Genetic Signatures of Positive Selection in Human Populations Adapted to High Altitude in Papua New Guinea

Ram González-Buenfil [1], Sofía Vieyra-Sánchez [1], Consuelo D. Quinto-Cortés [1], Stephen J. Oppenheimer [2], William Pomat [3], Moses Laman[3], Mayté C. Cervantes-Hernández [1], Carmina Barberena-Jonas [1], Kathryn Auckland [4], Angela Allen [5], Stephen Allen [6], Maude E. Phipps [7], Emilia Huerta-Sanchez [8,9], Alexander G. Ioannidis [10,11], Alexander J. Mentzer [4], Andrés Moreno-Estrada [1,*]

[1]Advanced Genomics Unit (UGA), Center for Research and Advanced Studies of the National Polytechnic Institute (Cinvestav), Irapuato, Guanajuato, Mexico

[2]School of Anthropology, University of Oxford, Oxford, UK

[3]Vector-Borne Diseases Unit, Papua New Guinea Institute of Medical Research, Madang, Papua New Guinea

[4]The Centre for Human Genetics, University of Oxford, Oxford, UK

[5]Department of Molecular Haematology, MRC Weatherall Institute of Molecular Medicine, Headley Way, Headington, Oxford, OX3 9DS, UK

[6]Department of Clinical Sciences,Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, L3 5QA, UK

[7]Jeffrey Cheah School of Medicine and Health Sciences, Monash University Malaysia, Subang Jaya 47500, Selangor, Malaysia

[8]Center for Computational Molecular Biology, Brown University, Providence, RI 02912, USA

[9]Department of Ecology, Evolution and Organismal Biology, Brown University, Providence, RI 02912, USA

[10]Department of Biomolecular Engineering, University of California Santa Cruz, Santa Cruz, CA, USA

[11]Department of Biomedical Data Science, Stanford Medical School, Stanford, CA, USA

*Corresponding author: E-mail: andres.moreno@cinvestav.mx.

Accepted: July 09, 2024

Abstract

Papua New Guinea (PNG) hosts distinct environments mainly represented by the ecoregions of the Highlands and Lowlands that display increased altitude and a predominance of pathogens, respectively. Since its initial peopling approximately 50,000 years ago, inhabitants of these ecoregions might have differentially adapted to the environmental pressures exerted by each of them. However, the genetic basis of adaptation in populations from these areas remains understudied. Here, we investigated signals of positive selection in 62 highlanders and 43 lowlanders across 14 locations in the main island of PNG using whole-genome genotype data from the Oceanian Genome Variation Project (OGVP) and searched for signals of positive selection through population differentiation and haplotype-based selection scans. Additionally, we performed archaic ancestry estimation to detect selection signals in highlanders within introgressed regions of the genome. Among highland populations we identified candidate genes representing known biomarkers for mountain sickness (*SAA4*, *SAA1*, *PRDX1*, *LDHA*) as well as candidate genes of the Notch signaling pathway (*PSEN1*, *NUMB*, *RBPJ*, *MAML3*), a novel proposed pathway for high altitude adaptation in multiple organisms. We also identified candidate genes involved in oxidative stress, inflammation, and angiogenesis, processes inducible by hypoxia, as well as in components of the eye lens and the immune response. In contrast, candidate genes in the lowlands are mainly related to the immune response (*HLA-DQB1*, *HLA-DQA2*, *TAAR6*, *TAAR9*, *TAAR8*, *RNASE4*, *RNASE6*, *ANG*). Moreover, we find two candidate regions to be also enriched with archaic introgressed segments, suggesting that archaic admixture has played a role in the local adaptation of PNG populations.

Significance

Early settlers of New Guinea reached the island around 50,000 years ago, eventually adapting to distinct highland and lowland ecoregions. The genetic basis of this local adaptation, however, remains understudied. Here we scan for signals of positive selection in populations from both ecoregions in Papua New Guinea (PNG). In highlanders, candidate genes under selection are consistent with the response to the physiological stress caused by increased altitude, with some being reported in other moderate and high-altitude populations. Conversely, lowlanders exhibit candidate genes mainly associated with the immune response, likely reflecting a stronger selective pressure imposed by pathogens in this region. These results enhance our knowledge of genetic adaptation in Pacific islanders, highlighting the impact of differential environments shaping human evolution.

Key words: Papua New Guinea, positive selection, high altitude, archaic introgression. population genomics, human adaptation.

Introduction

The peopling of the island of New Guinea, located in the Pacific directly north of Australia, represents one of the earliest human migrations out of Africa, with the first settlers arriving 50 thousand years ago (kya) (O'Connell et al. 2018). Upon settling, they rapidly spread across different regions, with evidence of ancient occupation in both the highlands and lowlands dating back to ~40 kya (Summerhayes et al. 2010). This early dispersal eventually led to genetic divergence between the inhabitants of the Highlands and Lowlands occurring approximately 10 to 20 kya (Bergström et al. 2017; Brucato et al. 2021). Today, Papua New Guinea (PNG) is renowned for its exceptional cultural and linguistic diversity, making it one of the most varied areas globally. This diversity was further enriched by subsequent migrations that included the arrival of ancestral Austronesian-speaking people around 3.5 kya, who primarily settled along the coastal regions (Kayser 2010; Skoglund et al. 2016). The latter part of the 19th century saw an increase in European settlers arriving to the island, also mainly in coastal areas, while populations in the Highlands remained relatively isolated from European contact until the early-mid 20th century. The Highlands and Lowlands of PNG are very distinct geographical areas. For example, the Highlands feature rugged terrain with some of the highest mountains in the Pacific, ranging between ~1,200 and 4,500 meters above sea level, and cooler temperatures, whereas the Lowlands are characterized by a more humid tropical environment with higher temperatures and denser rainforests. Such environments have influenced past settlement patterns, with malaria likely playing a significant role in driving populations to settle in the Highlands, away from the mosquito-dense Lowlands where the spread of malaria is facilitated due to ideal conditions for the *Plasmodium*-carrying Anopheles mosquito to thrive (Senn et al. 2010; Trájer 2022). It is possible, but uncertain, that other prevalent infectious diseases, such as tuberculosis (TB), were present to a degree before the first European contacts in the early 16th century, potentially affecting indigenous populations to some extent long before wider global interactions (Riley 1983; Comas et al. 2013; Trájer 2022). Importantly, historical reports indicate that the arrival of Europeans favored the spread of new and previously localized diseases across the island (Ley et al. 2014). Differences between the Highlands and Lowlands ecoregions, and the history of relative population isolation across PNG, might have led to different selective pressures acting in each one. Previous research has found that high altitude is a major factor contributing to the phenotypic variation observed between highlanders and lowlanders in PNG, suggesting this to be the result of an adaptive response to the selective pressure exerted by hypoxia and highlighting the need for further studies to investigate the genetic basis of such adaptation (André et al. 2021). A more recent study on two populations living at distinct altitude levels in PNG found candidate genetic signals related to cardiovascular phenotypes: in highlanders related to red blood cell count, and in lowlanders related to white blood cell count (André et al. 2024). Additionally, long-term exposure to high altitude can induce chronic mountain sickness (CMS), angiogenesis, increased oxidative stress, and cardiovascular complications (Pena et al. 2022). Moreover, populations from New Guinea have a history of genetic admixture with archaic hominins, exhibiting some of the highest proportions of Denisovan introgression (~4%) (Reich et al. 2010; Meyer et al. 2012; Vernot et al. 2016; Jacobs et al. 2019), and levels of Neanderthal introgression similar to those of other populations outside of Africa (~2%) (Prüfer et al. 2014). Archaic introgression has been identified to have adaptive effects on human populations (Racimo et al. 2015; Gittelman et al. 2016). For instance, Neanderthal introgression has contributed adaptive traits related to the immune response in modern humans (Enard and Petrov 2018), whereas Denisovan introgression has contributed to traits related to high altitude adaptation

in Tibetan populations (Huerta-Sánchez et al. 2014). Recent studies have shown that Denisovan introgression might have contributed to the immune system of populations from PNG (Vespasiani et al. 2022; André et al. 2024; Yermakovich et al. 2024), and that some signals of selection found in Pacific Islanders are located in Denisovan segments harbored by Papuan ancestry (Choin et al. 2021; Brucato et al. 2022). André et al further elucidated the role of archaic introgression in local adaptive processes among PNG populations, reporting an archaic haplotype potentially related to angiogenesis in highlanders, and another in lowlanders related to immunity (André et al. 2024). Whereas Yermakovich et al explored Denisova introgression and its potential adaptive roles in brain biology and immune response in PNG populations (Yermakovich et al. 2024). Despite these advancements, the full scope of adaptive archaic introgression continues to be an area for further investigation. The unique setting and history of PNG provide an opportunity to discover regions in the genome underlying human adaptation to contrasting environment conditions in the Highlands and the Lowlands, as well as to assess the involvement of archaic introgression in molecular signatures of selection.

To address this, we investigated signals of positive selection in 62 highlanders and 43 lowlanders sampled from 14 different locations across the main island of PNG using whole-genome genotype data gathered from the Oceanian Genome Variation Project (OGVP) (Quinto-Cortés et al. 2024) (Fig. 1a). Our selection scans included statistics suitable to detect both within and between population signals, followed by functional annotation clustering of candidate genes to identify potential biological processes under selection in the Highlands and Lowlands, and archaic ancestry estimation to detect signals possibly involving introgressed regions of the genome.

Results

Population Structure and Admixture

To explore patterns of global ancestry and population structure across the Oceanian region we performed Principal Component Analysis (PCA) and ADMIXTURE runs combining our PNG samples and the broader OGVP data set Quinto-Cortés et al. (2024) as well as the Human Genome Diversity Project (HGDP) Han, French, and Yoruba from (Bergström et al. 2020) as continental reference samples (supplementary fig. S1, Supplementary Material online). We then focused on populations from the main island of PNG for downstream analyses (Fig. 1b and c). Principal component 1 (PC1) separates African from non-African ancestries, whereas principal component 2 (PC2) separates PNG populations from African, European, and East Asian ancestries. Individuals from the Highlands form a tighter cluster than those from the Lowlands, which are more dispersed in PCA space, ranging from near the highlanders to the axis toward the Han cluster, reflecting the presence of other ancestries besides Papuan-related that are closer to the East Asian ancestry of the Han (Fig. 1b). In ADMIXTURE (Fig. 1c), we found the lowest cross-validation error at $K = 9$ (supplementary fig. S2, Supplementary Material online), identifying the following ancestry components: African, European, and East Asian, along with Papuan-related ancestry, and five other distinct ancestries mainly found in offshore PNG islands and other islands across the Bismarck Archipelago, the Solomon Islands, Vanuatu, and Fiji in the Southwest Pacific (Figure S3 in Quinto-Cortés et al. 2024). Individuals from the Lowlands show varying proportions of these Pacific-related ancestry components, consistent with other findings from the region (Choin et al. 2021), whereas individuals from the Highlands show a predominant Papuan-related ancestry. No European or African ancestries were observed in populations from the Highlands or Lowlands.

Candidate Regions Under Positive Selection

Since different environmental pressures might exist between the Highlands and Lowlands (oxygen availability, temperature variability, ultraviolet radiation, pathogen/s exposure, staple food types, etc.), we categorized samples from each province into two ecoregions (supplementary table S1, Supplementary Material online), and conducted four different tests for positive selection: the integrated haplotype score (iHS) (Voight et al. 2006), the number of segregating sites by length (nSL) (Ferrer-Admetlla et al. 2014), the cross-population number of segregating sites by length (XP-nSL) (Szpiech et al. 2021) and the population branch statistic (PBS) (Yi et al. 2010).

We then integrated results from compatible statistics into two Fisher's Combined Scores (FCS). We combined the ranked results of iHS and nSL as FCS_{within}, comprising signals obtained from *within*-population tests, and the ranked results of XP-nSL and PBS as $FCS_{between}$, comprising signals obtained from *between*-population tests (see Materials and Methods). While the FCS_{within} represents the integration of signals found in a given ecoregion without a reference population to contrast against, the $FCS_{between}$ represents the integration of signals identified in a given ecoregion when contrasting against the other. In the Highlands, $FCS_{between}$ denotes signals obtained from comparing against the Lowlands and conversely, $FCS_{between}$ in the Lowlands represents signals obtained from comparing against the Highlands.

We initially identified the top 50 candidate genomic regions putatively under positive selection in each ecoregion, per FCS computation, using an empirical outlier approach (see "Analyses of Genomic Signatures of Positive Selection" in Materials and Methods) (Fig. 2). This yielded

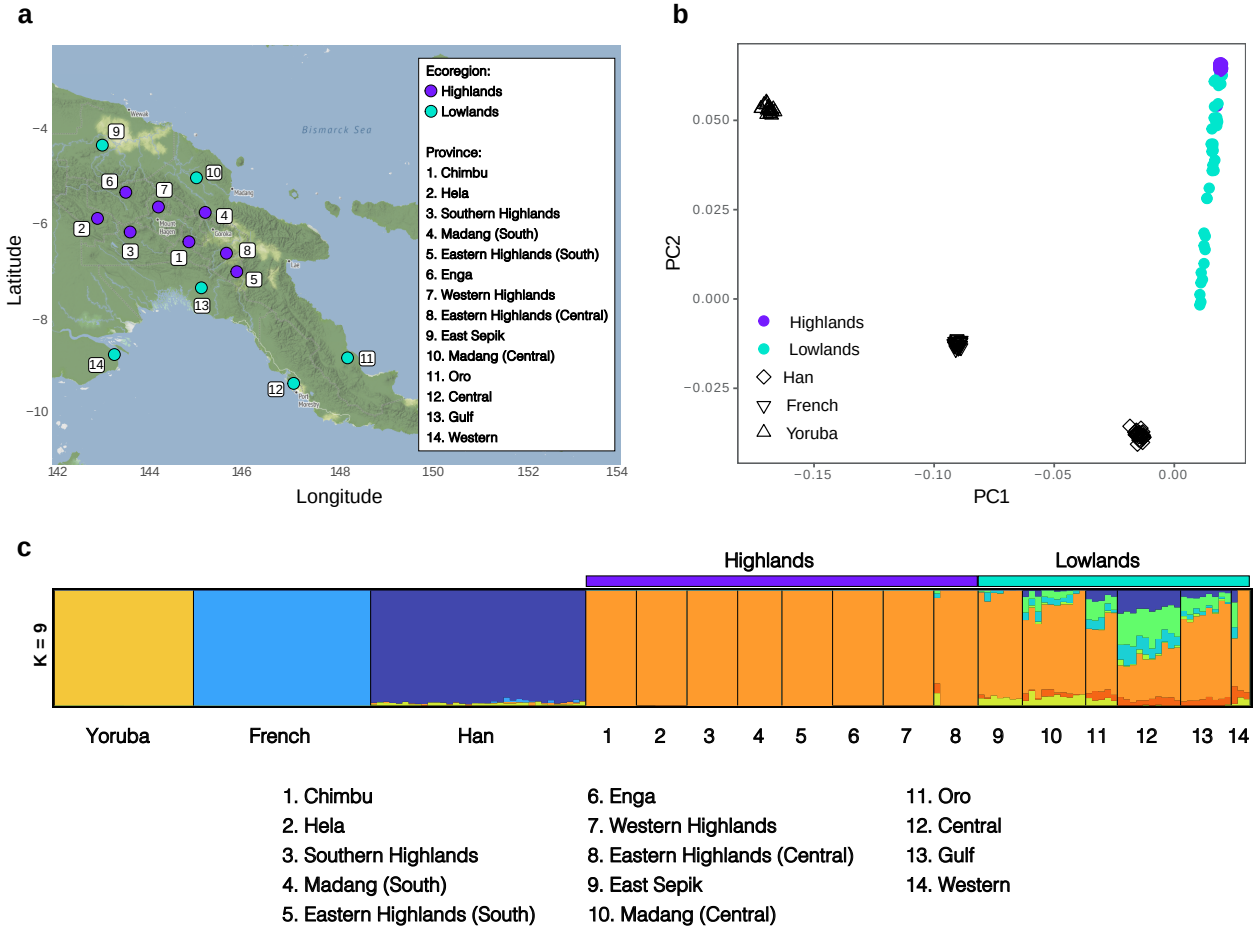

Fig. 1. Sampling locations and global ancestry patterns. a) Map of PNG indicating the locations, labeled as provinces, of sampled individuals. Data points are colored by ecoregion. b) PCA of individuals from mainland PNG colored by ecoregion, including continental reference populations from HGDP. c) Admixture plot of continental references and the subset of samples from OGVP used in this study at K=9, with samples from mainland PNG classified by province. Additional K values and ancestry components defined by OGVP samples outside of PNG are shown in SM Fig. 2. Map tiles from a) by Stamen Design, under CC BY 4.0. with data by OpenStreetMap, under ODbL.

two separate sets of candidate regions: 50 regions identified through FCS_{within} and another 50 through $FCS_{between}$, accounting for a total of 100 candidate regions in each ecoregion. Subsequently, we annotated all genes contained in these candidate regions using ANNOVAR (Wang et al. 2010) and noncoding variants in regulatory elements to target genes using GeneHancer (Fishilevich et al. 2017). Additionally, we used UniProt (The UniProt Consortium 2017) and GeneCards (Stelzer et al. 2016) to further characterize gene functions (supplementary table S2 and S3, Supplementary Material online). We considered a candidate region under positive selection to be unique to an ecoregion if it did not overlap in 50 kb or more with a candidate region from the other ecoregion, or if it was not entirely contained within another from the other ecoregion. The top 10 unique candidate regions in each ecoregion, per FCS computation, are described in Table 1 for the Highlands, and Table 2 for the Lowlands. We also identified shared candidate regions between ecoregions among the initial top 50 signals, per FCS computation, and separated them for individual analysis as a distinct group (supplementary table S4, Supplementary Material online).

In each ecoregion, we identified the top 50 strongest candidate regions based on FCS computations, amounting to 100 candidate regions per ecoregion (50 from FCS_{within} and 50 from $FCS_{between}$ for both Highlands and Lowlands). Out of these 200 signals (100 in each ecoregion), we found 23 to be common between the two ecoregions: 11 originating from the Highlands and 12 from the Lowlands. In these shared candidate regions, two were ranked among the top 10 in highlanders, and four among the top 10 in lowlanders (Table 3). After excluding these 23 shared candidate regions from the initial set of 100 in each ecoregion, we analyzed the remaining candidate regions, resulting in 41 regions from FCS_{within} and 48 from $FCS_{between}$ in the Highlands (supplementary table S2, Supplementary Material online),

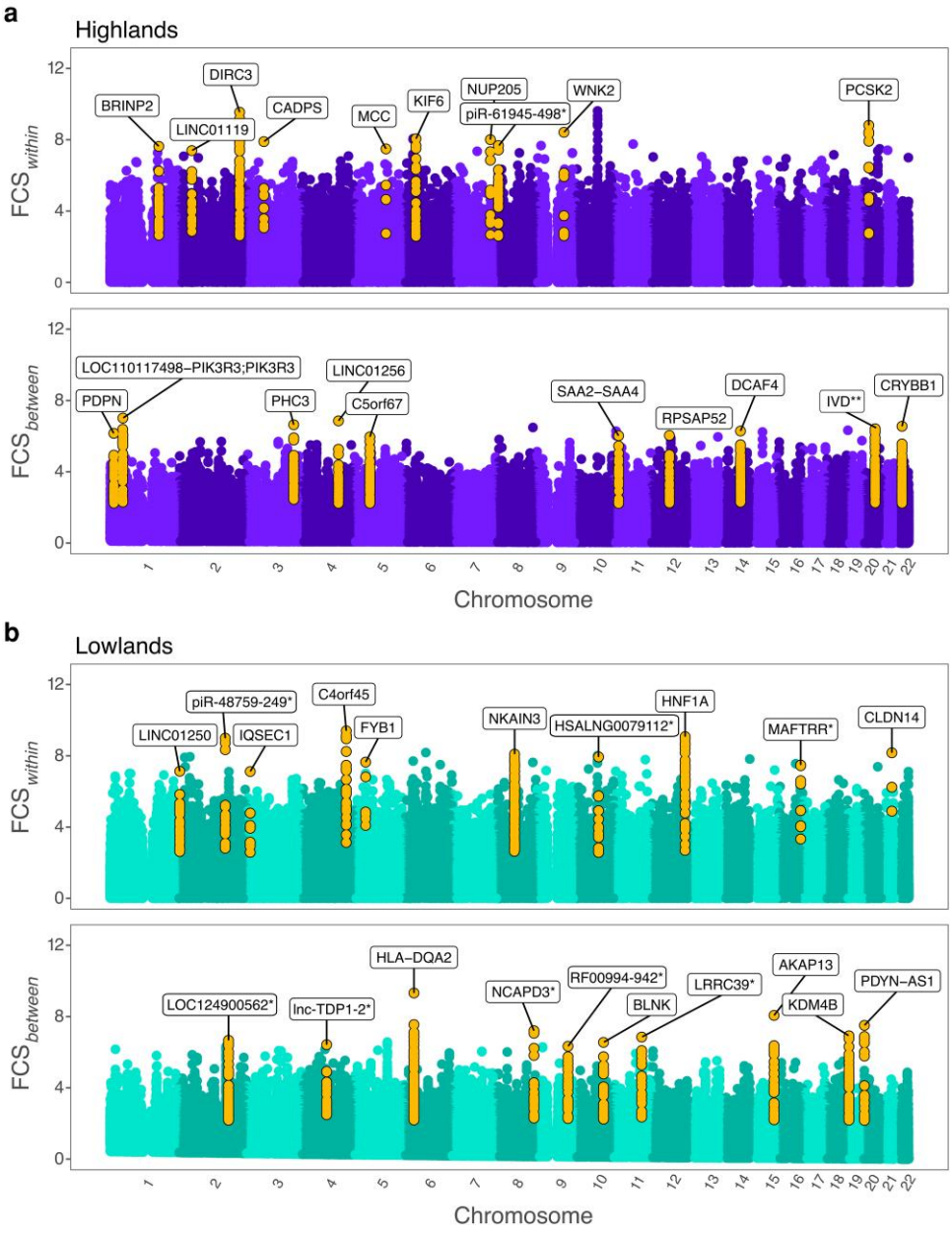

Fig. 2. Manhattan plots representing the FCS scores of within and between approaches in the Highlands a) and Lowlands b) of PNG. In yellow, the top 10 candidate regions identified in each test, unique to each ecoregion. Genomic regions are considered under selection exclusively in an ecoregion when they are not found within the top 50 signals obtained by tests performed on other ecoregions. The gene mapping to the top scoring SNP in every candidate region is indicated. A separator denoted by ";" indicates that the candidate SNP is located between two genes. Asterisks indicate putative enhancer (*) or putative promoter/enhancer (**) of a gene.

and 40 from FCS_{within} and 48 from $FCS_{between}$ in the Lowlands (supplementary table S3, Supplementary Material online).

Candidate Genes Under Selection in the Highlands

When analyzing the top 10 candidate regions under selection private to the Highlands (Table 1), we observed that both FCS approaches, FCS_{within} and $FCS_{between}$, identified genes associated with similar biological processes and pathways, albeit pinpointing different specific genes within these categories. The genes involved in glucose metabolism, *PCSK2* and *DIRC3* were identified by FCS_{within}, while $FCS_{between}$ identified *LDHA*, known to not only play a role in glucose metabolism but also in the Hypoxia-Inducible Factor (HIF) pathway (Taylor and Scholz 2022). Similarly,

Table 1 Top 10 candidate regions under selection per FCS computation exclusively found in PNG highlanders

Approach	Candidate region	Candidate genes	Top scoring SNP for the region	DAF
FCS_{within}				
	2:218050827-218462340	*HSALNG0022188***, ***DIRC3***, *DIRC3; DIRC3-AS1*	2:218335551-rs77469227-A>C	0.83
	20:17274061-17468983	***PCSK2***	20:17331246-rs11905775-C>A	0.34
	9:95994554-96125819	***WNK2***, *C9orf129, OAT ***	9:96068234-rs10992701-G>T	0.71
	6:39244959-39794653	*KCNK17, NHLRC3*, KCNK16,* ***KIF6****, DAAM2*	6:39608028-rs2466408-T>C	0.33
	7:135241045-135376893	***NUP205***, *STMP1, SLC13A4*	7:135293128-rs4294134-A>G	0.09
	3:62667103-62788035	***CADPS***	3:62733212-rs3843397-T>C	0.5
	8:5700776-5923150	***piR-61945-498*** ***	8:5793059-rs12546419-A>C	0.74
	1:177158631-177651729	***BRINP2***, *HSALNG0030894 *, LINC01645*	1:177165301-rs10753144-G>A	0.27
	5:112464953-112530716	***MCC***	5:112528038-rs76233603-C>T	0.33
	2:47065699-47249239	***LINC01119***, *MCFD2, MCFD2; TTC7A, TTC7A*	2:47090944-rs11125097-T>C	0.38
$FCS_{between}$				
	1:45804050-46633874	*MUTYH, TESK2, MMACHC, PRDX1, AKR1A1, NASP, CCDC17, GPBP1L1, TMEM69, IPP, IPP **, MAST2,* ***LOC110117498- PIK3R3; PIK3R3****, PIK3R3, LOC101929626*	1:46621740-rs7512966-G>A	0.99
	4:133474342-133769194	*CM034962-406 **,* ***LINC01256***	4:133524015-rs62311724-T>C	0.99
	3:169659270-170187431	*SAMD7, LOC100128164, SEC62, GPR160,* ***PHC3****, PRKCI, SKIL, CLDN11, SLC7A14, SLC7A14-AS1*	3:169900909-rs9866351-G>A	0.78
	22:26765167-27180156	*SEZ6L, ASPHD2, HPS4, SRRD, TFIP11, LOC100507599, TPST2,* ***CRYBB1****, CRYBA4, MIAT, MIATNB, PIGV **	22:27004964-rs4822748-G>A	0.97
	20:40573522-40904924	***IVD*** ***, PTPRT*	20:40697404-rs1078607-C>T	0.92
	14:73242095-74264624	*DPF3,* ***DCAF4****, ZFYVE1, RBM25, PSEN1, PAPLN, LOC101928123, NUMB, LOC101928143, RIOX1, HEATR4, ACOT2, ACOT4, ACOT6, DNAL1, PNMA1, ELMSAN1, LOC100506476*	14:73379091-rs12878887-C>T	0.75
	1:13863307-14154114	*LRRC38, MIR23AHG **,* ***PDPN****, PRDM2*	1:13913180-rs355028-C>T	0.97
	12:66018473-66327632	*LOC100507065, B4GAT1-DT **,* ***RPSAP52****, HMGA2, HMGA2- AS1*	12:66217345-rs6581658-A>G	0.95
	11:18228871-18423504	*SLC25A51P4, SAA2-SAA4; SAA4, SAA2; SAA2-SAA4,* ***SAA2- SAA4****, SAA1, HPS5, GTF2H1, LDHA*	11:18278136-rs57322649-C>T	0.87
	5:55665535-55959484	*piR-36770-015 *, LINC01948,* ***C5orf67***	5:55829831-rs30363-G>A	0.90

In bold, genes mapped to the highest scoring SNP within each candidate region. A separator denoted by ";" indicates that the candidate SNP is located between two genes. Asterisks indicate putative enhancer (*) or putative promoter/enhancer (**) of a gene. DAF is given for the top scoring SNP in each candidate region identified in PNG highlanders. Genomic coordinates are given for GRCh37.

genes participating in ion homeostasis and transport were detected (*WNK2*, *KCNK16*, *KCNK17*, and *SLC13A4* by FCS_{within}, and *LRRC38* by $FCS_{between}$). Notably, FCS approaches identified genes related to the Wnt pathway, a major developmental pathway that plays a role in hypoxia tolerance (Gersten et al. 2014) (*DAAM2*, *MCC*, *LINC01119* identified by FCS_{within} and *PNMA1* identified by $FCS_{between}$).

Genes in the top 10 candidate regions from the Highlands identified only by the FCS_{within} approach are related to cell structure and transport (*KIF6*, *NUP205*), neurotransmission and hormone release (*CADPS*), and functions related to the regulation of the nervous system and cell cycle control (*BRINP2*, *STMP1*). Through this approach, we also found *MCFD2*, a gene with a role in blood coagulation, and *TTC7A*, involved in the development of the intestinal epithelial and immune system.

The FCS approach designed to identify candidate regions under selection in the Highlands when compared against the Lowlands, $FCS_{between}$, identified genes linked to various biological processes, some of which could have implications for high-altitude adaptation. This is the case of candidate genes involved in oxidative stress and redox regulation (*MUTYH, PRDX1, ASPHD2, HEATR4*), heart and vascular cell function (*MIAT, MIATNB, DPF3*), Notch signaling (*PSEN1, NUMB*), fatty acid metabolism (*ACOT2, ACOT4, ACOT6*) and inflammatory response (*SAA2-SAA4, SAA4, SAA2, SAA1*). Other signals in the top 10 candidate regions are related to the immune system (*PIK3R3, ZFYVE1, PDPN*), epigenetic modifications and chromatin structure (*PHC3, RIOX1, ELMSAN1, PRDM2, RPSAP52, HMGA2, GTF2H1*), retina function and eye development (*SAMD7, CRYBB1, CRYBA4, SLC7A14*), translational modifications (*TPST2, DCAF4*) as well as post-transcriptional gene regulation (*TFIP11, RBM25*), cell migration and signal transduction (*SEC62, GPR160, CLDN11*), pigmentation (*HPS4, HPS5*), circadian rhythm (*SRRD*), and ciliary function (*DNAL1*).

Candidate Genes Under Selection in the Lowlands

Similarly, we analyzed the top 10 candidate regions with signatures of selection private to the Lowlands (Table 3)

Table 2 Top 10 selection signatures per FCS computation exclusively found in PNG lowlanders

Approach	Candidate region	Candidate genes	Top scoring SNP for the region	DAF
FCS_{within}				
	4:159802267-160151176	*FNIP2*, ***C4orf45***, *SLAMF8* *, *RAPGEF2*	4:159981640-rs581849-T>C	0.84
	12:121135508-121506154	*MLEC, SPPL3, XLOC_009911, HNF1A-AS1, GLI2* *, ***HNF1A***, *C12orf43, C12orf43; HNF1A, OASL*	12:121416988-rs2244608-A>G	0.88
	2:165769051-166281132	*SLC38A11,* ***piR-48759-249*** *, *SCN3A, SCN2A*	2:165926663-rs6432808-C>T	0.07
	21:37817906-37828862	***CLDN14***	21:37818305-rs9974322-A>G	0.91
	8:62851130-64548431	*lnc-RASGRP2-2* **, ***NKAIN3***, *NKAIN3-IT1, GGH, TTPA, YTHDF3-AS1, YTHDF3*	8:63830366-rs2061133-G>A	0.79
	10:80491647-80676558	***HSALNG0079112*** *	10:80540697-rs536394-T>C	0.15
	5:39125955-39290912	***FYB1***, *C9*	5:39217226-rs700191-C>A	0.17
	16:79681212-79832757	***MAFTRR*** *, *LINC01229*	16:79741871-rs62043309-C>T	0.10
	2:2773164-3125571	*C20orf141* **, ***LINC01250***	2:3012290-rs2031028-A>G	0.75
	3:13066535-13198052	***IQSEC1***, CCDC3 *	3:13106959-rs13061267-G>A	0.60
$FCS_{between}$				
	6:32627747-32720632	*HLA-DQB1-AS1, HSALNG0090275* **, *HLA-DQB1*, ***HLA-DQA2***, *MIR3135B*	6:32682174-rs3104404-C>A	0.73
	15:85903051-86085802	*DDAH1* *, ***AKAP13***	15:86003866-rs11074256-C>T	0.07
	20:1883939-1972460	*SIRPA,* ***PDYN-AS1***	20:1926275-rs6136501-C>T	0.92
	8:134586637-134654193	***NCAPD3*** *, *ST3GAL1*	8:134587451-rs12678754-C>A	0.82
	19:4883932-5191473	*ARRDC5, UHRF1,* ***KDM4B***, *PTPRS*	19:5109701-rs1017821-A>G	0.95
	11:100237927-100413065	***LRRC39*** *	11:100334324-rs7928162-C>T	0.93
	2:177331710-177483123	***LOC124900562*** *, *MIR1246, LINC01116*	2:177370519-rs7568419-A>C	0.64
	10:97962687-98226277	***BLNK***, *DNTT, OPALIN, TLL2*	10:98011364-rs3789930-A>G	0.84
	4:90199007-90335200	*GPRIN3,* ***lnc-TDP1-2*** *	4:90232381-rs12646308-C>T	0.59
	9:110944986-111075953	***RF00994-942*** *	9:110997199-rs12682814-T>C	0.70

In bold, genes or intergenic elements mapped to the highest scoring SNP within each candidate region. A separator denoted by ";" indicates that the candidate SNP is located between two genes. Asterisks indicate putative enhancer (*) or putative promoter/enhancer (**) of a gene. DAF is given for the top scoring SNP in each candidate region identified in PNG lowlanders. Genomic coordinates are given for GRCh37.

and found that both FCS approaches revealed genes linked to analogous biological processes and pathways. However, as in the Highlands, these signals pinpointed distinct genes within the identified categories. The strongest signals were detected in genes related to immune response: *FYB1, C9,* and *OASL* identified through FCS_{within}, and *HLA-DQB1*, *HLA-DQA2*, *SIRPA*, *ST3GAL1*, *BLNK,* and *DNTT* identified when comparing Lowlands against Highlands with $FCS_{between}$. Other genes were also identified to be associated in neuronal functions and ion transport: *NKAIN3*, *SCN3A*, *SCN2A* (FCS_{within}), *GPRIN3*, and *OPALIN* ($FCS_{between}$), as well as genes involved in spermatogenesis and sperm function (*C4orf45* obtained with FCS_{within} and *ARRDC5* obtained with $FCS_{between}$). Other biological processes identified through both approaches include genes playing a role in histone modification and gene expression (*HNF1A*, *FNIP2* with FCS_{within} and *UHRF1*, *KDM4B* with $FCS_{between}$). As well as cell adhesion and migration (*IQSEC1* with FCS_{within} and *MIR1246*, *LINC01116* with $FCS_{between}$).

When looking at genes identified only in the between population comparison of Lowlands against Highlands ($FCS_{between}$), we found *MIR3135B* and *PDYN-AS1* to be top candidates, although there is currently no assigned function or potential relationship to any biological process in Uniprot and Genecards for these two genes.

Regarding genes revealed only through the FCS_{within} approach, we found candidates involved in folate metabolism (*GGH*), vitamin E transport (*TTPA*), solute transport and cellular metabolism (*SLC38A11*), skeletal development (*TLL2*), RNA metabolism and modification recognition (*YTHDF3*), and serum urate control (*LINC01229*).

One of the strongest signals detected in both ecoregions with the FCS_{within} approach is the region containing genes *KAT6B, DUPD1, DUSP13,* and *SAMD8*. In the Highlands, these four genes are part of the highest-ranking candidate region, resulting from the FCS_{within} approach, which comprises 11 candidate genes in total. These four genes are also present in 7th strongest candidate region found in the Lowlands using the same FCS approach.

Among the 10 strongest shared signals in overlapping candidates under selection in each ecoregion, we found genes involved in gene regulation and chromatin structure (*KAT6B, ZBED9, ZNF311, ZSCAN12, ZSCAN23*), cell signaling and post translational modifications (*DUPD1, DUSP13, TRIM27, UBD*). Some also constitute olfactory receptors (*OR2B3, OR2J1, OR2J3*), antioxidant enzymes (*GPX5* and *GPX6*), and

Table 3 Shared candidate regions between highlanders and lowlanders found in the top 10 strongest signatures of any ecoregion

Candidate region in highlanders	Candidate region in lowlanders	Approaches (highlanders-lowlanders)	Ranking in highlanders	Ranking in lowlanders	Candidate genes
10:75868114-77128000	10:76581146-76919273	FCS_{within}-FCS_{within}	1	7	*VCL, AP3M1, ADK, LOC102723439,* **intergenic, *KAT6B, DUPD1, DUSP13, SAMD8,*** *VDAC2, COMTD1, ZNF503-AS1, LOC101929165*
6:29486065-29571609	6:28001610-29571609	FCS_{within}-FCS_{within}	5	9	Intergenic, *ZNF165, ZSCAN16-AS1, ZKSCAN8, ZNF192P1, TOB2P1, ZSCAN9, ZKSCAN4, ZSCAN26, PGBD1, ZSCAN31, ZKSCAN3, ZSCAN12, ZSCAN23, GPX6, GPX5, ZBED9, LINC00533, LINC01623, HCG14, TRIM27, LINC01556, HCG15, ZNF311, LOC100129636, OR2B3, OR2J1, OR2J3, OR2J2, LOC101929006, OR14J1, OR5V1, OR12D3, OR12D2, OR11A1, OR2H1, MAS1L,* ***LINC01015, UBD, SNORD32B, OR2H2, GABBR1***
6:28382262-29236722	6:28001610-29571609	FCS_{within}-FCS_{within}	18	9	***ZSCAN12, ZSCAN23,*** **intergenic, *GPX6, GPX5, ZBED9, LINC00533, LINC01623, HCG14, TRIM27, LINC01556, HCG15, ZNF311, LOC100129636, OR2B3, OR2J1, OR2J3, OR2J2, LOC101929006,*** *ZNF165, ZSCAN16-AS1, ZKSCAN8, ZNF192P1, TOB2P1, ZSCAN9, ZKSCAN4, ZSCAN26, PGBD1, ZSCAN31, ZKSCAN3, OR14J1, OR5V1, OR12D3, OR12D2, OR11A1, OR2H1, MAS1L, LINC01015, UBD, SNORD32B, OR2H2, GABBR1*
6:74853727-74944004	6:74612516-74982935	FCS_{within}-FCS_{within}	43	4	Intergenic, ***LOC101928516***
5:117397489-117566290	5:117331544-117880338	FCS_{within}-$FCS_{between}$	49	8	***LINC02147***, *LINC02148,* intergenic, *LINC02208*

We define a region to be shared if there is at least a 50 kb overlap between a candidate region in the top 50 of any FCS approach in the other ecoregion, or if one is contained within another. In bold, overlapping genes in shared candidate regions. Genomic coordinates are given for GRCh37.

some others are related to the immune response (*HCG14, HCG15*), lipid metabolism (*SAMD8*), and RNA processing (*SNORD32B*). We also found shared uncharacterized long noncoding RNAs (*LINC00533, LINC01623, LINC01556, LOC100129636, LOC101929006, LOC101928516, and LINC02147*).

Haplotype Structure in Candidate Regions for Positive Selection in the Highlands and Lowlands

To better understand the haplotype structure of the top 10 candidate regions under selection when comparing Highlands and Lowlands ($FCS_{between}$), we constructed haplotype networks and determined the decay of site-specific extended haplotype homozygosity (EHHS) between ecoregions around each candidate region using the R packages pegas (Paradis 2010) and rehh (Gautier et al. 2017), respectively. In haplotype networks, although haplotypes in a candidate region for the highlands tend to be more frequent in highlanders than in lowlanders (Fig. 3a, supplementary fig. S3, Supplementary Material online), and vice versa (Fig. 3c, supplementary fig. S4, Supplementary Material online), we did not observe a clear set of haplotypes exclusive to highlanders in any Highlands candidate region, nor exclusive to lowlanders in any Lowlands candidate region. This suggests that populations in the Highlands and Lowlands have not reached complete differentiation in these haplotypes, potentially indicating their shared ancestral genetic background or ongoing gene flow. Furthermore, it is likely that the selection observed is acting on standing genetic variation, meaning that the advantageous haplotype was already present in the low-altitude population before becoming selected for in the high-altitude environment.

For some candidate regions, we observe a clear structure between several haplotypes, evidenced by a large number of mutations separating haplotypes that are more frequent in highlanders than in lowlanders, and vice versa (Fig. 3c, supplementary fig. S3, Supplementary Material online, supplementary fig. S4, Supplementary Material online). This could indicate that, for a candidate region identified in one ecoregion, selection is not acting upon, or at least not as strongly, in the other ecoregion at the same genomic location. Furthermore, we find the decay of EHHS to be slower in highlander haplotypes compared to lowlander ones at candidate regions of the Highlands (Fig. 3b, supplementary fig. S5, Supplementary Material online), and a slower decay in lowlander haplotypes compared to highlander ones at candidate regions of the Lowlands (Fig. 3d, supplementary fig. S6, Supplementary Material online). Additionally, at certain candidate regions, haplotypes in both ecoregions decay at similar rates, although highlander haplotypes in candidate regions of the Highlands are present at a higher frequency than in the Lowlands, and vice versa (supplementary fig. S7, Supplementary Material online, supplementary fig. S8, Supplementary Material online). This is consistent with our $FCS_{between}$ approach, designed to capture regions with both EHHS and high-frequency differentiation between populations of each ecoregion.

Functional Annotation of Candidate Regions for Positive Selection

To functionally characterize the identified candidate regions to be evolving under selection, we conducted a gene-set over-representation analysis on the genes in candidate regions unique to each ecoregion, per FCS computation, using the functional annotation clustering tool from DAVID (Sherman et al. 2022). This allowed us to identify distinct trends of gene clustering whose functions shed light on which biological processes might underlie adaptive events in the Highlands compared to the Lowlands, and vice versa.

In the Highlands, genes obtained from our FCS_{within} analysis revealed clustering annotation of candidate genes associated with immune response (Allograft rejection), whereas candidate genes from $FCS_{between}$ against the Lowlands revealed clustering annotation of terms associated with inflammatory response (Acute phase response), constituents of the eye lens (Eye lens Protein), oxidative processes (hydrogen peroxide catabolic process), lipid metabolism (acyl-CoA hydrolase activity), virus transcription elongation (HIV transcription elongation), and cellular signaling and development (Notch signaling pathway) (Fig. 4a and supplementary table S5, Supplementary Material online).

Conversely, in the Lowlands, clustering annotation of candidate genes was primarily observed in terms such as RNA metabolism (RNA metabolic process), obtained with FCS_{within}, as well as neurotransmitter receptor activity (trace-amine receptor activity) with $FCS_{between}$ against the Highlands (Fig. 4b and supplementary table S6, Supplementary Material online).

Archaic Introgression in Candidate Genomic Regions Under Selection

Given the known history of archaic introgression in PNG, we looked for overlap between candidate genomic regions under selection and archaic segments identified in 25 whole genomes of PNG highlanders (Malaspinas et al. 2016) using IBDmix (Chen et al. 2020). As the method is only suited for whole-genome sequencing data, we used these genomes to identify regions of archaic introgression in PNG highlanders. Given that we include individuals from the same highland locations reported in Malaspinas et al. and have an overlap of 19 samples shared between

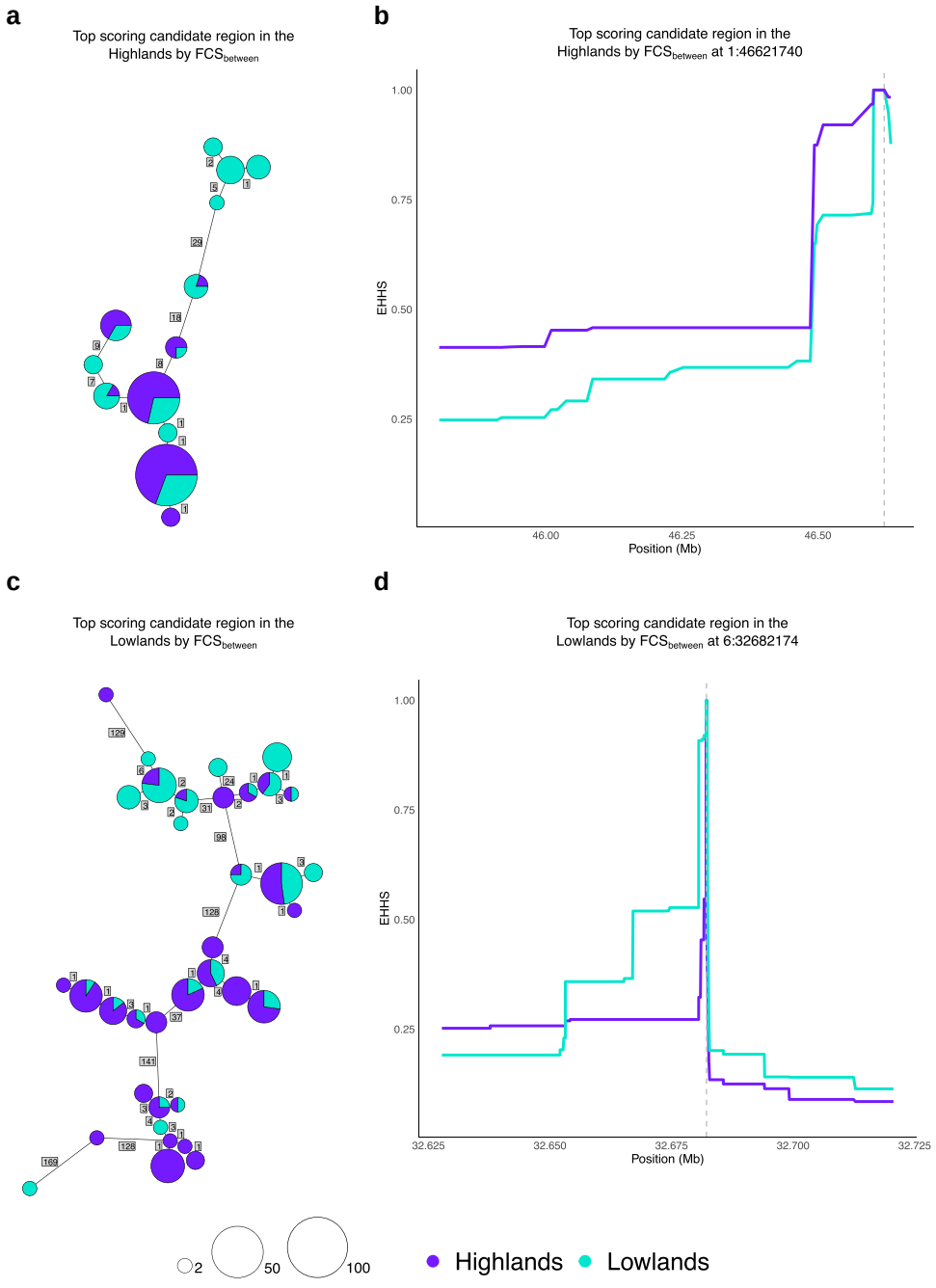

Fig. 3. Haplotype networks and site-specific decay of extended haplotype homozygosity (EHHS) at top-ranking candidate genomic regions by $FCS_{between}$. a) Haplotype network for the top-ranking candidate region in the Highlands when compared against the Lowlands. b) EHHS decay plot around the top scoring SNP in the top-ranking candidate region in the Highlands when comparing against Lowlands. c) Haplotype network for the top-ranking candidate region in the Lowlands when comparing against Highlands. d) EHHS decay plot around the top-scoring SNP in the top-ranking candidate region in Lowlands when comparing against Highlands. In haplotype networks, each pie chart represents one unique haplotype. The radius of the pie chart is proportional to the number of chromosomes with that haplotype. Pie proportions indicate the frequency of each haplotype amongst populations. The edges indicate the number of pairwise differences between the joined haplotypes. In EHHS plots, each colored line represents the extent of homozygosity across all sampled haplotypes of a given ecoregion starting from a focal marker. In haplotype networks and EHHS plots, each color is respective to an ecoregion.

data sets, we used the identified archaic regions as a proxy to compare with the selection results of highlanders obtained from the array data. Using the whole-genome data, for each candidate region under selection, we created bins along the genome of the same length and quantified the number of highlander individuals that had segments

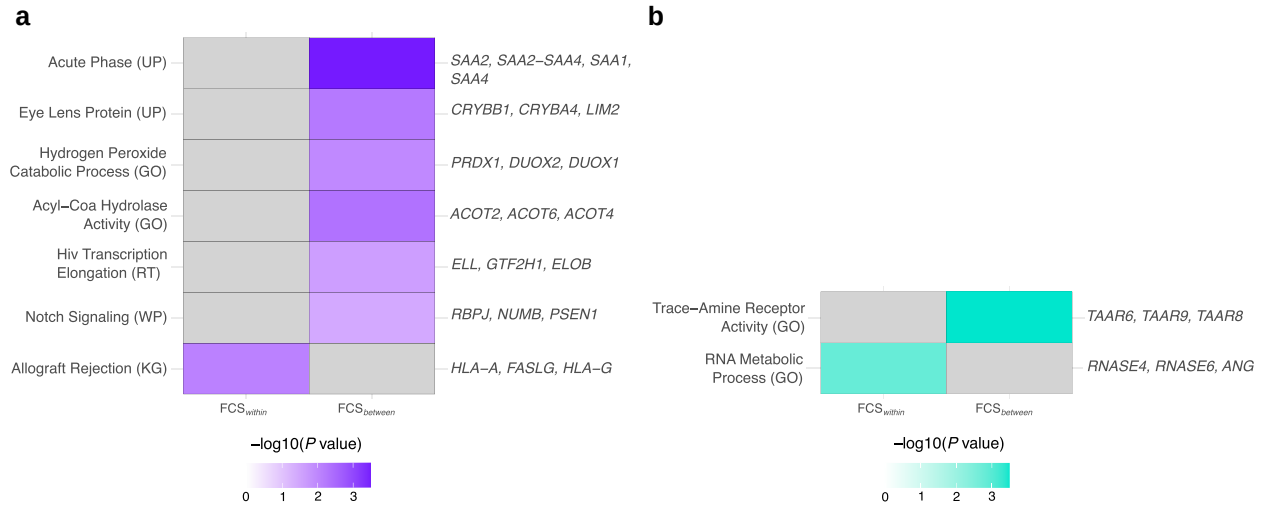

Fig. 4. Clustering of candidate genes in functional annotation terms using DAVID in a) PNG highlanders and b) PNG lowlanders for both FCS_{within}, left columns, and $FCS_{between}$, right columns. Functional annotation terms shown on the left *y* axis indicate the representative hit of each cluster (i.e. the functional term within each cluster with the best *P*-value). In parentheses, databases where the annotation term comes from: UniProt (UP), Gene Ontology (GO), Reactome (RT), WikiPathways (WP), KEGG (KG). Genes within functional annotation terms are indicated on the right *y* axis. Intensity of the color corresponds to the -log(*P*-value) of the clustering score. Higher color intensities indicate better clustering scores on the *P*-value scale. Grey slots indicate functional annotation terms absent in a given FCS computation.

of archaic introgression in each bin. We then used the extreme of the distribution (top 0.05) of these values to define a threshold for the significance of archaic introgression. Interestingly, two of our candidate regions for selection passed the overlap threshold of archaic introgression. The first significant result is an FCS_{within} hit (individuals with overlaps = 17; 0.05 threshold = 10 individuals) found within the HLA region in chromosome 6 (Fig. 5a). It comprises several genes and pseudogenes, with functions related to immunity (*HCG4, HLA-V, HLA-G, HLA-A, HCG9, HLA-J, TRIM31, TRIM31-AS1, TRIM10, IFITM4P, RNF39*) and the less characterized *ZNRD1ASP*. The second significant hit we identified in highlanders was in a candidate region shared with lowlanders (individuals with overlaps = 11; 0.05 threshold = 10 individuals) in chromosome 10 (Fig. 5b). This region includes mostly protein-coding genes related to a wide variety of functions including protein transport (*AP3M1*), adenosine metabolism (*ADK*), chromatin remodeling (*KAT6B*), phosphatase activity (DUPD1, DUSP13), lipid metabolism (*SAMD8*), and energetic exchange (*VDAC2*).

Discussion

In this study, we identified different genomic regions as candidates to be evolving under positive selection in populations from the Highlands and Lowlands of PNG through the analysis of genome-wide genotype data from 62 highlanders and 43 lowlanders from the OGVP.

PNG Highlanders Exhibit Signals Related to High Altitude Adaptation

Given the difference in altitude between populations from the Highlands and the Lowlands, we initially hypothesized that selection could have targeted genomic regions in loci relevant for high altitude adaptation in populations from the Highlands. Previous studies on high-altitude adaptation in Tibetans, Andeans and Ethiopians have implicated genes like *EPAS1*, *EGLN1*, and *DEC2* (Bigham et al. 2010; Huerta-Sánchez et al. 2013; Li et al. 2019) to be involved in the HIF pathway as the primary targets of selection. Although those genes were not among our candidates of selection in PNG highlanders, we found new candidates that can be playing a role under hypoxic conditions and in processes related to the body's response to high altitude.

The HIF pathway and hypoxia have several effects on different aspects of metabolism like glycolysis, anaerobic metabolism, oxidative stress, fatty acid metabolism and the inflammatory response (Eltzschig and Carmeliet 2011; Ge et al. 2015; McGarry et al. 2018; O'Brien et al. 2020; Taylor and Scholz 2022). One of the top candidate genes under selection found in the Highlands is *LDHA* ($FCS_{between}$). Its encoded enzyme, lactate dehydrogenase A, is integral in glycolysis and plays a crucial role in the metabolism of lactate, while also being particularly active in anaerobic metabolism. Moreover, the transcription factor HIF1 is known to enhance the expression of *LDHA* (Firth et al. 1995; Semenza et al. 1996), facilitating a transition from oxidative to glycolytic metabolism during periods of hypoxia (Taylor and Scholz 2022). *LDHA* is also a known

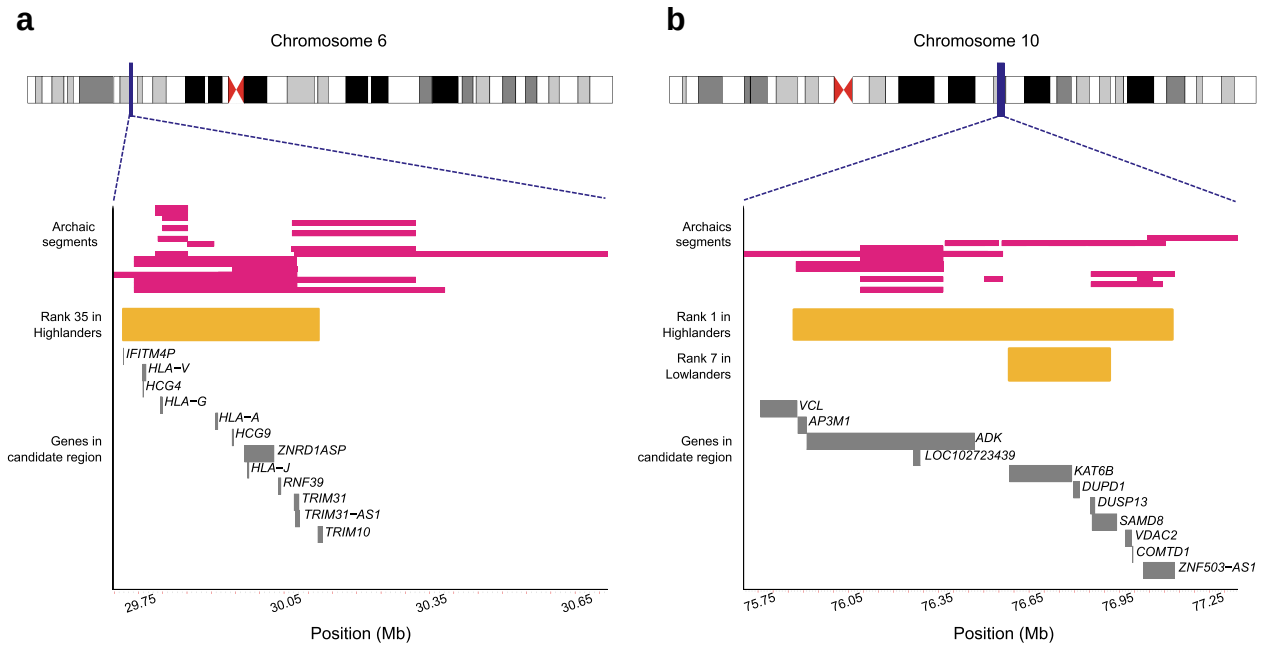

Fig. 5. Candidate regions overlapping with archaic segments and reaching statistical significance. a) Overlap of archaic segments with a candidate region in chromosome 6 obtained by FCS_{within} (rank 35) in PNG highlander populations. b) Overlap of archaic segments with a candidate region in chromosome 10 in PNG highlander populations (rank 1, FCS_{within}), shared with a candidate region in PNG lowlander populations (rank 7, FCS_{within}). The chromosome location of the overlap is depicted at the top of each panel. On the *y* axis, archaic segments identified in PNG highlanders by IBDmix are colored in pink, candidate region(s) under selection are colored in yellow, and genes contained within the candidate region(s) are colored in gray. On the *x* axis, numbers represent the genomic positions expressed in megabases (Mb).

biomarker for acute mountain sickness (AMS), showing increased levels of lactate dehydrogenase A in nonaffected individuals (Yang et al. 2022). In high-altitude populations from Peru, expression levels of *LDHA* have been suggested to predict the presence of CMS as well (Xing et al. 2008). A study on a high-altitude Tibetan population found that increased lactate concentration is positively correlated with high-altitude adaptation, while also suggesting a possible decrease in the activity of fatty acid oxidation at high altitude (Ge et al. 2012).

Hypoxic conditions also favor oxidation processes leading to an increase in reactive oxygen species (ROS), which in turn aids in promoting angiogenesis (Kim and Byzova 2014), the process of new blood vessel growth, which has been found to play a key adaptive role in other high altitude populations (Droma et al. 2022; Ferraretti et al. 2023). Within our top candidate regions under selection, we found the dual oxidase genes *DUOX2*, *DUOX1,* and *DUOXA1* in chromosome 15 as part of the 13th strongest candidate region (supplementary table S2, Supplementary Material online) and the antioxidant enzyme *PRDX1* in chromosome 1 as part of the strongest candidate region ($FCS_{between}$) (Table 1) forming an enriched cluster related to hydrogen peroxide catabolic process in our functional annotation clustering approach (Fig. 4a). *PRDX1*, primarily involved in peroxide elimination accumulated during metabolism, has been identified as a candidate gene under selection in an Andean population residing at intermediate altitudes (Eichstaedt et al. 2015). Moreover, *DUOX2*, which is expressed mainly in the lung and blood vessels and plays a role in the hydrogen peroxide pathway and ROS, has also been identified as a candidate gene under selection in highland Andeans (Jacovas et al. 2018; Borda et al. 2020), underscoring the importance of the interplay between ROS management and angiogenesis in populations adapted to high altitude.

Inflammation is interdependent with the production of ROS and hypoxia (McGarry et al. 2018). Within the top 10 candidate regions under selection in the Highlands we found the serum amyloid A (SAA) genes *SAA1, SAA2-SAA4, SAA4,* and *SAA2* in chromosome 11 ($FCS_{between}$) to be also present in a functional cluster related to the acute phase response (Fig. 4a), a series of events that occur as a response to inflammation. *SAA4 and SAA1* have been identified as biomarkers to predict AMS in humans (Julian et al. 2014; Lu et al. 2016). Additionally, SAA is also implicated in the cardiovascular system (Shridas and Tannock 2019), which can be affected by hypoxia and is targeted by selection in other high-altitude populations (Mallet et al. 2021). Moreover, we find the gene *BRINP2* among our top 10 candidate regions under selection ($FCS_{between}$). *BRINP2* is an homolog of *BRINP3* (Kawano

et al. 2004), a gene associated with cardiovascular phenotypes in humans and found to be under selection in high-altitude Andean populations (Crawford et al. 2017). The high homology between their coded proteins suggests they might perform a similar or interchangeable function (Berkowicz et al. 2016), therefore playing a possibly similar role in high altitude adaptation in PNG highlanders. Furthermore, we also found *SANBR* (aka *KIAA1841*) among our candidate genes under selection ($FCS_{between}$) (supplementary table S2, Supplementary Material online), a newly identified gene to be under selection in high-altitude Tibetan populations and suggested to play a role in the adaptation of cardio-pulmonary functions (Zheng et al. 2023).

Within the candidate regions under selection in the Highlands we identified four genes associated with the Notch signaling pathway (supplementary table S2, Supplementary Material online), three of them identified by $FCS_{between}$ (*NUMB, PSEN1, RBPJ*), found also to form a cluster using DAVID (Fig. 4a), and one of them identified through FCS_{within}, *MAML3* (SM table S2). Of these, *NUMB* and *PSEN1* are notably among the top 10 candidate regions under selection (Table 1). The Notch signaling pathway is a key modulator in cell fate determination in the development of vertebrate and invertebrate species, a regulator of blood vessel structure and angiogenesis, and participates in a cross-talk with the HIF pathway in the cellular response to low oxygen conditions (Landor and Lendahl 2017). It has been proposed as a candidate pathway for high altitude adaptation in multiple organisms (O'brien et al. 2022), including human Tibetan and Andean populations (Bigham et al. 2009; Xin et al. 2020). The *NUMB* gene is involved in cell fate determination and has been shown to enhance Notch signaling by ubiquitination processes (Luo et al. 2020). *NUMB* has been identified as a key differentially expressed gene involved in high-altitude adaptation in Tibetan chickens when compared against lowland breeds (Zhang et al. 2020). Although *PSEN1* has been primarily studied in the context of Alzheimer's disease, the protein encoded by *PSEN1* is essential for the cleavage of Notch receptors during cell signaling and has also been suggested to play a part in the regulation of the oxygen sensing pathway in mice (Kaufmann et al. 2013; Newman et al. 2014). *RBPJ* is an essential transcription factor in the Notch signaling pathway that is also involved in angiogenesis (Wang et al. 2009; Lake et al. 2014; Ramasamy et al. 2014). It was identified as a candidate gene to be under selection in a study of a high-altitude macaque population and also as a potential novel regulator of hypoxia adaptation in Tibetan pigs (Jia et al. 2016; Szpiech et al. 2021). The gene *MAML3* is also a transcriptional mediator of the Notch signaling pathway (Kitagawa 2015; Yamasaki et al. 2020). Although not much research has been done on this gene in the context of adaptation, it has been shown to be upregulated, along with *RBPJ*, by hypoxic conditions in pancreatic adenocarcinoma cells (Onishi et al. 2016). Moreover, *MAML3* and *ASPHD2*, another candidate gene identified by $FCS_{between}$ that is among our top 10 candidate regions under selection, are also upregulated together in glioblastoma cells under hypoxic conditions (Seidel et al. 2010).

The acyl-CoA thioesterase (ACOT) gene family encodes enzymes, located primarily in the mitochondria and peroxisomes, that play an important role in lipid metabolism by maintaining the balance of free and activated fatty acids in the body through oxidation processes (Brocker et al. 2010). We find the ACOT genes *ACOT2, ACOT4,* and *ACOT6* to be within the genes in our top 10 candidate regions under selection obtained from $FCS_{between}$ (Table 1), and as part of a functional annotation cluster (Fig. 4a). Although not much has been studied on these particular gene family in the context of human adaptation, hypoxic conditions have been found to attenuate fatty acid oxidation in adapted high-altitude populations through targeting other genes involved in lipid metabolism (O'Brien et al. 2020).

We also found *MMACHC,* involved in vitamin B_{12} metabolism, in the candidate genomic region under selection in chromosome 1 showcasing the strongest signals ($FCS_{between}$) (Table 1). Under hypoxia, it has been shown to be significantly repressed by the transcription factor HIF1 (Kiessling et al. 2022).

Thus far our results suggest that, in contrast to other locations where human populations have adapted to high altitude, the HIF pathway has not been the primary target of natural selection related to increased high altitude in the Highlands of PNG. Instead, the observed molecular signatures suggest that these populations might be adapting to high altitude through pathways that mediate the response to inflammatory and oxidative stresses inducible by hypoxia as well as through the Notch signaling pathway, a regulator of the response to low oxygen conditions, with a potential link between these and angiogenesis. Furthermore, we note that ongoing or past selective pressures related to pathogens might have also given rise to signals in genes involved in inflammation and oxidative stress, as these too are processes affected by the systemic defense responses to pathogens (Tschopp 2011). Future research that can identify a causal distinction between the two is needed.

Signals in Genes Related to Eye Components in PNG Highlanders

In populations from the Highlands, we identify five candidate genes related to retina function and eye development: *SAMD7, CRYBB1, CRYBA4, SLC7A14,* and *LIM2* ($FCS_{between}$) (supplementary table S2, Supplementary

Material online). With *SAMD7, CRYBB1, CRYBA4,* and *SLC7A14* being found within our top 10 candidate regions under selection (Table 1). In our functional annotation clustering approach, we find *CRYBB1, CRYBA4,* and *LIM2* comprising an annotation cluster related to Eye lens proteins (Fig. 4a). *SAMD7* is a crucial component in the development of rod photoreceptor cells in the retina (Omori et al. 2017). *CRYBB1* and *CRYBA4* are two adjacent genes that form part of the crystallin gene cluster on chromosome 22, coding for major proteins of the eye that maintain the transparency and refractive index of the lens (Siggs et al. 2017). Previous studies have found that mutations in any of these two genes can lead to congenital cataracts (Mackay et al. 2002; Billingsley et al. 2006). *SLC7A14* is a cationic transporter that is involved in the development of the retina, with mutations associated with retinitis pigmentosa (Jin et al. 2014). Other candidate genes encoding solute carrier proteins (SLC) have also been previously identified as candidates under selection in indigenous populations of Malaysia (Deng et al. 2014). *LIM2* codes for an integral membrane protein of the lens with a potential role in cell junction communication (Louis et al. 1989; Tenbroek et al. 1992), and mutations on this gene have also been associated with congenital cataracts (Irum et al. 2016). A 2017 study (Lee et al. 2019) found that PNG has a high global prevalence of blindness in individuals over 50, primarily due to cataracts and uncorrected refractive errors (Lee et al. 2019). The prevalence is notably higher in the Highlands, suggesting health disparities. Although trachoma, an eye disease caused by *Chlamydia trachomatis*, is endemic in PNG (Bourne et al. 2013; Handley et al. 2018), it is not a primary cause of blindness, implying a potential genetic resistance among the PNG population (Ko et al. 2016). Research in a Gambian population identified the glycoprotein-encoding gene *TNC* as upregulated during trachoma infection (Natividad et al. 2010), which is also among the candidate genes under selection in PNG highlanders ($FCS_{between}$) (supplementary table S2, Supplementary Material online). A recent study revealed that the glycolysis pathway plays a role in adapting human lens epithelial cells to hypoxic conditions (Huang et al. 2023) with *LDHA*, described above, identified among other genes as upregulated under hypoxia. However, glycolysis-related genes were not prominent in our DAVID annotation clustering analysis. Acknowledging the complexity of eye diseases, which involve both social and demographic factors as well as care access, we concur with other authors that the variations in eye health in PNG may reflect broader health disparities across different regions. Moreover, these findings enhance the understanding of eye diseases in PNG from a genetic standpoint.

Signals Related to the Immune Response in PNG Highlanders

We identified genes linked to the immune system, consistent with evidence of pathogens being a primary force driving evolutionary adaptation across human populations (Fumagalli et al. 2011). In PNG, where infectious diseases like TB, malaria, and pneumonia are prevalent (World Health Organization [WHO] 2018), we found the candidate genes *ELL, GTF2H1*, and *ELOB* (supplementary table S2, Supplementary Material online) ($FCS_{between}$), in our clustering results in DAVID related to HIV Transcription Elongation (Fig. 4a). It is important to note that while these genes are currently annotated as being involved in HIV infection, given the relatively recent appearance of the virus, we believe that these identified selection candidates might be involved with viral infection response in a wider sense and not necessarily related to HIV. These identified candidate genes are associated with RNA elongation and are crucial in viral infection (Schaeffer et al. 1993; Aso et al. 1995; Kong et al. 2005; Bergeron et al. 2010; Jain et al. 2016; Liu et al. 2020). However, there is currently no conclusive evidence that genetic variations in any of them provide adaptive advantages against any infectious disease. Notably, *ELOB*, which interacts with HIF1A (Greer et al. 2012), could also be selected for high-altitude adaptation.

Allograft rejection, a process related to the immune response, was an annotation cluster formed by our candidate genes obtained from FCS_{within} in the Highlands that passed our cluster enrichment thresholds in DAVID. The cluster is comprised of the genes *HLA-A, HLA-G,* and *FASLG* (Fig. 4a). *FASLG*, prominently expressed in B cells and involved in apoptosis (Hahne et al. 1996), has been recognized for its role in immunity against *Mycobacterium tuberculosis* (Van Rensburg et al. 2017; Loxton and Van Rensburg 2021). In PNG, the incidence of TB has been observed to be higher in Western Province (McBryde 2012), included in this study as part of the Lowlands ecoregion. Moreover, *HLA-A* and *HLA-G* genes, central in the immune response, are of note: *HLA-G* potentially serves as a TB disease state biomarker (Saurabh et al. 2016), and *HLA-A* alleles have been linked to TB in a Mali population (Kone et al. 2019).

Signals in PNG Lowlanders are Mainly Related to the Immune Response

To put into contrast the pattern of signatures of selection found in PNG highlanders, we also scanned for signatures distinguishable in lowlanders. Given the generally warmer and humid conditions in the Lowlands ecoregion, we initially hypothesized that selection could have targeted genomic regions mainly related to the immune response in populations from this area compared to the Highlands.

The candidate genomic region with the strongest signal of selection in the Lowlands when compared against the Highlands, $FCS_{between}$, encompassed genes from the human leukocyte antigen (HLA), a crucial component of the human immune system, *HLA-DQB1* and *HLA-DQA2* (Fig. 2b,

Table 3). This suggests that pathogen-driven selection has likely acted particularly in the PNG Lowland populations as well. Moreover, we found the trace-amine-associated receptor (TAAR) genes *TAAR6, TAAR9,* and *TAAR8* obtained with $FCS_{between}$ (supplementary table S3, Supplementary Material online) to cluster together in terms related to the trace–amine receptor activity (Fig. 4b). TAARs belong to a family of transmembrane receptors and have been mainly studied for their involvement in functions of the central nervous system (Rutigliano and Zucchi 2020). TAARs have also been suggested to play a role in leukocyte function (Christian and Berry 2018). We find that functional annotation clustering of terms related to RNA metabolic processes comprises the genes *RNASE4, RNASE6* and *ANG* obtained with FCS_{within}. *RNASE4* and *RNASE6* code for proteins belonging to the ribonuclease A superfamily, with roles in immune modulation and inflammation processes (Lu et al. 2018). *RNase 6* shows high antimicrobial activity against Gram-positive and Gram-negative bacteria (Becknell et al. 2015; Pulido et al. 2016). Whereas *RNase 4* contributes to innate defenses in the kidney and urinary tract by showing antimicrobial activity (Bender et al. 2021). The gene *ANG* also codes for a ribonuclease, RNase 5, that is the same superfamily as RNase 4 and RNase 6. *ANG* is a mediator of new blood vessel formation through angiogenesis that shows both antimicrobial and antiviral capabilities as well as being involved in the inflammatory response (Hooper et al. 2003; Cocchi et al. 2012; Sheng and Xu 2016).

Shared Signals of Selection Between Ecoregions

Candidate regions under selection shared between the Highlands and Lowlands, among the top 10 signals of any ecoregion (Table 3), include genes related to cell signaling, post-translational modifications, and to a lesser extent antioxidant enzymes and lipid metabolism. Shared selection signals in these genes could hint at pressures that populations from both regions are potentially subjected to, or at adaptive benefits in both environmental contexts. We note that such patterns could also be the result of gene flow across ecoregions, suggesting that gene flow between PNG highlanders and lowlanders has been more significant than previously thought. Moreover, it is also possible that we are detecting selection signatures that were already present in an ancestral population before divergence of the Highlands and Lowlands. Further research is warranted to elucidate this.

Candidate Regions Under Selection in PNG Overlap with Archaic Introgressed Segments

Previous studies have identified an important role of archaic introgression in human adaptation (Racimo et al. 2015; Gittelman et al. 2016). In some cases, this adaptation appears to be related to the immune response directly (Abi-Rached et al. 2011; Urnikyte et al. 2023). Here, we find a proportion of candidate genes under selection within archaic introgressed regions in PNG highlanders (35th strongest signal with FCS_{within}) to be part of a region in the HLA system (*HLA-V, HLA-G, HLA-A, HLA-J, HCG4, HCG9, RNF39, TRIM31, TRIM31-AS1, TRIM10*), crucial in the immune response, aligning with previous findings on the role of archaic introgression in immunity in PNG populations (Vespasiani et al. 2022; Yermakovich et al. 2024). This observation provides evidence that adaptive signals in PNG highlanders could be partially attributed to archaic genetic introgression, possibly as a response to local pathogenic challenges or other environmental stressors. Alternatively, the observed patterns could be the result of shared ancestral variation maintained by balancing selection, which is hard to distinguish from introgression in such a complex region like the HLA. We find a second candidate region under selection that is also shared with PNG lowlanders and located within an archaic introgressed segment identified in PNG highlanders. This shared region constitutes the strongest ranking signal in PNG highlanders (FCS_{within}) and the 7th strongest signal in PNG lowlanders (FCS_{within}) (Table 3), implying again that some of the adaptive traits in both highland and lowland populations may have deep evolutionary roots originating from archaic hominins. This candidate region comprises the genes *VCL, AP3M1, ADK, KAT6B, DUPD1, DUSP13, SAMD8, VDAC2, COMTD1, and ZNF503–AS1*, which appear to not be directly centered on the immune response and share no common function among all of them. The candidate region identified in PNG lowlanders is entirely contained within that of the one identified in PNG highlanders (Fig. 5b). It is interesting to note that a majority of archaic segments stack within the ADK gene, in a section of the shared signal that seems exclusive to the highlanders, since it does not directly overlap with the limits of the region identified in lowlanders. *ADK* codes for the enzyme adenosine kinase, an important regulator of homeostatic and metabolic networks that is mainly involved in energy metabolism, neurodevelopment, and angiogenesis. Given the diverse functions of these genes within the shared candidate region, further research is warranted to better understand their potential contributions to adaptative processes. We also note that the results presented here are restricted to a comparison of candidate regions under selection (identified as particular to highlanders, or shared between them and lowlanders) with regions of introgression identified only in highlander populations, due to limited data availability where whole genomes are required (see "Archaic introgression in candidate regions under selection" in Materials and Methods). Further analyses on whole-genome data from lowlander populations may reveal new patterns of introgression.

Conclusions

Our work focused on studying genetic adaptation in populations from the highlands and lowlands of PNG. The signals observed in the highlands clustered in annotations of diverse functional terms related to the inflammatory response, oxidative processes, constituents of the eye lens, lipid metabolism, virus transcription elongation, and Notch signaling. We propose that even if highlanders of PNG are exposed to moderate altitude, adaptation to lower oxygen conditions might be occurring in genes and processes that respond to the physiological stress caused by increased altitude or by playing a role in modulating the HIF pathway response. Additionally, we identified signatures at genes known to serve as biomarkers for mountain sickness, hinting at the possibility of complex adaptive responses in high-altitude settings. It is interesting to note that some of the candidate genes detected here have also been identified as under selection in other human and animal populations residing at intermediate and high altitudes (Ge et al. 2012; Eichstaedt et al. 2015; Jia et al. 2016; Jacovas et al. 2018; Borda et al. 2020; Zhang et al. 2020; Szpiech et al. 2021; Zheng et al. 2023) strongly indicating convergent evolution in response to living at increased altitude and highlighting the significance of these genes across species and geographic locations. Furthermore, we also identify one candidate region in the Highlands composed by the genes *VAMP1*, *GAPDH*, *LPAR5*, *CHD4*, *IFFO1,* and *NCAPD2* that has been also identified by André et al in a highland PNG population, with an association to blood composition in the UK biobank André et al. (2024). We report this region as ranking 20th in our analyses and believe that discrepancies in our results stem mainly from differences in the surveyed populations, methodology, and type of genetic data. As for the Lowlands, we found signals mainly related to the immune response, likely reflecting a higher exposure to pathogens in this ecoregion.

While the observed selection signatures in PNG predominantly suggest distinct selective pressures across the two ecoregions, we also identified shared selection signals between the two. These shared signals, rather than pointing to a single set of functional implications, span a diverse array of biological processes. This complexity highlights the need for further investigation to fully understand the evolutionary dynamics and potential interplay of selective forces acting upon these regions. Finally, we identify selection signals within archaic introgressed segments to be mainly related to immune-associated genes. This finding indicates that archaic genetic contributions may have been pivotal in shaping the immune response of modern humans inhabiting islands in the Pacific. Further integration of past and recent demographic events experienced by populations of PNG into appropriate demography and selection models is required to enhance the understanding of adaptive events and their temporal dynamics. In addition, exploring analyses of polygenic selection could offer valuable insights into the genetics of adaptation of PNG populations. By characterizing signatures of positive selection in populations from PNG across different ecoregions, this work contributes to our understanding on the genetic basis of human adaptation in Pacific islanders.

Materials and Methods

Samples and Genotype Data

The genotype data analyzed here derives from the Oceanian Genome Variation Project (OGVP) in which 981 individuals from a broad range of diverse populations across Oceania were genotyped for 1.8 million SNPs with the Illumina Infinium Multi Ethnic Global Array (MEGA) and mapped to the GRCh37 (hg19) version of the human reference genome (Quinto-Cortés et al. 2024). This data was merged with reference continental populations from HGDP (Bergström et al. 2020): Yoruba from Nigeria in Africa, French from France in Europe, and Han from China in East Asia at overlapping sites with plink 1.9 (Chang et al. 2015), solving strand incompatibilities by flipping SNPs on the reverse strand to the forward using snpflip (binary retrieved January 4, 2022 from https://github.com/biocore-ntnu/snpflip).

Quality control on the merged data set was performed using Plink 1.9. We restricted analyses to keep individuals from continental references and the populations reported in this paper, as well as autosomal biallelic SNPs, removing all variants with a missing genotype rate higher than 5% and individual samples with missing genotype rates higher than 10%. We also filtered out SNPs with minor allele frequencies less than 1%. The final merged data set retained 780 individuals and 543,287 SNPs.

Population Structure and Subsetting of PNG Samples

PCA was performed on the OGVP-HGDP merged data set using Plink 1.9 (supplementary fig. S1, Supplementary Material online) and ancestry proportions were estimated using ADMIXTURE 1.3.0 (Alexander et al. 2009). In accordance with the scope of this study, we restricted our analyses to PNG. To reduce the potential confounding impact of admixture on the detection of signals of positive selection, which can introduce novel genetic variations and linkage disequilibrium patterns that may either distort or hide the hallmarks of a selective sweep (Souilmi et al. 2022), we limited our samples to individuals from the main island of PNG (Fig. 1a to c). This decision was informed by the substantial heterogeneity in admixture profiles observed among individuals from the offshore islands (Figure S3 in Quinto-Cortés et al. 2024). Therefore, individuals from the main island of PNG were subsetted, retaining a total of 105 individuals

from 12 provinces across the main island and 84 individuals from continental references.

We determined elevation data for populations in the main island of PNG using the geographic coordinates of sampling locations, obtaining elevation data in meters relative to the local mean sea level from the Google Maps Elevation API (Google Maps Elevation API. Retrieved June 7, 2023, from https://developers.google.com/maps/documentation/elevation/start). With this information, we define the ecoregions "Highlands" and "Lowlands" as regions above and below 1000 m relative to the local mean sea level (mLMSL), respectively. Since two provinces with two different sampling locations displayed different altitudes, we further divided them indicating the location (south and central) of each one, leading to a total of 14 different locations (supplementary table S1, Supplementary Material online). With this, we categorized samples from the provinces of Enga, Southern Highlands, Eastern Highlands, Chimbu (Simbu), Hela, Western Highlands, Madang and Eastern Highlands into the Highlands ecoregion, and categorized samples from the provinces of Gulf, Oro, Madang, Central, Western and East Sepik into the Lowlands. We detected two sampling locations in Madang with different altitudes, so we assigned each one to its corresponding ecoregion. In total, the Highlands comprised 62 samples and the Lowlands 43 samples. We note that the terms "PNG highlander" and "PNG lowlander" used throughout the text do not intend to reflect specific cultural, historical, social, ethnic, or linguistic groupings. Instead, they are meant to merely denote human groups living at distinct altitudes.

Analyses of Genomic Signatures of Positive Selection

We investigated signatures of positive selection through population differentiation and haplotype-based methods, characterizing signals particular to populations in each ecoregion. In total, 62 individuals from the Highlands and 43 individuals from the Lowlands were analyzed for signatures of positive selection.

After phasing the merged data set with SHAPEIT4 (Delaneau et al. 2019), haplotype-based statistics were computed using selscan (Szpiech and Hernandez 2014). To look for signals within populations, we computed the iHS (Voight et al. 2006), a statistic based on haplotype homozygosity used to identify areas of the genome that have undergone recent positive selection, and the nSL (Ferrer-Admetlla et al. 2014), which measures the decay of the number of segregating sites with distance from a core allele, in a similar fashion to the incorporation of the extended haplotype homozygosity (EHH) in iHS. Ancestral and derived allele information necessary for these statistics was obtained from 1KGP (The 1000 Genomes Project Consortium et al. 2015). To look for signals between populations we computed the XP-nSL, an implementation of nSL into a cross-population statistic that uses a reference population to compare against (Szpiech et al. 2021), and the PBS, based on differences in allele frequency between two closely related population groups in comparison to an outgroup (Yi et al. 2010). Genome-wide PBS values were calculated using an in-house R script, implementing F_{st} values (Weir and Cockerham 1984) obtained with PLINK 1.9 from the pairwise comparisons between Highlands, Lowlands and French, using the latter as an outgroup.

To improve the robustness of signals obtained from selection statistics, we integrated the results of compatible statistics into two FCS. With the FCS score being equal to the sum, over two statistics, of –log10 (rank of the SNP in a given statistic/total number of SNPs) following (Lopez et al. 2019). We integrated the ranked results of iHS and nSL as FCS_{within}, comprising signals obtained from within population statistics (supplementary tables S8 and S10, Supplementary Material online). And integrated the ranked results of XP-nSL and PBS as $FCS_{between}$, comprising signals obtained from between population statistics (supplementary tables S9 and S11, Supplementary Material online). Additionally, the integration of these two statistics into a FCS leverages the identification of regions with both EHH and high frequency differentiation between populations of each ecoregion. In the Highlands, results of $FCS_{between}$ represent signals obtained from comparing against the Lowlands and conversely, results of $FCS_{between}$ in the Lowlands represent signals obtained from comparing against the Highlands.

To reduce false positives, strongest candidate regions (i.e. regions showcasing the highest FCS values) were identified through an empirical outlier approach as in Caro-Consuegra et al. 2022. For each ecoregion and FCS computation, candidate genomic regions of positive selection were defined as those harboring SNPs above the top 5% of the empirical distribution, only if they also contained at least three other SNPs with scores above the top 1% in a surrounding ±100 Kb vicinity. We kept the initial top 50 strongest candidate regions in each ecoregion, per FCS, for downstream analysis.

We considered candidate regions of positive selection specific to each ecoregion as those not overlapping in 50 Kb or more a candidate region in the top 50 of the other ecoregion, or if one candidate region is entirely contained within another from the other ecoregion. If a shared candidate region was found within the top 50, we assigned that region as "shared" and excluded it from the analysis of both ecoregions, keeping the remaining strong candidate regions that are unique in each ecoregion. Shared candidate regions were analyzed as its own group (supplementary tables S4 and S12 to S15, Supplementary Material online).

Furthermore, SNPs within the strongest candidate regions specific to each ecoregion and those found to be shared between them were annotated with ANNOVAR (Wang et al. 2010). For the top 10 strongest candidate regions unique to an ecoregion, we determined putative candidate genes driving the signal by considering the location of the highest-scoring SNP in each of them. Intergenic signals in the strongest candidate regions were assigned to surrounding genes only if they were at a distance below 50 Kb. Intergenic SNPs that were not assigned to any gene using this criterion were annotated with GeneHancer, a data base of human regulatory elements (enhancers and promoters) and their inferred target genes widely used to annotate noncoding variants (Fishilevich et al. 2017) (supplementary tables S16 and S17, Supplementary Material online). The gene with the highest association score was assigned to each intergenic SNP and the putative type of regulatory element (Promoter, Enhancer, Promoter/Enhancer) was annotated. The function of candidate genes was characterized using UniProt (The UniProt Consortium 2017) and GeneCards (Stelzer et al. 2016). Derived allele frequencies (DAF) were obtained for the top-scoring SNP in the top 10 candidate regions of each FCS approach.

Haplotype Networks and Decay of Site-specific EHH

Haplotype networks were constructed including highlanders and lowlanders for each of the top 10 candidate regions, in the Highlands and lowlands, obtained by $FCS_{between}$ (supplementary figs. S3 and S4, Supplementary Material online). To limit the number of haplotypes to display, we restricted the networks to include only haplotypes with a minimum frequency of two chromosomes. We used the R package pegas (Paradis 2010) to build networks that connect haplotypes based on pairwise differences.

The site-specific EHHS measures the probability that any two randomly chosen chromosomes from a population are homozygous over a given surrounding chromosomal region of a focal marker (Sabeti et al. 2007; Tang et al. 2007). EHHS was calculated to compare the decay of homozygosity between Highlands and Lowlands around candidate regions obtained by $FCS_{between}$, using the R package rehh (Gautier et al. 2017). Both normalized (Tang et al. 2007) and unnormalized (Sabeti et al. 2007) EHHS comparisons between ecoregions were calculated for each of the top 10 candidate regions identified in the Highlands (supplementary figs. S5 and S7, Supplementary Material online), and for each of the top 10 candidate regions identified in the Lowlands (supplementary figs. S6 and S8, Supplementary Material online). Each EHHS calculation began at a focal marker set to the position of the top-scoring SNP in that candidate region. Normalized EHHS facilitates the direct comparison of the rate of decay of homozygosity between population haplotypes by standardizing the initial EHHS value at the focal marker to 1. This removes the influence of different baseline levels of homozygosity, which may vary due to differences in allele frequencies between populations. In contrast, the unnormalized version shows the decay in homozygosity while retaining differences in initial haplotype frequency at and around the focal marker. These differences can reflect genomic regions with EHH in one population that are also differentiated when compared to another, highlighting genomic regions that could have potentially been targeted by selection.

Functional Annotation Clustering

Analyses of functional annotation clustering were conducted separately on the candidate genes located within the remaining unique candidate regions under selection in the Highlands (41 FCS_{within}, 42 $FCS_{between}$), and the remaining unique candidate regions under selection in the Lowlands (40 FCS_{within}, 48 $FCS_{between}$) after filtering out shared signals between the two. For each FCS and ecoregion, these analyses utilized the functional annotation clustering tool provided by the Database for Annotation, Visualization and Integrated Discovery (DAVID) version 6.8 (Sherman et al. 2022), using the whole-genome as background and high classification stringency. We included OMIM (Amberger et al. 2015), GO (Harris et al. 2004), KEGG (Kanehisa et al. 2017), WikiPathways (Agrawal et al. 2023) and REACTOME (Fabregat et al. 2017) as query databases and only considered clusters with enrichment scores equal to and above 1.3, equivalent to a *P*-value of 0.05 (supplementary tables S5 and S6, Supplementary Material online).

Archaic Introgression in Candidate Regions Under Selection

Archaic introgression was extracted from whole-genome sequences of 25 Papuan highlander individuals (Malaspinas et al. 2016). For this purpose, we used IBDmix (Chen et al. 2020), a probabilistic method based on identity-by-descent (IBD) designed to identify segments of archaic introgression without need for an unadmixed modern human reference. The data was analyzed per chromosome, and only autosomes were used. The Papuan data was restricted only to bi-allelic single nucleotide variants (SNVs). For the Neanderthal and Denisovan individuals (Meyer et al. 2012; Prüfer et al. 2014), the data was restricted to SNVs, and multi allelic sites were removed. For each archaic hominin, IBDmix genotype data was created using the generate_gt command with the Papuan samples as the modern data. A mask was created for analysis with each of the Neanderthal and Denisova by combining the corresponding minimal filter mask (Prüfer et al. 2014) for each hominin with the 1000 Genomes Project

accessibility mask using bedtools (Quinlan and Hall 2010). IBDmix was then run on the data using the *ibdmix* command, with a minor allele count threshold (*-m*) of 2 and applying the combined mask for each archaic hominin (*-r*), and all other values set as default. We created summary files of our results using the summary.sh script provided by IBDmix, filtering with a minimum length of 50 kb and a minimum LOD score of five (supplementary table S7, Supplementary Material online).

Papuans are believed to have introgression from both the Neanderthal and the Denisova, and several regions were called as both Neanderthal and Denisovan. For this reason, our calls were combined into a single "archaic" category. Using the GenomicRanges R package (Lawrence et al. 2013), we looked for overlap between candidate regions under selection and archaic segments. For each candidate region under selection, the number of Papuan samples with archaic segments that overlap in at least 50 kb with the region was quantified. For each region with at least three hits, the genome was divided into bins of the same length as the region, and the distribution of the number of individuals with archaic segments overlapping at least 50 kb with each bin was calculated. The top 5% of this distribution was used as the threshold for significance of a region (supplementary fig. S9, Supplementary Material online).

Supplementary Material

Supplementary material online is available at *Genome Biology and Evolution* online.

Acknowledgments

We thank all sample donors who contributed to this study; Michael Alpers, former head of the Papua New Guinea Institute of Medical Research (IMR), and George Koki for valuable support from the IMR; Sarah Kaewert, Victor Moreno-Mayar, Diego Ortega del Vecchyo, Alexander de Luna Fors, and Rafael Montiel Duarte for assistance on methods discussion; Santiago Medina Muñoz, Daniela Orozco, and Karla Benítez for helpful feedback. Data generation for this work was supported by the Oxford-Stanford Big Data Initiative and we thank the Mexican Council for Science, Humanities, and Technology (CONAHCyT) for fellowship support awarded to R.G-B., S.V-S, M.C-H, and C.B-J (CVU numbers: 918462, 1242720, 1234769, and 873785, respectively).

Data Availability

Data availability is described in the source publication of the Oceanian Genome Variation Project (OGVP) (https://arxiv.org/abs/2405.09216). Genotype data for the subset of samples analyzed here are available upon request by contacting the corresponding author.

Ethics Statement

Samples were collected under supervision of the Papua New Guinea Institute of Medical Research (IMR) and have been stored with approval from the IMR for future research. Use of the samples for the Oceanian Genome Variation Project (OGVP) was approved by the Oxford Tropical Research Ethics Committee at the University of Oxford (approval reference: OXTREC 537-14). All samples were anonymised and coded at the time of collection according to the region of origin of the participant. Further sampling details and community engagement procedures are described in (Bergström et al. 2017).

Literature Cited

Abi-Rached L, Jobin MJ, Kulkarni S, McWhinnie A, Dalva K, Gragert L, Babrzadeh F, Gharizadeh B, Luo M, Plummer FA, et al. The shaping of modern human immune systems by multiregional admixture with archaic humans. Science. 2011:334(6052):89–94. https://doi.org/10.1126/science.1209202.

Agrawal A, Balcı H, Hanspers K, Coort SL, Martens M, Slenter DN, Ehrhart F, Digles D, Waagmeester A, Wassink I, et al. WikiPathways 2024: next generation pathway database. Nucleic Acids Res. 2023:52(D1):gkad960. https://doi.org/10.1093/nar/gkad960.

Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009:19(9):1655–1664. https://doi.org/10.1101/gr.094052.109.

Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. Nucleic Acids Res. 2015:43(D1):D789–D798. https://doi.org/10.1093/nar/gku1205.

André M, Brucato N, Hudjasov G, Pankratov V, Yermakovich D, Montinaro F, Kreevan R, Kariwiga J, Muke J, Boland A, et al. Positive selection in the genomes of two Papua New Guinean populations at distinct altitude levels. Nat Commun. 2024:15(1):3352. https://doi.org/10.1038/s41467-024-47735-1.

André M, Brucato N, Plutniak S, Kariwiga J, Muke J, Morez A, Leavesley M, Mondal M, Ricaut F-X. Phenotypic differences between highlanders and lowlanders in Papua New Guinea. PLoS One. 2021:16(7):e0253921. https://doi.org/10.1371/journal.pone.0253921.

Aso T, Lane WS, Conaway JW, Conaway RC. Elongin (sill): a multisubunit regulator of elongation by RNA polymerase II. Science. 1995:269(5229):1439–1443. https://doi.org/10.1126/science.7660129.

Becknell B, Eichler TE, Beceiro S, Li B, Easterling RS, Carpenter AR, James CL, McHugh KM, Hains DS, Partida-Sanchez S, et al. Ribonucleases 6 and 7 have antimicrobial function in the human and murine urinary tract. Kidney Int. 2015:87(1):151–161. https://doi.org/10.1038/ki.2014.268.

Bender K, Schwartz LL, Cohen A, Vasquez CM, Murtha MJ, Eichler T, Thomas JP, Jackson A, Spencer JD. Expression and function of human ribonuclease 4 in the kidney and urinary tract. Am J Physiol Renal Physiol. 2021:320(5):F972–F983. https://doi.org/10.1152/ajprenal.00592.2020.

Bergeron JRC, Huthoff H, Veselkov DA, Beavil RL, Simpson PJ, Matthews SJ, Malim MH, Sanderson MR. The SOCS-box of HIV-1

Vif interacts with ElonginBC by induced-folding to recruit its Cul5-containing ubiquitin ligase complex. PLoS Pathog. 2010: 6(6):e1000925. https://doi.org/10.1371/journal.ppat.1000925.
Bergström A, McCarthy SA, Hui R, Almarri MA, Ayub Q, Danecek P, Chen Y, Felkel S, Hallast P, Kamm J, et al. Insights into human genetic variation and population history from 929 diverse genomes. Science. 2020:367(6484):eaay5012. https://doi.org/10.1126/science.aay5012.
Bergström A, Oppenheimer SJ, Mentzer AJ, Auckland K, Robson K, Attenborough R, Alpers MP, Koki G, Pomat W, Siba P, et al. A neolithic expansion, but strong genetic structure, in the independent history of New Guinea. Science. 2017:357(6356):1160–1163. https://doi.org/10.1126/science.aan3842.
Berkowicz SR, Featherby TJ, Whisstock JC, Bird PI. Mice lacking Brinp2 or Brinp3, or both, exhibit behaviors consistent with neurodevelopmental disorders. Front Behav Neurosci. 2016:10:196. https://doi.org/10.3389/fnbeh.2016.00196.
Bigham A, Bauchet M, Pinto D, Mao X, Akey JM, Mei R, Scherer SW, Julian CG, Wilson MJ, López Herráez D, et al. Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. PLoS Genet. 2010:6(9):e1001116. https://doi.org/10.1371/journal.pgen.1001116.
Bigham AW, Mao X, Mei R, Brutsaert T, Wilson MJ, Julian CG, Parra EJ, Akey JM, Moore LG, Shriver MD. Identifying positive selection candidate loci for high-altitude adaptation in Andean populations. Hum Genomics. 2009:4(2):79–90. https://doi.org/10.1186/1479-7364-4-2-79.
Billingsley G, Santhiya ST, Paterson AD, Ogata K, Wodak S, Hosseini SM, Manisastry SM, Vijayalakshmi P, Gopinath PM, Graw J, et al. CRYBA4, a novel human cataract gene, is also involved in microphthalmia. Am J Hum Genet. 2006:79(4):702–709. https://doi.org/10.1086/507712.
Borda V, Alvim I, Mendes M, Silva-Carvalho C, Soares-Souza GB, Leal TP, Furlan V, Scliar MO, Zamudio R, Zolini C, et al. The genetic structure and adaptation of Andean highlanders and Amazonians are influenced by the interplay between geography and culture. Proc Natl Acad Sci U S A. 2020:117(51):32557–32565. https://doi.org/10.1073/pnas.2013773117.
Bourne RRA, Stevens GA, White RA, Smith JL, Flaxman SR, Price H, Jonas JB, Keeffe J, Leasher J, Naidoo K, et al. Causes of vision loss worldwide, 1990–2010: a systematic analysis. Lancet Glob Health. 2013:1(6):e339–e349. https://doi.org/10.1016/S2214-109X(13)70113-X.
Brocker C, Carpenter C, Nebert DW, Vasiliou V. Evolutionary divergence and functions of the human acyl-CoA thioesterase gene (ACOT) family. Hum Genomics. 2010:4(6):411–420. https://doi.org/10.1186/1479-7364-4-6-411.
Brucato N, André M, Hudjashov G, Mondal M, Cox MP, Leavesley M, Ricaut F-X. Chronology of natural selection in oceanian genomes. iScience. 2022:25(7):104583. https://doi.org/10.1016/j.isci.2022.104583.
Brucato N, André M, Tsang R, Saag L, Kariwiga J, Sesuki K, Beni T, Pomat W, Muke J, Meyer V, et al. Papua New Guinean genomes reveal the complex settlement of North Sahul. Mol Biol Evol. 2021: 38(11):5107–5121. https://doi.org/10.1093/molbev/msab238.
Caro-Consuegra R, Nieves-Colón MA, Rawls E, Rubin-de-Celis V, Lizárraga B, Vidaurre T, Sandoval K, Fejerman L, Stone AC, Moreno-Estrada A, et al. Uncovering signals of positive selection in Peruvian populations from three ecological regions. Mol Biol Evol. 2022:39(8):msac158. https://doi.org/10.1093/molbev/msac158.
Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015:4(1):7. https://doi.org/10.1186/s13742-015-0047-8.
Chen L, Wolf AB, Fu W, Li L, Akey JM. Identifying and interpreting apparent neanderthal ancestry in African individuals. Cell. 2020: 180(4):677–687.e16. https://doi.org/10.1016/j.cell.2020.01.012.
Choin J, Mendoza-Revilla J, Arauna LR, Cuadros-Espinoza S, Cassar O, Larena M, Ko AM-S, Harmant C, Laurent R, Verdu P, et al. Genomic insights into population history and biological adaptation in oceania. Nature. 2021:592(7855):583–589. https://doi.org/10.1038/s41586-021-03236-5.
Christian SL, Berry MD. Trace amine-associated receptors as novel therapeutic targets for immunomodulatory disorders. Front Pharmacol. 2018:9:680. https://doi.org/10.3389/fphar.2018.00680.
Cocchi F, DeVico AL, Lu W, Popovic M, Latinovic O, Sajadi MM, Redfield RR, Lafferty MK, Galli M, Garzino-Demo A, et al. Soluble factors from T cells inhibiting X4 strains of HIV are a mixture of β chemokines and RNases. Proc Natl Acad Sci U S A. 2012:109(14):5411–5416. https://doi.org/10.1073/pnas.1202240109.
Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, Parkhill J, Malla B, Berg S, Thwaites G, et al. Out-of-Africa migration and neolithic coexpansion of Mycobacterium tuberculosis with modern humans. Nat Genet. 2013:45(10):1176–1182. https://doi.org/10.1038/ng.2744.
Crawford JE, Amaru R, Song J, Julian CG, Racimo F, Cheng JY, Guo X, Yao J, Ambale-Venkatesh B, Lima JA, et al. Natural selection on genes related to cardiovascular health in high-altitude adapted Andeans. Am J Hum Genet. 2017:101(5):752–767. https://doi.org/10.1016/j.ajhg.2017.09.023.
Delaneau O, Zagury JF, Robinson MR, Marchini JL, Dermitzakis ET. Accurate, scalable and integrative haplotype estimation. Nat Commun. 2019:10(1):5436. https://doi.org/10.1038/s41467-019-13225-y.
Deng L, Hoh BP, Lu D, Fu R, Phipps ME, Li S, Nur-Shafawati AR, Hatin WI, Ismail E, Mokhtar SS, et al. The population genomic landscape of human genetic structure, admixture history and local adaptation in Peninsular Malaysia. Hum Genet. 2014:133(9):1169–1185. https://doi.org/10.1007/s00439-014-1459-8.
Droma Y, Hanaoka M, Kinjo T, Kobayashi N, Yasuo M, Kitaguchi Y, Ota M. The blunted vascular endothelial growth factor-A (VEGF-A) response to high-altitude hypoxia and genetic variants in the promoter region of the *VEGFA* gene in Sherpa highlanders. PeerJ. 2022:10:e13893. https://doi.org/10.7717/peerj.13893.
Eichstaedt CA, Antão T, Cardona A, Pagani L, Kivisild T, Mormina M. Genetic and phenotypic differentiation of an Andean intermediate altitude population. Physiol Rep. 2015:3(5):e12376. https://doi.org/10.14814/phy2.12376.
Eltzschig HK, Carmeliet P. Hypoxia and inflammation Schwartz. N Engl J Med. 2011:364(7):656–665. https://doi.org/10.1056/NEJMra0910283.
Enard D, Petrov DA. Evidence that RNA viruses drove adaptive introgression between neanderthals and modern humans. Cell. 2018:175(2):360–371.e13. https://doi.org/10.1016/j.cell.2018.08.034.
Fabregat A, Sidiropoulos K, Viteri G, Forner O, Marin-Garcia P, Arnau V, D'Eustachio P, Stein L, Hermjakob H. Reactome pathway analysis: a high-performance in-memory approach. BMC Bioinformatics. 2017:18(1):142. https://doi.org/10.1186/s12859-017-1559-2.
Ferraretti G, Abondio P, Alberti M, Dezi A, Sherpa PT, Cocco P, Tiriticco M, di Marcello M, Gnecchi-Ruscone GA, Natali L, et al. 2023. Archaic introgression contributed to shape the adaptive modulation of angiogenesis and nitric oxide induction in human high-altitude populations from the Himalayas. bioRxiv 89815.1. https://doi.org/10.7554/eLife.89815.1., 19 December 2023, preprint: not peer reviewed.

Ferrer-Admetlla A, Liang M, Korneliussen T, Nielsen R. On detecting incomplete soft or hard selective sweeps using haplotype structure. Mol Biol Evol. 2014:31(5):1275–1291. https://doi.org/10.1093/molbev/msu077.

Firth JD, Ebert BL, Ratcliffe PJ. Hypoxic regulation of lactate dehydrogenase A. J Biol Chem. 1995:270(36):21021–21027. https://doi.org/10.1074/jbc.270.36.21021.

Fishilevich S, Nudel R, Rappaport N, Hadar R, Plaschkes I, Iny Stein T, Rosen N, Kohn A, Twik M, Safran M, et al. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. Database (Oxford). 2017:2017:bax028. https://doi.org/10.1093/database/bax028.

Fumagalli M, Sironi M, Pozzoli U, Ferrer-Admetlla A, Pattini L, Nielsen R. Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. PLoS Genet. 2011:7(11):e1002355. https://doi.org/10.1371/journal.pgen.1002355.

Gautier M, Klassmann A, Vitalis R. Rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure. Mol Ecol Resour. 2017:17(1):78–90. https://doi.org/10.1111/1755-0998.12634.

Ge R, Simonson TS, Gordeuk V, Prchal JT, McClain DA. Metabolic aspects of high-altitude adaptation in Tibetans. Exp Physiol. 2015:100(11):1247–1255. https://doi.org/10.1113/EP085292.

Ge R-L, Simonson TS, Cooksey RC, Tanna U, Qin G, Huff CD, Witherspoon DJ, Xing J, Zhengzhong B, Prchal JT, et al. Metabolic insight into mechanisms of high-altitude adaptation in Tibetans. Mol Genet Metab. 2012:106(2):244–247. https://doi.org/10.1016/j.ymgme.2012.03.003.

Gersten M, Zhou D, Azad P, Haddad GG, Subramaniam S. Wnt pathway activation increases hypoxia tolerance during development. PLoS One. 2014:9(8):e103292. https://doi.org/10.1371/journal.pone.0103292.

Gittelman RM, Schraiber JG, Vernot B, Mikacenic C, Wurfel MM, Akey JM. Archaic hominin admixture facilitated adaptation to out-of-Africa environments. Curr Biol. 2016:26(24):3375–3382. https://doi.org/10.1016/j.cub.2016.10.041.

Greer SN, Metcalf JL, Wang Y, Ohh M. The updated biology of hypoxia-inducible factor: the updated biology of HIF. EMBO J. 2012:31(11):2448–2460. https://doi.org/10.1038/emboj.2012.125.

Hahne M, Renno T, Schroeter M, Irmler M, French L, Bornand T, MacDonald HR, Tschopp J. Activated B cells express functional Fas ligand. Eur J Immunol. 1996:26(3):721–724. https://doi.org/10.1002/eji.1830260332.

Handley BL, Roberts CH, Butcher R. A systematic review of historical and contemporary evidence of trachoma endemicity in the Pacific Islands. PLoS One. 2018:13(11):e0207393. https://doi.org/10.1371/journal.pone.0207393.

Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, et al. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res. 2004:32(90001):D258–D261. https://doi.org/10.1093/nar/gkh036.

Hooper LV, Stappenbeck TS, Hong CV, Gordon JI. Angiogenins: a new class of microbicidal proteins involved in innate immunity. Nat Immunol. 2003:4(3):269–273. https://doi.org/10.1038/ni888.

Huang Y, Ping X, Cui Y, Yang H, Bao J, Yin Q, Ailifeire H, Shentu X. Glycolysis aids in human lens epithelial cells' adaptation to hypoxia. Antioxidants (Basel). 2023:12(6):1304. https://doi.org/10.3390/antiox12061304.

Huerta-Sánchez E, Degiorgio M, Pagani L, Tarekegn A, Ekong R, Antao T, Cardona A, Montgomery HE, Cavalleri GL, Robbins PA, et al. Genetic signatures reveal high-altitude adaptation in a set of Ethiopian populations. Mol Biol Evol. 2013:30(8):1877–1888. https://doi.org/10.1093/molbev/mst089.

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

Irum B, Khan SY, Ali M, Kaul H, Kabir F, Rauf B, Fatima F, Nadeem R, Khan AO, Al Obaisi S, et al. Mutation in LIM2 is responsible for autosomal recessive congenital cataracts. PLoS One. 2016:11(11):e0162620. https://doi.org/10.1371/journal.pone.0162620.

Jacobs GS, Hudjashov G, Saag L, Kusuma P, Darusallam CC, Lawson DJ, Mondal M, Pagani L, Ricaut F-X, Stoneking M, et al. Multiple deeply divergent Denisovan ancestries in Papuans. Cell. 2019:177(4):1010–1021.e32. https://doi.org/10.1016/j.cell.2019.02.035.

Jacovas VC, Couto-Silva CM, Nunes K, Lemes RB, de Oliveira MZ, Salzano FM, Bortolini MC, Hünemeier T. Selection scan reveals three new loci related to high altitude adaptation in native Andeans. Sci Rep. 2018:8(1):12733. https://doi.org/10.1038/s41598-018-31100-6.

Jain S, Arrais J, Venkatachari NJ, Ayyavoo V, Bar-Joseph Z. Reconstructing the temporal progression of HIV-1 immune response pathways. Bioinformatics. 2016:32(12):i253–i261. https://doi.org/10.1093/bioinformatics/btw254.

Jia C, Kong X, Koltes JE, Gou X, Yang S, Yan D, Lu S, Wei Z. Gene co-expression network analysis unraveling transcriptional regulation of high-altitude adaptation of Tibetan pig. PLoS One. 2016:11(12):e0168161. https://doi.org/10.1371/journal.pone.0168161.

Jin Z-B, Huang X-F, Lv J-N, Xiang L, Li D-Q, Chen J, Huang C, Wu J, Lu F, Qu J. SLC7A14 linked to autosomal recessive retinitis pigmentosa. Nat Commun. 2014:5(1):3517. https://doi.org/10.1038/ncomms4517.

Julian CG, Subudhi AW, Hill RC, Wilson MJ, Dimmen AC, Hansen KC, Roach RC. Exploratory proteomic analysis of hypobaric hypoxia and acute mountain sickness in humans. J Appl Physiol (1985). 2014:116(7):937–944. https://doi.org/10.1152/japplphysiol.00362.2013.

Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 2017:45(D1):D353–D361. https://doi.org/10.1093/nar/gkw1092.

Kaufmann MR, Barth S, Konietzko U, Wu B, Egger S, Kunze R, Marti HH, Hick M, Müller U, Camenisch G, et al. Dysregulation of hypoxia-inducible factor by presenilin/γ-secretase loss-of-function mutations. J Neurosci. 2013:33(5):1915–1926. https://doi.org/10.1523/JNEUROSCI.3402-12.2013.

Kawano H, Nakatani T, Mori T, Ueno S, Fukaya M, Abe A, Kobayashi M, Toda F, Watanabe M, Matsuoka I. Identification and characterization of novel developmentally regulated neural-specific proteins, BRINP family. Brain Res Mol Brain Res. 2004:125(1-2):60–75. https://doi.org/10.1016/j.molbrainres.2004.04.001.

Kayser M. The human genetic history of oceania: near and remote views of dispersal. Curr Biol. 2010:20(4):R194–R201. https://doi.org/10.1016/j.cub.2009.12.004.

Kiessling E, Peters F, Ebner LJA, Merolla L, Samardzija M, Baumgartner MR, Grimm C, Froese DS. HIF1 and DROSHA are involved in MMACHC repression in hypoxia. Biochim Biophys Acta Gen Subj. 2022:1866(9):130175. https://doi.org/10.1016/j.bbagen.2022.130175.

Kim Y-W, Byzova TV. Oxidative stress in angiogenesis and vascular disease. Blood. 2014:123(5):625–631. https://doi.org/10.1182/blood-2013-09-512749.

Kitagawa M. Notch signalling in the nucleus: roles of mastermind-like (MAML) transcriptional coactivators. J Biochem. 2015:159(3):mvv123. https://doi.org/10.1093/jb/mvv123.

Ko R, Macleod C, Pahau D, Sokana O, Keys D, Burnett A, Willis R, Wabulembo G, Garap J, Solomon AW. Population-Based trachoma mapping in six evaluation units of Papua New Guinea. Ophthalmic Epidemiol. 2016:23(sup1):22–31. https://doi.org/10.1080/09286586.2016.1235715.

Kone A, Diarra B, Cohen K, Diabate S, Kone B, Diakite MT, Diarra H, Sanogo M, Togo ACG, Sarro YDS, et al. Differential HLA allele frequency in *Mycobacterium africanum* vs *Mycobacterium tuberculosis* in Mali. HLA. 2019:93(1):24–31. https://doi.org/10.1111/tan.13448.

Kong SE, Banks CAS, Shilatifard A, Conaway JW, Conaway RC. ELL-associated factors 1 and 2 are positive regulators of RNA polymerase II elongation factor ELL. Proc Natl Acad Sci U S A. 2005:102(29):10094–10098. https://doi.org/10.1073/pnas.0503017102.

Lake RJ, Tsai P-F, Choi I, Won K-J, Fan H-Y. RBPJ, the major transcriptional effector of notch signaling, remains associated with chromatin throughout mitosis, suggesting a role in mitotic bookmarking. PLoS Genet. 2014:10(3):e1004204. https://doi.org/10.1371/journal.pgen.1004204.

Landor SK-J, Lendahl U. The interplay between the cellular hypoxic response and notch signaling. Exp Cell Res. 2017:356(2):146–151. https://doi.org/10.1016/j.yexcr.2017.04.030.

Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ. Software for computing and annotating genomic ranges. PLoS Comput Biol. 2013:9(8):e1003118. https://doi.org/10.1371/journal.pcbi.1003118.

Lee L, D'Esposito F, Garap J, Wabulembo G, Koim SP, Keys D, Cama AT, Limburg H, Burnett A. Rapid assessment of avoidable blindness in Papua New Guinea: a nationwide survey. Br J Ophthalmol. 2019:103(3):338–342. https://doi.org/10.1136/bjophthalmol-2017-311802.

Ley SD, Riley I, Beck H-P. Tuberculosis in Papua New Guinea: from yesterday until today. Microbes Infect. 2014:16(8):607–614. https://doi.org/10.1016/j.micinf.2014.06.012.

Li C, Li X, Xiao J, Liu J, Fan X, Fan F, Lei H. Genetic changes in the *EPAS1* gene between Tibetan and Han ethnic groups and adaptation to the plateau hypoxic environment. PeerJ. 2019:7:e7943. https://doi.org/10.7717/peerj.7943.

Liu R, Chen C, Li Y, Huang Q, Xue Y. ELL-associated factors EAF1/2 negatively regulate HIV-1 transcription through inhibition of super elongation complex formation. Biochim Biophys Acta Gene Regul Mech. 2020:1863(5):194508. https://doi.org/10.1016/j.bbagrm.2020.194508.

Lopez M, Choin J, Sikora M, Siddle K, Harmant C, Costa HA, Silvert M, Mouguiama-Daouda P, Hombert J-M, Froment A, et al. Genomic evidence for local adaptation of hunter-gatherers to the African rainforest. Curr Biol. 2019:29(17):2926–2935.e4. https://doi.org/10.1016/j.cub.2019.07.013.

Louis CF, Hur KC, Galvan AC, TenBroek EM, Jarvis LJ, Eccleston ED, Howard JB. Identification of an 18,000-dalton protein in mammalian lens fiber cell membranes. J Biol Chem. 1989:264(33):19967–19973. https://doi.org/10.1016/S0021-9258(19)47205-0.

Loxton AG, Van Rensburg IC. Fasl regulatory B-cells during Mycobacterium tuberculosis infection and TB disease. J Mol Biol. 2021:433(13):166984. https://doi.org/10.1016/j.jmb.2021.166984.

Lu H, Wang R, Li W, Xie H, Wang C, Hao Y, Sun Y, Jia Z, et al. Plasma cytokine profiling to predict susceptibility to acute mountain sickness. Eur Cytokine Netw. 2016:27(4):90–96. https://doi.org/10.1684/ecn.2016.0383.

Lu L, Li J, Moussaoui M, Boix E. Immune modulation by human secreted RNases at the extracellular space. Front Immunol. 2018:9:1012. https://doi.org/10.3389/fimmu.2018.01012.

Luo Z, Mu L, Zheng Y, Shen W, Li J, Xu L, Zhong B, Liu Y, Zhou Y. NUMB enhances notch signaling by repressing ubiquitination of NOTCH1 intracellular domain. J Mol Cell Biol. 2020:12(5):345–358. https://doi.org/10.1093/jmcb/mjz088.

Mackay DS, Boskovska OB, Knopf HLS, Lampi KJ, Shiels A. A nonsense mutation in CRYBB1 associated with autosomal dominant cataract linked to human chromosome 22q. Am J Hum Genet. 2002:71(5):1216–1221. https://doi.org/10.1086/344212.

Malaspinas A-S, Westaway MC, Muller C, Sousa VC, Lao O, Alves I, Bergström A, Athanasiadis G, Cheng JY, Crawford JE, et al. A genomic history of aboriginal Australia. Nature. 2016:538(7624):207–214. https://doi.org/10.1038/nature18299.

Mallet RT, Burtscher J, Richalet J-P, Millet GP, Burtscher M. Impact of high altitude on cardiovascular health: current perspectives. Vasc Health Risk Manag. 2021:17:317–335. https://doi.org/10.2147/VHRM.S294121.

McBryde E. 2012. Evaluation of Risks of Tuberculosis in Western Province Papua New Guinea. https://api.semanticscholar.org/CorpusID:79069202.

McGarry T, Biniecka M, Veale DJ, Fearon U. Hypoxia, oxidative stress and inflammation. Free Radic Biol Med. 2018:125:15–24. https://doi.org/10.1016/j.freeradbiomed.2018.03.042.

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

Natividad A, Freeman TC, Jeffries D, Burton MJ, Mabey DCW, Bailey RL, Holland MJ. Human conjunctival transcriptome analysis reveals the prominence of innate defense in *Chlamydia trachomatis* infection. Infect Immun. 2010:78(11):4895–4911. https://doi.org/10.1128/IAI.00844-10.

Newman M, Wilson L, Verdile G, Lim A, Khan I, Moussavi Nik SH, Pursglove S, Chapman G, Martins RN, Lardelli M. Differential, dominant activation and inhibition of notch signalling and APP cleavage by truncations of PSEN1 in human disease. Hum Mol Genet. 2014:23(3):602–617. https://doi.org/10.1093/hmg/ddt448.

O'Brien KA, Murray AJ, Simonson TS. Notch signaling and cross-talk in hypoxia: a candidate pathway for high-altitude adaptation. Life (Basel). 2022:12(3):437. https://doi.org/10.3390/life12030437.

O'Brien KA, Simonson TS, Murray AJ. Metabolic adaptation to high altitude. Curr Opin Endocr Metab Res. 2020:11:33–41. https://doi.org/10.1016/j.coemr.2019.12.002.

O'Connell JF, Allen J, Williams MAJ, Williams AN, Turney CSM, Spooner NA, Kamminga J, Brown G, Cooper A. When did *Homo sapiens* first reach Southeast Asia and Sahul? Proc Natl Acad Sci U S A. 2018:115(34):8482–8490. https://doi.org/10.1073/pnas.1808385115.

Omori Y, Kubo S, Kon T, Furuhashi M, Narita H, Kominami T, Ueno A, Tsutsumi R, Chaya T, Yamamoto H, et al. Samd7 is a cell type-specific PRC1 component essential for establishing retinal rod photoreceptor identity. Proc Natl Acad Sci U S A. 2017:114(39):E8264–E8273. https://doi.org/10.1073/pnas.1707021114.

Onishi H, Yamasaki A, Kawamoto M, Imaizumi A, Katano M. Hypoxia but not normoxia promotes smoothened transcription through upregulation of RBPJ and mastermind-like 3 in pancreatic cancer. Cancer Lett. 2016:371(2):143–150. https://doi.org/10.1016/j.canlet.2015.11.012.

Paradis E. Pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics. 2010:26(3):419–420. https://doi.org/10.1093/bioinformatics/btp696.

Pena E, El Alam S, Siques P, Brito J. Oxidative stress and diseases associated with high-altitude exposure. Antioxidants (Basel). 2022:11(2):267. https://doi.org/10.3390/antiox11020267.

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

Pulido D, Arranz-Trullén J, Prats-Ejarque G, Velázquez D, Torrent M, Moussaoui M, Boix E. Insights into the antimicrobial mechanism of action of human RNase6: structural determinants for bacterial cell agglutination and membrane permeation. Int J Mol Sci. 2016:17(4):552. https://doi.org/10.3390/ijms17040552.

Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010:26(6):841–842. https://doi.org/10.1093/bioinformatics/btq033.

Quinto-Cortés CD, Jonas CB, Vieyra-Sánchez S, Oppenheimer S, González-Buenfil R, Auckland K, Robson K, Parks T, Moreno-Mayar JV, Blanco-Portillo J, et al. 2024. The genomic landscape of oceania. arXiv 09216. https://doi.org/10.48550/arXiv.2405.09216., 24 June 2024, preprint: not peer reviewed.

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

Ramasamy SK, Kusumbe AP, Wang L, Adams RH. Endothelial notch activity promotes angiogenesis and osteogenesis in bone. Nature. 2014:507(7492):376–380. https://doi.org/10.1038/nature13146.

Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, Viola B, Briggs AW, Stenzel U, Johnson PLF, et al. Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature. 2010:468(7327):1053–1060. https://doi.org/10.1038/nature09710.

Riley ID. Population change and distribution in Papua New Guinea: an epidemiological approach. J Hum Evol. 1983:12(1):125–132. https://doi.org/10.1016/S0047-2484(83)80017-7.

Rutigliano G, Zucchi R. Molecular variants in human trace amine-associated receptors and their implications in mental and metabolic disorders. Cell Mol Neurobiol. 2020:40(2):239–255. https://doi.org/10.1007/s10571-019-00743-y.

Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R, et al. Genome-wide detection and characterization of positive selection in human populations. Nature. 2007:449(7164):913–918. https://doi.org/10.1038/nature06250.

Saurabh A, Thakral D, Mourya MK, Singh A, Mohan A, Bhatnagar AK, Mitra DK, Kanga U. Differential expression of HLA-G and ILT-2 receptor in human tuberculosis: localized versus disseminated disease. Hum Immunol. 2016:77(9):746–753. https://doi.org/10.1016/j.humimm.2016.01.004.

Schaeffer L, Roy R, Humbert S, Moncollin V, Vermeulen W, Hoeijmakers JHJ, Chambon P, Egly J-M. DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor. Science. 1993:260(5104):58–63. https://doi.org/10.1126/science.8465201.

Seidel S, Wirta V, von Stechow L, Schänzer A, Meletis K, Wolter M, Sommerlad D, Henze A-T, Nistér M, Reifenberger G, et al. A hypoxic niche regulates glioblastoma stem cells through hypoxia inducible factor 2α. Brain. 2010:133(4):983–995. https://doi.org/10.1093/brain/awq042.

Semenza GL, Jiang B-H, Leung SW, Passantino R, Concordet J-P, Maire P, Giallongo A. Hypoxia response elements in the aldolase A, enolase 1, and lactate dehydrogenase A gene promoters contain essential binding sites for hypoxia-inducible factor 1. J Biol Chem. 1996:271(51):32529–32537. https://doi.org/10.1074/jbc.271.51.32529.

Senn N, Maraga S, Sie A, Rogerson SJ, Reeder JC, Siba P, Mueller I. Population hemoglobin mean and Anemia prevalence in Papua New Guinea: new metrics for defining malaria endemicity? PLoS One. 2010:5(2):e9375. https://doi.org/10.1371/journal.pone.0009375.

Sheng J, Xu Z. Three decades of research on angiogenin: a review and perspective. Acta Biochim Biophys Sin (Shanghai). 2016:48(5):399–410. https://doi.org/10.1093/abbs/gmv131.

Sherman BT, Hao M, Qiu J, Jiao X, Baseler MW, Lane HC, Imamichi T, Chang W. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Res. 2022:50(W1):W216–W221. https://doi.org/10.1093/nar/gkac194.

Shridas P, Tannock LR. Role of serum amyloid A in atherosclerosis. Curr Opin Lipidol. 2019:30(4):320–325. https://doi.org/10.1097/MOL.0000000000000616.

Siggs OM, Javadiyan S, Sharma S, Souzeau E, Lower KM, Taranath DA, Black J, Pater J, Willoughby JG, Burdon KP, et al. Partial duplication of the CRYBB1-CRYBA4 locus is associated with autosomal dominant congenital cataract. Eur J Hum Genet. 2017:25(6):711–718. https://doi.org/10.1038/ejhg.2017.33.

Skoglund P, Posth C, Sirak K, Spriggs M, Valentin F, Bedford S, Clark GR, Reepmeyer C, Petchey F, Fernandes D, et al. Genomic insights into the peopling of the Southwest Pacific. Nature. 2016:538(7626):510–513. https://doi.org/10.1038/nature19844.

Souilmi Y, Tobler R, Johar A, Williams M, Grey ST, Schmidt J, Teixeira JC, Rohrlach A, Tuke J, Johnson O, et al. Admixture has obscured signals of historical hard sweeps in humans. Nat Ecol Evol. 2022:6(12):2003–2015. https://doi.org/10.1038/s41559-022-01914-9.

Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, Stein TI, Nudel R, Lieder I, Mazor Y, et al. The GeneCards suite: from gene data mining to disease genome sequence analyses. Curr Protoc Bioinformatics. 2016:54(1):1.30.1–1.30.33. https://doi.org/10.1002/cpbi.5.

Summerhayes GR, Leavesley M, Fairbairn A, Mandui H, Field J, Ford A, Fullagar R. Human adaptation and plant use in Highland New Guinea 49,000 to 44,000 years ago. Science. 2010:330(6000):78–81. https://doi.org/10.1126/science.1193130.

Szpiech ZA, Hernandez RD. Selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. Mol Biol Evol. 2014:31(10):2824–2827. https://doi.org/10.1093/molbev/msu211.

Szpiech ZA, Novak TE, Bailey NP, Stevison LS. Application of a novel haplotype-based scan for local adaptation to study high-altitude adaptation in rhesus macaques. Evol Lett. 2021:5(4):408–421. https://doi.org/10.1002/evl3.232.

Tang K, Thornton KR, Stoneking M. A new approach for using genome scans to detect recent positive selection in the human genome. PLoS Biol. 2007:5(7):e171. https://doi.org/10.1371/journal.pbio.0050171.

Taylor CT, Scholz CC. The effect of HIF on metabolism and immunity. Nat Rev Nephrol. 2022:18(9):573–587. https://doi.org/10.1038/s41581-022-00587-8.

Tenbroek E, Arneson M, Jarvis L, Louis C. The distribution of the fiber cell intrinsic membrane proteins MP20 and connexin46 in the bovine lens. J Cell Sci. 1992:103(1):245–257. https://doi.org/10.1242/jcs.103.1.245.

The 1000 Genomes Project Consortium; Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, et al. A global reference for human genetic variation. Nature. 2015:526(7571):68–74. https://doi.org/10.1038/nature15393.

The UniProt Consortium. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 2017:45(D1):D158–D169. https://doi.org/10.1093/nar/gkw1099.

Trájer AJ. Late quaternary changes in malaria-free areas in Papua New Guinea and the future perspectives. Quat Int. 2022:628:28–43. https://doi.org/10.1016/j.quaint.2022.04.003.
Tschopp J. Mitochondria: sovereign of inflammation? Eur J Immunol. 2011:41(5):1196–1202. https://doi.org/10.1002/eji.201141436.
Urnikyte A, Masiulyte A, Pranckeniene L, Kučinskas V. Disentangling archaic introgression and genomic signatures of selection at human immunity genes. Infect Genet Evol. 2023:116:105528. https://doi.org/10.1016/j.meegid.2023.105528.
Van Rensburg IC, Kleynhans L, Keyser A, Walzl G, Loxton AG. B-cells with a FasL expressing regulatory phenotype are induced following successful anti-tuberculosis treatment. Immun Inflam Dis. 2017:5(1):57–67. https://doi.org/10.1002/iid3.140.
Vernot B, Tucci S, Kelso J, Schraiber JG, Wolf AB, Gittelman RM, Dannemann M, Grote S, McCoy RC, Norton H, et al. Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. Science. 2016:352(6282):235–239. https://doi.org/10.1126/science.aad9416.
Vespasiani DM, Jacobs GS, Cook LE, Brucato N, Leavesley M, Kinipi C, Ricaut F-X, Cox MP, Gallego Romero I. Denisovan introgression has shaped the immune system of present-day Papuans. PLoS Genet. 2022:18(12):e1010470. https://doi.org/10.1371/journal.pgen.1010470.
Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. PLoS Biol. 2006:4:e72. https://doi.org/10.1371/journal.pbio.0040072.
Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010:38(16):e164. https://doi.org/10.1093/nar/gkq603.
Wang L, Wang C-M, Hou L-H, Dou G-R, Wang Y-C, Hu X-B, He F, Feng F, Zhang H-W, Liang Y-M, et al. Disruption of the transcription factor recombination signal-binding protein-Jκ (RBP-J) leads to veno-occlusive disease and interfered liver regeneration in mice. Hepatology. 2009:49(1):268–277. https://doi.org/10.1002/hep.22579.
Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution. 1984:38(6):1358. https://doi.org/10.2307/2408641.
World Health Organization (WHO) CS and S (CSS). 2018. WHO country cooperation strategy brief: Papua New Guinea. [accessed 2024 July 5]. https://www.who.int/publications/i/item/WHO-CCU-18.02-PapuaNewGuinea.
Xin J, Zhang H, He Y, Duren Z, Bai C, Chen L, Luo X, Yan D-S, Zhang C, Zhu X, et al. Chromatin accessibility landscape and regulatory network of high-altitude hypoxia adaptation. Nat Commun. 2020:11(1):4928. https://doi.org/10.1038/s41467-020-18638-8.
Xing G, Qualls C, Huicho L, Rivera-Ch M, Stobdan T, Slessarev M, Prisman E, Ito S, Wu H, Norboo A, et al. Adaptation and Mal-adaptation to ambient hypoxia; Andean, Ethiopian and Himalayan patterns. PLoS One. 2008:3(6):e2342. https://doi.org/10.1371/journal.pone.0002342.
Yamasaki A, Yanai K, Onishi H. Hypoxia and pancreatic ductal adenocarcinoma. Cancer Lett. 2020:484:9–15. https://doi.org/10.1016/j.canlet.2020.04.018.
Yang J, Jia Z, Song X, Shi J, Wang X, Zhao X, He K. Proteomic and clinical biomarkers for acute mountain sickness in a longitudinal cohort. Commun Biol. 2022:5(1):548. https://doi.org/10.1038/s42003-022-03514-6.
Yermakovich D, André M, Brucato N, Kariwiga J, Leavesley M, Pankratov V, Mondal M, Ricaut F-X, Dannemann M. Denisovan admixture facilitated environmental adaptation in Papua New Guinean populations. Proc Natl Acad Sci U S A. 2024:121(26):e2405889121. https://doi.org/10.1073/pnas.2405889121.
Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZX, Pool JE, Xu X, Jiang H, Vinckenbosch N, Korneliussen TS, et al. Sequencing of fifty human exomes reveals adaptations to high altitude. Science. 2010:329(5987):75–78. https://doi.org/10.1126/science.1190371.Sequencing.
Zhang Y, Zheng X, Zhang Y, Zhang H, Zhang X, Zhang H. Comparative transcriptomic and proteomic analyses provide insights into functional genes for hypoxic adaptation in embryos of Tibetan chickens. Sci Rep. 2020:10(1):11213. https://doi.org/10.1038/s41598-020-68178-w.
Zheng W, He Y, Guo Y, Yue T, Zhang H, Li J, Zhou B, Zeng X, Li L, Wang B, et al. Large-scale genome sequencing redefines the genetic footprints of high-altitude adaptation in Tibetans. Genome Biol. 2023:24(1):73. https://doi.org/10.1186/s13059-023-02912-1.

Associate editor: Allie Graham