 approach to multiple testing. *J. R. Stat. Soc.* **57**, 289–300 (1995).
- 830 71. P. Danecek, *et al.*, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158
831 (2011).

- 832 72. L. Speidel, M. Forest, S. Shi, S. R. Myers, A method for genome-wide genealogy
833 estimation for thousands of samples. *Nat. Genet.* **51**, 1321–1329 (2019).
- 834 73. A. J. Stern, P. R. Wilton, R. Nielsen, An approximate full-likelihood method for inferring
835 selection and allele frequency trajectories from DNA sequence data. *PLoS Genet.* **15**,
836 e1008384 (2019).
- 837 74. S. Grote, GOfuncR: Gene ontology enrichment using FUNC. *R package version* (2018).
- 838 75. W. McLaren, *et al.*, The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).
- 839 76. J. Jumper, *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature*
840 **596**, 583–589 (2021).
- 841 77. D. Sehnal, *et al.*, Mol* Viewer: modern web app for 3D visualization and analysis of
842 large biomolecular structures. *Nucleic Acids Res.* **49**, W431–W437 (2021).

843 **Acknowledgements**

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

854 **Figure legends**

855

856 **Figure 1: Archaic DNA in PNG genomes.**

857 **(a)** Amount of diploid genome sequences reconstructed in PNG individuals based on archaic
858 haplotypes exhibiting higher sequence similarity with the Denisovan individual (left) and
859 Neandertals (right). Boxplots are employed to visualize the distributions, with the outer
860 whiskers indicating the minimum and maximum values. The recovered sequence amounts
861 are presented in megabases (Mb) for various categories, including total sequences (black),
862 sequences found in 1,000 Genomes Europeans (EUR, gray), East Asians (EAS, gray),
863 South Asians (SAS, gray), and the combined sequences from all three Eurasian populations
864 (All, orange). Additionally, the amount of genome sequences recovered using haplotypes
865 that were not detected in any of the Eurasian populations (Absent) is displayed in red. **(b-c)**
866 Empirical cumulative density distributions (eCDF, y-axis) are displayed for the sequence
867 similarity measures (in number of differences per 10^{-3} bases, x-axis) of archaic haplotypes in
868 PNG, specifically those exhibiting a greater sequence similarity with Neandertals than with
869 the Denisovan individual. Panels (b) and (c) show these distributions in comparison to the
870 Denisovan and three Neandertals, respectively. Haplotypes found in 1,000 Genomes
871 Eurasians are represented by dashed lines, while PNG-specific haplotypes are delineated
872 with solid lines. **(d)** Boxplots illustrating the amount of Denisovan-like (set 1 and 2) and
873 Neandertal-like diploid genome sequence in PNG highlanders (blue) and lowlanders
874 (orange). Outer whiskers represent the minimum and maximum values. **(e)** Boxplots
875 displaying the percentage of archaic (black/gray), Denisovan-like (set 1: green, set 2: dark
876 green) and Neandertal-like (blue) ancestry on the autosomes and X chromosome in the
877 PNG cohort. The outer whiskers indicate the 95% confidence interval borders. **(f-g)** The
878 proportion of Denisovan-like (set 2, purple) and Neandertal-like (blue) DNA within one-
879 megabase windows is depicted for chromosome 14 (f), which harbors the highest levels of
880 archaic DNA, and chromosome X (g), characterized by the least amount of archaic DNA.
881 Gray areas below the x-axis denote regions that were previously reported to be devoid of
882 archaic ancestry in present-day populations. Black areas indicate centromere regions.

883 **Figure 2: Frequency differences of archaic haplotypes between PNG highlanders and**
884 **lowlanders.**

885 **(a)** Manhattan plot visualizes F_{st} values, representing frequency differences between PNG
886 highlanders and lowlanders for Denisovan-like haplotypes (sets 1 & 2: circles, set 2 only:
887 squares) and Neandertal-like haplotypes (triangles). The Manhattan plot is symmetrically
888 divided across the x-axis, with data points displayed on both the upper and lower sides
889 conditioned on whether a haplotype is found at a higher frequency in highlanders or
890 lowlanders, respectively. **(b)** Distributions of the ratio of the mean F_{st} value for archaic
891 haplotypes compared to each mean F_{st} value obtained from 1,000 non-archaic background
892 sets (Methods). These distributions are presented for distinct categories of archaic
893 haplotypes: the full sets of Denisovan-like (set 1: left, set 2: middle) and Neandertal-like
894 (right) haplotypes are shown in black. Additionally, they include subsets of haplotypes having
895 higher frequency in highlanders (blue) or lowlanders (orange) for both sets of Denisovan-like
896 and Neandertal-like haplotypes. Whiskers in the boxplots represent the 95% confidence
897 interval for these distributions. Distributions that exhibit a significant deviation from a ratio of
898 1 ($P < 0.05$) are denoted by an asterisk. **(c-d)** Distribution of the ratio between F_{st} quantiles
899 for sets of Denisovan-like and Neandertal-like haplotypes displaying a higher frequency in
900 PNG highlanders (c) and PNG lowlanders (d) relative to the quantile F_{st} values derived from
901 1,000 matching non-archaic background sets. Distributions are smoothed using a cubic
902 spline smoothing algorithm. Quantiles for which the ratio significantly deviates from one are
903 shown with filled circles. A gray line is included to represent the neutral expectation of one.
904 **(e)** Spearman's correlation coefficients together with their 95% confidence intervals (y-axis)
905 calculated between F_{st} values and log-likelihood ratios derived from reconstructed allele
906 frequencies for sets of Denisovan-like and Neandertal-like haplotypes are shown.
907 Haplotypes are categorized into two sets: those exhibiting a larger archaic allele frequency in
908 PNG highlanders (blue) and in lowlanders (orange). Correlation coefficients that exhibit a
909 significant deviation from a value of 0 ($P < 0.05$) are denoted by an asterisk.

910 **Figure 3: Phenotypic implications of archaic admixture in PNG populations.**
911 **(a)** The percentage of aSNPs tested in two biobank cohorts, namely the UK Biobank and
912 Biobank Japan, is depicted. The relative distribution of aSNPs is categorized according to
913 the ancestral origin of the haplotype linked to them in PNG. Only aSNPs with a minor allele
914 frequency of at least 1% in both biobank cohorts are considered. **(b)** A three-dimensional
915 representation of the human Guanylate Binding Protein 7 (GBP7) is depicted on the right. A
916 distinct segment of the protein is emphasized in green. Two renditions of this protein
917 segment are displayed: one reflects the modern human reference protein sequence
918 (middle), while the other showcases an amino acid substitution introduced by the archaic
919 allele chr1:89,132,390:A>G, resulting in a Leucine to Proline alteration at position 559 (left).
920 **(c)** The genetic composition across 163 variable sites in the Denisovan and three Neandertal
921 individuals within the region of the introgressed Denisovan-like set 2 haplotype at
922 chr1:89,054,418-89,200,767 is depicted. Alleles are color-coded based on their prevalence
923 in the four archaic genomes (major allele in purple, minor allele in yellow, and heterozygous
924 sites in orange). The analysis only covers genetic positions with genotype information
925 available for all four archaic individuals. The bottom panel highlights positions that represent
926 aSNP locations in PNG (gray). **(d)** Scatterplot depicting the frequencies for sets of archaic
927 haplotypes spanning the Major Histocompatibility Complex (MHC) in PNG highlanders (x-
928 axis) and lowlanders (y-axis). Denisovan-like haplotypes are represented by circles (set 1 &
929 2) and squares (set 2), while Neandertal haplotypes are denoted by triangles. The symbol
930 sizes correspond to the number of trait associations associated with aSNPs segregating on
931 these haplotypes (Methods). A gray diagonal line has been included to represent the region
932 for haplotypes with identical frequencies in both populations. **(e)** Frequency distributions
933 illustrate the ratio of both the mean F_{st} value for genome-wide sets of archaic haplotypes
934 (right) and sets of haplotypes overlapping the MHC (left) compared to the mean F_{st} values
935 obtained from 1,000 non-archaic background sets genome-wide and within the MHC,
936 respectively (Methods). Whiskers in the boxplots represent the 95% confidence interval for
937 these distributions. Distributions with a significant deviation from one are highlighted by an
938 asterisk ($P < 0.05$).





