

Chronology of natural selection in Oceanian genomes

Nicolas Brucato, Mathilde André, Georgi Hudjashov, Mayukh Mondal, Murray Cox, Matthew Leavesley, François-Xavier Ricaut

▶ **To cite this version:**

Nicolas Brucato, Mathilde André, Georgi Hudjashov, Mayukh Mondal, Murray Cox, et al.. Chronology of natural selection in Oceanian genomes. iScience, 2022, 25 (7), pp.104583. 10.1016/j.isci.2022.104583 . hal-03753678

HAL Id: hal-03753678
https://cnrs.hal.science/hal-03753678

Submitted on 18 Aug 2022

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

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

Chronology of natural selection in Oceanian genomes

Nicolas Brucato[1,7,*], nicolas.brucato@univ-tlse3.fr, Mathilde André[1,2], Georgi Hudjashov[2], Mayukh Mondal[2], Murray P. Cox[3], Matthew Leavesley[4,5,6], François-Xavier Ricaut[1]

[1]Laboratoire Évolution et Diversité Biologique (EDB UMR 5174), Université de Toulouse Midi-Pyrénées, CNRS, IRD, UPS. 118 route de Narbonne, Bat 4R1, 31062 cedex 9 Toulouse, France

[2]Institute of Genomics, University of Tartu, Tartu, 51010 Tartumaa, Estonia

[3]School of Natural Sciences, Massey University, Palmerston North 4442, New Zealand

[4]Strand of Anthropology, Sociology and Archaeology, School of Humanities and Social Sciences, University of Papua New Guinea, PO Box 320, National Capital District 134, Papua New Guinea

[5]College of Arts, Society and Education, James Cook University, P.O. Box 6811, Cairns, QLD 4870, Australia

[6]ARC Centre of Excellence for Australian Biodiversity and Heritage, University of Wollongong, Wollongong, NSW 2522, Australia

*Corresponding author

[7]Lead contact

Published: June 30, 2022

Summary

As human populations left Asia to first settle in Oceania around 50,000 years ago, they entered a territory ecologically separated from the Old World for millions of years. We analyzed genomic data of 239 modern Oceanian individuals to detect and date signals of selection specific to this region. Combining both relative and absolute dating approaches, we identified a strong selection pattern between 52,000 and 54,000 years ago in the genomes of descendants of the first settlers of Sahul. This strikingly corresponds to the dates of initial settlement as inferred from archaeological evidence. Loci under selection during this period, some showing enrichment in Denisovan ancestry, overlap genes involved in the immune response and diet, especially based on plants. Pathogens and natural resources, especially from endemic plants, therefore appear to have acted as strong selective pressures on the genomes of the first settlers of Sahul.

Data and code availability

This paper analyzes existing, publicly available data. These accession numbers for the datasets are listed in the Key resources table.

Any additional information required to reanalyze the data reported in this paper is available from the Lead contact upon request.

Introduction

During the out-of-Africa migration, human populations crossed few ecological barriers as strong as the Wallace line (Wallace, 1869), which divides two ecozones between Asia and Oceania (Ficetola et al., 2017). 50,000 years before present (YBP), humans arrived on the shores of an ancient territory named Sunda resulting from the low sea levels and the junction of many current islands of west Island Southeast Asia. On the horizon, there was a vast area of islands separated by intense maritime currents, acting as a major ecological frontier for flora and fauna, named Wallacea (Ali and Heaney, 2021). Further away, an ancient continent, Sahul, yet unknown to humans, connected current Australia to New Guinea. It was not long before humans crossed Wallacea and reached Sahul, probably on multiple occasions (Bradshaw et al., 2021; Brucato et al., 2021). These initial attempts to settle the continent probably involved groups of a few hundred individuals (Bradshaw et al., 2019), which may not have all been successful (Clarkson et al., 2017). Sahul was a completely new environment to which humans had to adapt culturally and probably biologically. But undoubtedly, humans migrated across the continent rapidly, reaching the south of Australia around 49,000 YBP, eastern New Guinea around 47,000 YBP and the Bismarck Archipelago further to the east around 43,000 YBP (Summerhayes et al., 2017). They settled a wide range of environments, from the savannah dry lands in the south to the tropical forests in the north, deep inlands and the coasts. Currently the descendants of this exceptional chapter of human history have

among the highest genetic and cultural diversity in the world (Attenborough et al., 2005; Bergstrom et al., 2017; Malaspinas et al., 2016). Within their genetic diversity, almost 3–4% come from an extinct hominin, Denisovan (Reich et al., 2011), which might have settled Sahul before modern humans (Choin et al., 2021; Jacobs et al., 2019).

This unique genetic diversity likely helped adaptation of the first settlers to the wide range of environments in Sahul and is carried today in the genomes of their descendants. However, few examples of genetic adaptation are documented. The genomes of Indigenous Australians show signs of natural selection in genes related to thermoregulation, a physiological feature driven by the extreme temperatures in the Australian desert (Malaspinas et al., 2016). In Vanuatu, an archipelago located east of north Australia, genes related to metabolism and pregnancy were found to be under selection (Choin et al., 2021). Other signs of genetic adaptation relate to the Denisovan ancestry in modern genomes from New Guinea and the Bismarck Archipelago and possibly Australia (Sankararaman et al., 2016; Vernot et al., 2016; Zammit et al., 2019). These introgressed archaic loci include genes of the immune system (Vespasiani et al., 2020) involved in the interferon (IFN)-induced cell-autonomous host defense (MacMicking, 2012), such as *TNFAIP3* (Gittelman et al., 2016; Zammit et al., 2019) and the guanylate binding proteins (GBPs) locus (Vernot et al., 2016), and metabolism, such as *FADS1* (Jacobs et al., 2019) and *SUMF1* (Choin et al., 2021; Jacobs et al., 2019). These possible genetic adaptations suggest that the stark differences between the environments of Sunda and Sahul likely represented challenges to which the genomes of the first human settlers had to adapt (Ali and Heaney, 2021; Shipton et al., 2021).

To investigate the chronology of natural selection in Sahul in more detail, we analyzed the genomes of 239 modern Oceanians combining two different approaches: relative and absolute dating of selection events. Our relative dating relies on the demographic history previously described (Brucato et al., 2021), defining Wallacean populations as proxies for the ancestral population to both groups from New Guinea and the Bismarck Archipelago. Our absolute dating adopts a nonparametric approach to infer the time to the most recent common ancestor between individual genomes at every single genetic variation (Albers and McVean, 2020). For human populations from Wallacea, New Guinea and the Bismarck Archipelago, we performed multiple genomic scans for natural selection and dating of the full genetic diversity (i) to define periods of natural selection during the initial settlement of north Sahul, (ii) to characterize the contribution of Denisovan ancestry to Oceanian biological adaptations, and (iii) to identify biological functions under selection in Oceanian individuals.

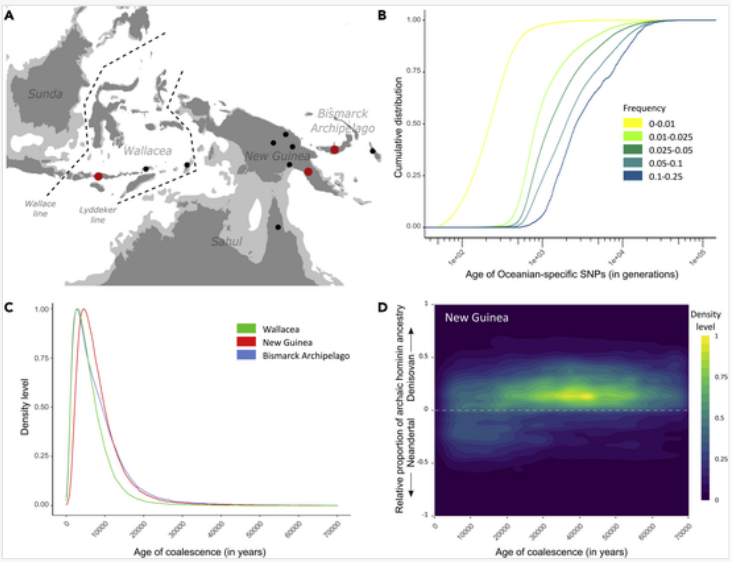
Results

Dating the Oceanian genetic diversity

The ages of coalescence for 13,459,472 SNPs were obtained using GEVA (Albers and McVean, 2020) on a dataset of 569 phased whole genome sequences previously described (Brucato et al., 2021) and representing a large worldwide panel of genetic diversity (Table S1). We used this approach as it requires no prior knowledge about demographic histories or selection events to infer the dates of genetic variants present in large-scale datasets. It is robust to the frequency and sequencing errors, but it relies on the assumptions of a single origin for each allele and a fixed recombination rate. In this absolute dating approach, our ages of coalescence are significantly correlated with the ages of coalescence previously obtained from genomes also included in our dataset (SGDP dataset: $\rho = 0.87$, $p < 10^{-16}$, Figure S1) (Albers and McVean, 2020; Mallick et al., 2016). In Oceanian populations (Figure 1A), we defined the ages of coalescence for 1,790,625 Oceanian-specific SNPs (Data S1). We defined Oceanian-specific SNPs as variants absent in the rest of the dataset, which includes several East Asian (n = 55) and West Island Southeast Asian genomes (n = 68, Table S1), to limit the influence of recent Asian genetic ancestry found in modern Oceanian genomes as seen in PCA and ADMIXTURE analyses (between 22 and 76%, Figures S2 and S3).

Images are optimised for fast web viewing. Click on the image to view the original version.

Figure 1



Distribution of ages of coalescence in Oceanian genomes

(A) Map of analyzed populations. Red dots localize the three studied Oceanian regional groups in Wallacea (Flores Island), New Guinea (Papua New Guinean southeast coast), and the Bismarck Archipelago (New Britain). Black dots represent other groups included in our dataset (Table S1).

(B) Cumulative distribution of ages of coalescence estimated with GEVA for Oceanian-specific SNPs binned by frequency.

(C) Density plot of ages of coalescence for SNPs specific to each Oceanian regional group.

(D) Density plot of ages of coalescence for SNPs present in archaic fragments in New Guinean genomes. The relative proportion of archaic ancestry, estimated with Skov's HMM method, is calculated as (Denisovan ancestry − Neanderthal ancestry)/(Denisovan ancestry + Neanderthal ancestry).

For each continental region, the cumulative distribution of ages of coalescence of variants within a given frequency bin were estimated, as in (Albers and McVean, 2020) (Figure S4). When considering only continent-specific variants, the age distribution reflects the demographic history of the continent as shown in (Albers and McVean, 2020) (Figure S4). In Oceania, the median age of SNPs restricted to the continent at a frequency of 0.25 is 78,795YBP (IQR: 47,383-225,075YBP, Figures 1B, S4, and S5). This median age is older than in other non-African groups, like Europeans (41,079 YBP, IQR: 26,068–72,236 YBP, Figure S4) and East Asians (38,171 YBP, IQR:26,322-70,503 YBP, Figure S4), which coincides with previous reports on the Oceanian genetic divergence from Africa (Brucato et al., 2021; Malaspinas et al., 2016; Pagani et al., 2016). However, as with other continental groups, the vast majority of the Oceanian-specific SNPs appeared between 10,000YBP and 20,000YBP (Figure S6). We focused on the genetic diversity in three Oceanian regions, Wallacea, New Guinea, and the Bismarck Archipelago (Figure 1A), chosen as representative groups for the major demographic movements between Sunda and Sahul that have been previously described (Brucato et al., 2021). These regional groups are related but genetically differentiated ($0.02 < Fst < 0.05$; Figure S7). SNPs exclusively present in one of the three Oceanian regions have a distribution of ages with a peak of diversity in the last 10,000 years (Figure 1C). No major differences in the distribution of ages are found between the three Oceanian regions of interest, although New Guinean diversity apparently increased earlier than in the two other groups (Figures 1C and S5).

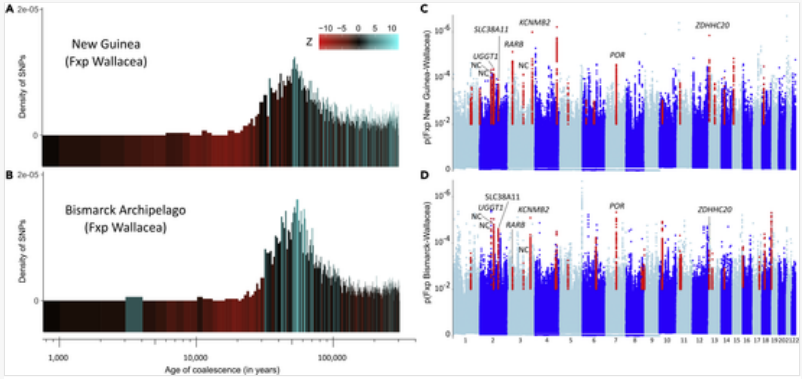
Focusing on the introgressed archaic fragments using an HMM method (Figure S8) (Skov et al., 2018), all three Oceanian groups show a higher proportion of Denisovan ancestry than Neanderthal ancestry (Figures 1D and S9) as expected (Jacobs et al., 2019). Remarkably, most SNPs present in Denisovan fragments have an age of coalescence between 35,000 and 45,000 YBP (Figures 1D and S9). Looking at SNPs only present in each of the three Oceanian groups, a high proportion of Denisovan ancestry is observed in New Guinean genomes around 42,000 YBP (Figure S9). However, in each of the three region groups most SNPs in archaic fragments have an age of coalescence of 10,000 years, suggesting that they are unlikely to be archaic SNPs (Figure S9). Noting that the analysis performed with GEVA is not optimal to determine the age of coalescence of introgressed variants (i.e., poorer knowledge of mutation rates in archaic genomes, population size of archaic hominins, no modeling of admixture processes), the genetic variants showing more recent ages than the assumed date of admixture with Denisovan (between 35,000 and 45,000 YBP) (Jacobs et al., 2019) probably corresponds to variants that emerged in the human genome after the admixture event with Denisovan rather than 'true' Denisovan SNPs.

Chronology of signals of selection

We detected 4,184 SNPs significantly under selection in New Guineans ($p < 0.01$) and 5,129 SNPs under selection in Bismarck Archipelago Islanders ($p < 0.01$, Figure 2), defining the Wallacean genomes as a reference and using a combined score (Fxp), as in (Choin et al., 2021), based on three different tests: XP-EHH, XP-SL (Szpiech and Hernandez, 2014), and PBS (Yi et al., 2010) (MAF≥0.05). The use of a reference for these tests implies a relative succession of events from Wallacea to north Sahul, which constitutes our relative dating approach.

Images are optimised for fast web viewing. Click on the image to view the original version.

Figure 2



Distribution of age of coalescence for SNPs under selection in genomes of New Guinea and the Bismarck Archipelago

(A and B) Density plot of ages of coalescence for all significant Fxp signals of selection in New Guinea (A) and (B) the Bismarck Archipelago, using Wallacea as reference ($p < 0.01$), binned by 1,000 years. The blue-red gradient represents the *Z* score of enrichment of SNPs with a significant Fxp signal of selection in comparison to the rest of the genome (resampled 1,000 times) for each 1,000 year bin.

(C and D) Manhattan plots of the Fxp p-values for each SNP in the analyses: (A) Fxp New Guinea vs. Wallacea and (B) Bismarck Archipelago vs. Wallacea. Red dots represent variants showing significant Fisher scores in 100kb windows enriched for significant scores (>0.7). Gene names represent loci including at least one SNP with an age of coalescence between 52,000 and 54,000YBP. All windows are detailed in Figure S14 and Table S2.

We crossmatched these Fxp results with our resource of estimated ages of coalescence to combine our absolute and relative chronologies. Both New Guinean and Bismarck Archipelago genomes show a significant enrichment of signals of selection for SNPs dating between 52,000 and 54,000 YBP ($Z > 6$, $p < 10^{-9}$, Figure 2A). This convergence of results is remarkable because the genetic diversity of New Guinea and the Bismarck Archipelago are highly divergent (Figure S7) (Brucato et al., 2021; Pedro et al., 2020). It suggests that a strong selective pressure acted on the genomes of the ancestors of both New Guineans and Bismarck Archipelago Islanders around these dates. The chronology of selection signals in the Bismarck Archipelago also shows enrichment around 42,000 and 43,000 YBP ($Z > 6$, Figure 2 B). This result is confirmed when using New Guinean genomes as a reference in the calculation of the Fxp score (Figure S10), indicating that a secondary strong selective force acted on Bismarck Islander genomes at this period. Our analyses also reveal that both New Guinean and the Bismarck Islander genomes have a significantly lower number of signals of selection for SNPs dating between 10,000 and 30,000 YBP ($Z < -6$, $p < 10^{-9}$, Figures 2A and 2B). These patterns are also detected by the Fehh scores, combining iHS (Voight et al., 2006), nSL (Ferrer-Admetlla et al., 2014), and iHH12 (Garud et al., 2015) indexes (with no reference population) for both New Guineans and Bismarck Archipelago Islanders, but not for the Wallaceans (Figure S11). It indicates that the obtained results are specific to the descendants of the first settlers of north Sahul.

We estimated the enrichment for Denisovan ancestry along the genomes of New Guineans and Bismarck Islanders and Wallaceans (Figure S12). The Fxp signals of selection are not enriched for Denisovan ancestry ($Z < 2$, $p > 10^{-2}$). However, a higher proportion of Denisovan ancestry is observed in SNPs under selection at certain time frames: 52,000–53,000 YBP in New Guineans and 39,000–40,000 YBP in Bismarck Islanders (Figure S13). These results suggest that the SNPs under selection appeared in Denisovan fragments at the time of the main phases of the settlement of the New Guinea region.

We further explored the signals of selection focusing on significant Fxp results present in both New Guinea and the Bismarck Archipelago and showing SNPs under selection with ages of coalescence between 52,000 YBP and 54,000 YBP. This highlighted nine loci of interest, overlapping both noncoding and coding sequences (Figures 2C, 2D, and Table S2). The coding sequences showing the most significant signals of selection overlap several genes known to be related to the immune system and/or the metabolic system. A strong Fxp signal of selection and a relative enrichment of Denisovan ancestry are detected in both populations in the gene *RARB* (Figures 2C, 2D, and S14). This gene encodes a crucial transmembrane receptor for retinoic acid, a derivative of vitamin A, which regulates intestinal adaptive immunity in response to diet (Bang et al., 2021; Gattu et al., 2019). Two other genes related to diet present strong signals of selection, but no enrichment for Denisovan ancestry: *KCMB2* encodes a regulator of neuronal activity, whose expression is influenced by a diet rich in folic acid and methionine (Ryan et al., 2018), and *POR* encodes a major regulator of the metabolism of polyphenols (Jensen et al., 2021) (Figure S14). One locus with a strong signal of selection and no enrichment for Denisovan ancestry overlaps a gene involved in the immune response (Figures 2C, 2D and S14): *ZDHHC20* is involved in the regulation of the antiviral response mediated by the type I interferon (IFN)

signaling pathway (McMichael et al., 2017; Rana et al., 2018). This locus is also detected in the Fehh scores in both New Guineans and Bismarck Archipelago Islanders but not in Wallaceans (Figure S15 and Table S3). Our results therefore indicate that the pathogenic environment and dietary resources might have acted as strong selective forces on the genomes of the first settlers of north Sahul.

Denisovan enrichment analyses, using both Skov's HMM and Q95 statistics, in New Guineans and Bismarck Archipelago Islanders showed significant results for genes of the immune system such as *TNFAIP3* (Zammit et al., 2019) and the GBP locus (Kim et al., 2016) and genes involved in metabolism such as *SUMF1* (Cosma et al., 2003) and *PPRAG* (Fritzius and Moelling, 2008) (Figure S12) as previously reported (Choin et al., 2021; Jacobs et al., 2019; Sankararaman et al., 2016; Vernot et al., 2016). However most of these genes are also enriched in Wallacean genomes resulting in low Fxp values (Figure S16), indicating that these adaptive introgressions were probably already selected before the settlement of north Sahul.

Discussion

Our study offers a chronology of events of natural selection on human genomes from Oceania. The Oceanian genetic diversity has a median age of coalescence of 78,795 YBP, older than other non-African continental groups. It is in agreement with previous studies on the genetic divergence of Oceanian genomes from African genomes (Brucato et al., 2021; Malaspinas et al., 2016; Pagani et al., 2016; Wohns et al., 2022). This in turn probably reflects the long demographic history of Oceanian groups, which have been relatively isolated from other continents (Malaspinas et al., 2016). However, as with other continental groups, the vast majority of Oceanian-specific SNPs appeared between 10,000 and 20,000 YBP, which corresponds to the end of the Last Glacial Maximum era and a demographic expansion of most human groups (Bergstrom et al., 2020). The relatively more recent ages of SNPs in Wallacea and the Bismarck Archipelago compared to New Guinea could be attributed to the population dynamics induced by the Austronesian influence 3,000 YBP (Hudjashov et al., 2017) or a possible expansion in New Guinea earlier than in other islands, perhaps related to its own development of farming practices. Overall, the ages of coalescence of SNPs in these three Oceanian groups follow an expected trajectory, as in other non-African groups, showing region-specific genetic diversity slowly emerging around 70,000 YBP, after the out-Of-Africa event, and reaching its peak around 10,000 YBP following the end of the Last Glacial Maximum era.

Within the unique genetic diversity of Oceania, Denisovan ancestry is remarkable by its high proportion (around 3–4%) (Reich et al., 2011). We found that most Oceanian SNPs in introgressed Denisovan tracks have an age of coalescence between 35,000 and 45,000 YBP. These inferred ages of coalescence are in the time frame of admixture events with Denisovans previously reported for the Oceanian region (30,000–46,000 YBP) (Jacobs et al., 2019; Malaspinas et al., 2016) and probably reflects a period shortly after the admixture event during which the Denisovan introgressed fragments were large enough to accumulate new genetic variants driven by modern human population dynamics in Oceania. We also detected a Denisovan signature around 42,000 YBP specific to New Guineans, which might correspond to a potential second admixture event with Denisovans (Choin et al., 2021; Jacobs et al., 2019), although this would need to be explored further.

Looking at enrichment of Denisovan ancestry and signs of selective force, most loci identified in all three Oceanian regional groups have already been reported, such as the immune genes involved in the type-I IFN pathway: *TNFAIP3*, *IFNGR1*, the GBP locus and the *SIGLEC* locus (Jacobs et al., 2019; Mendez et al., 2012; Natri et al., 2019; Sankararaman et al., 2016; Vernot et al., 2016). None of these genes show strong Fxp signals. It indicates that the signal of selection on Denisovan fragments likely predated the initial settlement of north Sahul, as shown by signals of adaptive Denisovan introgression in *TNFAIP3* in South Asian genomes (Jagoda et al., 2018).

However, we show that migration from one ecozone to another, from Wallacea to north Sahul, drove a significant pressure of selection on the human genetic diversity of the first settlers around 52,000–54,000 YBP. This remarkably corresponds to the date of the initial settlement based on archaeological data, around 50,000 YBP (Summerhayes et al., 2017). Our analytical approach combining both absolute and relative chronologies shows that focusing on loci influenced by natural selection could help to decipher the complexity of Oceanian genomes. Similarly, the genomes of Bismarck Archipelago Islanders showed a peak of signals of selection around 42,000–43,000, which coincides with the likely date of the first human settlement in the archipelago based on archaeological studies (Summerhayes et al., 2017). This indicates that these two major migratory events of the first settlers of north Sahul were followed by strong selective forces on variants appearing at this period. In parallel, we found a significantly low number of signals of selection between 10,000 and 30,000 YBP in both New Guineans and Bismarck Islanders, which corresponds to a period of a reduced demographic dynamics during the Last Glacial Maximum (18,000–30,000 YBP) (Summerhayes et al., 2017) as previously reported in (Pedro et al., 2020). Our study strikingly shows that the genomes of the descendants of the first settlers of the New Guinea region are marked by selective forces that acted on genetic variants emerging at key steps during their migration.

At the time of the initial settlement of north Sahul, the selective pressure mainly acted on genes related to the immune system and diet. One of the strongest signals in both New Guinean and Bismarck Archipelago groups overlaps the

gene *ZDHHC20*. This gene encodes an important regulator of the type-I IFN pathway to ensure a robust antiviral response (McMichael et al., 2017). Our result indicates a possible fine-tuning of the immune response by human-specific variants in *ZDHHC20*, through the IFN pathway, which includes several genes enriched for Denisovan ancestry like *TNFAIP3* and the GBP locus. On the other hand, *ZDHHC20* is co-opted by some viruses to potentiate their replication cycle as recently shown by its key role in the S-acylation of the spike protein of SARS-CoV-2 (Mesquita et al., 2021). The genetic variants of *ZDHHC20* in New Guinean and Bismarck Islander genomes might have been selected against an ancient viral pathogenic pressure, similarly to Asian populations (Souilmi et al., 2021), which could be of important consequence today given the dramatic impact of the COVID-19 disease on the populations of the region (Dong et al., 2020).

Another strong signal of selection overlaps the gene *RARB*, showing a relative enrichment of Denisovan ancestry. This gene encodes a receptor of retinol involved in intestinal adaptive immunity, through the development and recruitment of intestinal B and T cells (Bang et al., 2021; Mora et al., 2006). Reduced availability of retinol in the intestine was shown to directly increase susceptibility to infectious diseases (Bhaskaram, 2002). Retinol is a derivative of vitamin A which can be found in animal food sources (e.g., fish liver) and in vegetable food, in the form of some carotenes (e.g., pandanus fruit, yams) (Bang et al., 2021). Retinol is metabolized by the enzyme CYP26, which is regulated by the cytochrome P450 reductase encoded by the gene *POR* also showing a strong signal of selection (Pandey and Flück, 2013). Both signals of selection in *RARB* and *POR* strongly suggest a role of diet and the intake of Vitamin A as a major selective force. Another strong signal of selection overlaps the gene *KCNMB2*, whose expression is linked to the intake of folic acid (Ryan et al., 2018), a form of vitamin B that is present at high levels in green vegetables (Crider et al., 2011). It is remarkable that several of the most significant signals of selection in the genomes of New Guineans and Bismarck Islanders overlap genes related to food intake, especially from plants. The diet of modern New Guineans and Bismarck Islanders relies on a large proportion of local plant resources (Attenborough et al., 2005), among the richest of the world (Cámara-Leret et al., 2020), with agricultural practices independently developed around 8,000 years ago (Denham et al., 2003). Archaeological records revealed that yams and pandanus fruits were exploited since the first arrival of modern humans in New Guinea (47,000 YBP, Ivane Valley) (Summerhayes et al., 2010). The first settlers of north Sahul had to find edible resources in a continent unknown to humans and ecologically separated from the rest of the world for millions of years. This context must have been challenging and potentially drove genetic adaptations in genes related to diet to favor the use of local plant resources.

Our study provides a large set of estimated ages for genomic variants specific to Oceania, complementing previous work in other geographical regions (Albers and McVean, 2020), which will be valuable for future studies. The genomes of New Guineans and Bismarck Archipelago Islanders show a strong signature of natural selection in genes related to the immune response and diet, with variants emerging at the time of the initial settlement of north Sahul. Future genomic work with Indigenous Australians and ancient Oceanian genomics will complement this analysis on the adaptive strategies of the first settlers. Moreover, functional genomics analyses and genetic association with adaptive phenotypic traits will be necessary to validate the identified candidate genes (André et al., 2021; Dannemann and Romero, 2022). Our work reporting a chronology of natural selection in Oceanian genomes does however highlight the initial settlement of the north of Sahul as a major evolutionary PAGE in human history.

Limitations of the study

Genetic dating is a notoriously challenging task (Wohns et al., 2022), which we chose to approach combining a relative and an absolute chronology of events. Several absolute dating methodologies were recently developed (Albers and McVean, 2020; Speidel et al., 2019; Wohns et al., 2022), all limited by data quality (e.g., sequencing, variant calling, phasing) and by the difficulty to integrate demographic processes such as admixture events (e.g., archaic hominin introgression). Future ancient DNA data will be important to calibrate absolute genetic dating approaches (Wohns et al., 2022), but it still represents a major technical challenge in tropical areas such as in most of Oceania (Carlhoff et al., 2021; Oliveira et al., 2021). Our relative dating approach is limited by demographic movements between Wallacea and New Guinea, following the initial settlement of Sahul, which we previously described (Brucato et al., 2021; Purnomo et al., 2021). Population migrations out of New Guinea led to extensive gene flows, especially in Wallacea (Oliveira et al., 2021), which could influence our results, because we assigned our Wallacean data as the ancestral group. To mitigate these limitations, we focused on signals of selection present in both New Guinean and Bismarck Islander groups, less likely to be strongly biased by these secondary migratory waves in Wallacea.

Acknowledgments

This work was supported by the French Ministry of Research grant Agence Nationale de la Recherche (ANR PAPUAEVOL 20-CE12-0003-01), the French Ministry of Foreign and European Affairs (French Prehistoric Mission in Papua New Guinea), and the French Embassy in Papua New Guinea. We are grateful to the genotoul bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul, https://doi.org/10.15454/1.5572369328961167E12) for providing help and computing and storage resources. We acknowledge support from the Labex TULIP, France. MA was supported by the European Union through the European Regional Development Fund (Project No. 2014-2020.4.01.16-0030). MM

Q9 was supported by the European Union through the Horizon 2020 research and innovation program under grant no 810645 and through the European Regional Development Fund project no. MOBEC008.

Author contributions

N.B. designed the study. N.B. performed the analyses and wrote the paper based on input from all the other authors. All authors participated in the interpretation of the results and approved the submission.

Declaration of interests

The authors declare no competing interests.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2022.104583.

References

The corrections made in this section will be reviewed and approved by a journal production editor. The newly added/removed references and its citations will be reordered and rearranged by the production team.

Albers, P.K., McVean, G., 2020. Dating genomic variants and shared ancestry in population-scale sequencing data. PLoS Biol. 18, e3000586. doi:10.1371/journal.pbio.3000586.

Alexander, D.H., Novembre, J., Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19, 1655–1664. doi:10.1101/gr.094052.109.

Ali, J.R., Heaney, L.R., 2021. Wallace's line, Wallacea, and associated divides and areas: history of a tortuous tangle of ideas and labels. Biol. Rev. 96, 922–942. doi:10.1111/brv.12683.

Allen, J., O'Connell, J.F., 2020. A different paradigm for the initial colonisation of Sahul. Archaeol. Ocean. 55, 1–14. doi:10.1002/arco.5207.

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

Q10 Attenborough, R., Golson, J., Hide, R., 2005. Papuan Pasts: Cultural, Linguistic and Biological Histories of Papuan-Speaking Peoples. 572. Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Bang, Y.J., Hu, Z., Li, Y., Gattu, S., Ruhn, K.A., Raj, P., Herz, J., Hooper, L.V., 2021. Serum amyloid A delivers retinol to intestinal myeloid cells to promote adaptive immunity. Science 373, eabf9232. doi:10.1126/science.abf9232.

Bergström, A., McCarthy, S.A., Hui, R., Almarri, M.A., Ayub, Q., Danecek, P., Chen, Y., Felkel, S., Hallast, P., Kamm, J., et al., 2020. Insights into human genetic variation and population history from 929 diverse genomes. Science 367, eaay5012. doi:10.1126/science.aay5012.

Bergström, A., Oppenheimer, S.J., Mentzer, A.J., Auckland, K., Robson, K., Attenborough, R., Alpers, M.P., Koki, G., Pomat, W., Siba, P., et al., 2017. A Neolithic expansion, but strong genetic structure, in the independent history of New Guinea. Science 357, 1160–1163. doi:10.1126/science.aan3842.

Bhaskaram, P., 2002. Micronutrient malnutrition, infection, and immunity: an overview. Nutr. Rev. 60, S40–S45. doi:10.1301/00296640260130722.

Bradshaw, C.J.A., Norman, K., Ulm, S., Williams, A.N., Clarkson, C., Chadœuf, J., Lin, S.C., Jacobs, Z., Roberts, R.G., Bird, M.I., et al., 2021. Stochastic models support rapid peopling of Late Pleistocene Sahul. Nat. Commun. 12, 2440. doi:10.1038/s41467-021-21551-3.

Bradshaw, C.J.A., Ulm, S., Williams, A.N., Bird, M.I., Roberts, R.G., Jacobs, Z., Laviano, F., Weyrich, L.S., Friedrich, T., Norman, K., Saltré, F., 2019. Minimum founding populations for the first peopling of Sahul. Nat. Ecol. Evol. 3, 1057–1063. doi:10.1038/s41559-019-0902-6.

Brisbin, A., Bryc, K., Byrnes, J., Zakharia, F., Omberg, L., Degenhardt, J., Reynolds, A., Ostrer, H., Mezey, J.G., Bustamante, C.D., 2012. PCAdmix: Principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations. Hum. Biol. 84, 343–364. doi:10.3378/027.084.0401.

Brucato, N., André, M., Tsang, R., Saag, L., Kariwiga, J., Sesuki, K., Beni, T., Pomat, W., Muke, J., Meyer, V., et al., 2021. Papua New Guinean genomes reveal the complex settlement of north Sahul. Mol. Biol. Evol. 38, 5107–5121. doi:10.1093/molbev/msab238.

Cámara-Leret, R., Frodin, D.G., Adema, F., Anderson, C., Appelhans, M.S., Argent, G., Arias Guerrero, S., Ashton, P., Baker, W.J., Barfod, A.S., et al., 2020. New Guinea has the world's richest island flora. Nature 584, 579–583. doi:10.1038/s41586-020-2549-5.

Carlhoff, S., Duli, A., Nägele, K., Nur, M., Skov, L., Sumantri, I., Oktaviana, A.A., Hakim, B., Burhan, B., Syahdar, F.A., et al., 2021. Genome of a middle holocene hunter-gatherer from Wallacea. Nature 596, 543–547. doi:10.1038/s41586-021-03823-6.

Chen, L., Wolf, A.B., Fu, W., Li, L., Akey, J.M., 2020. Identifying and interpreting apparent neanderthal ancestry in African individuals. Cell 180, 677–687.e16.

Choin, J., Mendoza-Revilla, J., Arauna, L.R., Cuadros-Espinoza, S., Cassar, O., Larena, M., Ko, A.M.-S., Harmant, C., Laurent, R., Verdu, P., et al., 2021. Genomic insights into population history and biological adaptation in Oceania. Nature 592, 583–589. doi:10.1038/s41586-021-03236-5.

Clarkson, C., Jacobs, Z., Marwick, B., Fullagar, R., Wallis, L., Smith, M., Roberts, R.G., Hayes, E., Lowe, K., Carah, X., et al., 2017. Human occupation of northern Australia by 65, 000 years ago. Nature 547, 306–310. doi:10.1038/nature22968.

Cosma, M.P., Pepe, S., Annunziata, I., Newbold, R.F., Grompe, M., Parenti, G., Ballabio, A., 2003. The multiple sulfatase deficiency gene encodes an essential and limiting factor for the activity of sulfatases. Cell 113, 445–456. doi:10.1016/s0092-8674(03)00348-9.

Crider, K.S., Bailey, L.B., Berry, R.J., 2011. Folic acid food fortification-its history, effect, concerns, and future directions. Nutrients 3, 370–384. doi:10.3390/nu3030370.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al., 2011. The variant call format and VCFtools. Bioinformatics 27, 2156–2158. doi:10.1093/bioinformatics/btr330.

Dannemann, M., Gallego Romero, I., 2022. Harnessing pluripotent stem cells as models to decipher human evolution. FEBS J. 289, 2992–3010. doi:10.1111/febs.15885.

Delaneau, O., Marchini, J., Zagury, J.F., 2012. A linear complexity phasing method for thousands of genomes. Nat. Methods 9, 179–181. doi:10.1038/nmeth.1785.

Denham, T.P., Haberle, S.G., Lentfer, C., Fullagar, R., Field, J., Therin, M., Porch, N., Winsborough, B., 2003. Origins of agriculture at Kuk swamp in the highlands of new Guinea. Science 301, 189–193. doi:10.1126/science.1085255.

Dong, E., Du, H., Gardner, L., 2020. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect. Dis. 20, 533–534. doi:10.1016/s1473-3099(20)30120-1.

Durvasula, A., Sankararaman, S., 2020. Recovering signals of ghost archaic introgression in African populations. Sci. Adv. 6, eaax5097. doi:10.1126/sciadv.aax5097.

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

Ficetola, G.F., Mazel, F., Thuiller, W., 2017. Global determinants of zoogeographical boundaries. Nat. Ecol. Evol. 1, 0089. doi:10.1038/s41559-017-0089.

Fritzius, T., Moelling, K., 2008. Akt-and Foxo1-interacting WD-repeat-FYVE protein promotes adipogenesis. EMBO J. 27, 1399–1410. doi:10.1038/emboj.2008.67.

Garud, N.R., Messer, P.W., Buzbas, E.O., Petrov, D.A., 2015. Recent selective sweeps in North American Drosophila melanogaster show signatures of soft sweeps. PLoS Genet. 11, e1005004. doi:10.1371/journal.pgen.1005004.

Gattu, S., Bang, Y.J., Pendse, M., Dende, C., Chara, A.L., Harris, T.A., Wang, Y., Ruhn, K.A., Kuang, Z., Sockanathan, S., Hooper, L.V., 2019. Epithelial retinoic acid receptor beta regulates serum amyloid A expression and vitamin A-dependent intestinal immunity. Proc. Natl. Acad. Sci. USA 116, 10911–10916. doi:10.1073/pnas.1812069116.

Gittelman, R.M., Schraiber, J.G., Vernot, B., Mikacenic, C., Wurfel, M.M., Akey, J.M., 2016. Archaic hominin admixture facilitated adaptation to out-of-Africa environments. Curr. Biol. 26, 3375–3382. doi:10.1016/j.cub.2016.10.041.

Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al., 2012. GENCODE: the reference human genome annotation for the ENCODE Project. Genome Res. 22, 1760–1774. doi:10.1101/gr.135350.111.

Howie, B.N., Donnelly, P., Marchini, J., 2009. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 5, e1000529. doi:10.1371/journal.pgen.1000529.

Hudjashov, G., Karafet, T.M., Lawson, D.J., Downey, S., Savina, O., Sudoyo, H., Lansing, J.S., Hammer, M.F., Cox, M.P., 2017. Complex patterns of admixture across the Indonesian archipelago. Mol. Biol. Evol. 34, 2439–2452. doi:10.1093/molbev/msx196.

International HapMap, C., Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., et al., 2007. A second generation human haplotype map of over 3.1 million SNPs. Nature 449, 851–861. doi:10.1038/nature06258.

Jacobs, G.S., Hudjashov, G., Saag, L., Kusuma, P., Darusallam, C.C., Lawson, D.J., Mondal, M., Pagani, L., Ricaut, F.X., Stoneking, M., et al., 2019. Multiple deeply divergent denisovan ancestries in papuans. Cell 177, 1010–1021.e32. doi:10.1016/j.cell.2019.02.035.

Jagoda, E., Lawson, D.J., Wall, J.D., Lambert, D., Muller, C., Westaway, M., Leavesley, M., Capellini, T.D., Mirazón Lahr, M., Gerbault, P., et al., 2018. Disentangling immediate adaptive introgression from selection on standing introgressed variation in humans. Mol. Biol. Evol. 35, 623–630. doi:10.1093/molbev/msx314.

Jensen, S.B., Thodberg, S., Parween, S., Moses, M.E., Hansen, C.C., Thomsen, J., Sletfjerding, M.B., Knudsen, C., Del Giudice, R., Lund, P.M., et al., 2021. Biased cytochrome P450-mediated metabolism via small-molecule ligands binding P450 oxidoreductase. Nat. Commun. 12, 2260. doi:10.1038/s41467-021-22562-w.

Kim, B.H., Chee, J.D., Bradfield, C.J., Park, E.S., Kumar, P., MacMicking, J.D., 2016. Interferon-induced guanylate-binding proteins in inflammasome activation and host defense. Nat. Immunol. 17, 481–489. doi:10.1038/ni.3440.

Li, H., 2014. Toward better understanding of artifacts in variant calling from high-coverage samples. Bioinformatics 30, 2843–2851. doi:10.1093/bioinformatics/btu356.

Lipson, M., Loh, P.-R., Patterson, N., Moorjani, P., Ko, Y.-C., Stoneking, M., Berger, B., Reich, D., 2014. Reconstructing austronesian population history in Island Southeast Asia Preprint at bioRxiv. doi:10.1101/005603.

MacMicking, J.D., 2012. Interferon-inducible effector mechanisms in cell-autonomous immunity. Nat. Rev. Immunol. 12, 367–382. doi:10.1038/nri3210.

Malaspinas, A.S., Westaway, M.C., Muller, C., Sousa, V.C., Lao, O., Alves, I., Bergström, A., Athanasiadis, G., Cheng, J.Y., Crawford, J.E., et al., 2016. A genomic history of Aboriginal Australia. Nature 538, 207–214. doi:10.1038/nature18299.

Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., Tandon, A., et al., 2016. The Simons genome diversity project: 300 genomes from 142 diverse populations. Nature 538, 201–206. doi:10.1038/nature18964.

McMichael, T.M., Zhang, L., Chemudupati, M., Hach, J.C., Kenney, A.D., Hang, H.C., Yount, J.S., 2017. The palmitoyltransferase ZDHHC20 enhances interferon-induced transmembrane protein 3 (IFITM3) palmitoylation and antiviral activity. J. Biol. Chem. 292, 21517–21526. doi:10.1074/jbc.m117.800482.

Mendez, F.L., Watkins, J.C., Hammer, M.F., 2012. A haplotype at STAT2 Introgressed from neanderthals and serves as a candidate of positive selection in Papua New Guinea. Am. J. Hum. Genet. 91, 265–274. doi:10.1016/j.ajhg.2012.06.015.

Mesquita, F.S., Abrami, L., Sergeeva, O., Turelli, P., Qing, E., Kunz, B., Raclot, C., Paz Montoya, J., Abriata, L.A., Gallagher, T., et al., 2021. S-acylation controls SARS-CoV-2 membrane lipid organization and enhances infectivity. Dev. Cell 56, 2790–2807.e8. doi:10.1016/j.devcel.2021.09.016.

Mora, J.R., Iwata, M., Eksteen, B., Song, S.Y., Junt, T., Senman, B., Otipoby, K.L., Yokota, A., Takeuchi, H., Ricciardi-Castagnoli, P., et al., 2006. Generation of gut-homing IgA-secreting B cells by intestinal dendritic cells. Science 314, 1157–1160. doi:10.1126/science.1132742.

Natri, H., Bobowik, K.S., Kusuma, P., Darusallam, C.C., Jacobs, G.S., Hudjashov, G., Lansing, J.S., Sudoyo, H., Banovich, N.E., Cox, M.P., et al., 2019. Genome-wide DNA methylation and gene expression patterns reflect genetic ancestry and environmental differences across the Indonesian archipelago Preprint at bioRxiv. doi:10.1101/704304.

Oliveira, S., Nägele, K., Carlhoff, S., Pugach, I., Koesbardiati, T., Hübner, A., Meyer, M., Oktaviana, A., Takenaka, M., Katagiri, C., 2021. Ancient genomes from the last three millennia support multiple human dispersals into Wallacea. Nat. Ecol. Evol. 1–11. doi:10.1038/s41559-022-01775-2.

Pagani, L., Lawson, D.J., Jagoda, E., Mörseburg, A., Eriksson, A., Mitt, M., Clemente, F., Hudjashov, G., DeGiorgio, M., Saag, L., et al., 2016. Genomic analyses inform on migration events during the peopling of Eurasia. Nature 538, 238–242. doi:10.1038/nature19792.

Pandey, A.V., Flück, C.E., 2013. NADPH P450 oxidoreductase: structure, function, and pathology of diseases. Pharmacol. Ther. 138, 229–254. doi:10.1016/j.pharmthera.2013.01.010.

Patterson, N., Price, A.L., Reich, D., 2006. Population structure and eigenanalysis. PLoS Genet. 2, e190. doi:10.1371/journal.pgen.0020190.

Pedro, N., Brucato, N., Fernandes, V., André, M., Saag, L., Pomat, W., Besse, C., Boland, A., Deleuze, J.F., Clarkson, C., et al., 2020. Papuan mitochondrial genomes and the settlement of Sahul. J. Hum. Genet. 65, 875–887. doi:10.1038/s10038-020-0781-3.

Poplin, R., Ruano-Rubio, V., DePristo, M.A., Fennell, T.J., Carneiro, M.O., Van der Auwera, G.A., Kling, D.E., Gauthier, L.D., Levy-Moonshine, A., Roazen, D., et al., 2018. Scaling accurate genetic variant discovery to tens of thousands of samples Preprint at bioRxiv. doi:10.1101/201178.

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

Purnomo, G.A., Mitchell, K.J., O'Connor, S., Kealy, S., Taufik, L., Schiller, S., Rohrlach, A., Cooper, A., Llamas, B., Sudoyo, H., et al., 2021. Mitogenomes reveal two major influxes of papuan ancestry across Wallacea following the last glacial Maximum and austronesian contact. Genes 12, 965. doi:10.3390/genes12070965.

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

Rana, M.S., Kumar, P., Lee, C.J., Verardi, R., Rajashankar, K.R., Banerjee, A., 2018. Fatty acyl recognition and transfer by an integral membrane S-acyltransferase. Science 359, eaao6326. doi:10.1126/science.aao6326.

Reich, D., Green, R.E., Kircher, M., Krause, J., Patterson, N., Durand, E.Y., Viola, B., Briggs, A.W., Stenzel, U., Johnson, P.L.F., et al., 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature 468, 1053–1060. doi:10.1038/nature09710.

Q11 Reich, D., Patterson, N., Kircher, M., Delfin, F., Nandineni, M.R., Pugach, I., Ko, A.S., Ko, Y.C., Jinam, T.A., Phipps, M.E., et al., 2011. Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. Am. J. Hum. Genet. 89, 516–528. doi:10.1016/j.ajhg.2011.09.005.

Ryan, D.P., Henzel, K.S., Pearson, B.L., Siwek, M.E., Papazoglou, A., Guo, L., Paesler, K., Yu, M., Yu, M., Muller, R., et al., 2018. A paternal methyl donor-rich diet altered cognitive and neural functions in offspring mice. Mol. Psychiatr. 23, 1345–1355. doi:10.1038/mp.2017.53.

Sankararaman, S., Mallick, S., Patterson, N., Reich, D., 2016. The combined landscape of denisovan and neanderthal ancestry in present-day humans. Curr. Biol. 26, 1241–1247. doi:10.1016/j.cub.2016.03.037.

Shipton, C., O'Connor, S., Kealy, S., 2021. The biogeographic threshold of Wallacea in human evolution. Quat. Int. 574, 1–12. doi:10.1016/j.quaint.2020.07.028.

Skov, L., Hui, R., Shchur, V., Hobolth, A., Scally, A., Schierup, M.H., Durbin, R., 2018. Detecting archaic introgression using an unadmixed outgroup. PLoS Genet. 14, e1007641. doi:10.1371/journal.pgen.1007641.

Souilmi, Y., Lauterbur, M.E., Tobler, R., Huber, C.D., Johar, A.S., Moradi, S.V., Johnston, W.A., Krogan, N.J., Alexandrov, K., Enard, D., 2021. An ancient viral epidemic involving host coronavirus interacting genes more than 20,000 years ago in East Asia. Curr. Biol. 31, 3504–3514.e3509. doi:10.1016/j.cub.2021.07.052.

Speidel, L., Forest, M., Shi, S., Myers, S.R., 2019. A method for genome-wide genealogy estimation for thousands of samples. Nat. Genet. 51, 1321–1329. doi:10.1038/s41588-019-0484-x.

Summerhayes, G.R., Field, J.H., Shaw, B., Gaffney, D., 2017. The archaeology of forest exploitation and change in the tropics during the Pleistocene: the case of Northern Sahul (Pleistocene New Guinea). Quat. Int. 448, 14–30. doi:10.1016/j.quaint.2016.04.023.

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

Szpiech, Z.A., Hernandez, R.D., 2014. selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. Mol. Biol. Evol. 31, 2824–2827. doi:10.1093/molbev/msu211.

Vernot, B., Tucci, S., Kelso, J., Schraiber, J.G., Wolf, A.B., Gittelman, R.M., Dannemann, M., Grote, S., McCoy, R.C., Norton, H., et al., 2016. Excavating neandertal and denisovan DNA from the genomes of melanesian individuals. Science 352, 235–239. doi:10.1126/science.aad9416.

Vespasiani, D.M., Jacobs, G.S., Brucato, N., Cox, M.P., Romero, I.G., 2020. Denisovan introgression has shaped the immune system of present-day Papuans Preprint at bioRxiv. doi:10.1101/2020.07.09.196444.

Voight, B.F., Kudaravalli, S., Wen, X., Pritchard, J.K., 2006. A map of recent positive selection in the human genome. PLoS Biol. 4, e72. doi:10.1371/journal.pbio.0040072.

Q12 Wallace, A.R., 1869. The Malay Archipelago: the land of the oranguatan, and the bird of paradise. A narrative of travel, with studies of man and nature.. ~~The Malay Archipelago-Volume II.(Of II.)~~Macmillan.

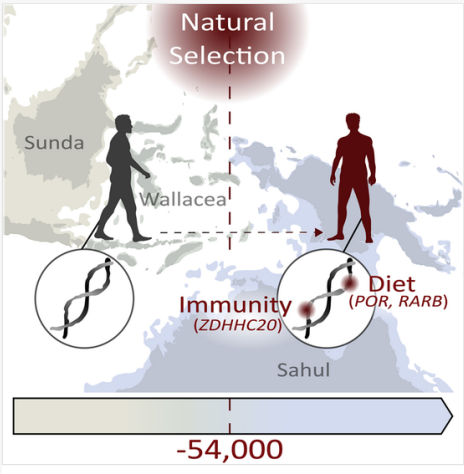
Wohns, A.W., Wong, Y., Jeffery, B., Akbari, A., Mallick, S., Pinhasi, R., Patterson, N., Reich, D., Kelleher, J., McVean, G., 2022. A unified genealogy of modern and ancient genomes. Science 375, eabi8264. doi:10.1126/science.abi8264.

Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z.X., Pool, J.E., Xu, X., Jiang, H., Vinckenbosch, N., Korneliussen, T.S., et al., 2010. Sequencing of 50 human exomes reveals adaptation to high altitude. Science 329, 75–78. doi:10.1126/science.1190371.

Zammit, N.W., Siggs, O.M., Gray, P.E., Horikawa, K., Langley, D.B., Walters, S.N., Daley, S.R., Loetsch, C., Warren, J., Yap, J.Y., et al., 2019. Denisovan, modern human and mouse TNFAIP3 alleles tune A20 phosphorylation and immunity. Nat. Immunol. 20, 1299–1310. doi:10.1038/s41590-019-0492-0.

Graphical abstract

Images are optimised for fast web viewing. Click on the image to view the original version.



Highlights

- 239 human genomes from both sides of the Wallacean ecogeographical barriers
- Signals of selection are dated between -54,000 to -52,000 in modern Oceanian genomes
- Genes related to immunity and diet were under strong selection
- Denisovan introgressions participated to the genetic adaptations present in Oceanians

eToc Blurb

Biological sciences; Genetics; Evolutionary biology; Genomics

STAR★Methods

Key resources table

The table layout displayed in this section is not how it will appear in the final version. The representation below is solely purposed for providing corrections to the table. To preview the actual presentation of the table, please view the Proof.

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
GATK	(Poplin et al., 2018)	https://github.com/broadinstitute/gatk/releases
BCFtools v. 1.4	(Li, 2014)	https://github.com/samtools/bcftools
SHAPEIT v. 2.r79033	(Delaneau et al., 2012)	https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html#download

IMPUTE2 v. 2.3.2	(Howie et al., 2009)	https://mathgen.stats.ox.ac.uk/impute/impute_v2.html#download
EIGENSOFT v.7.2.1	(Patterson et al., 2006)	https://github.com/DReichLab/EIG
GEVA v1.beta	(Albers and McVean, 2020)	https://github.com/pkalbers/geva
vcftools v.0.1.15	(Danecek et al., 2011)	https://github.com/vcftools/vcftools
Selscan v.1.3.0	(Szpiech and Hernandez, 2014)	https://github.com/szpiech/selscan
ADMIXTURE v1.3	(Alexander et al., 2009)	https://dalexander.github.io/admixture/download.html
PCAdmix v.1.0	(Brisbin et al., 2012)	https://sites.google.com/site/pcadmix/downloads/copyright_1-0
Skov's HMM Archaic introgression	(Skov et al., 2018)	https://github.com/LauritsSkov/Introgression-detection
Other		
Simons Genome Diversity Project	(Mallick et al., 2016)	https://ega-archive.org/studies/EGAS00001001959
Papua New Guinean Genome Diversity Project	(Brucato et al., 2021)	https://ega-archive.org/studies/EGAS00001005393
Indonesian Genome Diversity Project	(Jacobs et al., 2019)	https://ega-archive.org/studies/EGAS00001003054
Bismarck Archipelago genomes	(Vernot et al., 2016)	https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001085.v1.p1
Other Papua New Guinean genomes	(Malaspinas et al., 2016)	https://ega-archive.org/studies/EGAS00001001247
HapMap phase 2 genetic map	(International HapMap et al., 2007)	https://ftp.ncbi.nlm.nih.gov/hapmap/
Atlas of Variant Age database	(Albers and McVean, 2020)	https://human.genome.dating
Genecode	(Harrow et al., 2012)	https://www.gencodegenes.org/human/
High quality reads masks for Neandertal and Denisovan genomes	(Prufer et al., 2014)	http://cdna.eva.mpg.de/neandertal/Vindija/FilterBed/

Resource availability

Lead contact

Further information and requests should be directed to and will be fulfilled by the lead contact, Dr. Nicolas Brucato (nicolas.brucato@univ-tlse3.fr).

Materials availability

This study did not generate new unique reagents or genetic sequences.

Method details

Dataset

We compiled a dataset of 239 Oceanian whole genome sequences from publicly available data: the Simons Genome Diversity Project (SGDP) dataset (Mallick et al., 2016), data from New Guinea (Brucato et al., 2021; Jacobs et al., 2019; Malaspinas et al., 2016), data from Indonesia (Jacobs et al., 2019) and data from the Bismarck Archipelago (Brucato et al., 2021; Vernot et al., 2016) (Table S1). In the present study, the term 'Oceania' is used to include populations from Wallacea, New Guinea, Australia and the Bismarck Archipelago. Our dataset also includes genomes from other continental groups including Africa, East Asia, Europe, South Asia and America (Jacobs et al., 2019; Mallick et al., 2016) (n = 330, Table S1), to compare our results to a previous report (Albers and McVean, 2020). Base-calling and per-sample gVCF files were generated using GATK HaplotypeCaller (reads mapping quality ≥20) (Poplin et al., 2018). Multisample VCF files were obtained with CombineGVCFs and GenotypeGVCFs. Genotype calls were subsequently filtered for base depth (≥8x and ≤400x) and genotyping quality (≥30) using BCFtools v. 1.4 (Li, 2014). We removed sites within segmental duplications, repeats and low complexity regions based on publicly available masks (http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/genomicSuperDups.txt.gz; http://software.broadinstitute.org/software/genomestrip/node_ReferenceMetadata.html). Only biallelic sites with high quality variant calls in at least 99% of samples were retained, leading to a whole genome dataset of 569 individuals typed for 41,250,479 SNPs (Table S1). All whole genomes were phased using SHAPEIT v. 2.r79033 (Delaneau et al., 2012) with default settings, except for the number of iterations

(50) and states (200) and used the HapMap phase 2 genetic map (International HapMap et al., 2007) without a reference panel. Six continental groups of genomes were defined based on geography: Africa, Europe, East Asia, South Asia, America, and Oceania.

We defined three Oceanian regional groups to explore the signals of selection that emerged after the crossing Wallacea and the settlement of north Sahul (Allen and O'Connell, 2020): one group from Wallacea (Flores Island, n = 45), one group from Papua New Guinea (PNG Southeast lowlands, n = 54), and one group from the Bismarck Archipelago (New Britain Island, n = 50, Table S1). Because populations from Papua New Guinea and the Bismarck Archipelago diverged early after the initial settlement of north Sahul (Brucato et al., 2021), any common signal of selection when compared to Wallaceans would suggest an ancient selective pressure (or a convergent adaptation). The Papua New Guinean group includes previously generated imputed genomes (n = 27) to increase the number of individuals to 45–50, matching the number of whole genomes for Wallacea and the Bismarck Archipelago (Brucato et al., 2021). Briefly, genotyping data from Papua New Guinea (Bergstrom et al., 2017) were phased and imputed using all phased Papua New Guinean whole genomes as a reference panel (n = 166) with IMPUTE2 v. 2.3.2 (Howie et al., 2009). All SNPs with an imputation score above 0.4 were kept, leaving 12,168,294 SNPs. The three Oceanian populations were constituted based on previous analyses (Brucato et al., 2021) (Table S1). Genetic structure was checked using Principal Components Analyses (PCA) computed with *smartpca* from the EIGENSOFT v.7.2.1 package (Patterson et al., 2006).

Quantification and statistical analysis

Dating the genetic variation

The age of coalescence for each SNP was estimated using GEVA v1.beta, an absolute dating approach (Albers and McVean, 2020). This method does not require assumptions about the demographic or selective processes that shaped the underlying genealogy and is robust to the frequency and types of error found in modern whole-genome population sequencing studies. The age estimation process relies on the detection of haplotype segments, shared between hundreds of haplotype pairs, which are detected relative to a given target position. For each pair of chromosomes (concordant and discordant), a hidden Markov model is used to empirically calibrate an error model to estimate the region over which the MRCA does not change. From the inferred ancestral segment, an MRCA is estimated based on genetic distance and the number of mutations that have occurred on the branches. We used a mutation rate of 10^{-8} and an effective population size of 10,000 (Albers and McVean, 2020; International HapMap et al., 2007). All parameters were set to the defaults. The analysis was performed in batches of 5,000 SNPs. Results were further filtered using the provided *estimate.R* script with the scaling parameter set as 10,000. Results based on the 'joint clock model' were analyzed in our study. We performed Spearman's and Pearson's correlation tests between all of our estimated ages of coalescence and those provided on the Atlas of Variant Age database for the human genomes in the SGDP dataset (https://human.genome.dating).

For each of the six continental (Oceania, Africa, Europe, East Asia, South Asia, America) and three regional groups (Wallacea, Papua New Guinea, Bismarck Archipelago), we estimated the cumulative distribution of ages of coalescence of SNPs according to their frequency in the respective group, as in (Albers and McVean, 2020): below 1%, 2.5%, 5%, 10%, 25%, 50%, and 100%. Similar plots were generated for West Island Southeast Asian genomes (n = 68, Table S1), a subgroup of the East Asian continental group geographically close to Wallacea. To define group-specific SNPs, we used vcftools v.0.1.15 (Danecek et al., 2011) to find variants present in a group but absent in all other groups. A kernel density estimate was computed to visualize the distribution of ages of coalescence for each group.

Signals of selection

We computed three different genomic scans for signals of selection in a target population compared to a reference population. This constitutes our relative dating approach. The indexes XP-EHH and XP-SL were computed with Selscan v.1.3.0 (Szpiech and Hernandez, 2014) using the three regional groups, defining New Guinea or the Bismarck Archipelago as the target population and Wallacea as the reference population. Since human genetic diversity west of the Wallace line is largely dominated by Asian ancestry (Hudjashov et al., 2017; Lipson et al., 2014), the Wallacean population on Flores Island is the best proxy in our dataset of modern genomes for a genetic signature prior to the initial settlement of north Sahul. Previous reports showed that secondary gene flows from New Guinea to Wallacea occurred during the last 20,000 years (Brucato et al., 2021; Carlhoff et al., 2021; Oliveira et al., 2021; Purnomo et al., 2021) but they do not affect the ability to detect signals related to earlier phases of human history (Brucato et al., 2021; Jacobs et al., 2019). Results of these selection scans were normalized with the *norm* function of Selscan v.1.3.0 (Szpiech and Hernandez, 2014). The Population Branch Statistic (PBS) was computed (Yi et al., 2010) based on F_{ST} distances, estimated with vcftools v0.1.15 (Danecek et al., 2011), for each SNP with a minor allele frequency (MAF) above 0.05 between each pair of a trio of populations: populations in New Guinea or the Bismarck Archipelago as targets; populations in Wallacea as the parental population; the population from Africa as the outgroup. Since this analysis can be sensitive to recent admixture, recent Eurasian genetic ancestry was masked in admixed individuals from New Guinea, Wallacea and the Bismarck Archipelago using PCAdmix v.1.0 (Brisbin et al., 2012). Two parental metapopulations were defined by randomly choosing 100 Asian individuals and 100 New Guinean individuals. For the latter group, only unadmixed New Guinean whole genome sequences were chosen (Brucato et al., 2021). Posterior probabilities for non-Eurasian genetic ancestry above 0.9 were used to define SNPs with New Guinean ancestry. The result of the Asian ancestry masking was checked with ADMIXTURE v1.3 (Alexander et al., 2009). XP-EHH, XP-SL and PBS analyses were also performed for the Bismarck Archipelago using New Guinea as reference to determine pattern of genetic adaptations specific to the Bismarck Archipelago.

The indexes of signals of selection were combined into a Fisher score (Fxp), following the method previously described (Choin et al., 2021). For each SNP, the percentile rank of each of the indexes of the signal of selection was calculated. The Fisher score was computed as the sum of the -$\log_{10}$ of the percentile rank for each SNP. The outlier Fisher scores were defined as the top 1% of empirical p values. A sliding 100 kilobases (kb) window was defined to count the number of outlier Fisher scores every 50kb. Windows with less than 50 SNPs were discarded. Windows significantly enriched in outlier Fisher scores were defined as the top 1%. Only SNPs showing a significant Fisher score in a significantly enriched 100kb window were judged of interest in the context of this analysis. Each SNP was mapped to an Fxp score of selection, to an age of coalescence estimated with GEVA (Albers and McVean, 2020) and to gene coordinates extracted from Genecode (Harrow et al., 2012). To estimate the significance of enrichment of signals of selection in each time frame, a distribution of SNP ages of coalescence was computed with a 1000x resampling of 5,000 variants, which corresponds to the number of detected signals of selection in New Guinea and the Bismarck Archipelago. Only variants with a frequency above 5% were considered in the resampling to correspond to the criteria used for the analyses of selection performed with Selscan (Szpiech and Hernandez, 2014). A Z score was computed for each time frame of 1,000 years between the actual distribution of signals of selection and the computed distribution.

Similarly to Fxp, we computed another index (Fehh) based on three other selection scans iHS (Voight et al., 2006), nSL (Ferrer-Admetlla et al., 2014) and iHH12 (Garud et al., 2015) obtained for Wallaceans, New Guineans and the Bismarck Islanders (MAF>0.01), using Selscan v. 1.3.0 (Szpiech and Hernandez, 2014) and the *norm* function. Since these analyses do not use a reference population, they are less powerful to detect signals specific to the descendants of the first settlers of north Sahul, but they serve as in internal control for our approach (eg. Fxp results of interest should not appear significant in Fehh Wallacea).

Archaic genetic ancestry

Archaic genetic introgression in individual genomes of each of the three regional groups (Wallacea, New Guinea, the Bismarck Archipelago) was estimated using Skov's HMM method (Skov et al., 2018). This does not use an archaic genome as a reference but instead defines introgressed segments that contain a high density of variants found in the target genome but are absent in an unadmixed modern human reference. African genomes from our dataset (n = 38) were used as the unadmixed modern human reference set, following recommended guidelines (Skov et al., 2018), although we note that recent studies have found archaic genetic introgression in some Sub-Saharan African genomes (Chen et al., 2020; Durvasula and Sankararaman, 2020). Maximum likelihood estimates of the parameters were computed using the Baum-Welch algorithm and introgressed segments were identified using posterior decoding. Archaic segments with a posterior probability higher than 0.8 and a length above 100kb were analyzed further. Each derived allele exclusively found in fragments defined as 'archaic' were selected and matched to the Neandertal (Prufer et al., 2014) or Denisovan (Reich et al., 2010) genomes, both masked for high quality reads only (http://cdna.eva.mpg.de/neanderthal/Vindija/FilterBed/). Segments with fewer than 10 alleles that could be matched to both archaic genomes were excluded. The match proportion was calculated as the proportion of putative archaic-specific alleles in a segment that match the given archaic genome. For each of the three regional groups, a frequency of Denisovan and Neandertal ancestry was calculated for each archaic SNP. The top 1% of SNPs was defined as enriched for archaic ancestry. Each archaic SNP was assigned to an age of coalescence and an Fxp score of the signal of selection.

Additionally, we performed an adaptive introgression analysis using the Q*95* statistics (Racimo et al., 2015), following the guidelines previously described (Choin et al., 2021).

Supplemental information

Multimedia Component 1

Document S1. Figures S1–S16 and Table S1

Multimedia Component 2

Table S2. Loci enriched for significant Fxp scores (>0.7) in genomes from New Guinea or the Bismarck Archipelago compared to genomes from Wallacea, and for genomes from the Bismarck Archipelago compared to New Guinea, related to Figure 2

Multimedia Component 3

Table S3. Loci enriched for significant Fehh scores (>0.7) in genomes from New Guinea, the Bismarck Archipelago and Wallacean genomes, related to Figure 2

Multimedia Component 4

Data S1. Ages of coalescence of all Oceanian-specific SNPs estimated with GEVA, related to Figure 1

Queries and Answers