Science
AAAS

Supplementary Materials for

Ancient DNA reveals five streams of migration into Micronesia and matrilocality in early Pacific seafarers

Yue-Chen Liu *et al*.

Corresponding authors: Yue-Chen Liu, yuechen_liu@hms.harvard.edu; Rosalind Hunter-Anderson, rozinabq@gmail.com; Ron Pinhasi, ron.pinhasi@univie.ac.at; David Reich, reich@genetics.med.harvard.edu

Science **377**, 72 (2022)
DOI: 10.1126/science.abm6536

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

Materials and Methods

Ethics Approval

Permission for using present-day data from Island Southeast Asia (EGAS00001003054) was provided by the Indonesian Genome Diversity Project (IGDP) review committee. Permission for using present-day data from Papuans (EGAD00010001326) was provided by a review committee at the Wellcome Trust Sanger Institute under the Data Access Agreement #11803. Ethical approvals of collection, research, and publish permits of our newly reported ancient and present-day DNA data are described case-by-case in **Supplementary Text S1**.

Ancient DNA Wet Laboratory Work

In dedicated clean rooms at the University of Vienna, specimens of cochlea were separated from petrous bones using a dental sandblaster (*41*), milled into fine powder, and shipped to Harvard Medical School. In dedicated clean rooms at Harvard Medical School, DNA samples were extracted following previously published protocols (*42–44*). Double- or single-stranded libraries were constructed from the DNA extracts (*45*, *46*). A 7-bp index sequence (single-stranded libraries) or 7-bp in-line barcode sequence (double-stranded libraries) was appended to either end to prevent contamination after adding barcodes or indices (*47*) and allow multiplex sequencing of the libraries. Libraries were prepared with partial treatment using the enzyme uracil-DNA glycosylase (UDG) to decrease errors associated with ancient DNA induced by C-to-U deamination (*48*, *49*). Because the UDG enzyme USER (New England Biolabs) is inefficient at cutting terminal uracils, this protocol makes it possible to detect characteristic ancient DNA damage at the terminal positions of the sequences to confirm the presence of authentic ancient DNA (see below). Libraries were enriched for the mitochondrial genome and ~1.2 million targeted single nucleotide polymorphisms (SNPs) in the nuclear genome ("1240k") (*50–53*), through two rounds of in-solution target capture; or 'Twist Ancient DNA', a custom probe panel synthesized by Twist Biosciences (*54*). The Twist Biosciences custom panel targets the very same ~1.2 million SNPs as well as additional SNPs and tiling regions (Twist probes targeting the mitochondrial genome were spiked in) and was performed for only one round of enrichment using reagents and buffers provided by Twist Biosciences. Successful libraries were sequenced on an Illumina HiSeqX10 instrument with paired-end reads or an Illumina NextSeq 500 machine with 76-base paired-end reads. For this study, we restricted all our analysis to ~1.2 million in common between

1240k and Twist Ancient DNA, as well as the mitochondrial genome. More details for ancient samples and libraries are summarized in **tables S2** and **S38**.

Genotyping of Modern Samples

Ten out of fifty DNA samples collected at the Pacific Basin Medical Officer Training Program were sent to the Children's Hospital of Philadelphia Core Facility at the laboratory at the University of Pennsylvania for genome-wide genotyping using the Affymetrix Human Origins SNP genotyping array (HO). DNA from 27 present-day CHamoru from Guam was extracted from saliva samples and genotyped using the GenoChip 2.0 microarray (*55*) at Family Tree DNA. DNA samples from 75 individuals from Guam, Palau, Pohnpei and Chuuk were obtained randomly from a larger set anonymously donated by 500 participants spanning nine populations across Micronesia. Informed consent for studies of population history and human health was obtained at the time of study enrollment (*56*). Blood was collected, DNA was extracted using phenol-chloroform isolation, anonymized, and stored in batches analyzable only by geographic location of participant origin. DNA was standardized and genotyped on the Multi-Ethnic Global Array (MEGA) and data were provided by the Oceanic Genome Variation Project Consortium. The present-day data newly reported in this study are summarized in **tables S3** and **S4**.

Bioinformatic Processing of DNA Sequences

Sequencing reads of ancient DNA samples were assigned to libraries according to their barcodes and indices, allowing at most 1-bp mismatch of the barcodes for each pair of reads. After trimming barcodes and adapters, we used SeqPrep (https://github.com/jstjohn/SeqPrep) to merge forward and reverse reads of each pair, requiring a minimum of 15-bp overlapping sequence and a maximum 1-bp mismatch. For all the positions covered by multiple sequences, the sequence with the highest quality at that position was used. All merged reads were then mapped to the mitochondrial reference genome RSRS (*57*) and to the nuclear reference genome (GRCh37 / hg19) using BWA v. 0.6.1 (the bwa 'samse' command) with default parameters (*58*). Before variant calling, we removed sequences aligned to identical outer coordinates, eliminated duplicates produced by PCR amplification during the process of library construction, and only kept the sequences uniquely mapping to the reference. We applied filtering with a threshold of mapping quality no less than 10. For each duplication cluster, we kept the sequence with the highest quality. To minimize potential errors due to deamination damage of ancient DNA, nucleotides within two

base pairs from the end of each sequence were not used in analysis. Pseudo-haploid genotypes used for genomic analyses were generated by assigning the observed base from one randomly chosen sequencing read covering each targeted nucleotide position (SNP).

Ancient DNA Authenticity and Haplogroup Calling

After bioinformatic processing, we assessed the authenticity of ancient DNA data through a combination of several criteria (summarized in **tables S2** and **S38**). We measured damage rates for the terminal positions of the sequences to confirm the presence of authentic ancient DNA, with a threshold of 3% required for partially UDG-treated double-stranded libraries and 10% for UDG-treated single-stranded libraries (*49*). We only used libraries for which DNA content was well above the background level in extraction and library controls. For all ancient DNA data, we inferred genetic sex by testing whether the ratio of reads mapping to X and Y chromosomes was consistent with the sequences deriving entirely from a male or a female (*59*). Potential contamination from external DNA was assessed by (a) computing the percentage of mitochondrial fragments matching to the reference sequence of the consensus mitochondrial haplogroup for each individual (*50*, *60*, *61*), and (b) measuring the rate of heterozygous sites on the X chromosome for males (assumed to have only one X chromosome) compared to the expectation for a female (*62*).

We determined mitochondrial haplogroups using *HaploGrep2* (*63*), and Y-chromosomal haplogroups by comparing SNPs (using all reads) to the International Society of Genetic Genealogy Y-tree (http://www.isogg.org). For individuals genotyped on the Human Origins array, we do not have the information about mitochondrial SNPs, while for individuals genotyped on the MEGA array, we do not have the information about Y-chromosomal SNPs, because these two arrays do not effectively target mitochondrial DNA and Y-chromosomal variation, respectively.

Radiocarbon Dating

Thirty-one new direct AMS radiocarbon dates were generated for this study, and 30 of them were on the same samples used for ancient DNA analysis (**table S5**), including twelve Unai individuals sampled from the Naton Beach Site on Guam (I16817, I16819, I16820, I16821, I16822, I16998, I18445, I18446, I18447, I18461, I18495, and I18496), six Latte individuals sampled from the Naton Beach Site on Guam (I24238, I24259, I24517, I24525, I24529, and I24610), six Latte individuals from the Anaguan site on Saipan (I24496, I24497, I24501, I24508, I24510, and I24606), three late prehistoric individuals sampled from Na Island, Pohnpei (I7111, I7112 and

I7113), three late prehistoric individuals sampled from the Nan Madol site, Pohnpei (I30476, I30477, and I30480), and one Unai individual (I16815) for which we failed to get aDNA from the Naton Beach Site on Guam.

Samples were prepared using ultrafiltration or amino acid hydrolysis in the Human Paleoecology and Isotope Geochemistry Laboratory at the Pennsylvania State University and AMS radiocarbon dating was performed in the Accelerator Mass Spectrometer Radiocarbon Laboratory at the Pennsylvania State University (lab code: PSUAMS) following procedures described in Narasimhan *et al.* (*64*). Stable C and N isotope and %C and %N measurements were performed on prepared samples by EA-IRMS at the Yale YASIC lab or the University of New Mexico Center for Stable Isotopes. All samples had atomic C:N ratios between 3.20 and 3.46, indicating good collagen preservation (*65*). Calibrated ages were calculated using OxCal v.4.4 (*66*), accounting for variations in the proportion of marine protein in individual diets as estimated by measured stable carbon isotope values ($\delta^{13}C$) and revised estimates of the local marine reservoir correction (ΔR, see immediately below). We use the standard definition of 1950 Common Era (CE) as 0 cal BP. The ages of other samples without direct AMS radiocarbon dating were estimated based on archaeological context and genetic clustering.

ΔR Calculation

The radiocarbon age of the global ocean (R) is older than the atmosphere on the order of hundreds of ^{14}C years. Prior to the recent revision of the marine mixing model and calibration curve, Marine20 (*67*), R was estimated to be ~400 ^{14}C years, e.g., Marine04 (*68*), Marine09 (*69*), and Marine13 (*70*) based on ^{14}C and U/Th measurements of foraminifera in ocean cores, submerged corals and other marine proxies. The revised marine model estimates R to be roughly 150 ^{14}C years greater than the prior estimate; thus, the updated estimate is ~550 ^{14}C years. There are also secular differences in the apparent ^{14}C age of marine organisms due to local offsets (ΔR) caused by upwelling and variable atmospheric CO_2 uptake in the mixed upper 75m ocean layer (*71*). ΔR estimates are typically derived by comparing the ^{14}C content of known age samples, e.g., pre-bomb museum shell specimens, to the marine model age in a marine calibration curve, where the difference is ΔR. In archaeological contexts, where long-term changes in ocean and atmospheric circulation could cause variations in the local offset at time *t*, a $\Delta R(t)$ can be calculated using paired, contemporaneous marine and terrestrial samples found in association. The calibrated age of the terrestrial sample establishes *t*, and $\Delta R(t)$ is calculated as the difference between the apparent

^{14}C age of the marine sample and the marine model age at *t*, with the errors on each propagated in quadrature (*71*). The process can be modeled in such programs as OxCal (*72*), and has been automated in an online interface available at http://calib.org/JS/JSdeltar20/ (*73*); the latter was used in the analysis that follows.

Away from the continental margins that are subject to upwelling and other circulation processes, Pacific Basin surface waters are generally better mixed with the atmosphere than the average global ocean (R), leading to negative values for ΔR. Two recently published ΔR data sets on associated marine and terrestrial samples from archaeological sites on Saipan, Tinian and Guam were calculated using the 2013 terrestrial and marine curves: ΔR=-55 ± 27 ^{14}C year (*13*) and ΔR=-49 ± 61 ^{14}C year (*74*). These estimates are recalculated below with the IntCal20 and Marine20 curves (**table S6**). The weighted mean of the 10 estimates is -181 ± 17 ^{14}C year following reference (*75*), with no outliers identified by a χ^2 test (df = 9; χ^2=4.49; T=16.82).

Marine Diet Contribution and Burial Dates Calibration

Calibrations of the direct AMS dates on the burials from Guam, Saipan, and Pohnpei were made assuming variable contributions of marine carbon in individual diets based on the measured δ^{13}C in processed collagen using two linear mixing models outlined by reference (*76*). Each uses linear interpolation between two diet end members representing 100% terrestrial and 100% marine diet contributions. According to the formula below for these models:

$$\% \text{ marine} = [(\delta^{13}C_{col} - \delta^{13}C_{terr}) / (\delta^{13}C_{mar} - \delta^{13}C_{terr})] \times 100, \text{ where:}$$

(1) for Diet 1 $\delta^{13}C_{terr}$ = -20‰, $\delta^{13}C_{mar}$ = -10‰, and

(2) for Diet 2 $\delta^{13}C_{terr}$ = -21‰, $\delta^{13}C_{mar}$ = -12‰.

Overall, Diet 1 estimates less marine contribution than Diet 2, and Diet 2 estimates larger proportions of marine diet at greater (less negative) $\delta^{13}C_{col}$ (e.g., $\delta^{13}C_{col}$ = -20‰, 0% marine vs. 11.1% marine; $\delta^{13}C_{col}$ = -12‰, 80% marine vs. 100% marine). Estimates for each individual using both models are presented in **table S5**.

Calibrations for each burial date were performed in OxCal v.4.4 (*66*) as mixtures of terrestrial (IntCal20) and marine (Marine20) with an uncertainty in the mix of ±10% marine and using the recalculated ΔR of -181 ± 17 ^{14}C year. A calibration assuming no marine contribution to diet (IntCal20 alone) is also included for reference. The ΔR estimate is derived from archaeological materials directly pertinent in time and space to the Guam samples, although this estimate and the diet models are also applied to the younger samples from Pohnpei for lack of better estimates of

local values. The calibrated dates for thirteen Unai Period individuals from Guam are quite similar to each other, as are those for the twelve Latte Period individuals from Guam (Naton) and Saipan (Anaguan). Six late prehistoric individuals from Pohnpei also have similar dates.

Newly Reported DNA Data

We screened 191 samples from five archaeological sites for ancient DNA (**table S38**), including 41 Unai samples and 84 Latte samples from the Naton Beach Site on Guam, 5 Latte samples from the Haputo Site on Guam, 50 samples from the Anaguan Site, Garapan, on Saipan, and 11 samples from Na Island and the nearby Nan Madol site in Pohnpei's protected lagoon. For each of these samples, we constructed one to nine libraries, performed bioinformatic processing, and calculated a series of metrics to evaluate the authenticity and the data quality of endogenous ancient DNA. A total of 27 samples failed to yield genome-wide ancient DNA data passing our authenticity data, while we fully report the genome-wide aDNA data for the remaining 164 samples.

All 28 Unai individuals had a damage rate in the final nucleotide consistent with genuine ancient DNA (>3% for double-stranded libraries and >10% for single-stranded libraries) (**table S2**). All individuals had a ratio of sequences matching the Y chromosome to sequences matching both of the sex chromosomes of either <3% (consistent with all female sequences) or >40% (consistent with all male sequences). The one male (I18495) with enough data to perform an estimate of contamination based on polymorphism rate on the X chromosome had a 95% confidence interval (CI) for contamination of 0.0-5.8%. The inferences of contamination in mtDNA sequences were more variable. Among the 17 individuals with estimates, two individuals had an upper bound of mtDNA matching rate to the consensus of ≥96%, indicating (nearly) no contamination. Six individuals had more marginal levels of contamination, with the upper bounds ranging from 92.3-95.8%. The remaining nine individuals had more concerning estimates, ranging from upper bound of 70-89%. For two individuals (I18446 and I18959) with the highest level of ancient DNA damage and evidence of mitochondrial DNA contamination, we restricted analysis to sequences with characteristic ancient DNA deamination. Nevertheless, we initially included all of the samples in our analyses at first because (1) all the individuals with more than 12,000 autosomal SNPs clustered together in principal component analysis (**Fig. 1B** and **fig. S2A**), (2) other authenticity metrics were encouraging, and (3) it is known that mitochondrial-to-nuclear ratios are often extremely low for petrous bone samples and thus mitochondrial DNA

contamination estimates often greatly overestimate autosomal contamination levels for DNA from petrous bones (*60*).

All 125 Latte individuals from Guam and Saipan had a damage rate higher than 3% for all double-stranded libraries in the final nucleotide, consistent with authentic ancient DNA (**table S2**). All individuals had a ratio of sequences matching the Y chromosome to sequences matching both sex chromosomes of either <3% (consistent with all female sequences) or >40% (consistent with all male sequences). Sixty-two male individuals with enough data to perform estimates of contamination based on polymorphism rate on the X chromosome had upper bounds of the 95% CI for contamination lower than 0.8%. For these individuals, the lower bounds of 95% CI for the mtDNA matching rate to the consensus were all higher than 93%. Thus, all of the above parameters indicated good data quality from the Latte samples. For the nine individuals (I24238, I24239, I24240, I24250, I24256, I24262, I24612, I24616, and I24252) with the upper bound of 95% CI of the mtDNA matching rate to the consensus lower than 98%, we restricted to sequences with characteristic ancient DNA deamination to call SNPs.

All 11 late prehistoric individuals from Pohnpei had a damage rate higher than 3% for all double-stranded libraries in the final nucleotide, consistent with authentic ancient DNA (**table S2**). Four female individuals had a ratio of sequences matching the Y chromosome to sequences matching the combined sex chromosomes equal to 1.0-1.7%, while seven male individuals had ratios equal to 42.1-44.2%. Seven male individuals with enough data to perform estimates of contamination based on polymorphism rate on the X chromosome had upper bounds of 95% CI for contamination no higher than 0.8%. All 11 individuals had a lower bound of mtDNA matching rate to consensus of the 95% CI higher than 95%. Thus, all the above parameters indicated good data quality for the late prehistoric Pohnpei samples.

For the archaeological and physical anthropology context of all the above 164 newly reported prehistoric samples of this study, see the detailed descriptions in **Supplementary Text S2**.

One hundred and twelve present-day samples from Guam, Palau, Pohnpei, Chuuk, and other islands of Micronesia were genotyped on either the Human Origins array, the MEGA array, and the GenoChip 2.0 array (**table S3**). Seventy-five individuals were genotyped on the MEGA array, including 20 individuals from Guam, 19 from Palau, 19 from Chuuk, and 17 from Pohnpei. The MEGA array contains 1,895,184 sites, including 1,835,687 autosomal sites, 405 mitochondrial sites, and 59,092 sites on the sex chromosomes. Twenty-seven individuals who were self-identified CHamoru from Guam were genotyped on the GenoChip 2.0 array. The GenoChip 2.0 array

contains 160,832 sites, including 126,306 autosomal sites, 17,112 mitochondrial sites, and 17,414 sites on the sex chromosomes. Ten individuals for whom we have information on their four grandparents were genotyped on the Human Origins array (**tables S3** and **S4**). The Human Origins array contains 597,573 sites, including 593,124 autosomal sites, and 4,449 sites on the sex chromosomes.

Merging with Published Data

We compared our newly generated data with data from published studies. These datasets included 352 individuals with whole-genome resequencing data for which we used the entire genome in some analyses (*77–81*); an additional 356 individuals with whole-genome resequencing data for which all analyses were restricted to sites that overlap with the 1240K panel (*79*, *82–85*); 95 prehistoric individuals genotyped on the 1240K panel (*6–10*, *28*); 329 present-day individuals genotyped on the MEGA array (*86*); and 605 present-day individuals genotyped on the Affymetrix Human Origins genotyping array (*7*, *8*, *87*, *88*). All the published datasets we used in this study are summarized in **table S7**.

To merge data from different arrays, we used the UCSC tool LiftOver first to bring all data onto the same genome assembly (GRCh37 / hg19). We subsequently constructed a dataset containing the union of sites among different arrays (a total of 2,719,061 autosomal sites), after removing the ambiguous sites (non-biallelic sites or sites whose strand definitions did not match across different arrays). We detected and corrected strand flips using Plink v.1.9. All A/T and C/G SNPs were removed because of concern about whether we would be using the same or different allele strand across datasets. This removal resulted in the loss of ~12% of SNPs, and retention of 2,390,980 autosomal sites for downstream analyses. The number of sites in our working datasets varied for different analyses. See **table S36** for details on the datasets used for different analyses.

Pairs of Samples from the Same Individual and Identification of Close Relatives

We searched for pairs of samples that were genetically clearly from the same individual and close relatives (up to third-degree) among all newly reported prehistoric individuals in our study following the method described in the reference (*89*). Briefly, this method compares the mean mismatch rate of all autosomal SNPs with at least one sequence between individuals (randomly selecting one sequence if coverage is higher than one-fold at a particular SNP for a given individual). The mismatch rate is then used to estimate the relatedness coefficient (r). This method

can provide accurate estimates up to third-degree relatives. Within 28 Unai individuals from the Naton Beach site on Guam, we identified four two-member families, including one pair with a first-degree relationship, and three pairs with second- or third-degree relationships. Within 75 Latte individuals from Naton, we identified seven families, including two to thirty family members in each. Among them, we found three pairs of individuals with mother-daughter relationships, three individuals that had father-son-mother relationship, four pairs of individuals that were siblings, and three pairs of individuals with a first-degree relationship. Additionally, we found a large number of pairs (N=82) of Latte individuals from Naton with second- or third-degree relationships. Within four Latte individuals from the Haputo Beach site on Guam, we identified a family containing three members. One pair of individuals had a first-degree relationship, and another pair of individuals had a second- or third-degree relationship. Among 46 Latte individuals from the Anaguan site in Garapan, Saipan, we identified six families, including two to six members in each. A family of five included a mother and her two sons; a pair of individuals had a mother-son relationship; and there were also two families consisting of pairs of siblings. We identified 17 pairs of Latte individuals from the Anaguan site with second- or third-degree relationships. We did not identify close relatives among separate sites within Guam or across different islands. Within eight individuals from the Nan Madol site in Pohnpei's protected lagoon, we found a pair of individuals that were siblings. Unless otherwise specified, we use only one individual with the highest number of SNPs from each first-degree relative pair. The results of our kinship analyses are summarized in **table S33**. The locations of the Unai and Latte individuals within these families in the cemetery are shown in **fig. S17**.

PCA and ADMIXTURE

Principal Component Analysis (PCA) was conducted using *smartpca* in EIGENSOFT (*90*) v.7.2.1 using default parameters with three options: "lsqproject: YES", "autoshrink: YES", and "inbreed: YES". A total of 1292 prehistoric and present-day individuals were included in the PCA (the 1016 published individuals used in the PCA are specified in **table S7**). Axes were defined by whole-genome sequencing data (2,390,980 sites) of four present-day populations (Papuans from the Eastern Highlands and the middle Sepik region of mainland New Guinea, Nasioi speakers from Bougainville Island in the Solomons archipelago, and Dai people from Southern China) to minimize the effects of large genetic drift across Oceanian groups. All the other prehistoric and present-day individuals genotyped on different arrays were projected onto the first two principal

components (PCs) to create a two-dimensional plot. For our newly reported prehistoric and present-day individuals shown in **Fig. 1B**, we removed close relatives (retaining one individual for each first-degree relative pair), individuals who had low coverage (<12,000 autosomal SNPs), and outliers. We show the results for all individuals labeled by population in **figs. S2** and **S3**.

Unsupervised clustering analysis was conducted using ADMIXTURE (*91*) v.1.23 with default parameters restricting to the autosomal SNPs of the 1240K panel after removing all sites in CpG dinucleotides (this left ~825,000 SNPs). A total of 1265 prehistoric and present-day individuals were included in the analysis (the 989 published individuals used in the ADMIXTURE analysis are specified in **table S7**). For those individuals genotyped on the MEGA array, the Human Origins array, and the GenoChip 2.0 arrays, we analyzed the data focusing on their intersections with the 1240K sites. We restricted to 1240K SNPs because we wanted to investigate the relationships among Micronesians, First Remote Oceanians (FRO), and Papuans to see if Micronesians were admixed between FROs and Papuans. The data for all the FRO individuals represented by Lapita individuals from Vanuatu and Tonga, and the Unai individuals from the Naton Beach site on Guam, was generated using the 1240K panel.

We conducted five independent runs. In each run, we predefined the number of genetic clusters (K) from 2 to 12 and implemented 10-fold cross-validation (CV) for each K (**fig. S5**). Across five independent runs, K=8 (**fig. S6**) consistently yielded the lowest CV errors. However, we do not show the K=8 plot in the main manuscript because among the major ancestry components in Oceania, only three show a clear correspondence to specific archaeological cultures. These three components include East Asian-related ancestry maximized in Unai and Latte individuals from Guam and Saipan (medium gray), and Papuan-related ancestry maximized in New Guinea Highlanders (green) and Baining speakers from New Britain (blue). The fourth ancestry component (light gray) widely seen in Oceania is not geographically maximized. K=7 and K=9 alternately yielded the second or the third lowest CV errors with very close values, suggesting they may fit the data equally well. When K=9 (**fig. S7**), the division of different ancestry components in Oceania is much clearer than K=8 and more similar to an ADMIXTURE plot published in Lipson *et al.* (*8*) which facilitates comparison across studies. With K=9 (**fig. S7**), we observe five different ancestry components in Oceania, including two East Asian-related ancestry maximized in the Unai / Latte individuals from Guam and Saipan (medium gray) and Lapita individuals from Vanuatu and Tonga (light gray), respectively; and three Papuan-related ancestry maximized in the Highlanders from mainland New Guinea (green), Baining speakers from New

Britain (blue), and Nasioi speakers from the Solomon Islands (orange). Since these components also correspond well to those delineated in our *f*-statistic analyses, we chose K=9 for our main figure. K=10 (**fig. S8**) always produced higher CV errors than K=7 to 9, indicating that K=10 may not be as good a fit to the data. In conclusion, we selected and displayed the results of K = 9 clusters in the main paper, because K = 9 consistently exhibits low CV error and clearly separates Africans (one cluster), Europeans (one cluster), Native Americans (one cluster), East Asians (one cluster), Papuans (three clusters), and FROs (two clusters), showing a strong correspondence to meaningful genetic groups.

For the plot of newly reported prehistoric and present-day individuals shown in **Fig. 2**, we removed close relatives (retaining one individual for first-degree relative pair), low coverage ones (< 12,000 autosomal SNPs), and outliers. We show the results for all individuals in **figs. S6** to **S8**.

Detecting Papuan Admixture in Micronesians Using f_4-Statistics

f_4-statistics were computed using *qpDstats* in ADMIXTOOLS (*92*) v.7.0 with the parameters "f4mode: YES" and "printsd: YES". Standard errors were computed using a block jackknife. To examine potential Papuan admixture for prehistoric individuals from Guam, Saipan, and Pohnpei, we computed f_4(X, Kankanaey Igorot; New Guinea Highlander, Dai), using New Guinea Highlanders as a representative population with unmixed Papuan ancestry and Kankanaey Igorot as a representative population related to the ancestors of East Asians without Papuan admixture. We chose to use a relatively distant population, the East Asian group "Dai" (from Southern China), in the fourth position. Since our null hypothesis for this f_4 symmetry test was that the Unai and Latte individuals from Guam and Saipan have unmixed FRO ancestry, using the large sample size "Dai" in the fourth position of the f_4-statistic provided enhanced statistical power.

For present-day individuals, we detect evidence of European and Native American ancestry likely caused by admixture in the last few hundred years. To reduce confounding effects introduced due to this ancestry, we replaced "Dai" with "Australian" in the fourth position of the f_4 statistic and computed f_4(X, Kankanaey Igorot; New Guinea Highlander, Australian). Since New Guinea Highlanders and Australians are a (deep) clade, this statistic is not expected to deviate from zero unless population X has Papuan- or Australian-related admixture (European or Native American admixture are not expected to cause a deviation) Because we had whole-genome shotgun data for Kankanaey Igorot, New Guinea Highlanders, Australians, and Dai people, we could use the maximum number of SNPs for each f_4-statistic. We summarize results in **table S9**.

Curation of Data

The Unai individuals show extremely high levels of characteristic ancient DNA damage. A number of them also have evidence of substantial mitochondrial DNA contamination; while this does not necessarily imply substantial nuclear DNA contamination (because petrous bones typically have very low mitochondrial DNA to nuclear DNA ratios), it does document the presence of contamination in the dataset. We therefore performed a series of analyses to distinguish which samples were most likely to provide robust genome-wide analysis; nevertheless, we report all samples as a resource for the scientific community.

1) We first removed from the analysis dataset individual I18959, which was a close relative of individual I18462 with higher coverage.

2) We removed from the analysis dataset seven individuals (I16817, I16818, I16821, I18446, I18502, I18504, and I18958) with small numbers of autosomal SNPs (<12,000) and that also did not cluster tightly with the primary Unai group in PCA (**fig. S2A**). The small number of SNPs available for these individuals means they are not expected to contribute much statistical power to analyses.

3) Based on ADMIXTURE (**fig. S7**), we removed from the analysis dataset seven individuals (I16819, I16997, I16998, I18440, I18445, I18447, and I18494) that seemed to contain unexpected ancestry components, e.g., Native American-related (I16997, I18440, and I18447), European-related (I16997, I16998, and I18494), and additional East Asian-related components (I16819 and I18445).

4) We computed the statistic f_4(X, Latte Naton Guam; CDX.SG, GBR.SG) for the remaining 13 individuals (**table S37**), based on the hypothesis that the main source of contamination was either European or additional East Asian ancestry. The goal here was to determine if some of them showed significant allele sharing with East Asians (represented by Dai people of Southern China from the 1000 Genomes Project) or Europeans (represented by English people from the 1000 Genomes Project) compared to high quality FRO data from the Latte Period. We excluded I18461 (Z = -2.635) from the analysis dataset after performing this analysis.

5) For the remaining 12 Unai individuals, we computed the statistic f_4(X1, X2; CDX.SG, GBR.SG) to perform pairwise comparisons among these individuals and thus to determine if any outliers could be detected. We implemented the *qpfstats* and *qpfmv* approach, recently released in ADMIXTOOLS to allow higher resolution analysis of low-coverage data, to perform this analysis. We further excluded I18498 from the analysis dataset, because it had suggestive evidence of being an outlier compared to the other Unai individuals (**fig. S21**), and it had the smallest number of autosomal SNPs (24,666 SNPs) compared with other remaining individuals implying that its removal would have minimal effect on the analyses.

6) Lastly, we applied *qpfstats* and *qpfmv* framework to test the homogeneity of the remaining 11 Unai individuals regarding to 12 world-wide outgroups: Australian, French, Pima, Onge, Japanese, Dai, Ami, Kankanaey Igorot, New Guinea Highlanders, Baining, Nasioi, and Hawaiian. The *P*-value under Hotelling's T^2 test estimated by *qpfmv* was 0.41, providing no evidence of ancestry heterogeneity for the remaining individuals.

After this procedure, the 11 individuals we used for our downstream analyses are I16820, I16822, I18442, I18443, I18444, I18450, I18462, I18495, I18496, I18501, and I18503. All of them had more than 30,000 autosomal SNPs covered on the 1240K panel.

For the Lapita individuals, we restricted our downstream analyses to seven individuals (I1368, I1369, I1370, I5266, I5951, Sk10, and TON002) with the best quality data to represent $FRO_{SouthwestPacific}$.

For the Latte individuals from the Naton Beach site on Guam, we removed three individuals (I24250, I24256, and I24616) with low coverage of data (<40,000 autosomal SNPs), one individual with evidence of Papuan admixture (I24579), and eleven close relatives of other higher coverage or similar coverage individuals in the dataset (I24243, I24247, I24252, I24255, I24261, I24263, I24518, I24525, I24576, I24581 and I24596). Finally, we retained 60 out of 75 individuals for primary analyses.

For the Latte individuals from the Haputo Beach site of Guam, we excluded one individual (I25188) from the analyses due to being a first-degree relative of I25191 and having relatively low coverage data (~75,000 SNPs).

For the Latte individuals from Saipan, we excluded from our primary analysis dataset close relatives (I24339, I24569, and I24605) and individuals with evidence of Papuan admixture (I24297, I24341, I24499, and I24503). We retained 39 out of 46 individuals for downstream analyses.

For the prehistoric individuals from the Nan Madol site in Pohnpei, we excluded one individual (I30481) from the analyses due to being the brother of I30480.

For present-day Micronesians, we removed PCA and ADMIXTURE outliers of each group (**figs. S3** and **S7**) to avoid confounding effects on the analyses that follow from more recently introduced European, Native American, and East Asian ancestry.

Allele-sharing Estimates Based on Outgroup f_3-Statistics

We computed outgroup f_3-statistics of the form $f_3(X_1, X_2;$ Mbuti) using the *qp3pop* in ADMIXTOOLS with default parameters. The goal of this analysis was to investigate the amount of genetic drift separating X_1 (Unai and Latte individuals and groups from Guam and Saipan) and X_2 (various groups from mainland and island Southeast Asia, as well as Near and Remote Oceania). Higher values always indicate greater allele sharing and lower genetic distance. We used the autosomal SNPs either on the 1240K panel or the intersection between the 1240K and the MEGA SNP array to compute the statistics. Results are shown in **fig. S4** and **table S8**.

f_4-Regression Analyses

We used an f_4-based regression approach to distinguish differential sources that contributed to the ancestry of Micronesians. All f_4-statistics were computed using *qpDstats* in ADMIXTOOLS with the parameter "f4mode: YES" and "printsd: YES". The regression lines (linear equations / models) were computed via inverse-variance-weighted least-squares method using R v.3.6.2. Residuals of each data point to the regression line and their standard errors estimated by block jackknife resampling were computed using *qp4diff* in ADMIXTOOLS. Residual Z scores were computed as the residuals divided by their standard errors. See details in **Supplementary Text S3**.

We performed the analyses on two separate datasets. One dataset consisted of approximately 398,000 autosomal SNPs from the Human Origins array, after excluding all sites at CpG dinucleotides, as well as a set of SNPs (~40,000 SNPs) with high missing rates across various present-day Oceanian groups genotyped on the array. The other dataset was comprised of approximately 169,000 autosomal SNPs common to both the 1240K and the MEGA array, after

excluding all sites at CpG dinucleotides. For the groups used in this part of the analyses and the results, see **Fig. 3**, **figs. S9**, **S10**, **S11**, and **S18**, as well as **tables S10** to **S21**, **S34**, and **S35**.

Phylogenetic Tree Reconstruction

To investigate the phylogenetic relationships among FRO and Latte groups and to detect if there was population structure among the Latte population in Guam and Saipan, we reconstructed a phylogenetic tree based on the genetic distance matrix from outgroup f_3-statistics. We computed the statistic $f_3(X_1, X_2;$ Mbuti), where X_1 and X_2 were Latte individuals from Guam and Saipan, Lapita individuals from Vanuatu (merged as a single group) and Tonga (merged as a single group), Unai individuals from the Naton Beach site on Guam (merged as a single group), Kankanaey Igorot (merged as a single group), and Australians (merged as a single group). We then used the reciprocal of the statistic as the genetic distance to manually generate a genetic distance matrix. A Neighbor-Joining (NJ) tree was reconstructed using the R package *phangorn* v.2.5.5. The tree shown in **fig. S19** was displayed and modified using Figtree v.1.4.4 by setting Australian as the outgroup.

Clade Testing and Inference of Admixture Proportions Using *qpWave* and *qpAdm*

We used *qpfstats*, *qpWave* (*93*) and *qpAdm* (*52*) in ADMIXTOOLS to address the following questions. (1) We investigated whether the FRO and Latte groups consistently form a clade using *qpWave*. (2) We investigated whether 10 present-day Micronesian Islanders genotyped on the Human Origins array and without precisely known geographic origins could be genetically clustered using *qpWave*. (3) We inferred proportions of Papuan ancestry in FRO and Latte groups using *qpAdm* (in general inferring the proportion to be consistent with zero). (4) We tested and fit two-way or three-way mixture models for prehistoric and present-day Micronesian Islanders using *qpAdm*. (5) We examined the genetic continuity and homogeneity of late prehistoric and present-day Pohnpei and Chuuk using *qpWave*. See details in **Supplementary Text S4**.

When running analyses with *qpWave* and *qpAdm*, we first ran *qpfstats* using "allsnps: YES". We used the option "inbreed: YES" if some groups to the left (target and source populations) or the right (outgroups) contained more than one ancient individual. We used the output of *qpfstats* as the input of *qpWave* and *qpAdm* to perform the modeling. If some groups to the left or to the right had only one individual, we changed the parameter to "inbreed: NO". For all *qpWave* and *qpAdm* models, *P*-values higher than 0.05 were typically interpreted as good fits. All results are summarized in **Fig. 4**, **fig. S20**, and **tables S22** to **S25**.

Admixture Graph Reconstruction and Integration

We used the *qpfstats* and *qpGraph* framework in ADMIXTOOLS to fit admixture graphs. We first ran *qpfstats* using "allsnps: YES" and "inbreed: YES", and then used the output of *qpfstats* as the input of *qpGraph*. The results are shown in **figs. S12** to **S15**, and we discuss them in detail in **Supplementary Text S5**. We finally integrated admixture graphs with previous *qpAdm* and *qpWave* models to generate a schematic shown in **Fig. 4C**.

Simulations of Matrilineal Drift between FRO Lineages

Simulations of matrilineal drift between FRO lineages were performed under the classic Wright-Fisher model of genetic drift (see **Supplementary Text S6** for details). Haplogroup frequency distributions were counted and plotted using in R v.3.6.2. Empirical *P*-values were calculated manually. The results are summarized in **fig. S16** and **table S28**.

F_{ST} Calculation

We computed F_{ST} using *smartpca* in EIGENSOFT, specifying parameters "fstonly: YES" and "fastmode: YES". We set "inbreed: YES", which directs the software to treat each individual as haploid rather than diploid, correctly accounting for the fact that each individual is represented by a single randomly sampled allele at each location. This means that we do not need to call heterozygous positions using low coverage ancient DNA data in order to compute F_{ST}.

ROH and IBD Segment Identification

We identified runs of homozygosity (ROH) within ancient DNA samples using the Python package hapROH v.0.3a4 (https://pypi.org/project/hapROH/), following a previously described method (*94*). As recommended by this method for the 1240K data, we only included prehistoric individuals with at least 400,000 covered SNPs. We ran the method to identify ROH >4 centimorgan (cM) using default parameters. We separated the inferred ROH into four categories of length: >4 cM, >8 cM, >12cM, and >20 cM. We report the total sum of ROH for each category in **table S30**. To estimate effective population size (*Ne*) from ROH, we applied a maximum-likelihood inference method incorporated in hapROH and fit the lengths of all genome-wide ROH in the range of 4–20 cM. The *Ne* estimates are summarized in **table S32**.

We also identified segments of Identity by Descent (IBD) on the X chromosome between each pair of male individuals, using a method incorporated in hapROH. Specifically, we paired pseudo-haploid data of two X chromosomes and identified ROH for each artificial diploid individual. We report the total sum of IBD for pairs of male individuals in **table S31**. We also inferred *Ne* from IBD using the same maximum-likelihood approach incorporated in hapROH. The *Ne* estimates are summarized in **table S32**.

Admixture Date Estimation

We used the method Distribution of Ancestry Tracts of Evolutionary Signals (DATES v.3530) (*64*) to estimate the admixture dates of Latte individuals from Guam and Saipan, as well as present-day admixed individuals from Micronesia. We merged two Kankanaey Igorot individuals from the Philippines and three Ami / Atayal individuals from Taiwan as the reference population for East Asian-related ancestry. We did this because all five of these individuals had whole-genome data. We could therefore use the maximum number of SNPs for each run (either all autosomal SNPs of the 1240K panel or all autosomal SNPs of the MEGA array) to increase the power. These groups were also the closest groups to the FRO lineages. For Latte individuals from Guam and Saipan, we used Unai individuals from the Naton Beach site of Guam as the other reference population. For present-day individuals from Pohnpei, Chuuk, and Palau, we used New Guinea Highlanders as the other reference population. We ran DATES with 'maxdis: 1.0' to ensure that we could detect even relatively long blocks of linkage disequilibrium (LD) for recent admixture events. The results are summarized in **table S27**.

Estimation of Sex-biased Admixture

To estimate the proportion of female and male ancestors for a given population, we used the method first described in Skoglund et al. (*7*). Using the same *qpAdm* two-way model, we estimate the ancestry proportions (FRO and Papuan) for the X chromosome and autosomes and then obtain point estimates for the proportions of female and male ancestors (**table S29**).

Supplementary Text

S1. Ethics Statement

Genetic studies of population history are important not only for providing insights into the origin of human genetic variation and past migrations, but also because of the significance of this information for present-day peoples. In this study, we took a case-by-case approach with the goal of ensuring that the work was informed by the perspectives of local people who might have connections to the individuals we have analyzed.

Permission to sample, transport and conduct minimally destructive analysis of human skeletal remains from the Naton Beach Site (GHPI Site 66-01-2488), housed at the museum repository in Hagåtña, Guam, was granted on October 17, 2014, by Raymond F.Y. Blas, Director of the Guam Department of Parks and Recreation with concurrence by Guam Historic Preservation Officer (HPO) Lynda B. Aguon. Permission to expand the scope of the Naton research to include the Pinhasi and Reich laboratories was granted by HPO Aguon on September 25, 2018. Permission to select and transport samples from Unai Period burials from the Naton Beach site collection was granted on March 8, 2019, by HPO Aguon with the concurrence of Anna Marie Arceo, Director of the Department of CHamoru Affairs and oversight from the Guam Museum. On March 8, 2020, permission to select and transport Naton Beach site Latte Period samples was granted through a Research Agreement overseen by the KOSAS Advisory Board and Leona Young, Administrator, Guam Museum, with concurrence of Guam HPO Patrick Q. Lujan. Samples from the Haputo Beach Site House 16 were stewarded by Richard Olmo, University of Guam, Co-Principal Investigator, Haputo Excavations; in a letter dated March 8, 2019, he requested that the samples be analyzed for ancient DNA. Permission to transport and study Latte Period petrous bones from the Anaguan Site (SP-1-0762), Garapan, Saipan, curated at the Best Sunshine storage facility in Lower Base, Saipan, was granted on March 13, 2020, by CNMI HPO Rita Chong-Dela Cruz with concurrence of Principal Investigator Michael Dega, Scientific Consultant Services, Inc., Honolulu, Hawai‘i. The Unai samples were returned to the Guam Museum repository for re-accession on March 9, 2020. The Latte samples were returned to their original repositories on Guam and Saipan on May 25, 2022 and May 27, 2022, respectively.

Permission to remove and study the ~500 BP skeletons from Na Island, Pohnpei was granted to two Catholic Diocese of Honolulu priests (Fathers Louis H. Yim and Joseph Matheis) in 1977

by the District Administrator for Pohnpei District, Trust Territory of the Pacific Islands (now the Federated States of Micronesia), Resio S. Moses, and the Nanmwarki (chief) of Madolenihmw Municipality, which includes Na Island. The skeletons from Pohnpei were studied by co-author Michael Pietrusewsky and his students at the University of Hawai‘i at Mānoa, Honolulu, where they are currently stored. Permission to proceed with submission of the samples for analysis and reporting the results was granted on September 6, 2020, in a letter from Jason Barnabas, archaeologist at the Pohnpei HPO.

The skeletons from Nan Madol, Pohnpei were studied in 1984/85 by co-authors Michael Pietrusewsky and Michele Toomay Douglas at the University of Hawai‘i-Mānoa, and are now housed at the International Archaeological Research Institute, Inc., Honolulu, Hawai‘i. Permission to proceed with submission of the samples from Nan Madol, Temwen Island, Pohnpei for analysis and reporting the results was granted on September 6, 2020, in a letter from Jason Barnabas, archaeologist at the Pohnpei HPO.

DNA samples from 50 present-day Micronesian individuals were collected by Micronesian volunteers who drew each other’s blood in a class that was part of the Pacific Basin Medical Officer Training Program held on Pohnpei, and a randomly selected subset (N = 10) of these individuals was genotyped on the Affymetrix Human Origins SNP array for this study (**tables S3** and **S4**). As a part of this course, anonymized ancestry forms were filled out by participants, who provided informed consent to have their DNA analyzed for research into population history as reviewed by the Human Investigations Committee of Yale University [HIC #8711001387]. These forms contained information on the birth locations of each volunteer’s four grandparents, revealing that the 200 grandparents had ancestry from a wide variety of locations in Micronesia, with large groupings being: (1) 36% from the High islands of the Caroline Islands including Chuuk, Pohnpei and Kosrae; (2) 25% from Palau; (3) 17% from Polynesian and Polynesian outlier communities; (4) 12% from the Marshall Islands; (5) 5% from Yap; and (6) 6% from assorted other or admixed ancestries (*95*). We cannot link the self-reported information on grandparental origin to individual samples due to the protocol of human projects aiming to preserve anonymity of research subjects. However, the empirical ancestry inferences which are highly variable and span the range of ancestry we detect over Micronesian and Polynesian islands are consistent with this wide distribution of geographic origins.

Saliva samples of present-day CHamoru from Guam were collected for the purpose of DNA extraction and analysis with individual and group informed consent by co-authors Miguel Vilar

and Frank Camacho. The consent form and sampling strategy were approved by the University of Pennsylvania IRB #8 and by Genographic Project protocol #803115, coordinated by co-authors Theodore G. Schurr and Miguel Vilar. These researchers also collected genealogical information extending back three generations from each participant to help interpret the genetic data generated from the samples. DNA samples were genotyped using the GenoChip 2.0 microarray at Family Tree DNA. Data from 27 self-identified CHamoru individuals was used in this study (**table S3**).

DNA samples from 75 present-day individuals from Guam, Palau, Pohnpei and Chuuk were randomly identified from a wider set of anonymized samples donated from 500 participants spanning nine populations across Micronesia. Following informed consent for studies of population history and human health consistent with the standards prevailing at the time of study enrollment, blood was collected anonymously, extracted using phenol-chloroform DNA isolation, and stored in batches analyzable only by the geographic origins of participants. DNA was standardized and genotyped on the Multi-Ethnic Global Array (MEGA). Collection and use of the samples was approved by (1) the Oxford Tropical Research Ethics Community at the University of Oxford with a formal approval letter dated July 2, 2014 (OXTREC Reference: 537-14); (2) Mr Eliuel Pretrick (Chief Division of Health Services, Government of Federated States of Micronesia at the time of collection); (3) Dr M. Kumangai (Director of Health Services, McDonald Memorial Hospital, Koro, Republic of Palau at the time of sample collection); and (4) Dr Arthur Loerzel (Pathologist, Guam Memorial Hospital, Guam at the time of collection).

On October 15, 2020, the preliminary results of this study were presented by a team including five authors to 17 CHamoru community members from Guam. The feedback from the discussion was integrated before submission. An example of how we integrated this feedback is the terminology we used to discuss the indigenous CHamoru and the relationships of the CHamoru to past peoples who lived in the Mariana Islands and elsewhere: we use the spelling “CHamoru” rather than “Chamorro” out of respect for the preference of the Guam CHamoru Language Commission, communicated to us following the October 2020 engagement meeting by participants in the meeting. The 17 community members included Yvette Paulino, Tristan Paulino, Joe Quinata, Leonard Iriarte, Rlene Santos Steffy, the Honorable Senator Sabina Perez, Vincent Diego, Rosa Palomo, Ronald Laguana, Jillette Leon Guerrero, Kenneth Gofigan Kuper, Frank Rabon, Laura Torres Souder, Judy Amesbury, Darlene Moore, the Honorable Senator Therese Terlaje, and Manny Cruz. Before submission, we circulated this manuscript for review and comment to all co-authors, which includes an indigenous Micronesian (Frank Camacho), and to a broader set of

contacts in Guam, Saipan and Pohnpei with whom we have been engaged. We solicited feedback not only about the scientific content but also about the language used to describe inferred relationships between present-day and prehistoric people described in this paper. While it is often not possible to obtain feedback from everyone who may feel connected to an ancient culture, the approach of consulting a diversity of groups is valuable for shaping the narrative of a final paper and ensuring sensitivity to local perspectives.

S2. Archaeological and Physical Anthropology Context

I: Burials from Guam and Saipan

Overview

Material traces of human activities in the Mariana Islands, likely left by mobile marine foragers from Island Southeast Asia (ISEA) or Wallacea, provide unequivocal evidence for human presence in the Mariana Islands and possibly in all of Remote Oceania ~3500-3000 BP (*3*, *11*). Fewer than ten campsites are known from Guam, Rota, Tinian, and Saipan; these sites are deeply buried and located >100 meters inland from present shorelines (*96–98*). Sparsely distributed within calcareous sand deposits are marine shell beads, some unfinished, and other shell ornaments, worn sea urchin files used to make them (*97*), shell and stone adzes, shell fishhooks, chert flakes and cores, and numerous sherds of thin-walled, red-slipped pottery, rarely decorated, called Marianas Redware (*15*), similar to contemporary ceramics at coastal sites in northern Luzon (Philippines), western Borneo, the northern tip of the Sulawesi, and Kayoa Island of the northern Moluccas (*11*, *99*). Faunal remains include bones of turtle, pelagic and inshore fish, birds, fruit bat, and shellfish, mainly bivalves.

The campsites of the first occupants of the Mariana Islands have yielded no evidence for cultigens or domesticated animals. Also absent are remains of the Pacific rat (*Rattus exulans*), a human commensal often found in early sites in the Pacific Islands. Taken together, these findings fail to support the hypothesis of a demic expansion model for human colonization of Remote Oceania via the Mariana Islands, which proposes that population pressure on Neolithic farmers in the northern Philippines prompted a migration to the unoccupied Mariana Islands at this time (*11*, *100*). Cultural niche dynamics in ISEA provide a plausible explanation for why people bearing no cultural items relating to horticulture first sailed to these small, isolated islands but did not settle them permanently: *territorial range expansion* by economically vulnerable marine foragers. Range expansion can be understood as adaptive response to late Holocene environmental and social conditions, in particular sea level decline in the western Pacific (*101*, *102*). Sea level decline since its mid-Holocene peak had created extensive and abundant reef-lagoon systems in the southern Mariana Islands that could be harvested, albeit at great costs in transportation across the open sea. In a time of intense competition for raw materials, range expansion would enable small, more vulnerable marine foraging groups to continue to participate in the complex trade networks

of ISEA during the Southeast Asian Bronze Age (*103*). The Mariana marine foraging era ended with the onset of the Southeast Asian Iron Age, when newly configured trade networks shifted away from marine shell valuables and other trade items from these marginal groups. Iron Age socio-economic conditions may have severely constricted the ISEA marine forager niche, eliminating the most vulnerable: those who had occasionally visited the southern Mariana Islands during the previous thousand years. By 2500 BP, continued sea level decline had increased shoreline width in the Mariana Islands, creating arable coastal land and a new subsistence option: permanent settlement supported by lowland horticulture, foraging and fishing.

The shift from marine foraging and trading to a Mariana fisher-farmer cultural system, termed the Unai Period (3500-1600 BP), is recognized in the archaeological chronology used in this paper (**table S1**). The three Unai Periods span the time when land use patterns emphasized coastal living. The Huyong Period (1600-1100 BP) marked a settlement change to include use of interior habitats under conditions of low population density. During the Latte Period (1100-300 BP), clustered residential sites marked by *latte* were located near freshwater sources, and diverse site types and habitats were occupied throughout the archipelago. CHamoru phrases and English equivalents corresponding to the archaeological terms were supplied by Rosa S. Palomo and Laura M. T. Souder of the Guam CHamoru Language Commission. It should be noted that the spelling, CHamoru, for the indigenous language and ethnic group, is legal and preferred on Guam. However, elsewhere in the Mariana Islands, the older spelling, Chamorro, has been retained.

Unai sites have been found along the former shores of Tumon Bay on Guam's lee (western) side (*104*). The mortuary practices of the new cultural system are manifest at the Naton Beach Site, in multiple human interments. Possibly the descendants of earlier transient groups that had camped along the emerging small, intermittent beaches in the southern Mariana Islands, the Unai people responded culturally to new social and environmental conditions, especially continuing sea level decline, which now favored a mixed subsistence economy and supported a small, permanent population. Their social and demographic relationships with groups in ISEA have yet to be determined, yet it seems unlikely that Unai populations were entirely isolated and lacked occasional migration from outside, given the Mariana sub-optimal geographic conditions for horticulture (*105*).

Lower sea level had exposed more flat land and increased the complexity of coastal habitats. This made permanent near-shore settlement possible, supported in part by imported tropical cultigens (e.g., coconut, taro, breadfruit). Unai features and artifacts indicate regular and varied

residential activities. These include post-mold alignments from substantial pole-and-thatch structures; hearths and earth-oven deposits; domestic trash pits; basalt and marine shell tools of various shapes and sizes used in different tasks, especially wood working and processing of plant materials; clustered and single human burials located away from residential contexts, many individuals (all ages and both male and female) with a variety of elaborate grave goods not limited to shell ornaments; and a distinctive ceramic complex, including large, flat-bottomed, shallow, straight-sided, thick-walled, oval or round "pans". Charred taro starch residue has been identified on robust pan fragments found in earth-oven feature deposits (*106*).

On Guam, the Huyong Period (1600-1100 BP) succeeded the Unai and is archaeologically marked by limited inland settlement expansion. In addition to small coastal residential settlements, there were foraging and gardening sites in the island interior. The relatively low density of Huyong sites suggests a smaller population responding to a deterioration in agricultural productivity, an effect of cooler and drier climate at this time (*107*, *108*). The robust pan form was no longer used, indicating that cooking or eating contexts had changed again. Pot shapes were more suitable for above-ground cooking, although earth-ovens were used. The earliest known example of Mariana cave painting, in Mahlac Cave in southern Guam's interior, has been directly radiocarbon-dated to the late Huyong Period (*109*).

The archaeology of the Latte Period (1100-300 BP) is well-studied on Guam and Saipan due to the relatively large number of sites that have survived, although often heavily disturbed from historical and modern activities (*110*). A prominent trait of this last period in Marianas prehistory is the paired stone post-and-capstone two-row configurations (*latte* in CHamoru), a unique innovation within the Indo-Pacific megalithic tradition and unusual in its use in domestic architecture as house supports. Other innovations include large stone mortars, often embedded as site furniture near one end of the house or isolated in non-residential locales possibly related to dryland rice cultivation, a prehistoric practice unique to the Mariana Islands in the Pacific Islands (*111*). Artifact assemblages include an abundance of stone pounders, grinders, abraders, anvils, and adzes, dense accumulations of chipped stone, shell adzes, and bone and shell fishing gear appropriate to inshore and pelagic fishing (including spears and harpoons), and fragments of large, robust ceramic pots and jars appropriate for cooking and storage, especially after ~650 BP. The settlement system involved coastal and inland settlements, *latte*-supported structures throughout the archipelago, primary and secondary burial within or near houses at coastal villages and as isolated inland burials in caves — all evidence of a shift away from frequent and elaborate grave

goods as seen in the Unai Period. Adult Latte burials show evidence of betel nut chewing and deliberate teeth staining and incising, as well as post-mortem recovery of human skulls and long bones from which spears and harpoons were made. Evidence for endemic feuding is reflected in skeletal injuries and in the buried caches of sling stones at *latte* sites and scattered in open areas; inter-village rivalry was witnessed in 1565 by the Legazpi expedition.

The Naton Beach Site GHPI66-01-2488

The Naton Beach Site is situated in Gogna Cove at the north end of Tumon Bay (**fig. S22**), where a limestone headland juts into the sea, forming the northern terminus of the Tumon embayment. Elevations here range from 3-61 m above present sea level, and the land slopes steeply toward the sandy beach. Northern Tumon Bay once had a series of shallow lagoons and marshes that infilled as sea level declined from its mid-Holocene peak (*101*). A narrow strip of beach eventually formed between 3000-2000 BP (*112*, *113*), providing a stable land surface for Unai settlements. Freshets flowing over the sand during low tide may have been a source of fresh water in this limestone substrate area that lacks streams.

From 2006 to 2008, archaeological investigations at the Okura Hotel renovation project at the Naton Beach site resulted in the recovery and analysis of 367 sets of human remains described in an unpublished report prepared by SWCA Environmental Consultants, Guam. Some of the findings are summarized here and in the burial spatial distribution figure (**fig. S17**). Two temporally and culturally distinct groups were defined, 212 individuals associated with the Latte Period and early European contact era and 155 individuals associated with the Unai Period. Unai burials contrast with the Latte burials in morphology, pathology, and behavior. Descriptive and metric analyses revealed significant differences in size and dimensions of cranial bones, postcranial bones, and dentition. Frequencies of non-metric dental traits also differed between the Unai and Latte burials. Tooth dimensions decreased by 7% from the Unai to the Latte Period. The Unai individuals had a higher frequency of caries, antemortem tooth loss, and dental calculus, while the Latte individuals had a higher frequency of linear enamel hypoplasia (LEH), abscesses, and attrition. Nearly one-half of dental non-metric traits were significantly different, and dental modification (abrasion, incising) were also systematically different. Both groups suffered trauma, gout, periostitis, and degenerative joint disease (DJD), but the locations of DJD were different, suggesting differences in wear-and-tear activities. Yaws was not identified in the Unai individuals

but commonly occurred among the Latte Period individuals. Betel nut dentition staining, common in the Latte Period adults, was almost absent in the prior Unai population.

Ninety-six percent of the Unai burials were in a fully extended supine position, compared to 56% of the Latte Period burials. Both the Unai and the Latte burials were oriented predominately along on an east-west (E-W) axis, perpendicular to the shore. The Unai E-W orientation was 63%. The Unai skeletal assemblage had few subadults relative to adults. Adults represented 82%, with nearly an equal number of females and males. The relatively few Unai subadults may reflect different burial loci or preservation conditions. The Latte group was more evenly distributed relative to the numbers of subadults and adults, with adults comprising 70% of the total. Differential burial treatment by sex or age in either group was not observed. The Unai burials had a greater quantity and variety of grave goods than the Latte burials. Post-mortem bone harvesting and fabrication of human bone into tools, common in the Latte Period, was not found for the Unai burials.

The Unai burials were interred deeply in white calcareous sand, overlain by the dark brown, culturally enriched soils in which the Latte Period individuals were buried. The latter (Latte Period) burial context indicated interment in residential cultural deposits, contrasts with separate placement of the Unai burials away from residential deposits. Unai burials were located to the north of the Latte burials and were also clustered (**fig. S17a**). The Unai burial clusters may mark family mortuary spaces, whereas the Latte burial clusters may mark the location of houses, commonly observed at coastal Latte Period sites. The Unai burials were found approximately 50 meters inland of the current shoreline along the toe of a steep slope, four meters above mean high tide, which may account for their preservation, along with the well-drained, alkaline sand in which they were interred.

Haputo Beach Site GHPI66-08-0009

This site is situated in Haputo Bay, a small, sandy cove on Guam's northwest coast (**fig. S22**). It was occupied for about 1,800 years, from the Unai Period into the Historic Period (1700 BP-present), as indicated by 15 radiocarbon dates. The Latte Period village was abandoned when Spanish colonizers forcibly relocated the residents to the capital, Hagåtña. The Haputo Beach site was listed on the National Register of Historic Places in 1974, and in 1984 the Haputo Beach embayment was designated as an ecological reserve, covering 252 acres of coral reef and limestone forest. Haputo Bay was a traditional canoe landing place for visitors coming from the coral islands

of the Carolines. In 2013, several house floors were excavated by R. Olmo and O. Kataoka, and several burials were exposed. DNA from four teeth from individuals from one house was obtained during this study. The teeth were not directly radiocarbon-dated but are believed to come from Latte Period or early Historic Period individuals. Osteological descriptions of the excavated human remains are unavailable.

Ritidian Beach Cave GHPI66-08-0012

This site is located west of Ritidian Point in northern Guam. The site is positioned approximately 200 meters from the present shoreline. Excavations at Ritidian revealed multiple occupations starting ~3500 BP (*28*). Two sets of human remains were found buried deeply, side by side in an extended position near the mouth of a limestone cave. The two skeletons were highly fragmentary and had been disturbed by large pits dug into the original burial pit; torso and skull were missing from both individuals. In a separate study whose genetic data we reanalyze here, a tarsal bone from each individual was analyzed for DNA and a bone from one of the skeletons was radiocarbon-dated to 2180 ± 30 BP (*28*). Osteological analysis was not undertaken, and the remains were left *in situ*.

Anaguan Site SP-1-0762

This site, known traditionally and most accurately as Anaguan, has been also referred to in various documents as Best Sunshine, Imperial Pacific, *etc*. It is located in the town of Garapan on the western coastal plain on the west coast of Saipan, CNMI (**fig. S23**). Anaguan is situated within 200 meters of the present shoreline and less than two meters above present-day sea level. The site was first occupied on an unstable sandy beach area as early as 100–200 CE, and by 800-900 CE people were living in a more stable back beach area. These are approximate dates based on a regional sea-level history (*101*, *102*) in conjunction with dates from geoarchaeological testing of coastal site settings in the Mariana Islands (*11*, *96*, *114*).

Archaeological excavations at the Anaguan site were conducted in 2015 prior to construction of the Imperial Pacific International (IPI) casino, hotel, and supporting facilities. The 2015 excavations recovered a minimum of 416 individuals. Previous excavations were conducted from 1996 to 1999. A total of 676 individuals have been recovered from this location. Radiocarbon dating showed continuous occupation of this coastal zone from the 11th century through the 17th century, with the Latte Period occupation being densest in the 13th through the 16th centuries.

Analysis of the excavations in 2015 revealed three major burial clusters. An unpublished report prepared by Scientific Consultant Services, Honolulu on excavations conducted at Anaguan in 2015 was consulted for descriptions of the human remains and their spatial distribution in three major burial clusters (**fig. S17**). Only a few of the burials were fully intact, and many were represented by skeletal fragments. All but one of the burials were associated with the Latte Period. Multiple burials had incised and betel nut-stained dentitions but generally lacked grave goods. Long bones and skulls were often missing due to post-mortem harvesting for tool-making, typical of Latte Period mortuary customs.

II: Burials from Pohnpei

Pohnpeian Mortuary Behavior and Archaeology

Pohnpei possessed one of the most complex sociopolitical organizations of central Micronesia (*115*). A centralized sociopolitical hierarchy is often reflected by the differential treatment of the deceased during mortuary activities. Upon death, high-ranking individuals were typically afforded more elaborate burials and grave goods of greater value than were provided for those of lesser rank. Elaborate platform and vaulted stone structures were reserved for the highest-ranking Pohnpeians (*116*). Although there is little information available for describing the mortuary behavior associated with unranked (or commoner) individuals of Pohnpei, they were likely afforded simpler interments with relatively less prestigious and fewer grave goods. The lack of elaborate stone platforms and vaulted structures on Na Island suggests that the burials from Na Island sampled in this study do not represent individuals of high social rank.

Skeletons from Na Island

In early 1977, two Catholic priests from the Diocese of Honolulu, Fathers Louis H. Yim and Joseph Matheis, led an expedition to Pohnpei, Federated States of Micronesia, to search for the remains of Alexis Bachelot, the first Prefecture Apostolic of the Sandwich Islands (Kingdom of Hawai‘i). After being exiled from Hawai‘i on November 23, 1837, Bachelot, who was accompanied by Louis Désiré Maigret, set sail for the Philippines on board the Notre Dame de Paix. Already in poor health, Bachelot died at sea on December 5, 1837; he was 41 years old at the time of death. With the permission of the Nahnmwarki (Chief) of the Madolenihmw District, Pohnpei, Bachelot’s body was buried on the Island of Na, a small sandy motu (reef islet) found within the extensive reef that

encircles Pohnpei Island, eastern Caroline Islands. Na Island is located at the mouth of Madolenihmw Harbor slightly less than two miles (~2800 m) southeast of Pohnpei Island, not far from the monumental Nan Madol archaeological site. In the following year Maigret, with the assistance of local Pohnpeians, erected a cross 16 feet tall and a small chapel near the gravesite.

Although the cross and small chapel, or any other above-ground traces of Bachelot's grave did not survive, the organizers of the expedition anticipated that they would be able to locate Bachelot's remains, which they had planned to transport to Honolulu for interment at the Cathedral of Our Lady of Peace. Na Island, which has been uninhabited since the beginning of the early 20th century, measures only two miles long and several hundred yards wide. At the time of the expedition, Na Island was covered in coconut, breadfruit, and Pandanus trees and surrounded by mangroves.

With the assistance of an experienced local guide and more than 40 volunteer students from the Pohnpei Agriculture and Trade School, human skeletal remains were found at a depth of 1-2 feet below the surface and then exhumed over the course of six days. The skeletal remains, which were in a commingled state, came from a confined area near the center of the island. Except for some rocks and other surface rubble, above-ground indications of a gravesite were not observed. A few shell beads similar to those associated with burials at the Nan Madol nearby, were associated with some of the remains. Because none of the excavators was trained in archaeological techniques, no other information is available regarding the amateur recovery of the skeletons from Na Island. In the lab, the remains displayed varying degrees of completeness and preservation and most of the bones exhibited extensive postmortem damage. Climate, weather conditions (especially typhoons), coconut crabs, and other taphonomic factors likely contributed to the poor condition of the remains.

After the skeletons recovered from Na Island were transported to the University of Hawai‘i at Mānoa, Michael Pietrusewsky (a co-author of this paper) and his students analyzed them. Following a detailed study, which included a determination of the number of individuals represented and the recording of metric and nonmetric observations, it was concluded that none of them represented Bachelot who was of French ancestry. Ten individuals were identified, including two children (aged 2-5 years), two adult females, and six adult males ranging in age from young (20-35 years old) to old (more than 50 years old) age. Estimates of adult stature, cranial and infracranial morphology, dental pathology, and the presence of treponemal (yaws) infection in at least two individuals are consistent with Pohnpeian ancestry. Petrous bone samples from Burial

Nos. 2 (young to middle-aged male), 5 (middle-age to old female), and 6 (young adult female) were used to obtain ancient DNA and radiocarbon dates in this study. More detailed information about the Pohnpei skeletons is provided in previous publications (*117*).

Skeletons from Nan Madol

Nan Madol is a prehistoric megalithic complex of constructed mostly rectilinear islets located on a reef flat adjacent to Temwen Island on the eastern side of Pohnpei. The complex is distinctive for its abundant use of columnar basalt transported from at least several exposures around Pohnpei to construct walls and enclosures on the islets (*118*, *119*). Large basalt boulders were also used for the construction of some walls (especially the giant seawall). The complex encompasses 80 hectares and approximately 100 islets and several sea walls (*120*, *121*). It is notable that at least a portion of Nan Madol overlies a much earlier period occupation (*19, 120*). Construction of low platforms and the use of coral rubble fill began appearing around AD 900 – 1100. This was quickly followed by megalithic construction starting about AD 1200. Megalithic construction marked the beginning of the Sandeleur reign, which represented the development of a complex chiefdom society with hegemony (in some sense) over the entire island of Pohnpei (*119*, *122*). The Nan Madol complex was divided into three spatially identifiable sectors – a chiefly residential area with larger platforms and foundations, a priestly/ceremonial area with more numerous, smaller, proximate islets, and high-walled mortuary enclosures with burial vaults (*lolong)*, suitable for multiple interments; these were sometimes on platforms (*120*, *123*).

Nan Madol, despite its size, was not an urban location or town, but primarily functioned as an elite ritual center. Burials are common at Nan Madol, many of which are simple interments found under stone pavings, in caches between stones in walls, and beneath house floors on small platforms, and as noted, there are the much more elaborate *lolong* structures, which presumably were reserved for the highest-ranking individuals.

Saudeleur rule abruptly ended around AD 1600 for reasons that are not entirely clear (though there are oral accounts concerning the demise of the Saudeleur rule by an outside force). At this time, Pohnpei became politically and socially divided into four (and later 5) independent *wehi*, or chiefdoms, with defined territories and a new political order (*124*).

Human skeletal remains were recovered from nine of the Nan Madol islets during archaeological investigations led by J. Stephen Athens (a co-author of this paper) in 1984, which involved both islet mapping and excavations. The remains were submitted to Michael

Pietrusewsky and Michele Toomay Douglas (co-authors of this paper) at University of Hawai‘i-Mānoa in the spring of 1984 and examined during the 1984-85 academic year. The skeletal elements were cleaned in water, dried, and numbered using the islet number, NM as the designation for Nan Madol, followed by the catalog number assigned by the excavators. Reconstruction of the more important fragments was attempted utilizing adjacent bags or categories to estimate a minimum number of individuals. Sex, age-at-death, and measurements were recorded when possible. Preservation was generally fair to poor, with fragmentation, weathering, and loss of elements because of disturbance in part due to the extremely wet and humid tropical conditions of Pohnpei, and in some cases, exposure to sea water. Skeletal completeness varied from portions of the cranium, loose teeth, and diaphyseal portions of postcranial elements to more complete reconstructable long bones and crania. The entire assemblage represents 35 individuals, predominately adults (27), 13 females, five males, and nine of indeterminate sex. Estimation of stature in six individuals suggests short stature (5’ to 5’5”).

Eight individual petrous bones from four different islets at Nan Madol were submitted for ancient DNA extraction. A single unit was excavated in a mortuary structure (a basalt paving) on Islet 102 –Peinioar (we note that Pohnpei orthography is highly variable, but here we follow islet name spellings published in the reference (*124*), using a list prepared by the Pohnpeian anthropologist, Rufino Mauricio), and ten different bone groups were identified and collected from very shallow depths. Bone fragments from all the groups overlapped and contributed to the reconstruction of elements. Nine individuals, seven adults and two children were identified in all. Four adults are included in this analysis. These consisted of the cranium and long limb bones of an adult male (102-NM-106), skull fragments and long bone fragments of a male (102-NM-108A), and two individuals from bone group 102-NM-120. Bone group 102-NM-120 contained cranial fragments of an adult male (102-NM-120), and a middle-aged female (102-NM-120A), as well as postcranial fragments that could not be assigned.

Two excavation units on Islet 110 – Pahndauwas, sampled a house platform (Feature 9) and a square mortuary platform 60 cm high made of coral fill (Feature 4). Feature 9 revealed scattered fragmentary human bone (MNI three adults). Four bone groups were recovered from Feature 4 (193, 113, 114, 118). Bone group 110-NM-193 contained a nearly complete skull and other postcranial bone fragments that were not articulated but which were consistent with an adult female. The remaining bone clusters represented incomplete skeletons of an additional four adults.

A platform on Islet 114 – Pohndauwas, yielded the skeletons of three individuals identified from three bone caches hidden in openings in the north coral wall of the platform. These caches each contained a cranium and long bone shafts consistent with secondary burial of a single individual. These caches contained one adult female (114-NM-BC1) not utilized in the present analysis, one adult male (114-NM-BC2), and an adolescent less than 15 years of age (114-NM-BC3); aDNA analysis later identified the individual as a male aDNA.

Investigations at Islet 129 – Lemenkou, documented a mortuary platform structure (Feature 18) in which a single excavation yielded multiple disturbed and mixed bone groupings, including cranial and postcranial elements, representing one adult male and two adult females. The adult male (129-NM-147) was associated with shaft fragments of the femora, right tibia, and portions of the left parietal and temporal bones.

S3. f_4-regression Analyses to Elucidate the Ancestry Sources of Micronesians

Overview

To investigate the different ancestry sources of Micronesians, we construct linear regression models for pairs of f_4-statistics to test the degree of allele-sharing with different populations.

This method leverages the idea that if a series of candidate populations can be modeled as a mixture of the same two ancestry components with different proportions, then the corresponding f_4-statistic pairs can be modeled as a simple linear equation. Alternatively, if the sources of some test populations are different, these populations would show deviations from the linear model. By examining the extent to which the deviations between the observed and expected values match (produce significant residuals), we can infer whether test populations fit the model (or not) and find more appropriate sources of ancestry. The statistical significance of deviations (residual Z-scores) can be evaluated by block jackknife resampling.

In our analysis, we always used the statistic f_4(X, New Guinea Highlanders; Kankanaey Igorot, Australian) as the independent variable, as it is proportional to East Asian / FRO ancestry proportion (x-axis; higher values indicate more East Asian / FRO ancestry). This statistic allows us to correct for differential levels of East Asian-associated ancestry in the test populations "X". For the dependent variable (y-axis), we use a variety of f_4-statistics depending on the signals that we are attempting to isolate: Papuan sources, FRO sources, or Polynesian sources.

Test for Differences among the Papuan Sources of Remote Oceanian groups

To test for differential Papuan sources, we computed f_4(X, Dai; Nasioi, New Guinea Highlanders) and plotted the statistics against the independent variable proportional to FRO ancestry. Higher values indicate greater allele sharing with Nasioi, and lower values indicate greater allele sharing with New Guinea Highlanders, allowing identification of differential relatedness to specific Papuan ancestry groups after controlling for FRO ancestry proportion (orange cline in **Fig. 1B**).

We calculated the statistics using two separate datasets. First, we restricted to ~397,000 autosomal SNPs on the Human Origins array (**fig. S9A**). We calculated the f_4-statistics for late prehistoric Pohnpei (merging 11 individuals as one data point), 10 present-day Micronesians (the statistics were computed separately), and 106 prehistoric and present-day Oceanian populations (**table S10**). A linear regression model for the Polynesian-Vanuatu-New Britain cline (Eq.1 in **fig. S9A**) was constructed using all prehistoric and present-day populations from Vanuatu and

Polynesia, except for the Lapita groups from Vanuatu and Tonga. A linear model for the Micronesia-New Guinea cline (Eq.2 in **fig. S9A**) was constructed using all prehistoric and present-day populations from Micronesia and New Guinea, except for the Unai and the Latte groups from Saipan and Guam.

Second, we restricted our analysis to ~169,000 autosomal SNPs that form the intersection between the 1240K and the MEGA array, as shown in **Fig. 3A**. We calculated the statistics for late prehistoric and present-day Pohnpei, present-day Palau, Chuuk, and Guam, as well as 40 prehistoric and present-day Oceanian populations (**table S11**). A linear regression model for the Polynesian-Vanuatu-New Britain cline (Eq.1 in **Fig. 3A**) was constructed using all prehistoric and present-day populations from Vanuatu and Polynesia, except for the Lapita groups from Vanuatu and Tonga. A linear model for the Micronesia-New Guinea cline (Eq.2 in **Fig. 3A**) was constructed using late prehistoric and present-day groups from Pohnpei, Chuuk, Palau, and New Guinea.

Tests for Differences among the FRO Sources of Remote Oceanian groups

We computed f_4(X, New Guinea Highlanders; Lapita, Unai) (higher values indicate greater allele sharing with Lapita, and lower values indicate greater allele sharing with Unai). We estimated this statistic indirectly using *qp4diff* in ADMIXTOOLS by computing the difference between f_4(X, New Guinea Highlanders; Dai, Unai) and f_4 (X, New Guinea Highlanders; Dai, Lapita). This approach allowed us to gather information about non-overlapping SNPs and to increase the power of f_4-statistics when comparing two sets of samples with low sequencing coverage. We used present-day populations to perform these analyses to avoid confounding effects when involving present-day and ancient DNA data on both sides of the f_4-statistics at the same time (*125*). Nevertheless, to document the genetic continuity in Guam and Central Micronesia, we computed the statistics for the late prehistoric Pohnpei and the Latte groups from Guam and Saipan and then plotted them in the figure.

We calculated the statistics using two separate datasets. For the Human Origins dataset, as shown in **fig. S9B**, we calculated the f_4-statistics for late prehistoric Pohnpei (merging 11 individuals as one data point), three Latte groups from Guam (Naton, Haputo) and Saipan (Anaguan), 10 present-day Micronesians without precise knowledge of their geographic origin (the statistics were computed separately), and 83 present-day Oceanian populations (see **table S12**). A linear regression model (Eq.3 in **fig. S9B**) was constructed using all present-day populations except those from Micronesia. For the intersection between the 1240K and the MEGA

array, as shown in **Fig. 3B**, we calculated the statistics for late prehistoric and present-day Pohnpei, present-day Palau, Chuuk, and Guam, three Latte groups, as well as 17 present-day Oceanian populations (**table S13**). A linear regression model (Eq.3 in **Fig. 3B**) was constructed using all populations.

Test for Differential Polynesian Sources

For testing differential relatedness to Polynesians, we computed f_4(X, Tolai; Kankanaey Igorot, Polynesian) (higher values indicate greater allele-sharing with Kankanaey Igorot, and lower values indicate greater allele-sharing with Polynesians), where Polynesian includes populations from Polynesia (Tongan, Samoan, Tahiti, Rapa Nui) and Polynesian Outliers (Tikopia, Ontong Java, Rennell and Bellona) (**tables S14** to **S21**). As the choice of the second left population of the dependent f_4-statistics only affects the intercept of the regression model, for tests of differential Polynesian influence we chose Tolai (from New Britain) in that position as justified by previous work (*9*). We calculated the statistics using the Human Origins dataset, as shown in **figs. S10** and **S11**. Linear models were constructed using all the prehistoric and present-day populations from Near Oceania, including New Guinea, New Britain, New Ireland, and Solomon Islands.

Test for Additional Southeast Asian Sources

For testing differential relatedness to Southeast Asians comparing with the FRO lineages, we computed f_4(X, New Guinea Highlanders; Unai, Kankanaey Igorot) (higher values indicate greater allele-sharing with the $FRO_{Marianas}$ lineage, and lower values indicate greater allele-sharing with Southeast Asians represented by Kankanaey Igorot). We restricted the analysis to data from present-day populations (including late prehistoric Pohnpei) to avoid confounding affects when involving present-day and ancient DNA data on both sides of the f_4-statistics at the same time (*125*). We calculated the statistics using two separate datasets. For the Human Origins dataset, as shown in **fig. S18A**, we calculated the f_4-statistics for late prehistoric Pohnpei (merging 11 individuals as one data point), 10 present-day Micronesians (statistics computed separately), and 83 present-day Oceanian populations (**table S34**). We constructed the linear regression model using all the populations. For the intersection between the 1240K and the MEGA array (**fig. S18B**), we calculated the statistics for late prehistoric and present-day Pohnpei, present-day Palau, Chuuk, and Guam, as well as 17 other present-day Oceanian populations (**table S35**). We constructed the linear regression model using all the populations.

S4. Genetic Modeling by *qpWave* and *qpAdm*

Overview

We applied *qpfstats* in ADMIXTOOLS to obtain maximally precise estimates of *f*-statistics prior to all *qpWave* and *qpAdm* analyses; *qpfstats* addresses the complication that different populations are genotyped on different sets of SNPs. We set "allsnps: YES" for all analyses. For the models involving ancient samples we set "inbreed: YES" to let the software know that it should treat each individual's data at each position as a randomly sampled allele (pseudo-haploid data). If any group involved in the analysis (either in the left or in the right) was represented by only a single individual, the option of "inbreed: YES" cannot be applied, and we changed the parameter to "inbreed: NO".

Phylogenetic Relationships among FRO and Latte Lineages

We used *qpWave* to test whether two FRO lineages (Lapita Vanuatu and Tonga vs. Unai Naton Beach Guam) and three Latte lineages (Naton Beach Guam, Haputo Beach Guam, and Anaguan Saipan) form a single clade (**table S22**). We first used nine populations as the outgroups, which included Aboriginal Australian (deeply Papuan-related), French (western Eurasian), Pima (Native American), Onge (deep eastern Eurasian), Japanese (northern East Asians), Dai (southern East Asians), Ami (indigenous Taiwanese), Kankanaey Igorot (indigenous people in the Philippines), and New Guinea Highlanders (Papuans), with phylogenetically distinct positions relative to our target populations. Our target populations first consist of Unai Guam (Naton Beach), Latte Guam (Naton Beach), Latte Saipan (Anaguan), Latte Guam (Haputo), and Lapita (Vanuatu and Tonga). Our hypothesis was that the model would pass when rank = 0, because all of the target populations were derived from East Asian lineages. However, the results showed that the model failed when rank = 0 and passed when rank = 1. We next removed French, Pima, and New Guinea Highlanders from the outgroups in turn to determine which one collapsed the one-way model. The one-way model (rank = 0) succeeded when we removed French from the outgroups. When we added present-day Sardinian (western Eurasian) to the target populations and used the nine-population outgroup with French on the other side, the two-way model (rank=1) continued to be consistent with the data. This result meant that the European-related population in the outgroups collapsed the one-way model.

We highlight two possible explanations for this observation that Europeans share an attraction to some of the ancient Oceanian individuals more than to others.

The first explanation we considered is that there is a small fraction of contaminating sequences in the Lapita and/or Unai individuals that is of West Eurasian ancestral origin. Because many archaeologists and ancient DNA specialists have European ancestry, this is a plausible type of contamination, and it is not surprising to see it differentially represented in the lower quality (and hence more easily contaminated) data from the Unai and Lapita individuals, and in the higher quality (and hence less easily contaminated) data from the Latte individuals. Our direct estimates of contamination are low (on the order of a percent or less), but it is important to always keep in mind that contamination in ancient DNA data is never zero.

The other possible explanation for the observation of excess of allele sharing of some ancient Oceanians with Europeans relates to the fact that the sequences for both the Lapita and Unai individuals show extraordinarily high levels of ancient DNA damage, which could lead to artifactual excess allele sharing between damaged ancient samples and some deeply divergent outgroups with genetic variants that just happen to match errors due to damage (Africans and Europeans might be expected to show such signals more than East Asians) (*125*). Supportive evidence for this scenario comes from the fact that the one-way model (rank = 0) succeeds when only restricting to transversion polymorphisms.

Regardless of the cause, this analysis suggests that inclusion of Europeans in the outgroups for analyses of Remote Oceanian population history produces complexity in the inference that may not be reflecting real population structure in Remote Oceanians. We therefore removed French from the outgroups in all our subsequent modeling. This analysis is based on approximately 825,000 autosomal SNPs of the 1240K panel after removing sites in CpG dinucleotides.

Estimating Papuan Ancestry Proportions for FRO and Latte Lineages

We applied *qpAdm* to estimates their Papuan ancestry proportions in Remote Oceanian groups (**table S23**). We used Kankanaey Igorot as a proxy for unadmixed East Asian-related ancestry, and New Guinea Highlanders as a proxy for unadmixed Papuan-related ancestry. We used seven populations as outgroups, comprising Aboriginal Australian, French, Pima, Onge, Japanese, Dai, and Ami. We performed this analysis based on ~825,000 autosomal SNPs in the 1240K panel after removing sites in CpG dinucleotides. The results showed that the Unai and Latte groups were consistent with having no Papuan ancestry at all, except for four Latte individuals from Saipan, who we estimated to have around ~5% Papuan ancestry. The Lapita individuals from Vanuatu and

Tonga were inferred to have small but non-zero proportion of Papuan ancestry, with point estimates around 2% and 5% on average, respectively.

The Latte individuals included a small fraction (5 out of 62 males) of Y-chromosomal haplogroups (M and S) common in people of Papuan ancestry. It is plausible that these paternal lineages are related to the significant proportions of Papuan ancestry we detect in genome-wide analysis of four Latte individuals (point estimates of 4.0-5.8%), despite the overall low proportion of Papuan ancestry in the Latte group (95% CI of 0-3.1%). If this is true, then the Papuan ancestry could have arrived in the Latte group in a male sex-biased manner, potentially reflecting a cultural bias of local females to mate with high-prestige males with Papuan ancestry.

Testing for Genetic Continuity and Homogeneity in Pohnpei and Chuuk

To test for genetic evidence of continuity and homogeneity in Pohnpei and Chuuk, we used *qpWave* (**table S25**). We used 11 populations as the outgroups, including Aboriginal Australians, French, Pima, Onge, Japanese, Dai, Ami, Kankanaey Igorot, New Guinea Highlanders (Papuan lineage in mainland New Guinea), Nasioi (Papuan lineage in the Solomon Islands), and Baining (a Papuan lineage in the Bismarck Archipelago). We performed this analysis based on either the autosomal SNPs of the MEGA array or the intersection between the 1240K and the MEGA array. The results showed that late prehistoric and present-day Pohnpei and Chuuk were consistent with forming a single clade relative to the outgroups, as expected if they were genetically homogeneous. We found that present-day Chuuk were consistent with being an outgroup for late prehistoric and present-day Pohnpei, consistent with genetic continuity on Pohnpei.

Pairwise Comparisons among Micronesians Genotyped on the Human Origins Array

Because we did not have precise geographic information for 10 present-day Micronesians genotyped on the Human Origins array, we sought to use *qpfmv* and *qpWave* to determine if we could cluster them into genetic groups. We used 12 populations as outgroups, including Aboriginal Australians, French, Pima, Onge, Japanese, Dai, Ami, Kankanaey Igorot, New Guinea Highlanders, Nasioi, Baining, and Tongan (Polynesians). The results show that the ancestry of these individuals was extremely diverse ($P << 0.01$, Hotelling's T^2 test estimated by *qpfmv*, also revealed by *qpWave* shown in **fig. S20**), except for Jk2815-Jk2817 and Jk2813-Jk2814 who genetically consistent with being a clade. We thus modeled these 10 individuals separately (see immediately below).

Genetic Modeling of Prehistoric and Present-day Micronesians

We used *qpAdm* to fit two-way or three-way genetic models of late prehistoric and present-day Pohnpei, Chuuk, Palau, Guam, and other Micronesians (**Fig. 4B** and **table S24**). To find appropriate models for prehistoric and present-day Micronesian Islanders, who have already been confirmed to harbor Papuan and East Asian ancestry, our models always used New Guinea Highlanders as the proxy for their Papuan source, while five FRO and Latte groups were in-turn used as the proxies for their East Asian source. Outgroups for the models involved at least nine populations (basic outgroup set), including Aboriginal Australian, Pima, Onge, Japanese, Dai, Ami, Kankanaey Igorot, Nasioi, and Baining. We include multiple Papuan ancestry groups among the outgroups to distinguish the different Papuan lineages that contributed to Micronesia (New Guinea-related) and the Southwest Pacific / Polynesia (New Britain-related).

For late prehistoric individuals from Pohnpei, we first ran our models on the autosomal SNPs of the 1240K panel (~825,000 SNPs) using the basic outgroup set. We then added Hawaiian as a proxy for Polynesians in the outgroups to determine whether the late prehistoric individuals from Pohnpei had any evidence of Polynesian ancestry. We also ran our models on the autosomal SNPs of the Human Origins array (~397,000 SNPs) with Tongan (a population in western Polynesia) added to the outgroups. Across the models we tested, we found that the Lapita individuals from Vanuatu and Tonga were a better proxy for East Asian ancestry in late prehistoric individuals from Pohnpei than the Unai and the Latte groups from Guam and Saipan. Using the Human Origins dataset, we further investigated if some New Guinea-related groups were better proxies for the Papuan ancestry of Micronesia than New Guinea Highlanders. Specifically, we tested Manus, Central, and Gulf as the Papuan proxies, and models succeeded with New Guinea Highlanders in outgroups, suggesting that groups from the Admiralties and coastal area of mainland New Guinea were potentially better Papuan sources.

We modeled ten Micronesian Islanders genotyped on the Human Origins array individually because their genetic backgrounds were extremely diverse (see above). We included Tongan or Samoan (when Tongan was used as a source on the left) among the outgroups to determine if there was evidence of Polynesian ancestry. Three individuals (Jk2815, Jk2817, and Jk2824) were consistent with deriving their East Asian ancestry from the $FRO_{Marianas}$ lineage, whereas the models failed when changing to the Lapita individuals as a source, suggesting they were likely from islands in western Micronesia. Four individuals (Jk2810, Jk2813, Jk2814, and Jk2818) could be modeled using New Guinea Highlanders and the Lapita individuals as sources instead, suggesting they were

likely from islands in central Micronesia like Pohnpei and Chuuk, although they also had additional European ancestry (also observed in **fig. S7**). For individual Jk2812 who had genetic affinities to Polynesians, we could fit a model only by including a third, Polynesian-associated source, such as Tikopia or Tonga. This individual also harbored East Asian ancestry more related to the Unai and Latte individuals. We found again that groups from the Admiralties and the coastal area of mainland New Guinea (such as Manus, Central, and Gulf) were consistent with the better Papuan sources. There were two individuals (Jk2820 and Jk2822) for whom we were not able to identify a well-fitting model with no more than three sources, although our analyses suggested they had at least FRO (likely $FRO_{SouthwestPacific}$ in Jk2820 and $FRO_{Marianas}$ in Jk2822, **fig. S9B**) and Papuan ancestry, indicating a more complicated admixture process probably caused by additional migrations and/or high regional mobility (**fig. S18A** and **table S34**).

For present-day individuals from Pohnpei, Chuuk, and Palau, we performed analyses using the intersection between the 1240K and the MEGA array (~169,000 autosomal SNPs). We used Hawaiians in the outgroups to determine if present-day individuals from Pohnpei, Chuuk, and Palau had Polynesian ancestry. We also replaced New Guinea Highlanders with ten New Guinea-related groups from the coastal area of mainland New Guinea and the Admiralty Islands, including Manus, Central, Gulf, Western, East Sepik, Middle Sepik, Madang, Morobe, Northern, and Milne Bay, to determine if they were better proxies for the Papuan source of Micronesia. We found that Pohnpei and Chuuk harbored East Asian ancestry more closely related to the Lapita individuals, while Palau harbored East Asian ancestry more closely related to the Unai and Latte individuals from Guam and Saipan. We also noted that some groups from the northern fringe of mainland New Guinea and the Admiralty Islands, such as Manus, Madang, Morobe, Northern, and Milne Bay, were consistent with being better proxies for Papuan sources than New Guinea Highlanders.

For present-day CHamoru from Guam, we used Aboriginal Australians, Onge, Japanese, Dai, Ami, Kankanaey Igorot, New Guinea Highlanders, Cambodian (mainland Southeast Asia), Mentawai, Dieng (Java), Maanyan (Borneo), SangliatDol (Tanimbar), and Rampasasa (Flores) as outgroups to distinguish different lineages from Mainland and Island Southeast Asia. We performed the analyses on two datasets, one comprised of SNPs targeted in both 1240K enrichment and the MEGA array, and the other consisting of SNPs targeted in both the 1240K and GenoChip2.0 array. All present-day CHamoru from Guam could be successfully fitted using a three-way model, involving East Asian, Native American, and European ancestry. However, we could not distinguish whether the Lapita individuals from Vanuatu and Tonga or the Unai and

Latte individuals from Guam were a better proxy for the East Asian source. Despite all of these models succeeding, we judged the East Asian source of present-day CHamoru to be more related to the Unai and the Latte individuals from Guam based on the evidence from our other analyses (e.g., **Fig. 3**).

When calculating the total proportions of East Asian and Papuan ancestry across models, we ignored the possible contribution of a very small proportion of Papuan-related ancestry to the Lapita individuals. This simplifying assumption has the potential to lead to higher estimates of East Asian-ancestry proportions and lower estimates of Papuan-ancestry proportions for some Micronesians. This is likely to be a very small bias given the very low estimates of Papuan ancestry in the Lapita individuals.

S5. Admixture Graph Construction using *qpGraph*

Overview

We applied *qpfstats* in ADMIXTOOLS for all *qpGraph* analyses. We used "allsnps: YES" and minimized the influence of missing data across populations using *qpfstats*. For our main analyses designed to distinguish different sources that contributed to the ancestry of Micronesia, we systematically constructed a variety of models and assessed their fits to the data by comparing the f_4-statistics predicted by the models with the empirical statistics (**figs. S12** to **S15**). Those f_4-statistics with |Z|>3 between empirical and predicted statistics were considered to be outliers and gave us evidence for unmodeled shared genetic drift between populations.

Populations Involved in the Models

We involved a variety of prehistoric and present-day populations that were primarily from East Asia and Oceania to fit the models:

1) Outgroup: Mixe (Native American)
2) Aboriginal Australians (deeply Papuan-related)
3) Unadmixed Papuans: New Guinea Highlanders (Eastern Highlanders)
4) Admixed Papuans: Manus (the biggest island in the Admiralties and to the north of mainland New Guinea)
5) Indigenous Taiwanese without Papuan ancestry: Ami
6) Philippines population without Papuan ancestry: Kankanaey Igorot
7) First Remote Oceanians: Unai people from the Naton Beach site on Guam ($FRO_{Marianas}$); Lapita people from Vanuatu and Tonga ($FRO_{SouthwestPacific}$)
8) Latte people: Latte people from the Naton Beach site on Guam; Latte people from the Haputo site on Guam; Latte people from the Anaguan site in Garapan, Saipan
9) Late prehistoric and Present-day Micronesians: late prehistoric and present-day Pohnpei and Chuuk (Central), and Palau (Western).

We use Mixe (a present-day Native American population from present-day Mexico) as an outgroup in our models; this is valid because even though there is a known deep admixture of western and eastern Eurasian lineages in Native Americans, they can be collapsed into a single

lineage with respect to the populations in our model as their history is symmetrical relative to the other populations in our modeling.

Basic Skeleton Graph Fitting the Ancient DNA Data

Following published admixture graph modeling of populations in Near and Remote Oceania (*7*–*9*), we first constructed a simple tree-like skeleton graph without admixture events including Mixe (a present-day Native American group from Mexico, outgroup), Aboriginal Australians (deeply Papuan-related), Eastern Highlanders from mainland New Guinea (unadmixed Papuans), Ami (indigenous Taiwanese without Papuan ancestry), and Kankanaey Igorot (a population from the Philippines without Papuan ancestry). We did not include Baining and Nasioi in the admixture graphs, because they did not contribute to the Papuan ancestry in Micronesians as suggested by our previous modeling by *qpAdm* (**table S24**).

To fit our ancient and present-day DNA data, we added two First Remote Oceanian groups, three Latte groups, and the late prehistoric Pohnpei group into the graph as shown in **fig. S12**. We could not obtain successful models with no residuals |Z|>3 by involving only one (**fig. S12A**, max residual |Z| = 4.126) or two admixture events (**fig. S12B**, max residual |Z| = 3.269), but were able to do so by adding three admixture events to the graph. In the three admixture event model (**fig. S12C-F**), (1) individuals from Pohnpei were admixed with a Papuan source related to New Guinea Highlanders (contributing ~27% of the ancestry) and an East Asian source related to Lapita individuals (contributing ~73% of the ancestry); (2) Lapita individuals from Vanuatu and Tonga harbored ~1% Papuan ancestry related to New Guinea Highlanders which is within the range of the *qpAdm* estimate; and (3) Latte individuals from Guam and Saipan were admixed with the Unai lineage and an additional East Asian source, contributing ~16% ancestry to all of the Latte groups.

We further investigated the source of the additional East Asian ancestry that contributed to the Latte individuals (the FRO_{Palau} lineage, see next section). The evidence that a FRO_{Palau} lineage is necessary in the model arises not only due to the indications from f_4 outliers when we tried the simple models shown in **fig. S12A** and **B**, but also because the statistics f_4(Latte, Unai; Lapita, East Asians) consistently yielded negative significant Z values, with a maximum |Z| = 4.182 (**table S26**). We explicitly tested where the FRO_{Palau} lineage came from by reconstructing different models as shown in **fig. S12C-F**. Although the three models shown in **fig. S12D-F** have comparable worst residuals (|Z| all less than 3) to the model shown in **fig. S12C** (max residual |Z| = 2.959), models in **fig. S12D-F** became extremely poor when we add modern Palauans into the

admixture graph and allow no more than one additional admixture event (worst |Z| all greater than 8 for **fig. S13B-G**, see next section). In contrast, the model of **fig. S12C** continues to fit well when Palauans are added with one admixture event, so we favor this model (worst residual |Z|=2.762 for **fig. S13A**, see next section).

All the graphs shown in **fig. S12** were constructed using autosomal sites of the 1240K panel after removing sites in CpG dinucleotides (this left ~825,000 autosomal SNPs).

Adding Present-day Palauans to the Graph to Identify the Origin of the FRO_{Palau} Lineage

We added present-day Palauans to the graph. Because present-day Palauans were genotyped on the MEGA array, we constructed the graphs using SNPs at the intersection between the 1240K and the MEGA array after removing sites in CpG dinucleotides (this left ~169,000 autosomal SNPs).

To differentiate the origin of the non-Unai-related East Asian-associated ancestry in the Latte individuals and to investigate the origin of East Asian-associated ancestry in the present-day Palauans, we explicitly constructed different models as shown in **fig. S13**. All the models shown in **fig. S13B-G** have extremely poor fits (worst |Z| all greater than 8). We could only model present-day Palauans as the mixture of a Papuan source related to New Guinea Highlanders and a non-Unai-related East Asian-associated source that also contributed to the ancestry of Latte individuals (**fig. S13A**, max residual |Z| = 2.762). We call this third East Asian-associated lineage as FRO_{Palau} because the proportion of this lineage is maximized in modern Palauans (point estimate of 62%).

Adding Manus to the Graph

The *qpAdm* modeling (**table S24**) suggests that people of Manus from the Admiralty Islands may be a better proxy for the Papuan source of Micronesia than New Guinea Highlanders. We therefore tried to fit Manus in the graph. According to the results from f_4-based regression analyses (**Fig. 3B** and **fig. S9B**), we know that present-day Manus derive their East Asian ancestry from a lineage closely related to the $FRO_{SouthwestPacific}$. Corroborating our previous findings by *qpWave* and *qpAdm* modeling, Papuan related ancestry similar to that in Manus is shared in many Micronesian Islanders including Pohnpei (**figs. S14A** and **S15B**), Chuuk (**fig. S15B**), and Palau (**fig. S14B**). Because present-day Manus were genotyped on the MEGA array, we constructed the graphs using SNPs at the intersection between the 1240K and the MEGA array after removing sites in CpG dinucleotides (this left ~169,000 autosomal SNPs).

Testing Genetic Homogeneity and Continuity in Pohnpei and Chuuk

We tested for evidence of genetic homogeneity and continuity of Pohnpei and Chuuk populations by constructing admixture graphs (**fig. S15**). The graphs were consistent with our previous findings by *qpWave* and *qpAdm* that late prehistoric and present-day Pohnpei and Chuuk formed a single clade. Genetic structure between the islands of Pohnpei and Chuuk was detected, as Chuuk was an outgroup for late prehistoric and present-day Pohnpei. The model displayed in **fig. S15C** includes two *f*-statistics with a |Z|>3 difference between their observed and expected values: f_4(Kankaney Igorot, Lapita; present-day Chuuk, late prehistoric Pohnpei) with Z = 3.228, and f_3(present-day Chuuk; late prehistoric Pohnpei, Lapita) with Z = 3.017. Taking into account the large number of *f*-statistics tested for consistency between theoretical prediction and observed values without multiple hypothesis correction, we do not see these Z-scores as strong evidence against our model. This is especially the case as some of the statistics are of the form f_4(present-day DNA, ancient DNA; present-day DNA, ancient DNA), where data from two populations analyzed using the same methodology—for example two populations assessed by ancient DNA analysis as opposed to modern DNA analysis—can appear more similar to each other than is in fact theoretically expected based on the population history (*125*).

S6. Simulations to Assess the Significance of the Signal of Matrilocality

For the Unai samples from Guam and Lapita individuals from Vanuatu and Tonga, we observed complete mitochondrial haplogroup differentiation, with only mitochondrial haplogroups E1 or E2 being present in Unai individuals and only B4a1a1 in the Lapita individuals (*7–10*). This observation raised the possibility of accelerated maternal-line genetic drift compared to the expectation from the autosomes during the ancestral divergence between the Lapita and Unai lineages and the early radiation of FRO ancestry. We sought to formally test this possibility using a simulation-based approach. We asked: what is the probability that a single mitochondrial macrohaplogroup (like B in Lapita Vanuatu and Tonga vs. E in Unai Guam) would take over the whole population simply based on random genetic drift measured on the autosomes, with an appropriate conversion to adjust for differences in the effective population size of the mitochondrial DNA compared to the autosomes? If fewer than 5% of simulations of the mitochondrial drift process produced such extreme differentiation in a matched sample, we could reject the null hypothesis of no accelerated maternal line drift at $\boldsymbol{P} << 0.05$.

The autosomal F_{ST} between Lapita and Unai groups was estimated at 0.083 ± 0.004, using data from seven Lapita individuals and nine Unai individuals. Here, we restricted our analysis to these individuals because they had relatively high coverage, weak signals of contamination, and a low level of ancient DNA damage. Under the Wright-Fisher genetic drift model, the expected F_{ST} can be calculated as:

$$F_{ST} = 1 - e^{-\frac{T}{2Ne}}$$

where $\boldsymbol{T}$ is the number of generations since separation, and $\boldsymbol{Ne}$ is the effective size of the ancestral population. To simplify our simulations, we made the following assumptions: (1) *Ne* is constant among branches of the tree through time; (2) the numbers of males and females within the population are equal and mitochondrial DNA is only carried by females so the relevant effective population size is half of *Ne*; and (3) mutations do not modify the macrohaplogroup calls, which is the expectation under the infinite sites model of molecular evolution.

We started our modeling with a genetic pool of half the *Ne* mitochondrial sequences in each of the two lineages, corresponding to the number of females in the population. For each subsequent generation, we randomly resampled *Ne* / 2 mitochondrial sequences from the genetic pool of the last generation. After resampling *T–1* times, we randomly sampled (with replacement) seven mitochondrial sequences for the simulated Lapita population (because only seven Lapita

individuals have reliable mitochondrial haplotype calls) and nine mitochondrial sequences for the simulated Unai population in the T^{th} generation to simulate what we have observed in reality. Many replicates (e.g., 1,000,000 replicates) of this produce a distribution of mitochondrial haplogroup frequencies for the simulated Lapita and Unai populations (**fig. S16**). This procedure allows us to compute an empirical probability (*P*-value) for how often we would expect to observe a scenario as extreme as that we observed under the null hypothesis, in which all Lapita individuals have the same mitochondrial macrohaplogroup, and all Unai individuals have different haplogroups (complete differentiation).

A previous study (*6*) showed that the indigenous Iron Age population of Taiwan (~2000 BP) was consistent with being a clade with the common ancestral population of Lapita and Unai groups before they separated. Consistent with this, in a large sample of published ancient DNA from Iron Age Taiwanese, we observe all of the mitochondrial haplogroups seen in the Lapita and Unai samples as well as other haplogroups not present in First Remote Oceanians. The approximate frequencies of all different mitochondrial macrohaplogroups are: B, 13.3%; D, 4.4%; E, 28.9%; F, 31.1%; M, 6.7%; R, 15.6%. Those assigned to M or R likely belong to a derived branch of these two macrohaplogroups (e.g., M7 and R9) that occur in South-East Asia.

We used these as the starting frequencies for each mitochondrial macrohaplogroup in the initial sampling. Assuming the split time between the ancestral populations of Lapita and Unai groups was ~4000 BP and the generation interval was ~28 years per generation, we calculated the number of generations after separation. Note that the exact number of generations (since the split) used for the simulation is unlikely to matter, as a larger T would also translate to a larger Ne. We tried a small empirical experiment to test this assumption (see below). To address the fact that the Lapita individuals are dated to up to 3000 BP (*7–10*), and the Unai group individuals are ~2500 BP, we adjusted the date difference (~18 generations) in our simulated drift process.

We performed our simulations as follows. We allowed the ancestral effective population sizes of the Lapita and Unai populations to be different. By adopting the phylogeny shown in **fig. S16A**, the accumulations of genetic drift on branches ***a***, ***b***, and ***c*** can be measured using the autosomal differentiation between Iron Age Taiwan and Lapita Vanuatu / Tonga groups: $F_{ST} = 0.055 \pm 0.001$. The accumulation of genetic drift on branches ***a***, ***b***, and ***d*** can also be measured using the autosomal differentiation between Iron Age Taiwan and Unai Guam groups ($F_{ST} = 0.097 \pm 0.002$), and the accumulation of genetic drift on branches ***c*** and ***d*** can be measured using the autosomal differentiation between Lapita Vanuatu / Tonga and Unai Guam groups ($F_{ST} = 0.083 \pm 0.004$).

Since F_{ST} is not additive, we used a logarithmic scale for F_{ST} to represent the drift accumulations on branches (***a***, ***b***, ***c***, and ***d***), i.e., -ln(1-F_{ST}), which is equivalent to T / 2*Ne* according to the Wright-Fisher model. After resolving ***c*** and ***d*** using the equations:

$$\begin{cases} a + b + c = -ln\left(1 - F_{ST\_TaiwanIA\_vs\_Lapita}\right) \\ a + b + d = -ln\left(1 - F_{ST\_TaiwanIA\_vs\_LateUnai}\right) \\ c + d = -ln\left(1 - F_{ST\_Lapita\_vs\_LateUnai}\right) \end{cases}$$

we can resolve the genetic drift on each branch, and also make inferences about *Ne* if we assume split times.

By first assuming the split time between Lapita and Unai populations was ~4000 BP, we randomly sampled 36 generations to simulate the Lapita population and 54 generations to simulate the Unai population. We replicated the sampling process 1000000 times and then obtained haplogroup frequency distributions for two simulated populations as shown in **fig. S16B** and **C**. We then calculated an empirical probability that the two simulated populations had completely different mitochondrial macrohaplogroups.

Our primary interest was the probability that N = 7 Lapita and N = 9 Unai samples ended up with completely different macrohaplogroups (as designated by the first letter of the haplogroup name). For example, if in one simulation run, Lapita had all haplogroup R and Unai had all haplogroup F, this result would also contribute to the total observed probability. Thus, we divided the mitochondrial haplogroups present in the indigenous Iron Age Taiwan population into six different categories based on different letter type (B, D, E, F, M, and R). After 1000000 replicates, we obtained probabilities (*P*-values) that would allow us to answer our question. The exact value of the putative split time (e.g., 4000 BP, 5000 BP, 6000 BP, 7000 BP, and 8000 BP) and number of simulations (500,000; 1,000,000; 2,000,000) did not vary the probability very much (**table S28**). To investigate when the empirical probability would be larger than 5%, we assumed a series of *Ne* for Lapita and Unai populations, and then calculated the empirical *P* for each pair of *Ne*. We found that the empirical *P* would be larger than 5% only if both the Lapita group and the Unai group had relatively small effective population sizes of the maternal lineages (**fig. S16D**), pointing to matrilocal population genetic structure.

Mitochondrial haplogroup B is not very common in the indigenous Iron Age Taiwan individuals (only 13.3%). For this maternal lineage to have become so common in the Lapita individuals and widely dispersed across Oceania is extraordinary, and provides a particularly strong example of high matrilineal genetic drift.

These results support a scenario of accelerated maternal line drift compared to the expectation from the autosomes. The most reasonable scenario is that the ancestral populations of the Lapita and Unai individuals were matrilocal, with grooms moving to join the extended families of brides and females moving much less often between islands than men, a process that is well documented in some but not all Oceanic populations including those of Micronesia (*31*). This supports findings from ethnography and analysis of modern genetic data of matrilocal social practice (*34*, *35*, *126*), and provides the first direct evidence that it was the prevalent process in early Remote Oceanians.

We show an example of how we calculated the empirical *P*-values after receiving the simulation results below. As shown in **fig. S16B** and **C**, we obtained the following *P*-values for fixation of each mitochondrial haplogroup in the simulated $FRO_{Marianas}$ and $FRO_{SouthwestPacific}$ populations:

In simulated Unai Guam ($FRO_{Marianas}$) population:

M_P_B = 10014 / 1,000,000; M_P_D = 1905 / 1,000,000

M_P_E = 44432 / 1,000,000; M_P_F = 50082 / 1,000,000

M_P_M = 3844 / 1,000,000; M_P_R = 14307 / 1,000,000

In simulated Lapita ($FRO_{SouthwestPacific}$) population:

V_P_B = 589 / 1,000,000; V_P_D = 40 / 1,000,000

V_P_E = 6621 / 1,000,000; V_P_F = 8554 / 1,000,000

V_P_M = 112 / 1,000,000; V_P_R = 1032 / 1,000,000

The final *P* was calculated using this formula:

$$Final\ P = \sum_{i=B,D,E,F,M,R} M\_P_i \times \sum_{i=B,D,E,F,M,R} V\_P_i - \sum_{i=B,D,E,F,M,R} M\_P_i \times V\_P_i = 1.368 \times 10^{-3}$$

We note that two (I18501 and I18503) of 27 Unai individuals from the Naton Beach site are nominally noted as having B4a1a1 mtDNA haplogroups, which is identical to the haplogroup of the Lapita individuals, while another individual (I18450) is noted as having the R mtDNA haplogroup. However, all these individuals are among the individuals with evidence of mitochondrial DNA contamination. Although the nuclear DNA data from these individuals is consistent with being uncontaminated, nuclear-to-mitochondrial DNA ratios are highly variable across ancient DNA extracts (by orders of magnitude)—with the ratios typically being very high for petrous bones—and thus data can be contaminated in mitochondrial DNA and largely uncontaminated in nuclear DNA (and vice versa). For these three individuals, the upper bounds of

95% CI of the mtDNA matching rate to the consensus were lower than 82%, and the lower bounds were 50%-60%, suggesting the possibility of erroneous haplogroup determinations.

To minimize the bias of mtDNA haplogroup calling introduced by contamination, we also explored the mitochondrial haplogroup of these three individuals restricting to sequences with a damage signature indicative of genuine ancient DNA. For individuals I18450 and I18503, we could not obtain reliable haplogroup calls when restricting to the damaged reads because the mean coverage of their mtDNA is very low: ~0.30× and 1.08×, respectively. For individual I18501, we obtained a haplogroup call B4a1a1 with a mean coverage of ~2.50×. However, the 95% CI of the mtDNA matching rate to the consensus estimated to 63.1% to 81.3%, which is still questionable. A possible explanation for the outstanding high contamination estimates even after damage restriction is the relatively high level of damage rate of this individual (**table S2**). For individual I18501, we have merged two sequencing libraries together, in which one is the UDG-treated library with a damage rate of 0.350, and the other is the non-UDG treated library with a damage rate of 0.513. Thus, it is quite possible that the contamination is cross-contamination from an ancient DNA sample which cannot be corrected by damage-restriction. We simply do not have clean data showing non-E1/E2 haplogroups in the Unai individuals, and thus, we do not include these calls in our primary analyses of haplogroup distributions.

If there were non-E1/E2 haplogroups in the Unai individuals, this would of course reduce our estimate of maternal-line genetic drift between the $FRO_{SouthwestPacific}$ and the $FRO_{Marianas}$ lineage. However, even if these haplogroups were real the magnitude of maternal-line drift would still be large suggesting accelerated matrilocal population structure: for example, we still do not see any E1 or E2 haplogroups in Lapita individuals and B4a1a1 is still rare in Unai individuals.

S7. Estimates of Effective Population Size during the Latte Period

We identified ROH of over 4 centimorgan (cM) for 113 Latte individuals with high enough quality data (> 400,000 autosomal SNPs) to allow this inference (**table S30**). Only four individuals had more than 50 cM of their genome in blocks of ROH over 20 cM in length, indicating that Latte people avoided close kin unions. By contrast, 14 out of 67 Latte individuals from Guam (66 from Naton and one from Haputo) and eight out of 46 individuals from Saipan had at least one ROH over 20 cM, suggesting that mating pairs composed of close relatives such as the second or third cousins at both islands were common on both islands during the Latte Period. Shorter ROH segments (4-8 cM) were also abundant, implying a high level of background parental relatedness. We estimated *Ne* for the Latte population using the length distribution of all ROH in blocks of 4–20 cM. The *Ne* for populations in Guam and Saipan during the Latte Period were similar: 334 (95% CI of 315–356) in Guam, and 390 (95% CI of 361–424) in Saipan (**table S32**). We caution that effective population sizes are typically at least three times smaller than census sizes as they only reflect the reproductively active population (and thus do not reflect the presence of children and elderly individuals). They also may be biased low relative to the true census size if there has been a recent founder event, which is plausible for the Mariana Islands in the Latte Period. Nevertheless, these extremely small numbers suggest a highly restricted population.

We considered the possibility that, instead of generating the Mariana-wide population estimates, we were calculating *Ne* for isolated groups of people, and thus the actual *Ne* could be higher than the above estimates using ROH. We therefore analyzed long shared segments of IBD between the X chromosomes of male pairs across islands (one from Guam and the other from Saipan) to estimate *Ne*. We detected 149 pairs of individuals who shared IBD segments of at least 8 cM across the islands (**table S31**), revealing that the Latte people in Saipan and Guam shared common ancestors within hundreds of years before they came to live on those islands. A comparison between the Latte populations in Guam and Saipan gave an estimate of *Ne* = 1420 (95% CI of 1203–1712) (**table S32**). Since restricted migration reduces the rate of relatives and IBD sharing across islands, and the individuals were not all sampled from the same moment in time which means that we are effectively sampling a larger population (the combined population over multiple generations), the actual *Ne* of the joint Latte population across the Mariana Islands would be lower than the estimates based on IBD blocks. Thus, based on the above estimates, the *Ne* of the Mariana Islands during the Latte Period was between 356-1298 individuals.

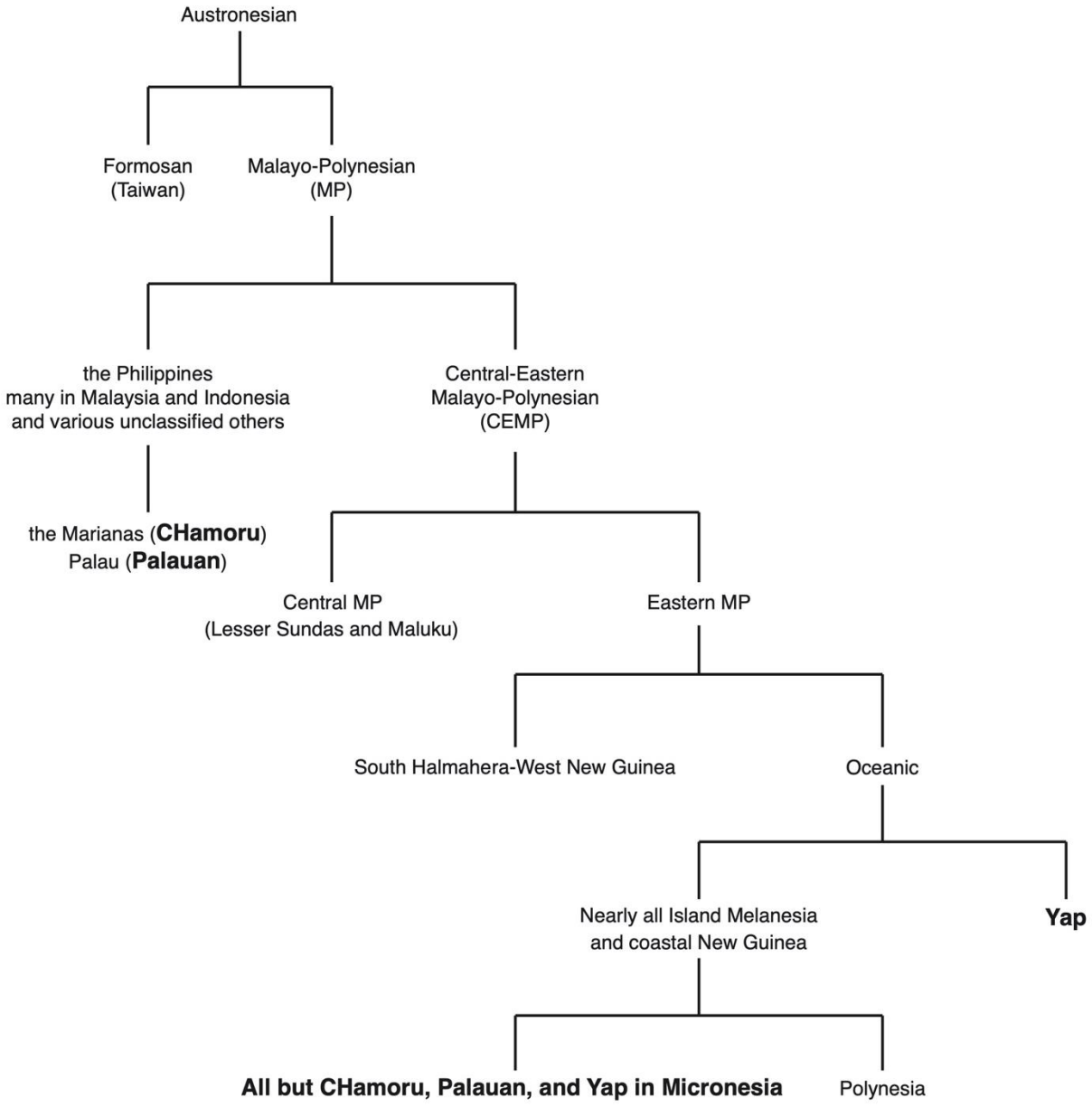


Fig. S1. Schematic of Austronesian Linguistic Groups in Micronesia. Information is based on the reference (*127*), and other literature cited in the text.

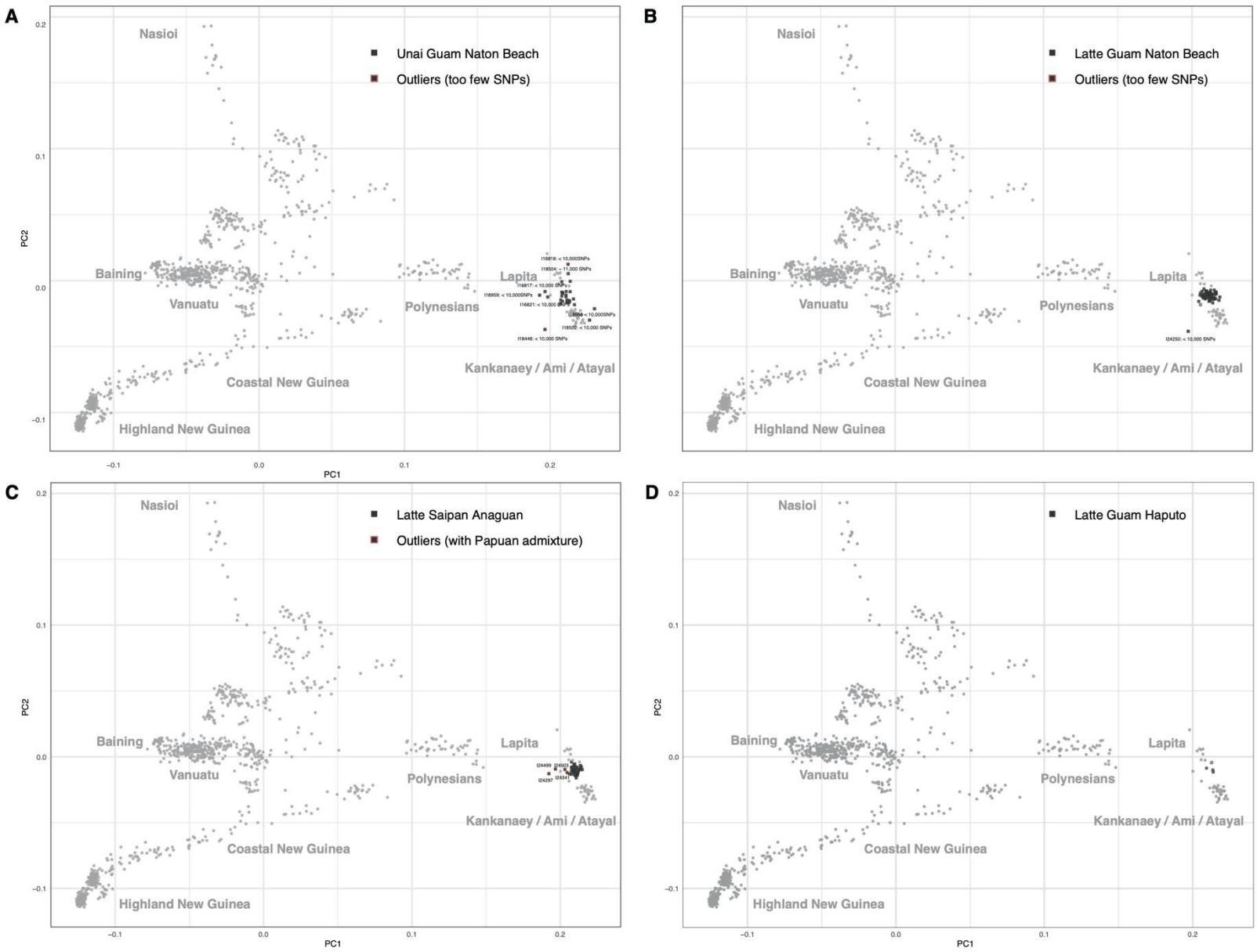

A
Nasioi
Unai Guam Naton Beach
Outliers (too few SNPs)
Baining
Vanuatu
Lapita
Polynesians
Kankanaey / Ami / Atayal
Coastal New Guinea
Highland New Guinea
PC1
PC2
B
Nasioi
Latte Guam Naton Beach
Outliers (too few SNPs)
Baining
Vanuatu
Lapita
Polynesians
Kankanaey / Ami / Atayal
Coastal New Guinea
Highland New Guinea
C
Nasioi
Latte Saipan Anaguan
Outliers (with Papuan admixture)
Baining
Vanuatu
Lapita
Polynesians
Kankanaey / Ami / Atayal
Coastal New Guinea
Highland New Guinea
PC1
PC2
D
Nasioi
Latte Guam Haputo
Baining
Vanuatu
Lapita
Polynesians
Kankanaey / Ami / Atayal
Coastal New Guinea
Highland New Guinea
PC1
PC2

Fig. S2. Principal Component Analysis of Unai and Latte Individuals from Guam and Saipan. Axes were computed using whole-genome shotgun data of three present-day populations—Dai (East Asians from Southern China), Nasioi (Solomon Islands; top), and Papuans (from eastern Highland and middle Sepik area of New Guinea; bottom left). Other present-day and prehistoric individuals shown in the figure were projected onto the first two PCs. (A) Unai individuals from the Naton Beach site on Guam; (B) Latte individuals from the Naton Beach site on Guam; (C) Latte individuals from the Anaguan site in Garapan, Saipan; and (D) Latte individuals from the Haputo Beach site on Guam. Outliers for each group are labeled with a red border in each panel, and are mostly low coverage suggesting that their outlier status is due to statistical noise or contamination which affects lower coverage samples more easily.

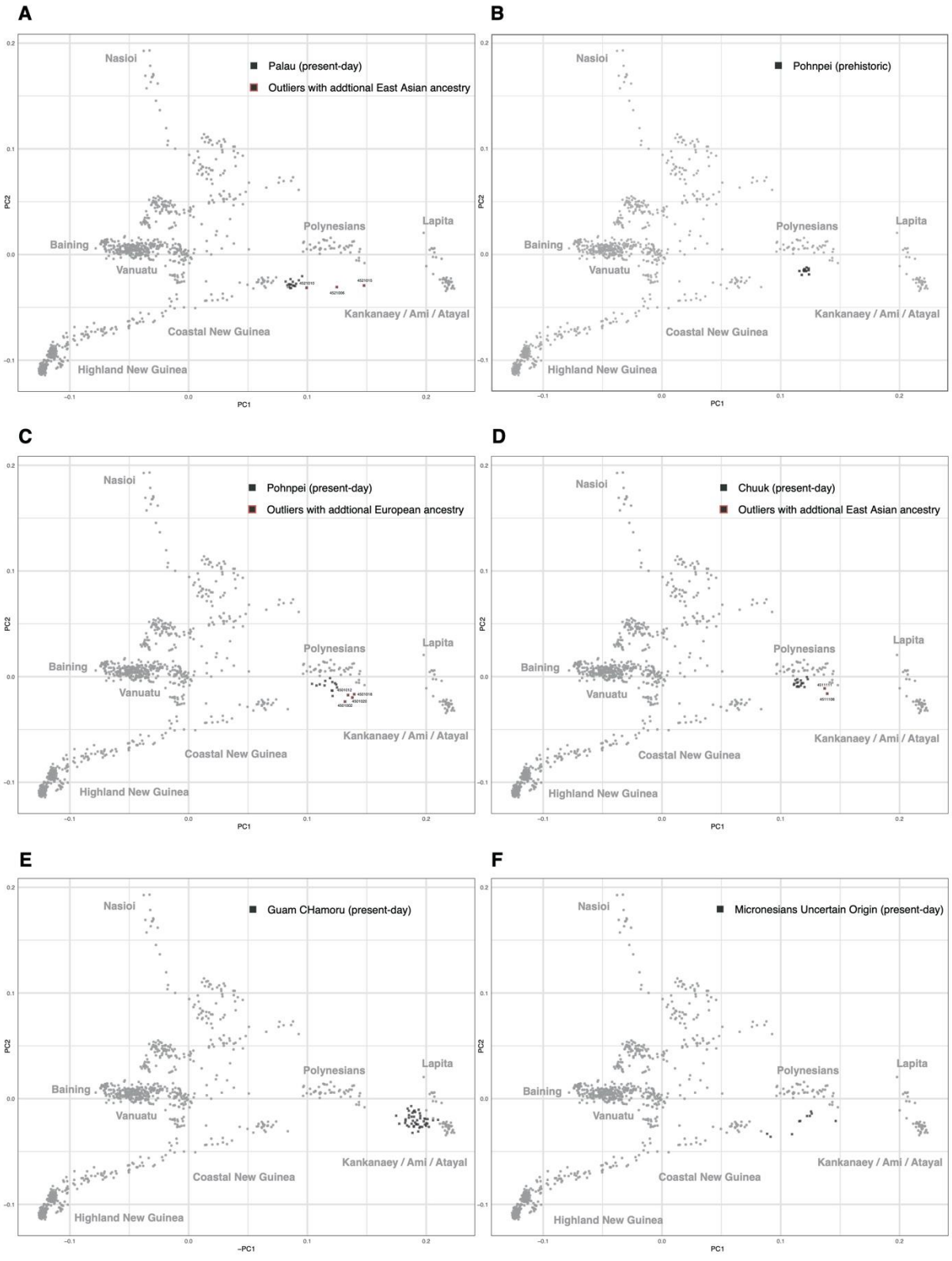

A
Palau (present-day)
Outliers with addtional East Asian ancestry
B
Pohnpei (prehistoric)
C
Pohnpei (present-day)
Outliers with addtional European ancestry
D
Chuuk (present-day)
Outliers with addtional East Asian ancestry
E
Guam CHamoru (present-day)
F
Micronesians Uncertain Origin (present-day)
Nasioi
Baining
Vanuatu
Polynesians
Lapita
Kankanaey / Ami / Atayal
Coastal New Guinea
Highland New Guinea
PC1
PC2

Fig. S3. Principal Component Analysis of Late Prehistoric Pohnpei and Present-day Micronesians. Axes were computed using whole-genome shotgun data of three present-day populations: Dai (East Asians from Southern China), Nasioi (Solomon Islands; top), and Papuans (from eastern Highland and middle Sepik area of New Guinea; bottom left). Other present-day and prehistoric individuals shown in the figure were projected onto the first two PCs. (A) present-day individuals from Palau; (B) late prehistoric individuals from Pohnpei; (C) present-day individuals from Pohnpei; (D) present-day individuals from Chuuk; (E) present-day individuals from Guam (CHamoru); and (F) other present-day Micronesian Islanders (geographic origin unknown). Outliers for each group are labeled with a red border in each panel.

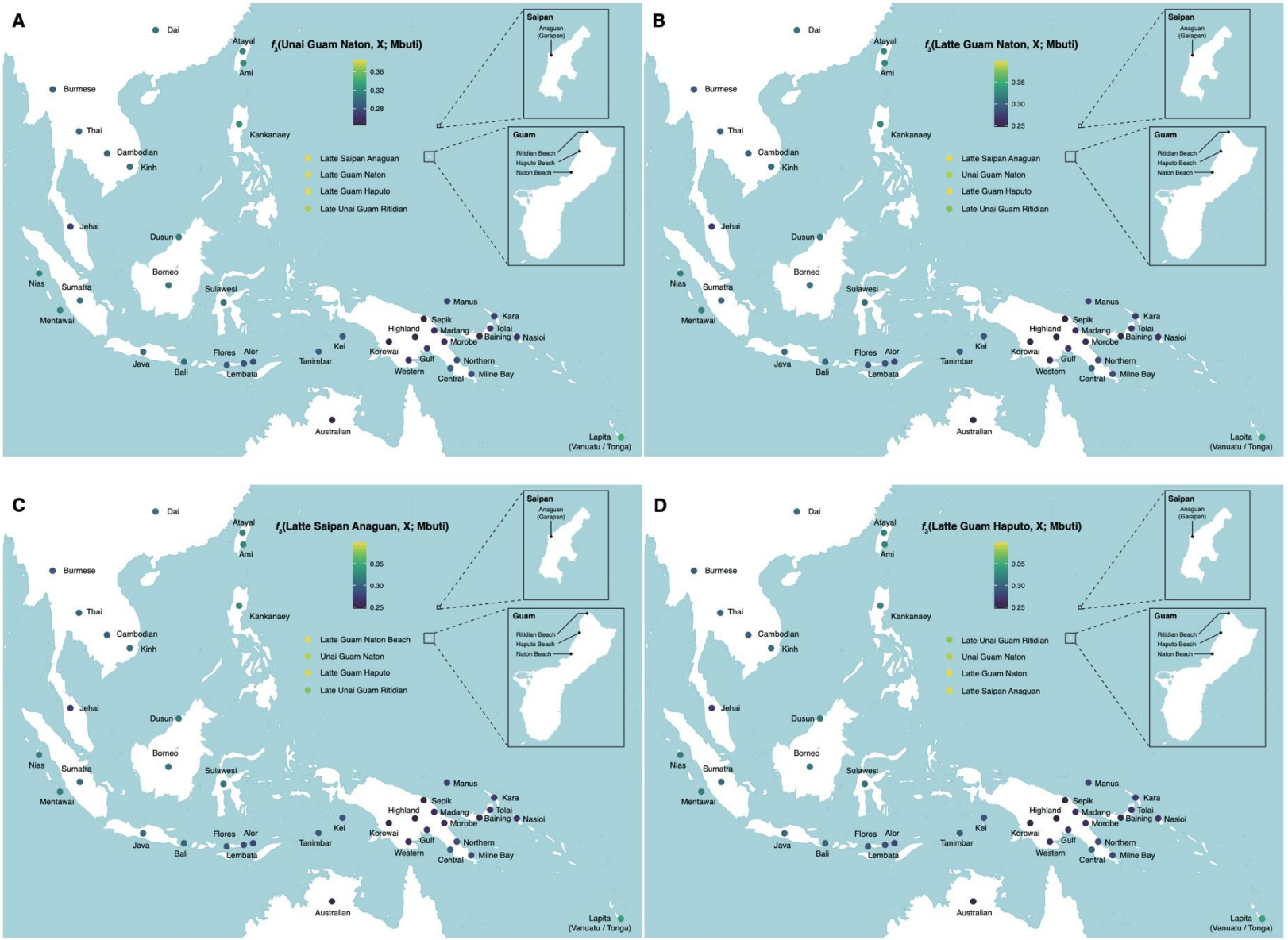

A
f_3(Unai Guam Naton, X; Mbuti)
0.36
0.32
0.28
Latte Saipan Anaguan
Latte Guam Naton
Latte Guam Haputo
Late Unai Guam Ritidian
Saipan
Anaguan (Garapan)
Guam
Ritidian Beach
Haputo Beach
Naton Beach
Dai
Atayal
Ami
Burmese
Thai
Cambodian
Kinh
Kankanaey
Jehai
Dusun
Borneo
Nias
Sumatra
Mentawai
Sulawesi
Java
Bali
Flores
Alor
Lembata
Tanimbar
Kei
Manus
Kara
Sepik
Tolai
Highland
Madang
Baining
Nasioi
Morobe
Korowai
Gulf
Northern
Western
Central
Milne Bay
Australian
Lapita (Vanuatu / Tonga)
B
f_3(Latte Guam Naton, X; Mbuti)
0.35
0.30
0.25
Latte Saipan Anaguan
Unai Guam Naton
Latte Guam Haputo
Late Unai Guam Ritidian
C
f_3(Latte Saipan Anaguan, X; Mbuti)
0.35
0.30
0.25
Latte Guam Naton Beach
Unai Guam Naton
Latte Guam Haputo
Late Unai Guam Ritidian
D
f_3(Latte Guam Haputo, X; Mbuti)
0.35
0.30
0.25
Late Unai Guam Ritidian
Unai Guam Naton
Latte Guam Naton
Latte Saipan Anaguan

Fig. S4. Genetic Affinities Estimated by Outgroup f_3-statistics. Genetic affinities between test populations and populations (X) from Southeast Asia and Oceania were assessed by f_3(Test, X; Mbuti). Higher values of the statistics indicate greater genetic affinities between a test population and population X (highest in yellow; lowest in blue). In each panel, test populations are (A) Unai individuals from the Naton Beach site on Guam, (B) Latte individuals from the Naton Beach site on Guam, (C) Latte individuals from the Anaguan site in Garapan, Saipan, and (D) Latte individuals from the Haputo Beach site on Guam.

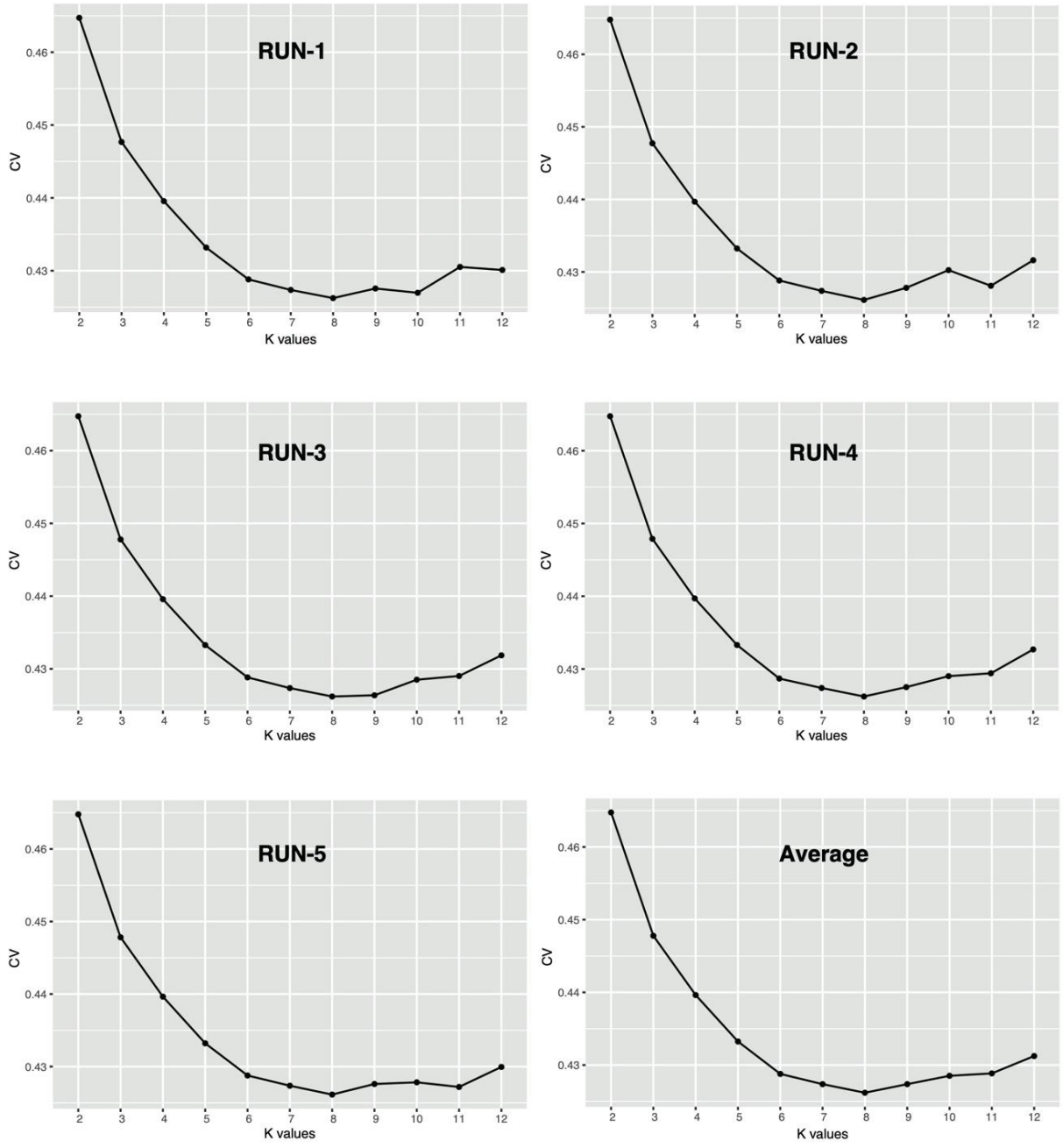


Fig. S5. Relationship between CV Errors and Pre-Defined K in ADMIXTURE. We conducted five independent runs. In each run, the number of genetic clusters (K) was predefined from 2 to 12, and we implemented 10-fold cross-validation (CV) for each K. Lower CV of a K value indicates a better fit for the data.

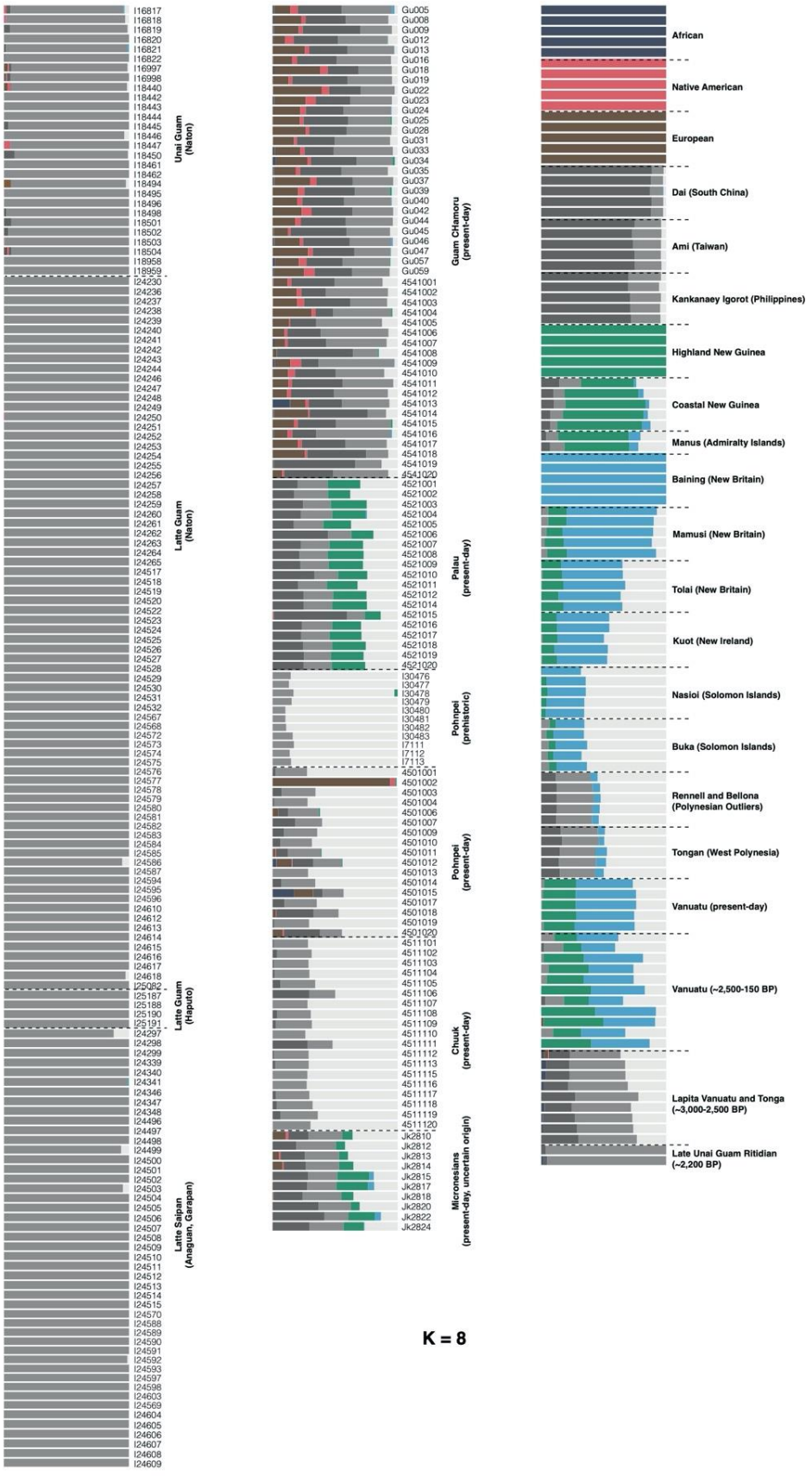

I16817
I16818
I16819
I16820
I16821
I16822
I16997
I16998
I18440
I18442
I18443
I18444
I18445
I18446
I18447
I18450
I18461
I18462
I18494
I18495
I18496
I18498
I18501
I18502
I18503
I18504
I18958
I18959
Unai Guam (Naton)
I24230
I24236
I24237
I24238
I24239
I24240
I24241
I24242
I24243
I24244
I24246
I24247
I24248
I24249
I24250
I24251
I24252
I24253
I24254
I24255
I24256
I24257
I24258
I24259
I24260
I24261
I24262
I24263
I24264
I24265
I24517
I24518
I24519
I24520
I24522
I24523
I24524
I24525
I24526
I24527
I24528
I24529
I24530
I24531
I24532
I24567
I24568
I24572
I24573
I24574
I24575
I24576
I24577
I24578
I24579
I24580
I24581
I24582
I24583
I24584
I24585
I24586
I24587
I24594
I24595
I24596
I24610
I24612
I24613
I24614
I24615
I24616
I24617
I24618
Latte Guam (Naton)
I25082
I25187
I25188
I25190
I25191
Latte Guam (Haputo)
I24297
I24298
I24299
I24339
I24340
I24341
I24346
I24347
I24348
I24496
I24497
I24498
I24499
I24500
I24501
I24502
I24503
I24504
I24505
I24506
I24507
I24508
I24509
I24510
I24511
I24512
I24513
I24514
I24515
I24570
I24588
I24589
I24590
I24591
I24592
I24593
I24597
I24598
I24603
I24569
I24604
I24605
I24606
I24607
I24608
I24609
Latte Saipan (Anaguan, Garapan)
Gu005
Gu008
Gu009
Gu012
Gu013
Gu016
Gu018
Gu019
Gu022
Gu023
Gu024
Gu025
Gu028
Gu031
Gu033
Gu034
Gu035
Gu037
Gu039
Gu040
Gu042
Gu044
Gu045
Gu046
Gu047
Gu057
Gu059
Guam CHamoru (present-day)
4541001
4541002
4541003
4541004
4541005
4541006
4541007
4541008
4541009
4541010
4541011
4541012
4541013
4541014
4541015
4541016
4541017
4541018
4541019
4541020
4521001
4521002
4521003
4521004
4521005
4521006
4521007
4521008
4521009
4521010
4521011
4521012
4521014
4521015
4521016
4521017
4521018
4521019
4521020
Palau (present-day)
I30476
I30477
I30478
I30479
I30480
I30481
I30482
I30483
I7111
I7112
I7113
Pohnpei (prehistoric)
4501001
4501002
4501003
4501004
4501006
4501007
4501009
4501010
4501011
4501012
4501013
4501014
4501015
4501017
4501018
4501019
4501020
Pohnpei (present-day)
4511101
4511102
4511103
4511104
4511105
4511106
4511107
4511108
4511109
4511110
4511111
4511112
4511113
4511115
4511116
4511117
4511118
4511119
4511120
Chuuk (present-day)
Jk2810
Jk2812
Jk2813
Jk2814
Jk2815
Jk2817
Jk2818
Jk2820
Jk2822
Jk2824
Micronesians (present-day, uncertain origin)
African
Native American
European
Dai (South China)
Ami (Taiwan)
Kankanaey Igorot (Philippines)
Highland New Guinea
Coastal New Guinea
Manus (Admiralty Islands)
Baining (New Britain)
Mamusi (New Britain)
Tolai (New Britain)
Kuot (New Ireland)
Nasioi (Solomon Islands)
Buka (Solomon Islands)
Rennell and Bellona (Polynesian Outliers)
Tongan (West Polynesia)
Vanuatu (present-day)
Vanuatu (~2,500-150 BP)
Lapita Vanuatu and Tonga (~3,000-2,500 BP)
Late Unai Guam Ritidian (~2,200 BP)
K = 8

Fig. S6. Unsupervised ADMIXTURE Plot (K = 8). All newly reported prehistoric and present-day individuals together with a subset of representative modern and prehistoric populations are shown. Individuals are represented by a thin horizontal bar and their ancestry is partitioned into eight colored segments (K = 8), representing their affiliation with each cluster. There are four major ancestry components inferred for Oceanians in this analysis: two Papuan components are maximized in Papuan New Guinea Highlanders (green) and Baining speakers in New Britain (blue), respectively; one East Asian-related component is maximized in Unai and Latte individuals from Guam and Saipan (middle gray); and the other ancestry (light gray) is widely seen in Oceania.

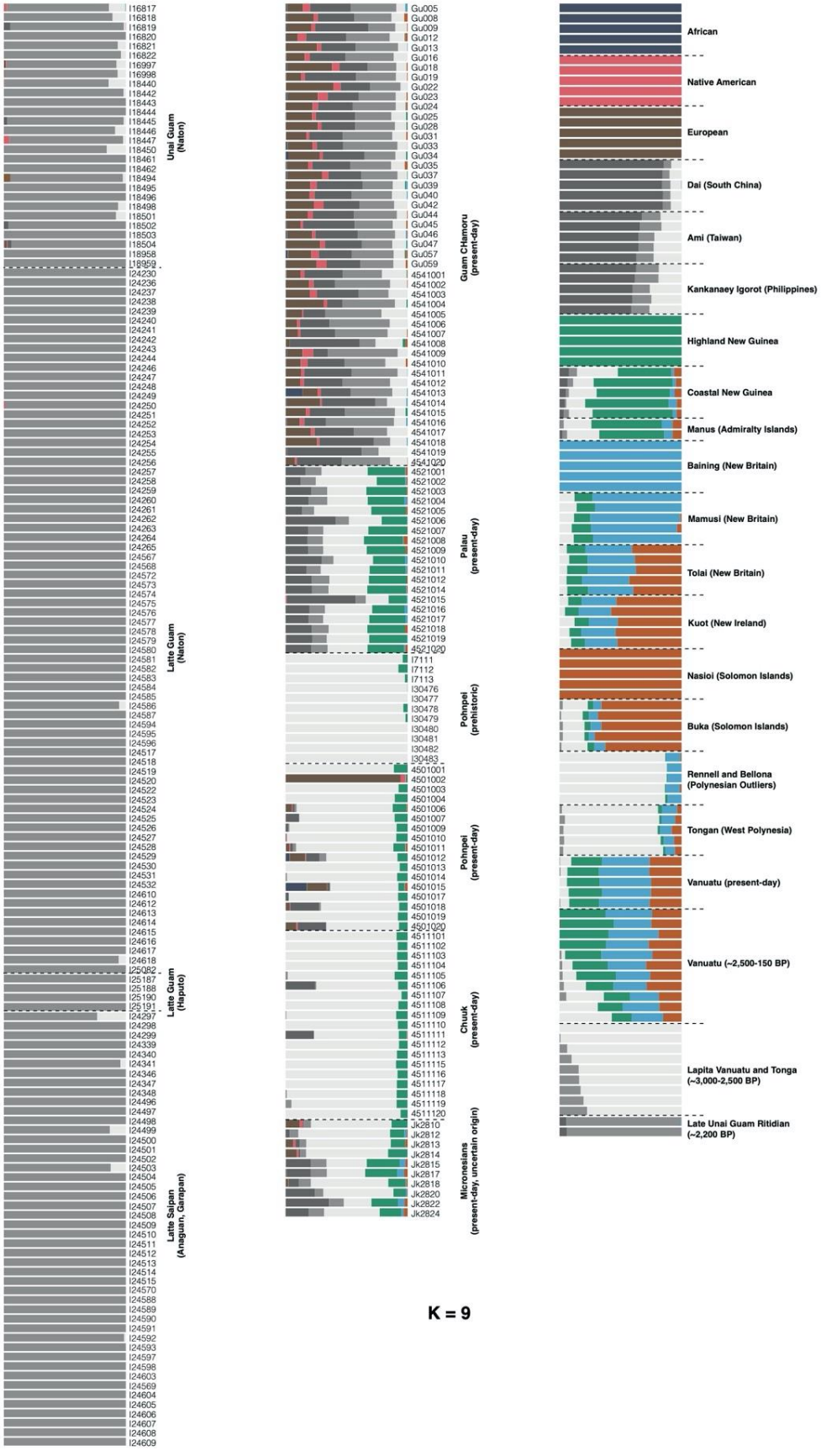
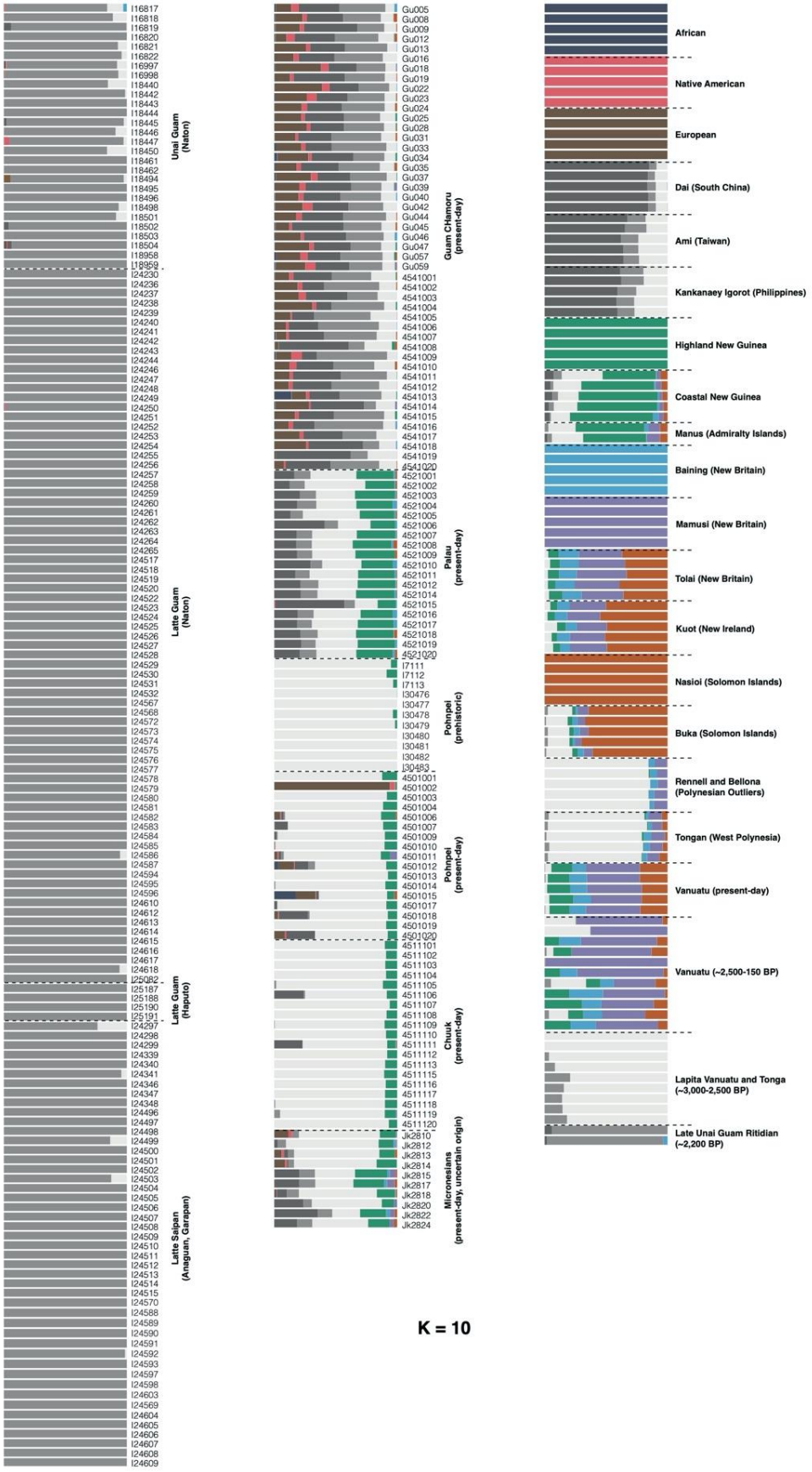

Unai Guam (Naton)
Latte Guam (Naton)
Latte Guam (Hapuhó)
Latte Saipan (Aniaguan, Garapan)
Guam CHamoru (present-day)
Palau (present-day)
Pohnpei (prehistoric)
Pohnpei (present-day)
Chuuk (present-day)
Micronesians (present-day, uncertain origin)
African
Native American
European
Dai (South China)
Ami (Taiwan)
Kankanaey Igorot (Philippines)
Highland New Guinea
Coastal New Guinea
Manus (Admiralty Islands)
Baining (New Britain)
Mamusi (New Britain)
Tolai (New Britain)
Kuot (New Ireland)
Nasioi (Solomon Islands)
Buka (Solomon Islands)
Rennell and Bellona (Polynesian Outliers)
Tongan (West Polynesia)
Vanuatu (present-day)
Vanuatu (~2,500-150 BP)
Lapita Vanuatu and Tonga (~3,000-2,500 BP)
Late Unai Guam Ritidian (~2,200 BP)
K = 9

Fig. S7. Unsupervised ADMIXTURE Plot (K = 9). All newly reported prehistoric and present-day individuals together with a subset of representative modern and prehistoric populations are shown in the figure. Individuals are represented by a thin horizontal bar and their ancestry partitioned into nine colored segments (K = 9), representing their affiliation with each cluster. The results suggested three primary Papuan components: one maximized in New Guinea (green); one maximized in New Britain (blue); and one maximized in the Solomon Islands (orange). Two other components were maximized in the Lapita individuals from Vanuatu and Tonga, as well as the Unai and the Latte individuals from Guam and Saipan, consistent with $FRO_{SouthwestPacific}$ (light gray) and $FRO_{Marianas}$ (dark gray) ancestries.


I16817
I16818
I16819
I16820
I16821
I16822
I16997
I16998
I18440
I18442
I18443
I18444
I18445
I18446
I18447
I18450
I18461
I18462
I18494
I18495
I18496
I18498
I18501
I18502
I18503
I18504
I18958
I18959
Unai Guam (Haton)
I24230
I24236
I24237
I24238
I24239
I24240
I24241
I24242
I24243
I24244
I24246
I24247
I24248
I24249
I24250
I24251
I24252
I24253
I24254
I24255
I24256
I24257
I24258
I24259
I24260
I24261
I24262
I24263
I24264
I24265
I24517
I24518
I24519
I24520
I24522
I24523
I24524
I24525
I24526
I24527
I24528
I24529
I24530
I24531
I24532
I24567
I24568
I24572
I24573
I24574
I24575
I24576
I24577
I24578
I24579
I24580
I24581
I24582
I24583
I24584
I24585
I24586
I24587
I24594
I24595
I24596
I24610
I24612
I24613
I24614
I24615
I24616
I24617
I24618
Latte Guam (Haputo)
I25082
I25187
I25188
I25190
I25191
Latte Saipan (Amaguan, Garapan)
I24297
I24298
I24299
I24339
I24340
I24341
I24346
I24347
I24348
I24496
I24497
I24498
I24499
I24500
I24501
I24502
I24503
I24504
I24505
I24506
I24507
I24508
I24509
I24510
I24511
I24512
I24513
I24514
I24515
I24570
I24588
I24589
I24590
I24591
I24592
I24593
I24597
I24598
I24603
I24569
I24604
I24605
I24606
I24607
I24608
I24609
Gu005
Gu008
Gu009
Gu012
Gu013
Gu016
Gu018
Gu019
Gu022
Gu023
Gu024
Gu025
Gu028
Gu031
Gu033
Gu034
Gu035
Gu037
Gu039
Gu040
Gu042
Gu044
Gu045
Gu046
Gu047
Gu057
Gu059
Guam CHamoru (present-day)
4541001
4541002
4541003
4541004
4541005
4541006
4541007
4541008
4541009
4541010
4541011
4541012
4541013
4541014
4541015
4541016
4541017
4541018
4541019
4541020
4521001
4521002
4521003
4521004
4521005
4521006
4521007
4521008
4521009
4521010
4521011
4521012
4521014
4521015
4521016
4521017
4521018
4521019
4521020
Palau (present-day)
I7111
I7112
I7113
I30476
I30477
I30478
I30479
I30480
I30481
I30482
I30483
Pohnpei (prehistoric)
4501001
4501002
4501003
4501004
4501006
4501007
4501009
4501010
4501011
4501012
4501013
4501014
4501015
4501017
4501018
4501019
4501020
Pohnpei (present-day)
4511101
4511102
4511103
4511104
4511105
4511106
4511107
4511108
4511109
4511110
4511111
4511112
4511113
4511115
4511116
4511117
4511118
4511119
4511120
Chuuk (present-day)
Jk2810
Jk2812
Jk2813
Jk2814
Jk2815
Jk2817
Jk2818
Jk2820
Jk2822
Jk2824
Micronesians (present-day, uncertain origin)
African
Native American
European
Dai (South China)
Ami (Taiwan)
Kankanaey Igorot (Philippines)
Highland New Guinea
Coastal New Guinea
Manus (Admiralty Islands)
Baining (New Britain)
Mamusi (New Britain)
Tolai (New Britain)
Kuot (New Ireland)
Nasioi (Solomon Islands)
Buka (Solomon Islands)
Rennell and Bellona (Polynesian Outliers)
Tongan (West Polynesia)
Vanuatu (present-day)
Vanuatu (~2,500-150 BP)
Lapita Vanuatu and Tonga (~3,000-2,500 BP)
Late Unai Guam Ritidian (~2,200 BP)

K = 10

Fig. S8. Unsupervised ADMIXTURE Plot (K = 10). All newly reported prehistoric and present-day individuals together with a subset of representative modern and prehistoric populations are shown. Individuals are represented by a thin horizontal bar and their ancestry partitioned into ten colored segments (K = 10), representing their affiliation with each cluster. The division of genetic clusters is the same with K = 9, except for an additional Papuan cluster that is identified and maximized in Mamusi speakers in New Britain (purple). However, K=10 consistently yields high CV indicating that it might not be a good fit.

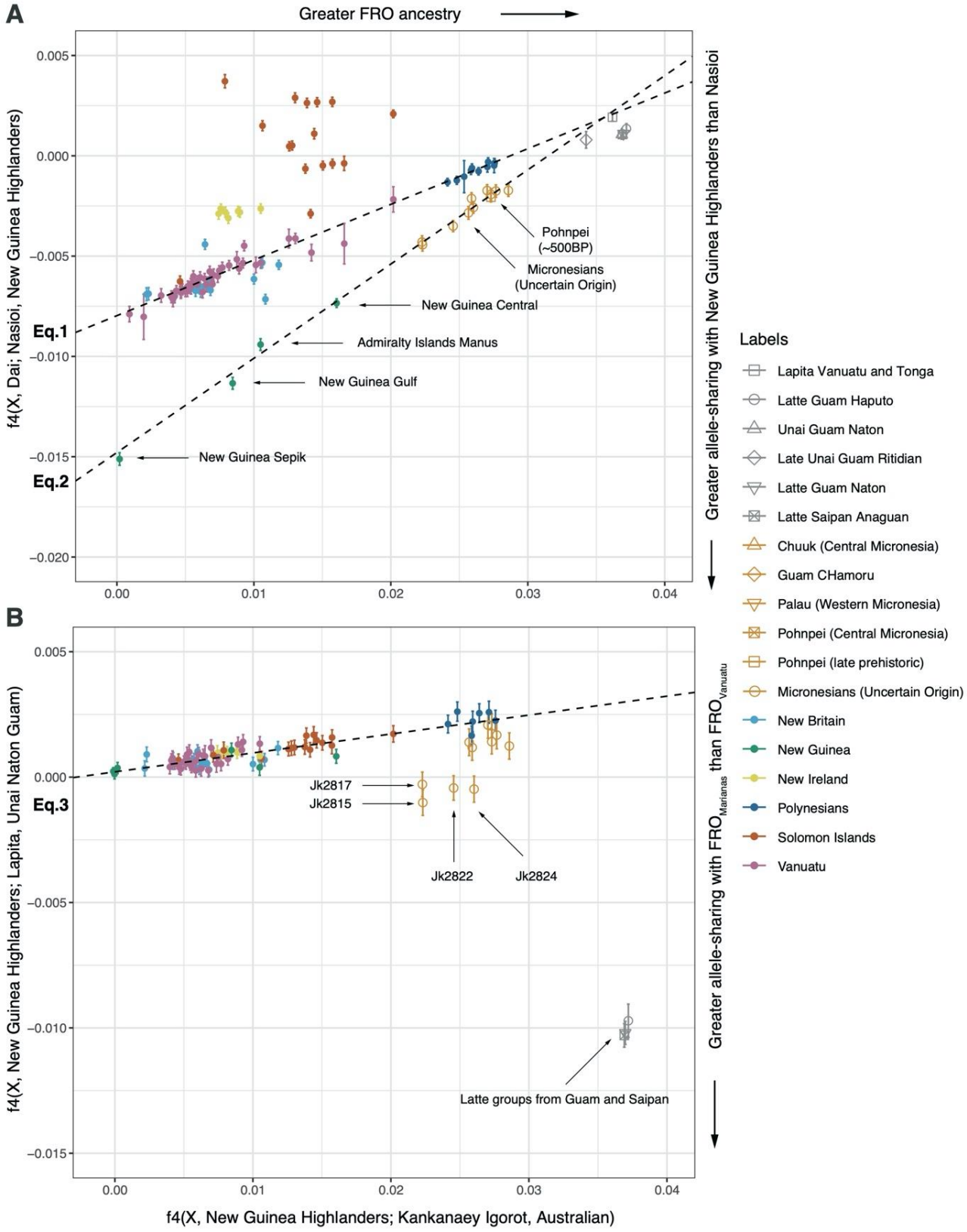

A
Greater FRO ancestry
f4(X, Dai; Nasioi, New Guinea Highlanders)
0.005
0.000
−0.005
−0.010
−0.015
−0.020
0.00
0.01
0.02
0.03
0.04
Eq.1
Eq.2
Pohnpei
(~500BP)
Micronesians
(Uncertain Origin)
New Guinea Central
Admiralty Islands Manus
New Guinea Gulf
New Guinea Sepik
Greater allele-sharing with New Guinea Highlanders than Nasioi
B
f4(X, New Guinea Highlanders; Lapita, Unai Naton Guam)
0.005
0.000
−0.005
−0.010
−0.015
Eq.3
Jk2817
Jk2815
Jk2822
Jk2824
Latte groups from Guam and Saipan
Greater allele-sharing with FRO Marianas than FRO Vanuatu
f4(X, New Guinea Highlanders; Kankanaey Igorot, Australian)
Labels
Lapita Vanuatu and Tonga
Latte Guam Haputo
Unai Guam Naton
Late Unai Guam Ritidian
Latte Guam Naton
Latte Saipan Anaguan
Chuuk (Central Micronesia)
Guam CHamoru
Palau (Western Micronesia)
Pohnpei (Central Micronesia)
Pohnpei (late prehistoric)
Micronesians (Uncertain Origin)
New Britain
New Guinea
New Ireland
Polynesians
Solomon Islands
Vanuatu

Fig. S9. f_4-Regression Analysis Partitioning FRO and Papuan Sources. In all panels, the statistic f_4(X, New Guinea Highlanders; Kankanaey Igorot, Australian) was calculated as the independent variable (x-axis, higher values indicate more East Asian ancestry) to correct for different ancestry proportions. (A) The statistic f_4(X, Dai; Nasioi, New Guinea Highlanders) (y-axis; higher values indicate greater allele-sharing with Nasioi-speaking people from the Bougainville in the Solomon Islands, and lower values indicate greater allele-sharing with New Guinea Highlanders) was plotted against the independent variable. Equation (Eq.) 1 was constructed using all prehistoric and present-day populations from Vanuatu and Polynesia, except for the Lapita groups from Vanuatu and Tonga. Eq.2 was constructed using all prehistoric and present-day populations from Micronesia and New Guinea, except for the Unai and the Latte groups from Saipan and Guam. (B) The statistic f_4(X, New Guinea Highlanders; Lapita, Unai Naton Guam) (y-axis; higher values indicate greater allele-sharing with Lapita, and lower values indicate greater allele-sharing with Unai Naton Guam) was plotted against the independent variable. Eq.3 was constructed using all present-day populations except those from Micronesia. The statistics were computed using the Human Origins dataset (~397,000 autosomal SNPs). Error bars show one standard error in each direction along the y-axis.

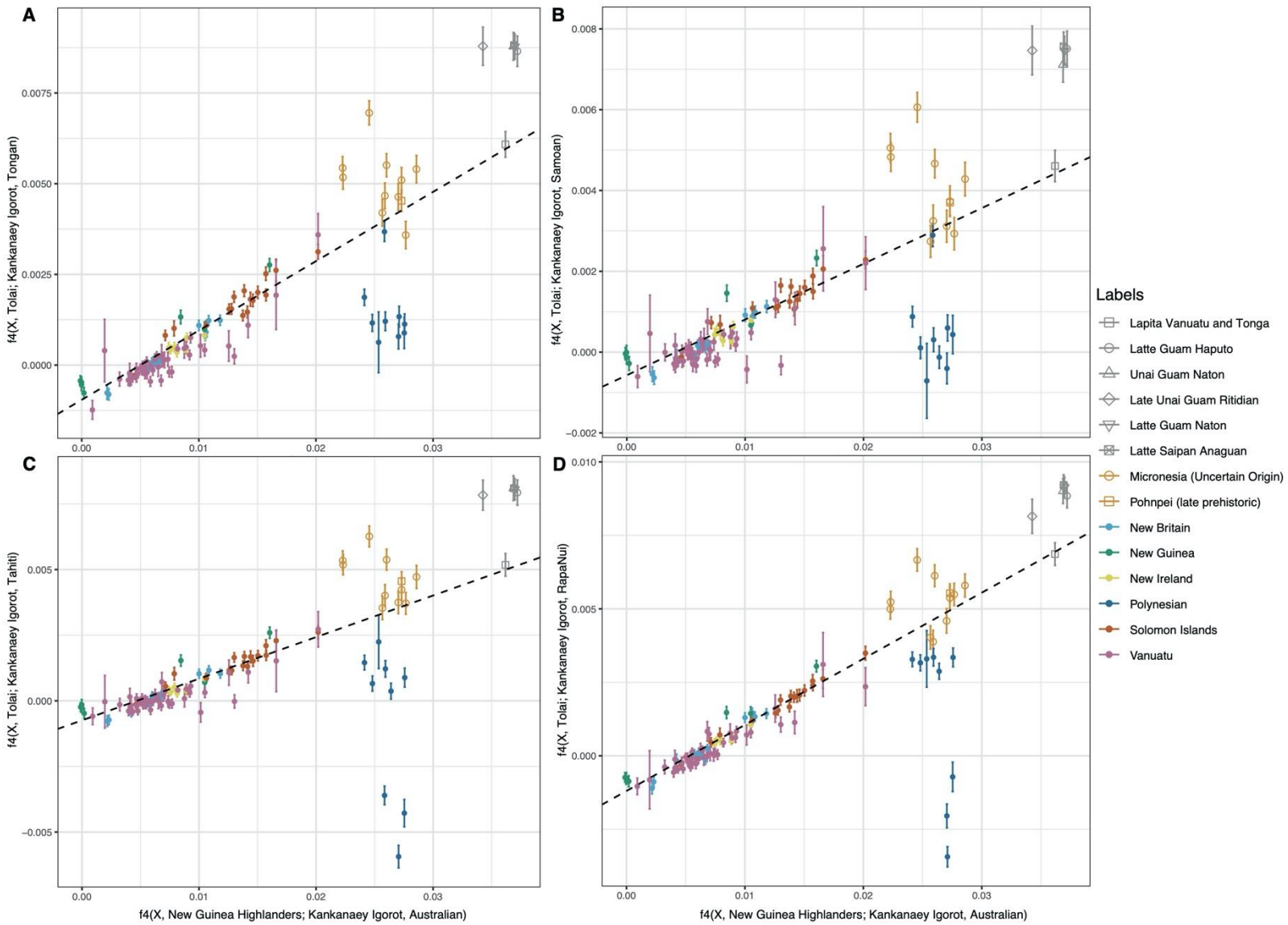

A
B
C
D
f4(X, Tolai; Kankanaey Igorot, Tongan)
f4(X, Tolai; Kankanaey Igorot, Samoan)
f4(X, Tolai; Kankanaey Igorot, Tahiti)
f4(X, Tolai; Kankanaey Igorot, RapaNui)
f4(X, New Guinea Highlanders; Kankanaey Igorot, Australian)
f4(X, New Guinea Highlanders; Kankanaey Igorot, Australian)
Labels
Lapita Vanuatu and Tonga
Latte Guam Haputo
Unai Guam Naton
Late Unai Guam Ritidian
Latte Guam Naton
Latte Saipan Anaguan
Micronesia (Uncertain Origin)
Pohnpei (late prehistoric)
New Britain
New Guinea
New Ireland
Polynesian
Solomon Islands
Vanuatu

Fig. S10. f_4-Regression Analysis Testing for Polynesian-associated ancestry (Based on Affinity to Polynesian populations). In all panels, the statistic f_4(X, New Guinea Highlanders; Kankanaey Igorot, Australian) was calculated as the independent variable (x-axis, higher values indicate more East Asian ancestry) to correct for different ancestry proportions. (A) The statistic f_4(X, Tolai; Kankanaey Igorot, Tongan) (y-axis; higher values indicate greater allele-sharing with Kankanaey Igorot, and lower values indicate greater allele-sharing with Tongan) was plotted against the independent variable. (B) The statistic f_4(X, Tolai; Kankanaey Igorot, Samoan) (y-axis; higher values indicate greater allele-sharing with Kankanaey Igorot, and lower values indicate greater allele-sharing with Samoan) was plotted against the independent variable. (C) The statistic f_4(X, Tolai; Kankanaey Igorot, Tahiti) (y-axis; higher values indicate greater allele-sharing with Kankanaey, and lower values indicate greater allele-sharing with Tahiti) was plotted against the independent variable. (D) The statistic f_4(X, Tolai; Kankanaey Igorot, Rapa Nui) (y-axis; higher values indicate greater allele-sharing with Kankanaey Igorot, and lower values indicate greater allele-sharing with Rapa Nui) was plotted against the independent variable. All of the regression lines were computed using present-day populations from Near Oceania (groups from New Guinea, New Britain, New Ireland, and Solomon Islands). All statistics were computed using the HO dataset (~397,000 autosomal SNPs). Error bars show one standard error in each direction along y-axis.

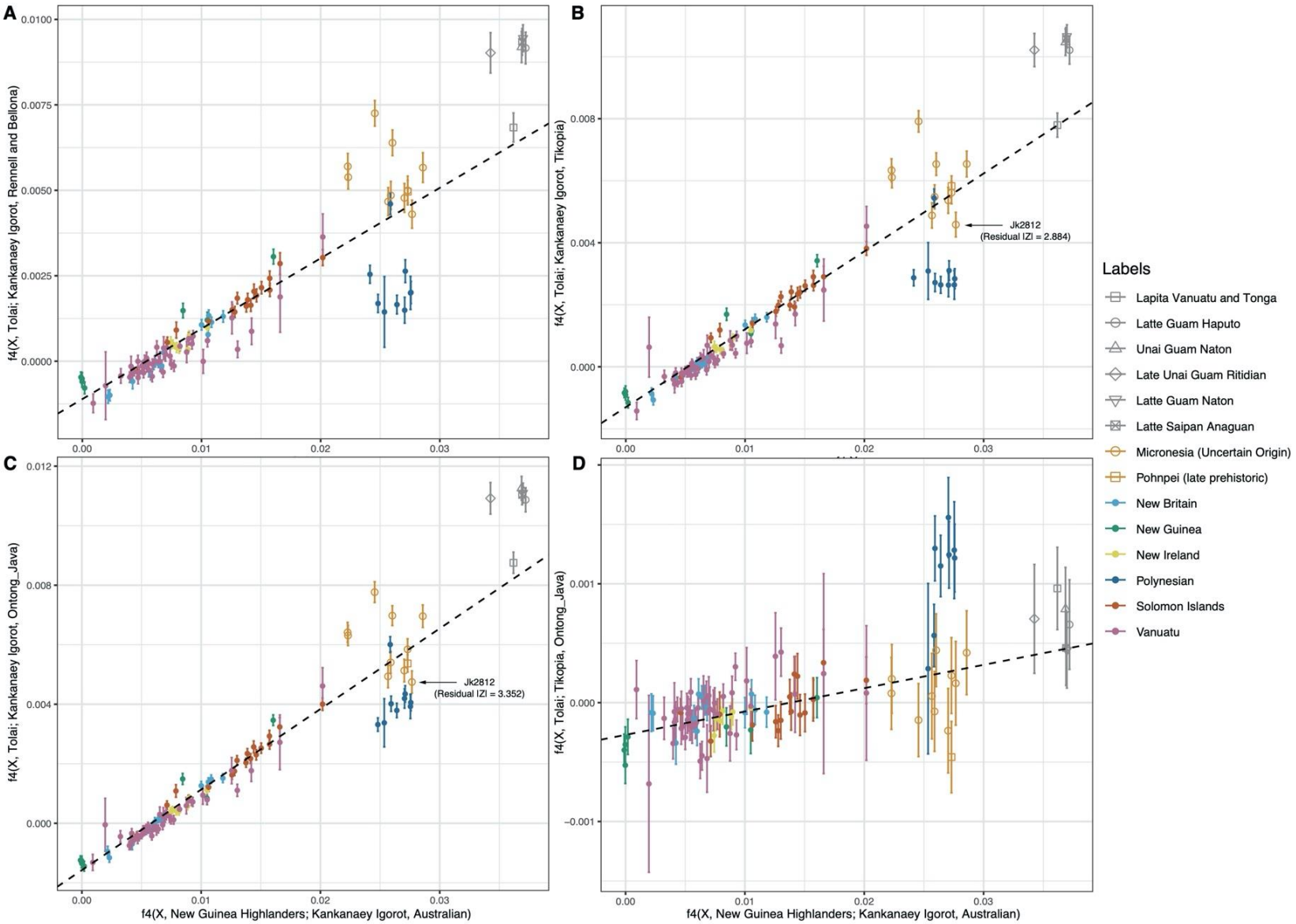

A
B
C
D
f4(X, Tolai; Kankanaey Igorot, Rennell and Bellona)
f4(X, Tolai; Kankanaey Igorot, Tikopia)
f4(X, Tolai; Kankanaey Igorot, Ontong_Java)
f4(X, Tolai; Tikopia, Ontong_Java)
f4(X, New Guinea Highlanders; Kankanaey Igorot, Australian)
Jk2812
(Residual IZI = 2.884)
Jk2812
(Residual IZI = 3.352)
Labels
Lapita Vanuatu and Tonga
Latte Guam Haputo
Unai Guam Naton
Late Unai Guam Ritidian
Latte Guam Naton
Latte Saipan Anaguan
Micronesia (Uncertain Origin)
Pohnpei (late prehistoric)
New Britain
New Guinea
New Ireland
Polynesian
Solomon Islands
Vanuatu

Fig. S11. f_4-Regression Testing for Polynesian-associated ancestry (Based on Affinity to Polynesia Outlier Populations). In all panels, the statistic f_4(X, New Guinea Highlanders; Kankanaey Igorot, Australian) was calculated as the independent variable (x-axis, higher values indicate more East Asian ancestry) to correct for different ancestry proportions. (A) The statistic f_4(X, Tolai; Kankanaey Igorot, Rennell and Bellona) (y-axis; higher values indicate greater allele-sharing with Kankanaey Igorot, and lower values indicate greater allele-sharing with Rennell and Bellona) was plotted against the independent variable. (B) The statistic f_4(X, Tolai; Kankanaey Igorot, Tikopia) (y-axis; higher values indicate greater allele-sharing with Kankanaey Igorot, and lower values indicate greater allele-sharing with Tikopia) was plotted against the independent variable. (C) The statistic f_4(X, Tolai; Kankanaey Igorot, Ontong Java) (y-axis; higher values indicate greater allele-sharing with Kankanaey Igorot, and lower values indicate greater allele-sharing with Ontong Java) was plotted against the independent variable. (D) The statistic f_4(X, Tolai; Tikopia, Ontong Java) (y-axis; higher values indicate greater allele-sharing with Tikopia, and lower values indicate greater allele-sharing with Ontong Java) was plotted against the independent variable. All of the regression lines were computed using present-day populations from Near Oceania (groups from New Guinea, New Britain, New Ireland, and Solomon Islands). All statistics were computed using the HO dataset (~397,000 autosomal SNPs). Error bars show one standard error in each direction along y-axis.

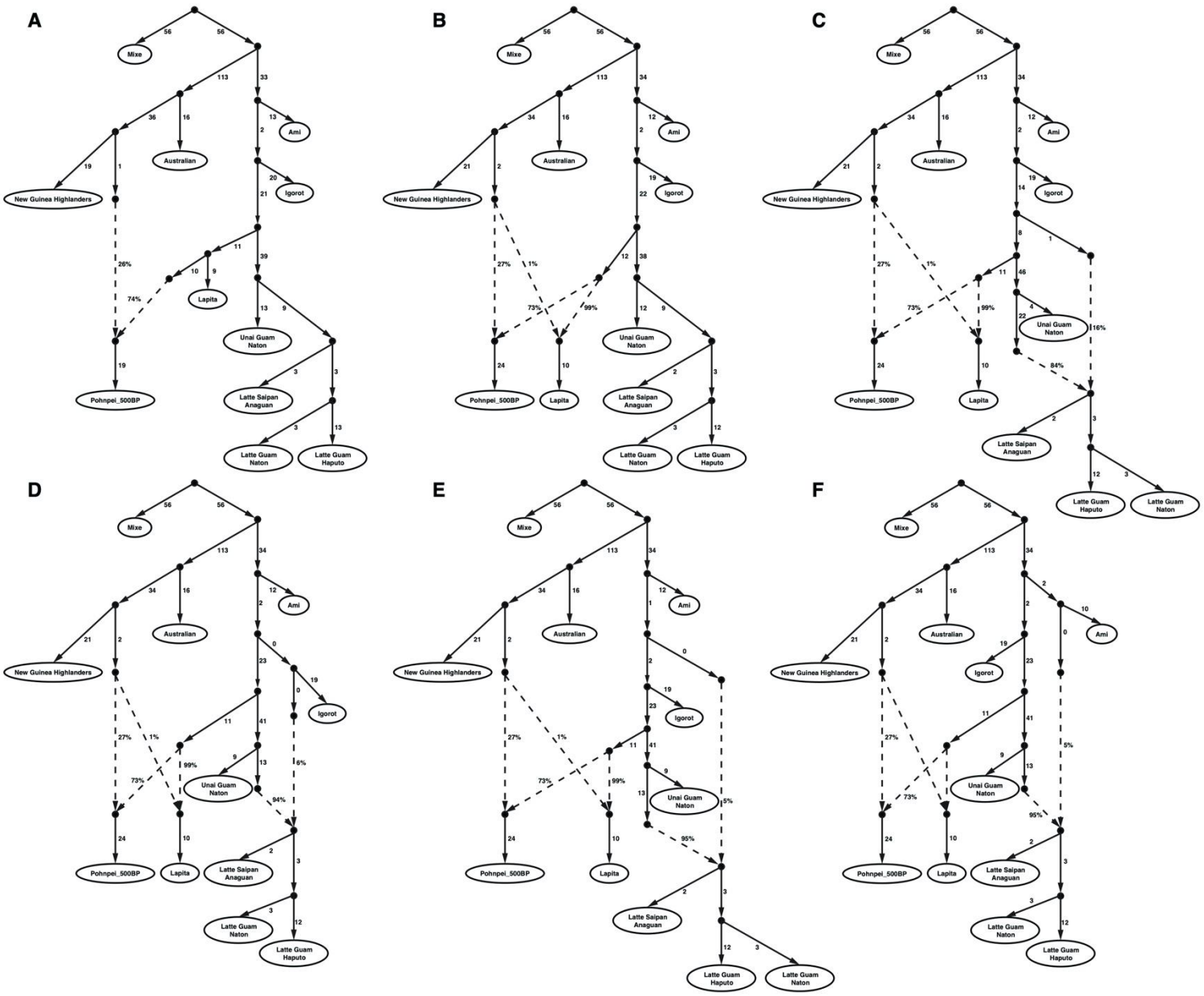

A
Mixe
Australian
Ami
Igorot
New Guinea Highlanders
Lapita
Pohnpei_500BP
Unai Guam Naton
Latte Saipan Anaguan
Latte Guam Naton
Latte Guam Haputo
B
Mixe
Australian
Ami
Igorot
New Guinea Highlanders
Lapita
Pohnpei_500BP
Unai Guam Naton
Latte Saipan Anaguan
Latte Guam Naton
Latte Guam Haputo
C
Mixe
Australian
Ami
Igorot
New Guinea Highlanders
Lapita
Pohnpei_500BP
Unai Guam Naton
Latte Saipan Anaguan
Latte Guam Haputo
Latte Guam Naton
D
Mixe
Australian
Ami
Igorot
New Guinea Highlanders
Lapita
Pohnpei_500BP
Unai Guam Naton
Latte Saipan Anaguan
Latte Guam Naton
Latte Guam Haputo
E
Mixe
Australian
Ami
Igorot
New Guinea Highlanders
Lapita
Pohnpei_500BP
Unai Guam Naton
Latte Saipan Anaguan
Latte Guam Haputo
Latte Guam Naton
F
Mixe
Australian
Ami
Igorot
New Guinea Highlanders
Lapita
Pohnpei_500BP
Unai Guam Naton
Latte Saipan Anaguan
Latte Guam Naton
Latte Guam Haputo

Fig. S12. Basic Skeleton Admixture Model explored by *qpGraph*. The skeleton graph of our *qpGraph* models includes 11 prehistoric and present-day populations. A) One admixture event was allowed, in which late prehistoric individuals from Pohnpei (~500 BP) were modeled as a mixture of ancestry related to New Guinea Highlanders and the $FRO_{SouthwestPacific}$ lineage. The maximum residual is |Z| = 4.126 for this graph, indicating a poor fit. B) Two admixture events were allowed, in which late prehistoric individuals from Pohnpei (~500 BP) were modeled as a mixture of ancestry related to New Guinea Highlanders and to the $FRO_{SouthwestPacific}$ lineage, and the Lapita individuals from Vanuatu and Tonga harbored ~1% Papuan ancestry. The maximum residual |Z| = 3.269 for this graph, indicating a plausible fit given the large number of hypotheses. C) Three admixture events were allowed, with late prehistoric individuals from Pohnpei (~500 BP) modeled as a mixture of ancestry related to New Guinea Highlanders and to the $FRO_{SouthwestPacific}$ lineage, the Lapita individuals from Vanuatu and Tonga harbored ~1% Papuan ancestry, and the Latte groups from Guam and Saipan modeled as a mixture of ancestry related to the $FRO_{Marianas}$ lineage (represented by the Unai individuals) and an additional East Asian lineage. The maximum residual |Z| = 2.959 for this graph. D) to F) Alternative models for the third non-Unai-related East Asian-associated lineage. All three models fit the data well, with the max residual |Z| equaled to 2.961, 2.960, and 2.959, respectively. However, these models do not pass when we add modern Palauans into the admixture graph with only a single additional admixture event allowed, whereas the model in C) does, leading us to prefer model C). See **fig. S13** for details. All these models were constructed on the autosomal sites of the 1240K panel after removing sites in CpG dinucleotides (~825,000 SNPs). Branch lengths in each figure are shown in units of average squared allele frequency divergence (multiplied by 1000, rounded to the nearest integer).

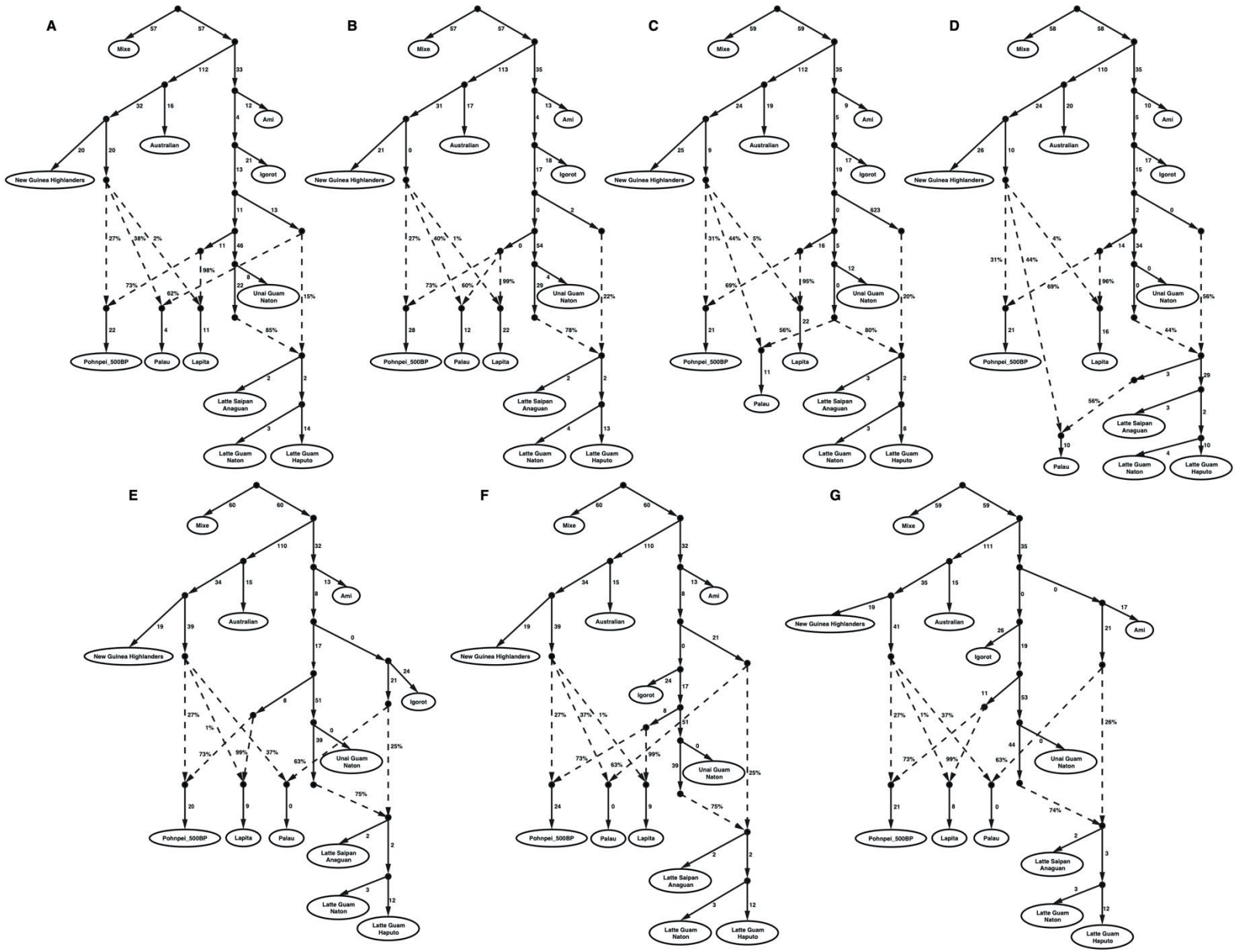

A
Mixe
Australian
Ami
Igorot
New Guinea Highlanders
Unai Guam Naton
Pohnpei_500BP
Palau
Lapita
Latte Saipan Anaguan
Latte Guam Naton
Latte Guam Haputo
B
Mixe
Australian
Ami
Igorot
New Guinea Highlanders
Unai Guam Naton
Pohnpei_500BP
Palau
Lapita
Latte Saipan Anaguan
Latte Guam Naton
Latte Guam Haputo
C
Mixe
Australian
Ami
Igorot
New Guinea Highlanders
Unai Guam Naton
Pohnpei_500BP
Palau
Lapita
Latte Saipan Anaguan
Latte Guam Naton
Latte Guam Haputo
D
Mixe
Australian
Ami
Igorot
New Guinea Highlanders
Unai Guam Naton
Pohnpei_500BP
Palau
Lapita
Latte Saipan Anaguan
Latte Guam Naton
Latte Guam Haputo
E
Mixe
Australian
Ami
Igorot
New Guinea Highlanders
Unai Guam Naton
Pohnpei_500BP
Palau
Lapita
Latte Saipan Anaguan
Latte Guam Naton
Latte Guam Haputo
F
Mixe
Australian
Ami
Igorot
New Guinea Highlanders
Unai Guam Naton
Pohnpei_500BP
Palau
Lapita
Latte Saipan Anaguan
Latte Guam Naton
Latte Guam Haputo
G
Mixe
Australian
Ami
Igorot
New Guinea Highlanders
Unai Guam Naton
Pohnpei_500BP
Palau
Lapita
Latte Saipan Anaguan
Latte Guam Naton
Latte Guam Haputo

Fig. S13. Fitting Present-day Palau. A) Palau was modeled as a mixture of ancestry related to New Guinea Highlanders and the non-Unai-related East Asian-associated lineage (FRO_{Palau}) that contributed to the Latte individuals. The max residual |Z| = 2.762 for this graph. B) Palau was modeled as a mixture of ancestry related to New Guinea Highlanders and the $FRO_{SouthwestPacific}$ lineage. The max residual |Z| = 9.889 for this graph. C) Palau was modeled as a mixture of ancestry related to New Guinea Highlanders and the $FRO_{Marianas}$ lineage. The max residual |Z| = 29.350 for this graph. D) Palau was modeled as a mixture of ancestry related to New Guinea Highlanders and the ancestral lineage of the Latte groups. The max residual |Z| = 26.396 for this graph. E) to G) Alternative models for the FRO_{Palau} lineage to fit present-day Palau. All three models in E) to G) cannot fit the data, with the max residual |Z| equal to 8.395, 8.395, and 8.159, respectively. All of these models were constructed based on the autosomal intersection between the 1240K and the MEGA array after removing sites in CpG dinucleotides (~169,000 SNPs). Branch lengths in each figure were shown in units of average squared allele frequency divergence (multiplied by 1000, rounded to the nearest integer).

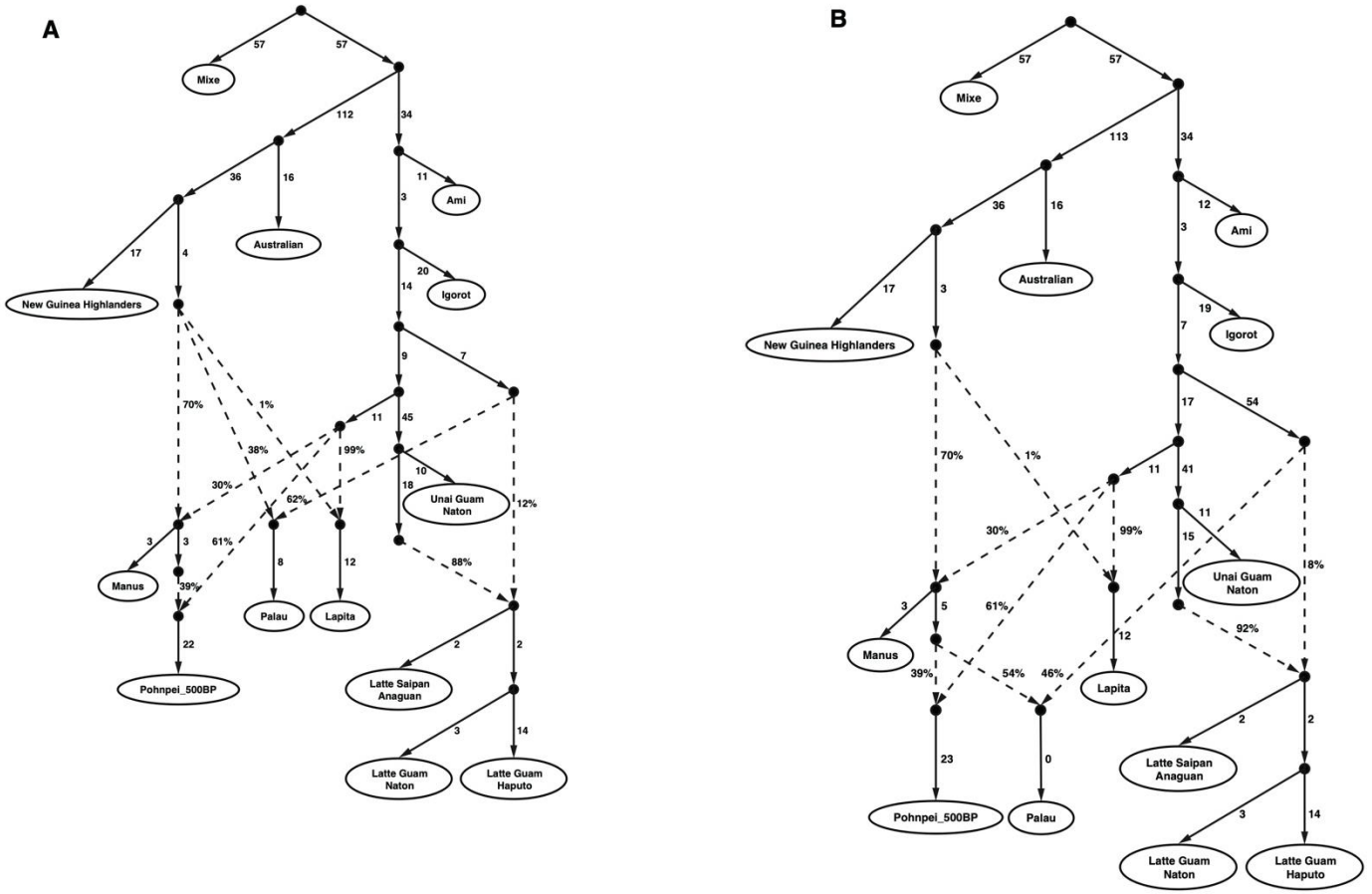


Fig. S14. Fitting Manus from the Admiralty Islands. All of these models were constructed on the autosomal intersection between the 1240K and the MEGA array after removing sites in the CpG dinucleotides (~169,000 SNPs). Branch lengths were shown in units of average squared allele frequency divergence (multiplied by 1000, rounded to the nearest integer). We modeled Manus as a mixture of $Papuan_{NewGuinea}$ and the $FRO_{SouthwestPacific}$ lineage. Late prehistoric Pohnpei (A) and present-day Palau (B) could be modeled using Manus as a representative population as well. The max residuals were |Z| = 2.974 and 2.979 for graphs A and B, respectively.

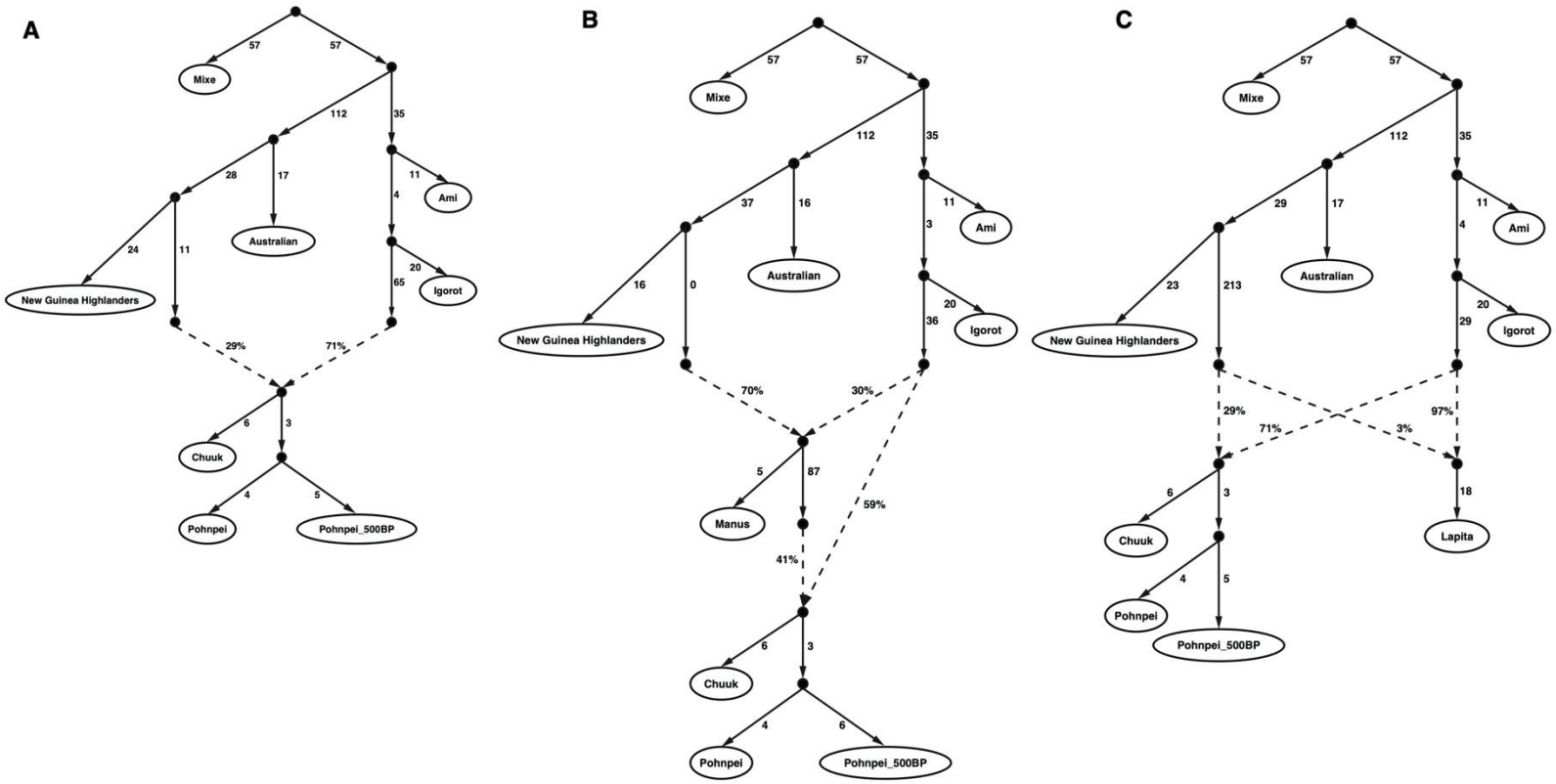


Fig. S15. Fitting Present-day Chuuk and Pohnpei. All models constructed using the overlap of autosomal sites between 1240K sites and the MEGA array after removing the sites in CpG dinucleotides (~169,000 SNPs). Branch lengths are shown in units of average squared allele frequency divergence (multiplied by 1000, rounded to the nearest integer). (A) Model for present-day Chuuk and Pohnpei. The Papuan ancestry of present-day Chuuk and Pohnpei could be modeled using a source related New Guinea Highlanders, and the East Asian ancestry could be modeled using a source related to the $FRO_{SouthwestPacific}$ lineage. The max residual |Z| = 1.287 for this graph. (B) Alternative model for present-day Pohnpei and Chuuk involving Manus. The Papuan ancestry of present-day Pohnpei and Chuuk could be instead modeled using a source related to Manus. The maximum residual |Z| is 2.534 for this graph. (C) Alternative model for present-day Pohnpei and Chuuk involving Lapita. The maximum residual |Z| is 3.288 for this graph (see **Supplementary Text S5**).

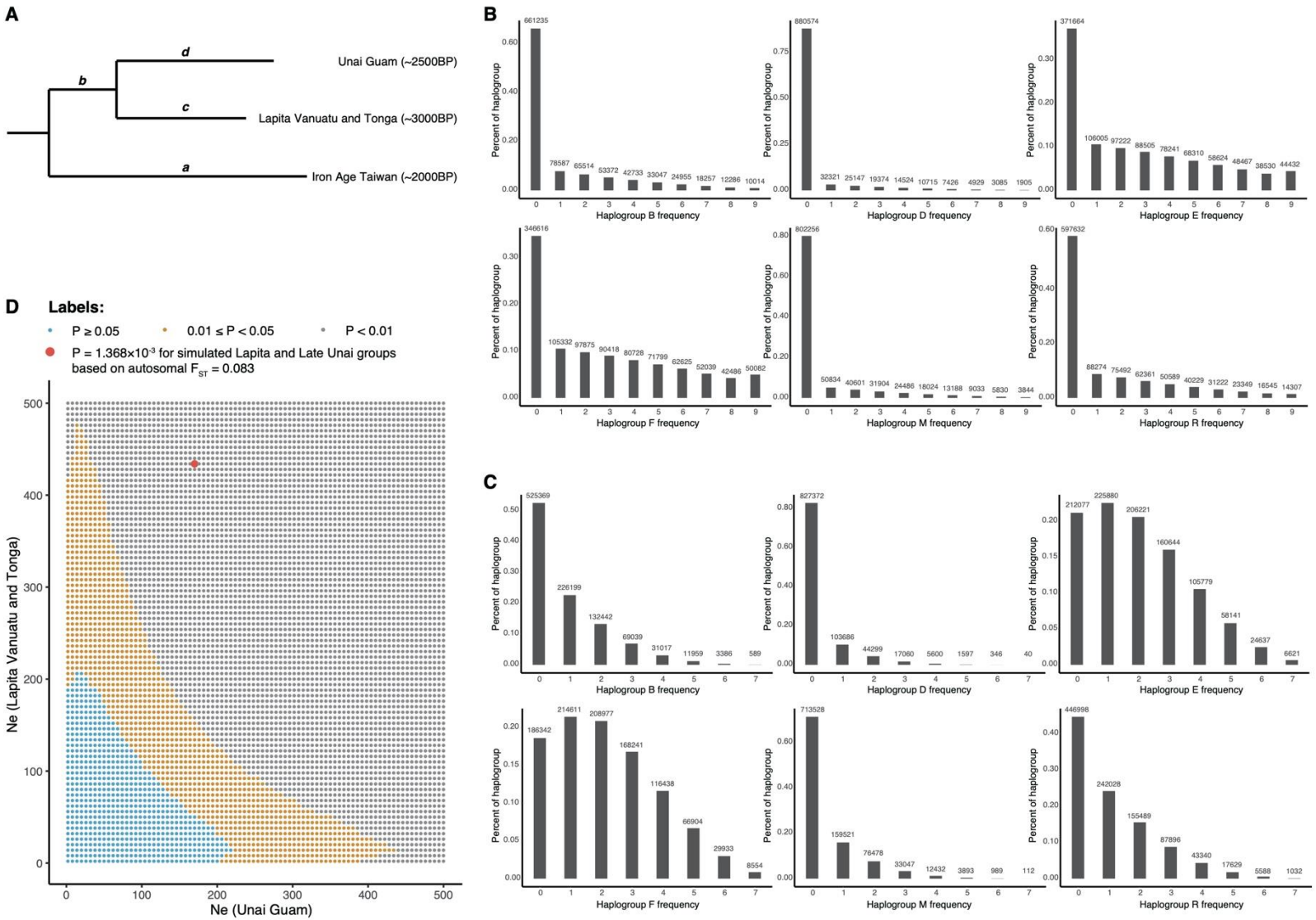

A
d
b
c
a
Unai Guam (~2500BP)
Lapita Vanuatu and Tonga (~3000BP)
Iron Age Taiwan (~2000BP)
B
Percent of haplogroup
Haplogroup B frequency
661235 78587 65514 53372 42733 33047 24955 18257 12286 10014
Haplogroup D frequency
880574 32321 25147 19374 14524 10715 7426 4929 3085 1905
Haplogroup E frequency
371664 106005 97222 88505 78241 68310 58624 48467 38530 44432
Haplogroup F frequency
346616 105332 97875 90418 80728 71799 62625 52039 42486 50082
Haplogroup M frequency
802256 50834 40601 31904 24486 18024 13188 9033 5830 3844
Haplogroup R frequency
597632 88274 75492 62361 50589 40229 31222 23349 16545 14307
C
Haplogroup B frequency
525369 226199 132442 69039 31017 11959 3386 589
Haplogroup D frequency
827372 103686 44299 17060 5600 1597 346 40
Haplogroup E frequency
212077 225880 206221 160644 105779 58141 24637 6621
Haplogroup F frequency
186342 214611 208977 168241 116438 66904 29933 8554
Haplogroup M frequency
713528 159521 76478 33047 12432 3893 989 112
Haplogroup R frequency
446998 242028 155489 87896 43340 17629 5588 1032
D
Labels:
P ≥ 0.05
0.01 ≤ P < 0.05
P < 0.01
P = 1.368×10-3 for simulated Lapita and Late Unai groups based on autosomal FST = 0.083
Ne (Lapita Vanuatu and Tonga)
Ne (Unai Guam)

Fig. S16. Simulations Establishing the Statistical Significance of the Signal of Matrilocality in Early Remote Oceanians. Simulations were performed under the Wright-Fisher genetic drift model. Ancestral effective population sizes of the Lapita and the Unai populations were computed based on autosomal F_{ST}, and the assumption of 36 generations since divergence for Lapita individuals, 54 generations since divergence for Unai individuals, and 28 years per generation. (A) Assumed phylogeny among Iron Age indigenous Taiwanese (~2000 BP), Unai Guam (~2500 BP), and Lapita (~3000 BP) individuals. (B) Simulated mitochondrial haplogroup distribution of Unai Guam (N = 9, $FRO_{Marianas}$), based on assuming that the divergence time between Lapita and Unai populations was ~4000 BP. The number above each column is the frequency of the mitochondrial haplogroup in the last generation after running 1,000,000 simulations. (C) Simulated mitochondrial haplogroup distribution of Lapita (N = 7, $FRO_{SouthwestPacific}$) under the assumption of identical male and female demography, assuming the divergence time between Lapita and Unai populations was ~4000 BP. The number above each column is the frequency of the mitochondrial haplogroup in the last generation after running 1,000,000 simulations. (D) Simulations of differential pairs of maternal N_e and corresponding empirical *P*-values for seeing complete macrohaplogroup dissociation as is observed empirically.

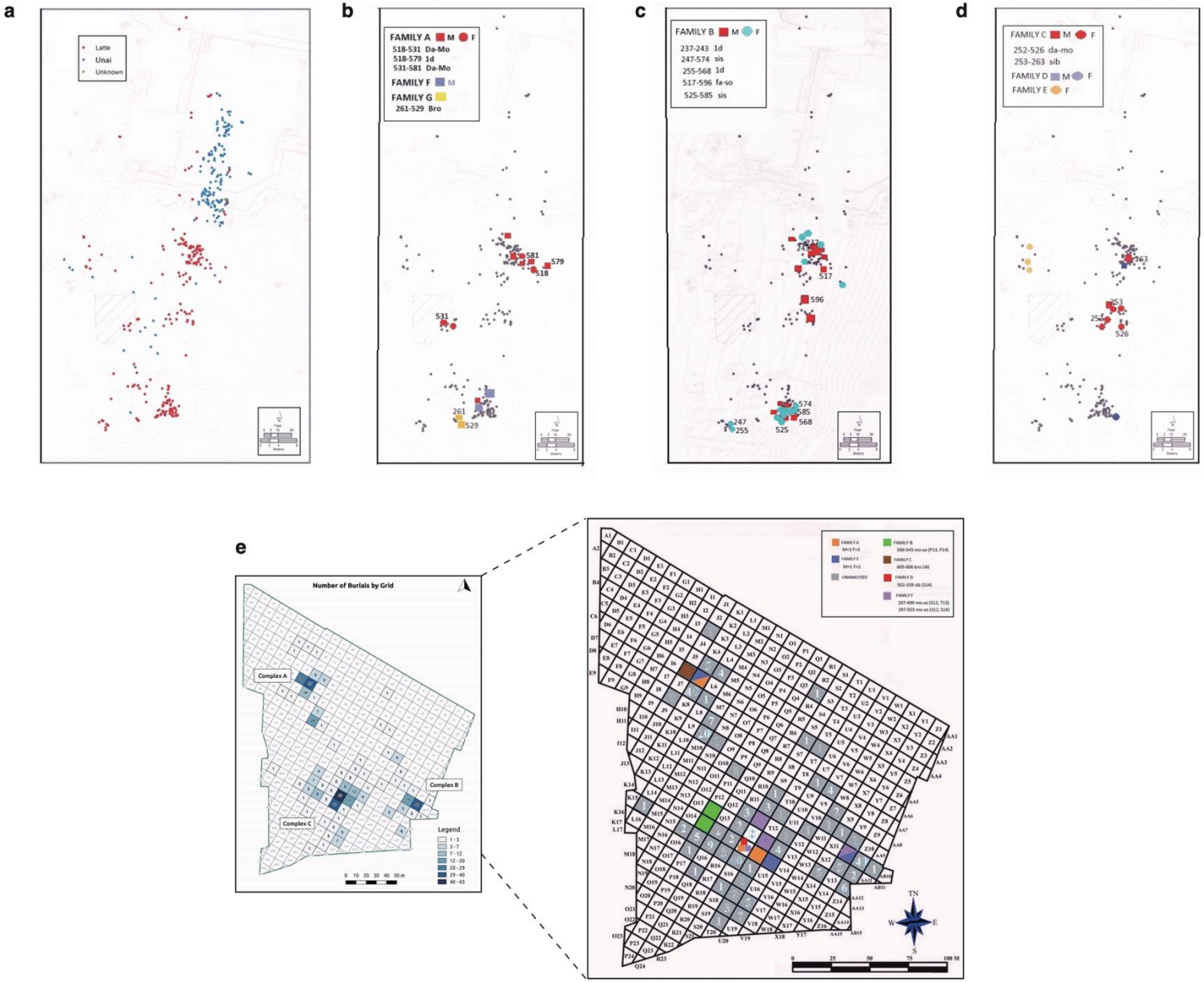

a
Latte
Unai
Unknown
b
FAMILY A M F
518-531 Da-Mo
518-579 1d
531-581 Da-Mo
FAMILY F M
FAMILY G
261-529 Bro
581
579
518
531
261
529
c
FAMILY B M F
237-243 1d
247-574 sis
255-568 1d
517-596 fa-so
525-585 sis
517
596
574
585
568
247
255
525
d
FAMILY C M F
252-526 da-mo
253-263 sib
FAMILY D M F
FAMILY E F
263
253
526
e
Number of Burials by Grid
Complex A
Complex B
Complex C
Legend
1 - 3
3 - 7
7 - 12
12 - 20
20 - 29
29 - 40
40 - 43
0 10 20 30 40 50 m
FAMILY A
FAMILY B
FAMILY C
FAMILY D
FAMILY E
FAMILY F
UNANALYZED
TN
W
E
S
0 25 50 75 100 M

Fig. S17. Distribution of Burials at Naton (Guam) and Anaguan (Saipan). a) spatial distribution of all Naton burials, Unai (blue) and Latte (red). b-d) Spatial distribution burial locations of seven Naton Latte families, with labels of relationships for first-degree relative pairs within these families. e) left, spatial distribution of Latte burials at Anaguan within original excavation grid showing number of burials by 10m² grid square; right, excavation grid with colored grid squares that contained individuals in Families A-F; gray squares mark presence of burials not analyzed in this project or that did not yield sufficient DNA. Base maps courtesy SWCA Environmental Consultants, Guam, and Scientific Consultants Services, Inc., Honolulu, Hawai‘i.

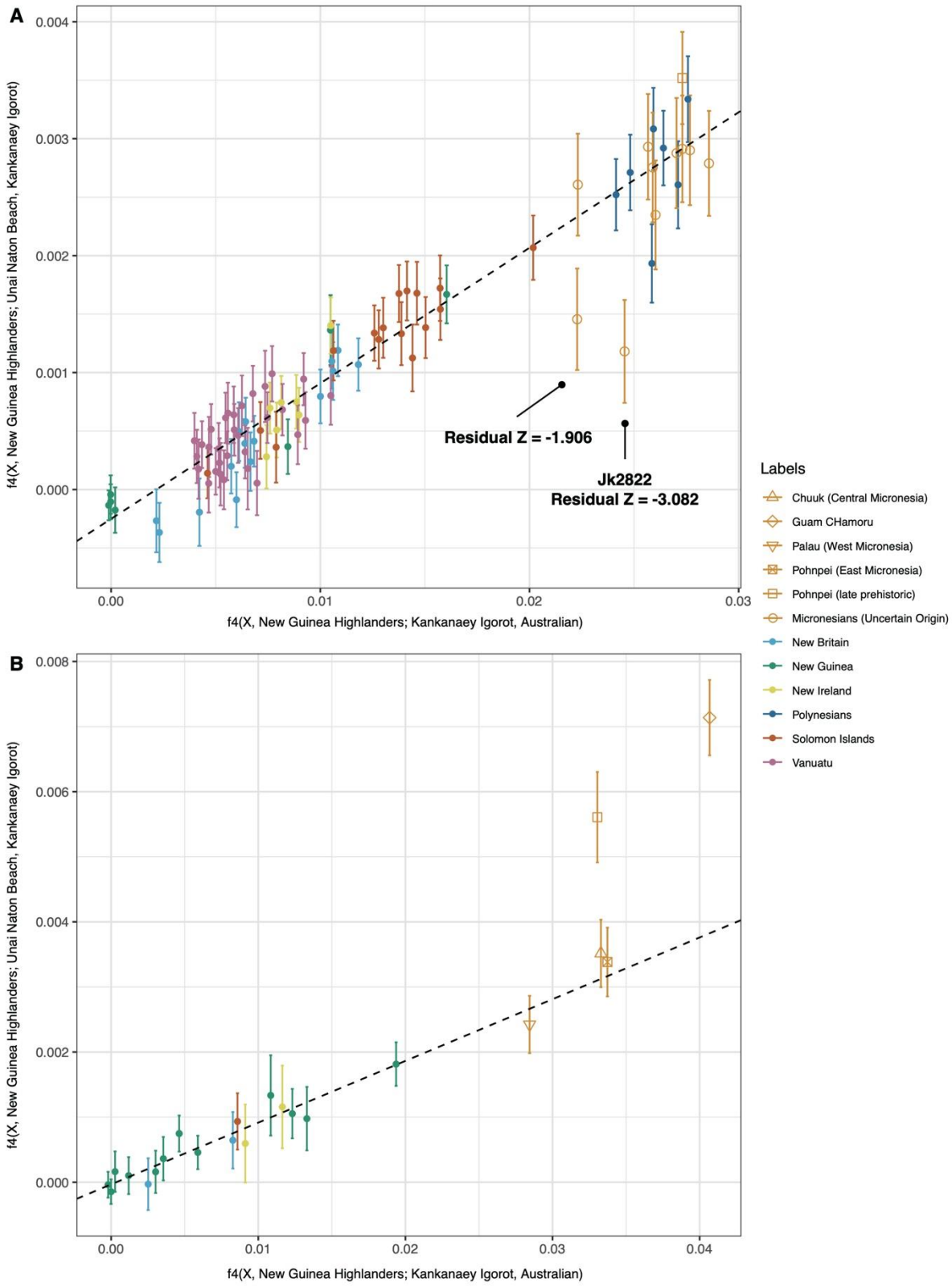

A
B
f4(X, New Guinea Highlanders; Unai Naton Beach, Kankanaey Igorot)
f4(X, New Guinea Highlanders; Kankanaey Igorot, Australian)
Residual Z = -1.906
Jk2822
Residual Z = -3.082
Labels
Chuuk (Central Micronesia)
Guam CHamoru
Palau (West Micronesia)
Pohnpei (East Micronesia)
Pohnpei (late prehistoric)
Micronesians (Uncertain Origin)
New Britain
New Guinea
New Ireland
Polynesians
Solomon Islands
Vanuatu

Fig. S18. f_4-Regression Analysis Testing for Allele Sharing with Unai Guam and Kankanaey Igorot. The statistic f_4(X, New Guinea Highlanders; Kankanaey Igorot, Australian) was the independent variable (x-axis, higher values indicate more East Asian ancestry). The statistic f_4(X, New Guinea Highlanders; Unai Guam, Kankanaey Igorot) (y-axis; higher values indicate greater allele-sharing with Unai individuals from Guam, and lower values indicate greater allele-sharing with Kankanaey) was plotted against the independent variable. The linear regression model was computed using present-day Oceanian groups except Micronesians. Error bars show one standard error in each direction along y-axis. (A) computed on the HO dataset (~397,000 autosomal SNPs); (B) computed on the intersection between the 1240K panel and the MEGA array (~169,000 autosomal SNPs). The present-day Micronesian individual Jk2822 showed a significant excess of allele-sharing with the Kankanaey Igorot, plausibly indicating an additional migration from Southeast Asia to Micronesia that brought another type of Southeast Asian-related ancestry into Micronesia different from that related to the Lapita and Unai individuals.

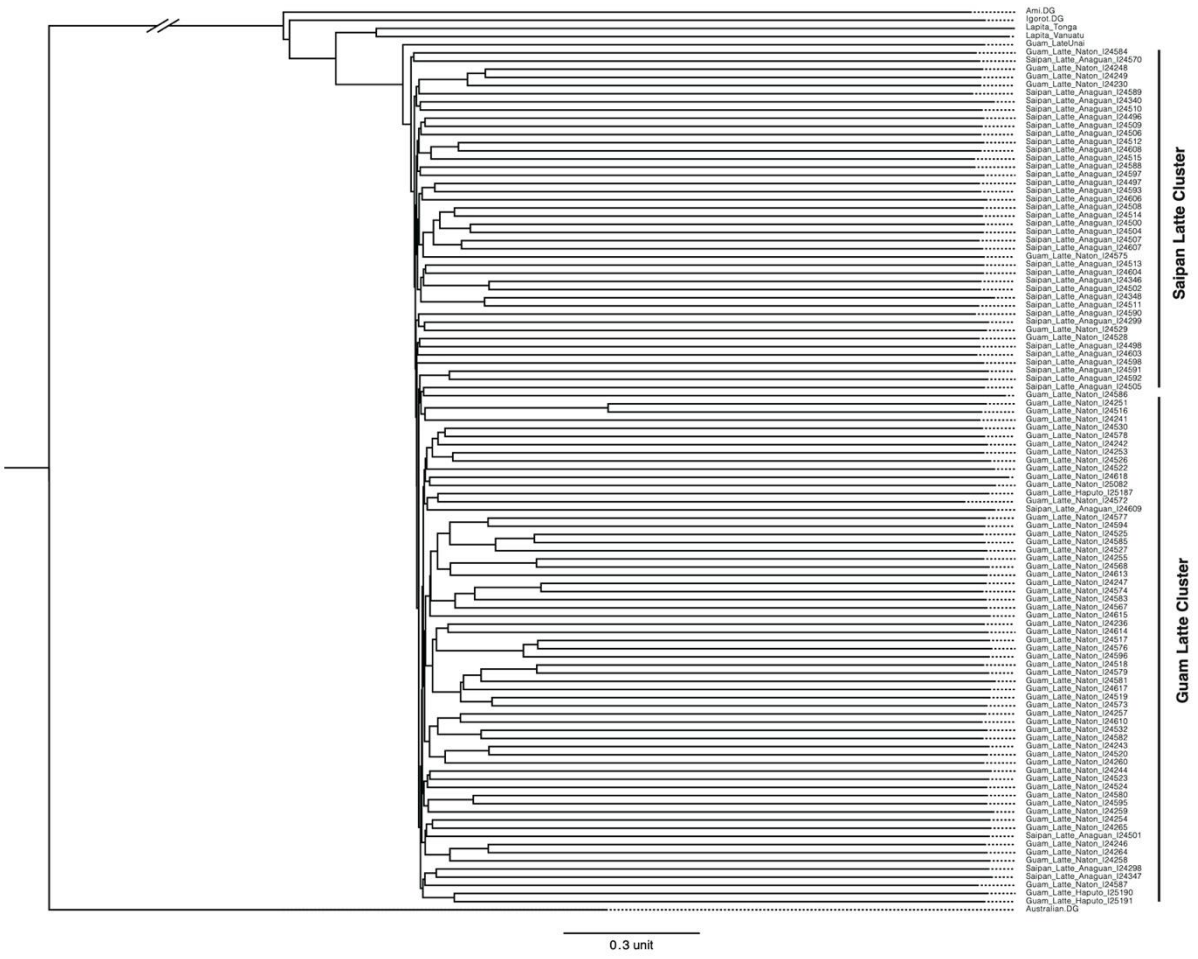


Fig. S19. Neighbor-Joining Tree of FRO and Latte Groups. We computed the statistic f_3(X1, X2; Mbuti), where X1 and X2 were the Latte individuals from Guam and Saipan, the Lapita individuals from Vanuatu (merged as one group) and Tonga (merged as one group), the Unai individuals from the Naton Beach site on Guam (merged as one group), Kankanaey Igorot (merged as one group), and Aboriginal Australians (merged as one group). We reconstructed a phylogenetic tree based on the genetic distance matrix transferred from outgroup f_3-statistics, using the reciprocal of the statistic as the genetic distance.

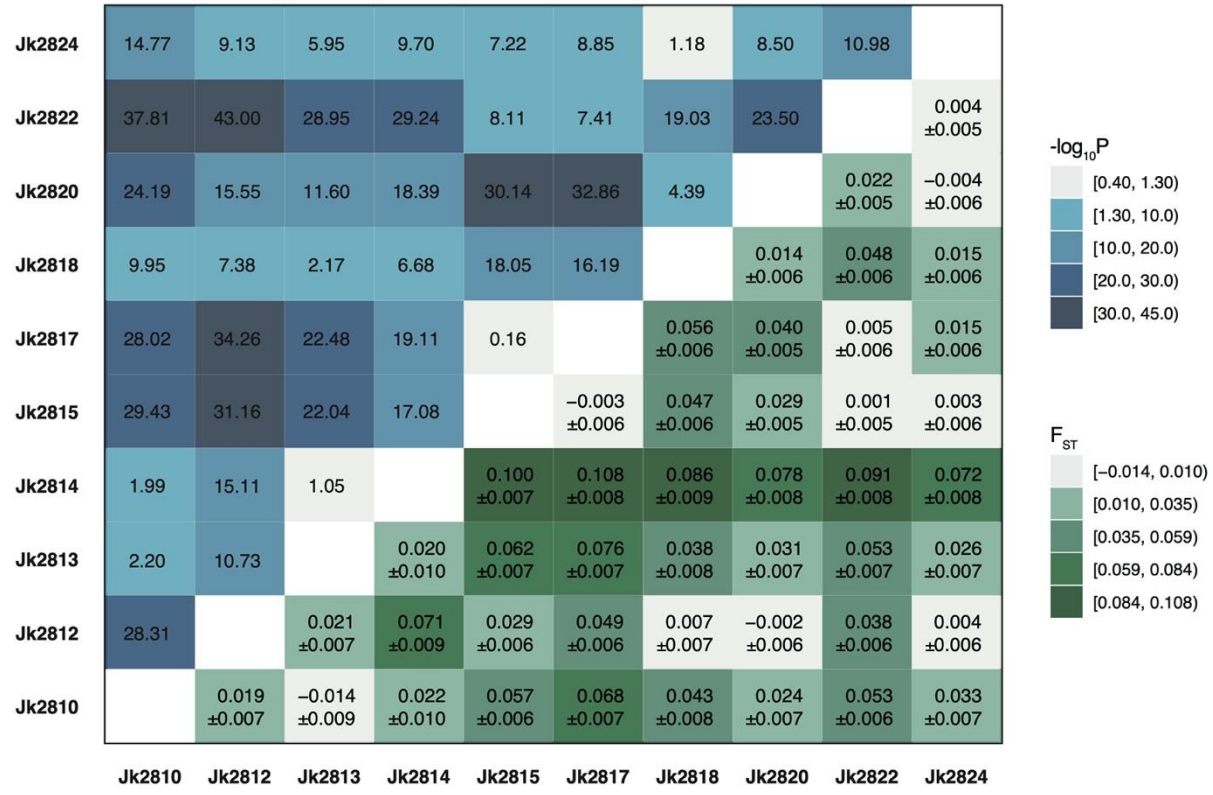


Fig. S20. Genetic Clustering of Present-day Micronesians Genotyped on the Human Origins Array. The upper triangular matrix displays -$\log_{10}$P of pairwise *qpWave* modeling (-$\log_{10}(0.05)\approx1.30$) for ten present-day Micronesian individuals, with twelve outgroups: Aboriginal Australian (deeply Papuan-related), French (western Eurasian), Pima (Native American), Onge (deep eastern Eurasian), Japanese (Northern East Asian), Dai (Southern East Asian), Ami (indigenous Taiwanese), Kankanaey Igorot (indigenous people in the Philippines), New Guinea Highlanders (Papuan lineage in mainland New Guinea), Nasioi (Papuan lineage in the Solomon Islands), Baining (Papuan lineage in the Bismarck Archipelago), Tongan (Polynesians). The lower triangular matrix reports the pairwise autosomal F_{ST} with one standard error.

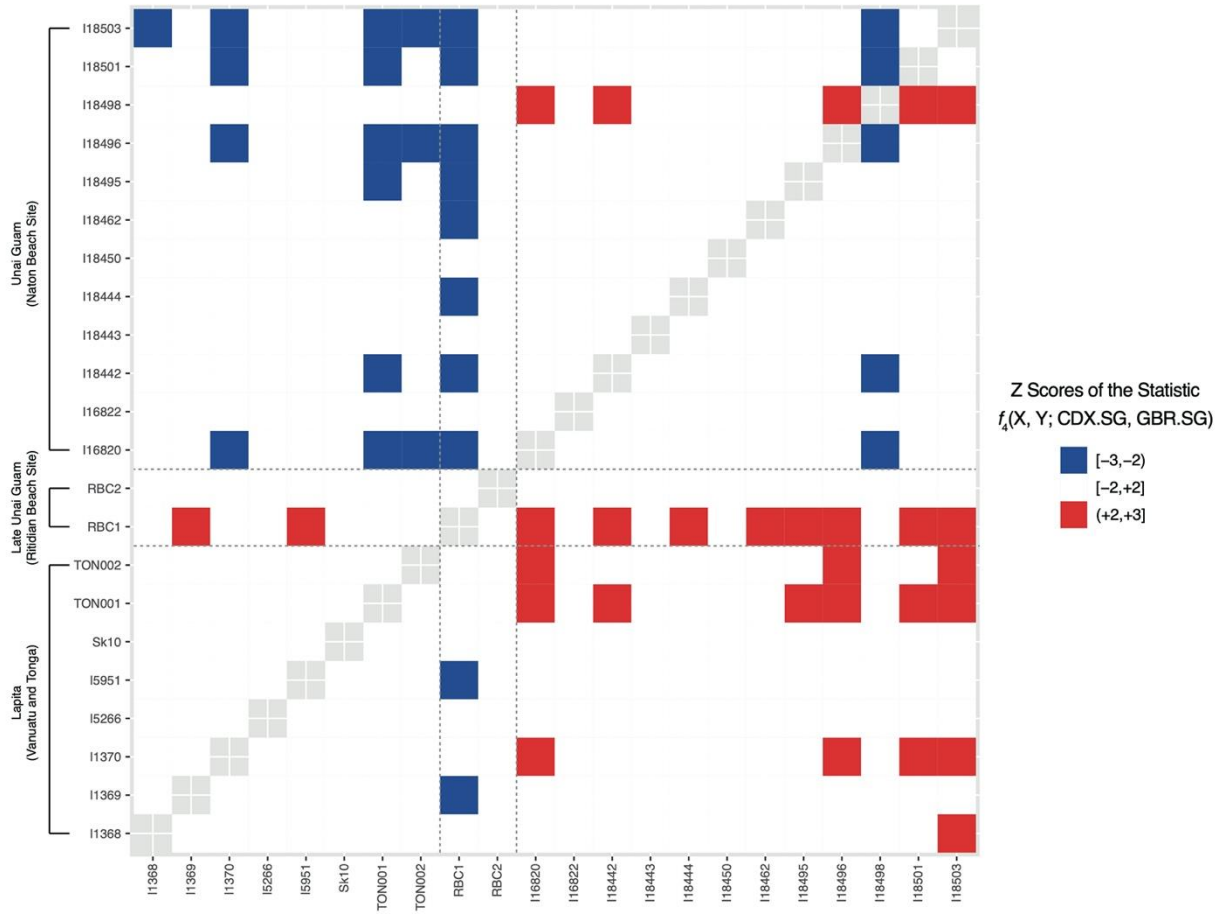


Fig. S21. In-depth Curations for Unai Samples. We computed the statistic f_4(X, Y; CDX.SG, GBR.SG) to identify outliers for each group, where X were the individuals labeled in the X axis, and Y were the individuals labeled in the Y axis. We indentified one individual as an outlier if a large fraction of statistics containing this individual within a group consistently yielded |Z| > 2. Here, there is evidence that I18498 was probably an outlier compared to other Unai individuals.

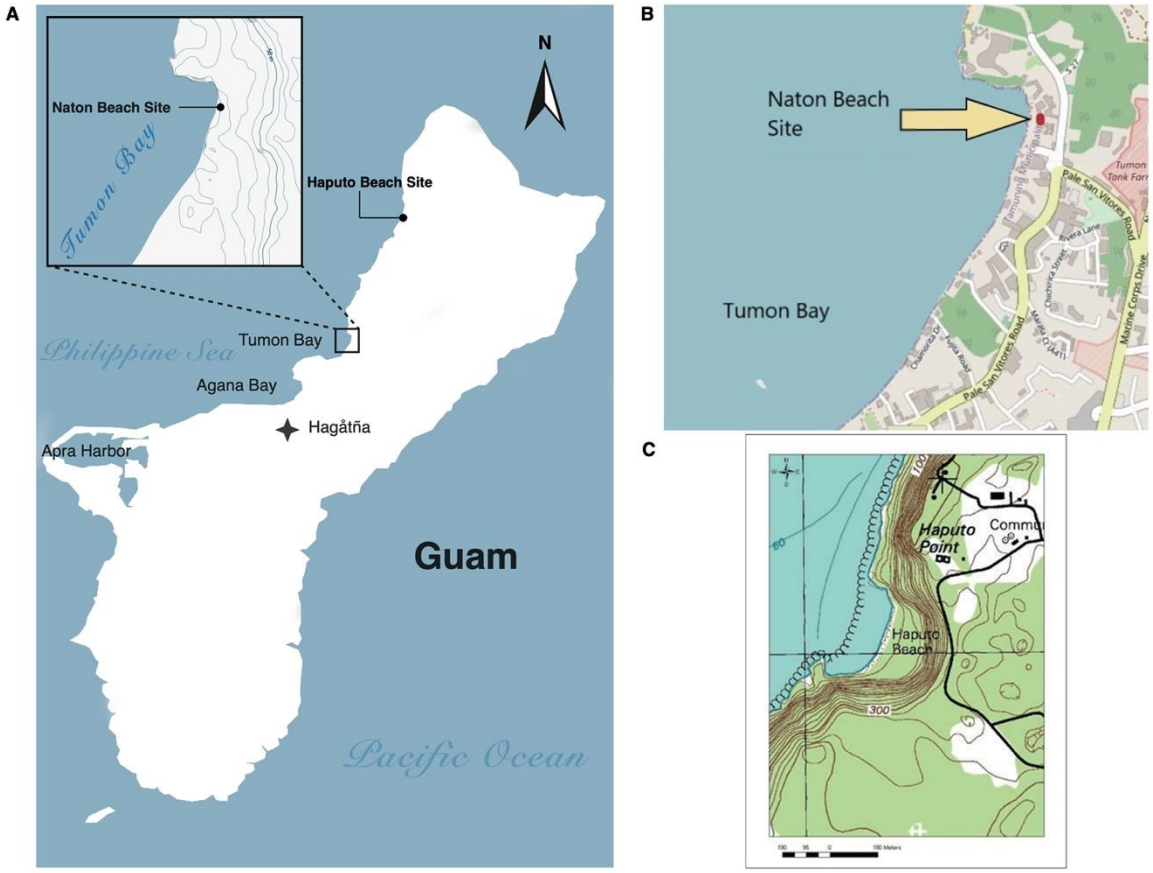


Fig. S22. Locations of Study Sites on Guam. (A) Location of the Naton site and Haputo sites on Guam. (B) Zoom-in map for the Naton site. The Naton Beach Site is situated in Gogna Cove at the north end of Tumon Bay, where a limestone headland juts into the sea, forming the northern terminus of the Tumon embayment. Protected bays were preferred settlement sites throughout the Mariana prehistory. (C) Zoom-in map for the Haputo site based on 1975 USGS Ritidian Quadrangle map, 1:24,000 (courtesy R. Olmo). Haputo Bay on Guam's northwest coast.

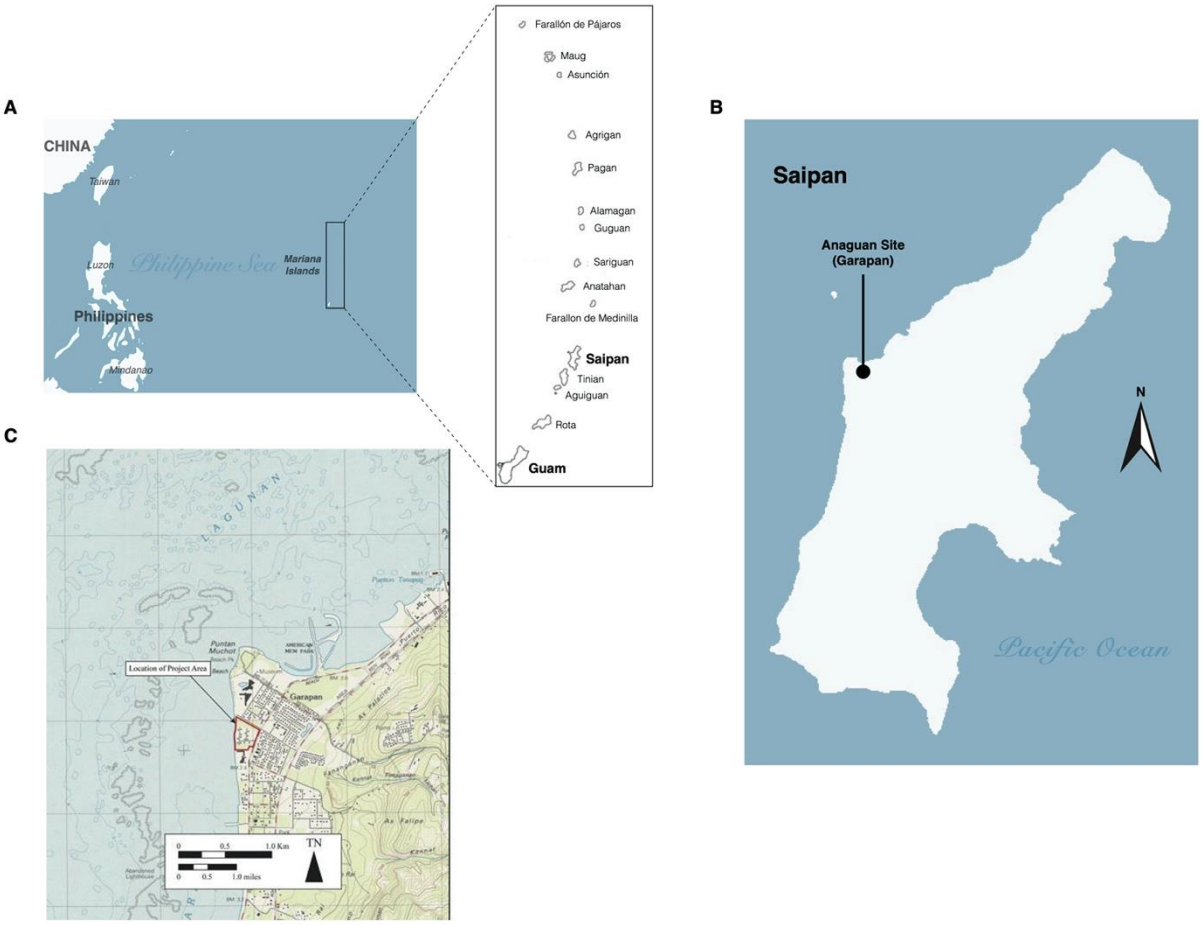


Fig. S23. Schematic map for the Mariana Islands and the Anaguan Site in Garapan on Saipan. (A) Map of the Mariana Islands. (B) Zoom-in map for Saipan and the location of the Garapan site. (C) Zoom-in map for the Garapan site based on the 1983 USGS Saipan map.

References and Notes

1. J. F. O'Connell, J. Allen, The process, biotic impact, and global implications of the human colonization of Sahul about 47,000 years ago. *J. Archaeol. Sci.* **56**, 73–84 (2015). doi:10.1016/j.jas.2015.02.020

2. S. Wickler, M. Spriggs, Pleistocene human occupation of the Solomon Islands. *Antiquity* **62**, 703–706 (1988). doi:10.1017/S0003598X00075104

3. T. M. Rieth, J. S. Athens, Late Holocene human expansion into Near and Remote Oceania: A Bayesian model of the chronologies of the Mariana Islands and Bismarck Archipelago. *J. Isl. Coast. Archaeol.* **14**, 5–16 (2019). doi:10.1080/15564894.2017.1331939

4. S. Bedford, M. Spriggs, Eds., *Debating Lapita: Distribution, Chronology, Society and Subsistence*, vol. 52 of *Terra Australis* (ANU Press, 2019).

5. M. A. Yang, X. Fan, B. Sun, C. Chen, J. Lang, Y.-C. Ko, C. H. Tsang, H. Chiu, T. Wang, Q. Bao, X. Wu, M. Hajdinjak, A. M.-S. Ko, M. Ding, P. Cao, R. Yang, F. Liu, B. Nickel, Q. Dai, X. Feng, L. Zhang, C. Sun, C. Ning, W. Zeng, Y. Zhao, M. Zhang, X. Gao, Y. Cui, D. Reich, M. Stoneking, Q. Fu, Ancient DNA indicates human population shifts and admixture in northern and southern China. *Science* **369**, 282–288 (2020). doi:10.1126/science.aba0909 Medline

6. C.-C. Wang, H.-Y. Yeh, A. N. Popov, H.-Q. Zhang, H. Matsumura, K. Sirak, O. Cheronet, A. Kovalev, N. Rohland, A. M. Kim, S. Mallick, R. Bernardos, D. Tumen, J. Zhao, Y.-C. Liu, J.-Y. Liu, M. Mah, K. Wang, Z. Zhang, N. Adamski, N. Broomandkhoshbacht, K. Callan, F. Candilio, K. S. D. Carlson, B. J. Culleton, L. Eccles, S. Freilich, D. Keating, A. M. Lawson, K. Mandl, M. Michel, J. Oppenheimer, K. T. Özdoğan, K. Stewardson, S. Wen, S. Yan, F. Zalzala, R. Chuang, C.-J. Huang, H. Looh, C.-C. Shiung, Y. G. Nikitin, A. V. Tabarev, A. A. Tishkin, S. Lin, Z.-Y. Sun, X.-M. Wu, T.-L. Yang, X. Hu, L. Chen, H. Du, J. Bayarsaikhan, E. Mijiddorj, D. Erdenebaatar, T.-O. Iderkhangai, E. Myagmar, H. Kanzawa-Kiriyama, M. Nishino, K. I. Shinoda, O. A. Shubina, J. Guo, W. Cai, Q. Deng, L. Kang, D. Li, D. Li, R. Lin, R. Nini, R. Shrestha, L. X. Wang, L. Wei, G. Xie, H. Yao, M. Zhang, G. He, X. Yang, R. Hu, M. Robbeets, S. Schiffels, D. J. Kennett, L. Jin, H. Li, J. Krause, R. Pinhasi, D. Reich, Genomic insights into the formation of human populations in East Asia. *Nature* **591**, 413–419 (2021). doi:10.1038/s41586-021-03336-2 Medline

7. P. Skoglund, C. Posth, K. Sirak, M. Spriggs, F. Valentin, S. Bedford, G. R. Clark, C. Reepmeyer, F. Petchey, D. Fernandes, Q. Fu, E. Harney, M. Lipson, S. Mallick, M. Novak, N. Rohland, K. Stewardson, S. Abdullah, M. P. Cox, F. R. Friedlaender, J. S. Friedlaender, T. Kivisild, G. Koki, P. Kusuma, D. A. Merriwether, F.-X. Ricaut, J. T. S. Wee, N. Patterson, J. Krause, R. Pinhasi, D. Reich, Genomic insights into the peopling of the Southwest Pacific. *Nature* **538**, 510–513 (2016). doi:10.1038/nature19844 Medline

8. M. Lipson, P. Skoglund, M. Spriggs, F. Valentin, S. Bedford, R. Shing, H. Buckley, I. Phillip, G. K. Ward, S. Mallick, N. Rohland, N. Broomandkhoshbacht, O. Cheronet, M. Ferry, T. K. Harper, M. Michel, J. Oppenheimer, K. Sirak, K. Stewardson, K. Auckland, A. V. S. Hill, K. Maitland, S. J. Oppenheimer, T. Parks, K. Robson, T. N. Williams, D. J. Kennett, A. J. Mentzer, R. Pinhasi, D. Reich, Population turnover in Remote Oceania shortly after initial settlement. *Curr. Biol.* **28**, 1157–1165.e7 (2018). doi:10.1016/j.cub.2018.02.051

Medline

9. M. Lipson, M. Spriggs, F. Valentin, S. Bedford, R. Shing, W. Zinger, H. Buckley, F. Petchey, R. Matanik, O. Cheronet, N. Rohland, R. Pinhasi, D. Reich, Three phases of ancient migration shaped the ancestry of human populations in Vanuatu. *Curr. Biol.* **30**, 4846–4856.e6 (2020). doi:10.1016/j.cub.2020.09.035 Medline

10. C. Posth, K. Nägele, H. Colleran, F. Valentin, S. Bedford, K. W. Kami, R. Shing, H. Buckley, R. Kinaston, M. Walworth, G. R. Clark, C. Reepmeyer, J. Flexner, T. Maric, J. Moser, J. Gresky, L. Kiko, K. J. Robson, K. Auckland, S. J. Oppenheimer, A. V. S. Hill, A. J. Mentzer, J. Zech, F. Petchey, P. Roberts, C. Jeong, R. D. Gray, J. Krause, A. Powell, Language continuity despite population replacement in Remote Oceania. *Nat. Ecol. Evol.* **2**, 731–740 (2018). doi:10.1038/s41559-018-0498-2 Medline

11. M. T. Carson, *First Settlement of Remote Oceania: Earliest Sites in the Mariana Islands*, vol. 1 of *Springer Briefs in Archaeology* (Springer, 2014).

12. F. Petchey, G. Clark, A marine reservoir correction value (ΔR) for the Palauan archipelago: Environmental and oceanographic considerations. *J. Isl. Coast. Archaeol.* **5**, 236–252 (2010). doi:10.1080/15564890903155935

13. F. Petchey, G. Clark, I. Lindeman, P. O'Day, J. Southon, K. Dabell, O. Winter, Forgotten news: Shellfish isotopic insight into changing sea-level and associated impact on the first settlers of the Mariana Archipelago. *Quat. Geochronol.* **48**, 180–194 (2018). doi:10.1016/j.quageo.2018.10.002

14. F. Petchey, G. Clark, Clarifying the age of initial settlement Horizon in the Mariana Islands and the Impact of Hard Water: A response to Carson (2020). *Radiocarbon* **63**, 905–913 (2021). doi:10.1017/RDC.2021.27

15. A. Spoehr, *Marianas Prehistory: Archaeological Survey and Excavations on Saipan, Tinian and Rota*, vol. 48 of *Fieldiana Anthropology* (Chicago Natural History Museum, 1957).

16. G. R. Clark, O. Winter, "The ceramic trail: Evaluating the Marianas and Lapita West Pacific connection" in *Debating Lapita: Distribution, Chronology, Society and Subsistence*, vol. 52 of *Terra Australis*, S. Bedford, M. Spriggs, Eds. (ANU Press, 2019), pp. 37–59.

17. T. Rieth, E. E. Cochrane, "The chronology of colonization in Remote Oceania" in *The Oxford Handbook of Prehistoric Oceania*, E. E. Cochrane, T. L. Hunt, Eds. (Oxford Univ. Press, 2018), pp. 133–161.

18. R. Shutler Jr., "The relationship of red-slipped and lime-impressed pottery of the southern Philippines to that of Micronesia and the Lapita of Oceania" in *The Pacific from 5000 to 2000 BP (Colonsation and Tranformations)*, J. C. Galipaud, I. Lilley, Eds. (IRD Editions, 1999), pp. 522–529.

19. J. S. Athens, Nan Madol pottery, Pohnpei. *Micronesica* **2**, 17–32 (1990).

20. See supplementary materials.

21. R. D. Gray, F. M. Jordan, Language trees support the express-train sequence of Austronesian expansion. *Nature* **405**, 1052–1055 (2000). doi:10.1038/35016575 Medline

22. R. D. Gray, A. J. Drummond, S. J. Greenhill, Language phylogenies reveal expansion pulses

and pauses in Pacific settlement. *Science* **323**, 479–483 (2009). doi:10.1126/science.1166858 Medline

23. A. D. Smith, The Western Malayo-Polynesian problem. *Ocean. Linguist.* **56**, 435–490 (2017). doi:10.1353/ol.2017.0021

24. J. J. Song, The migration pathways of the Proto Nuclear Micronesians: A linguistic trail. *Macrolinguistics* **3**, 26–66 (2009). doi:10.26478/ja2009.3.3.2

25. M. Ross, "Is Yapese oceanic?" in *Reconstruction, Classification, Description — Festschrift in Honor of Isidore Dyen*, vol. 3 of *Abera Network Asia-Pacific*, B. Nothofer, Ed. (Abera, 1996) pp. 121–166.

26. P. V. Kirch, The Polynesian outliers: Continuity, change, and replacement. *J. Pac. Hist.* **19**, 224–238 (1984). doi:10.1080/00223348408572496

27. P. V. Kirch, *On the Road of the Winds: An Archaeological History of the Pacific Islands Before European Contact* (Univ. California Press, 2017).

28. I. Pugach, A. Hübner, H. C. Hung, M. Meyer, M. T. Carson, M. Stoneking, Ancient DNA from Guam and the peopling of the Pacific. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2022112118 (2021). doi:10.1073/pnas.2022112118 Medline

29. W. L. Schurz, *The Manila Galleon*, vol. 40 of *Historical Conservation Society* (E. P. Dutton & Company, 1939).

30. R. Blust, Malaita-Micronesian once again. *Ocean. Linguist.* **49**, 559–567 (2010).

31. K.-I. Sudo, Social organization and types of sea tenure in Micronesia. *Senri Ethnol. Stud.* **17**, 203–230 (1984).

32. A. Mittnik, K. Massy, C. Knipper, F. Wittenborn, R. Friedrich, S. Pfrengle, M. Burri, N. Carlichi-Witjes, H. Deeg, A. Furtwängler, M. Harbeck, K. von Heyking, C. Kociumaka, I. Kucukkalipci, S. Lindauer, S. Metz, A. Staskiewicz, A. Thiel, J. Wahl, W. Haak, E. Pernicka, S. Schiffels, P. W. Stockhammer, J. Krause, Kinship-based social inequality in Bronze Age Europe. *Science* **366**, 731–734 (2019). doi:10.1126/science.aax6219 Medline

33. C. Fowler, I. Olalde, V. Cummings, I. Armit, L. Büster, S. Cuthbert, N. Rohland, O. Cheronet, R. Pinhasi, D. Reich, A high-resolution picture of kinship practices in an Early Neolithic tomb. *Nature* **601**, 584–587 (2022). doi:10.1038/s41586-021-04241-4 Medline

34. P. Hage, J. Marck, Matrilineality and the Melanesian origin of Polynesian Y chromosomes. *Curr. Anthropol.* **44**, S121–S127 (2003). doi:10.1086/379272

35. F. M. Jordan, R. D. Gray, S. J. Greenhill, R. Mace, Matrilocal residence is ancestral in Austronesian societies. *Proc. Biol. Sci.* **276**, 1957–1964 (2009). doi:10.1098/rspb.2009.0088 Medline

36. M. Spriggs, D. Reich, An ancient DNA Pacific journey: A case study of collaboration between archaeologists and geneticists. *World Archaeol.* **51**, 620–639 (2019). doi:10.1080/00438243.2019.1733069

37. J. H. Stone, "The bioarchaeology of initial human settlement in Palau, Western Micronesia," thesis, University of Oregon (2020).

38. M. T. Carson, H. Hung, G. Summerhayes, P. Bellwood, The pottery trail from Southeast

Asia to Remote Oceania. *J. Island Coast. Archaeol.* **8**, 17–36 (2013). doi:10.1080/15564894.2012.726941

39. J. S. C. Dumont D'Urville, Sur les Iles du Grande Océan. *Bull. la Société Géographie Paris* **105**, 1–21 (1832).

40. Y. C. Liu *et al*., Data for: "Ancient DNA Reveals Five Streams of Migration into Micronesia and Matrilocality in Early Pacific Seafarers". Harvard Dataverse (2022); https://doi.org/10.7910/DVN/63QFEC.

41. R. Pinhasi, D. M. Fernandes, K. Sirak, O. Cheronet, Isolating the human cochlea to generate bone powder for ancient DNA analysis. *Nat. Protoc.* **14**, 1194–1205 (2019). doi:10.1038/s41596-019-0137-7 Medline

42. J. Dabney, M. Knapp, I. Glocke, M.-T. Gansauge, A. Weihmann, B. Nickel, C. Valdiosera, N. García, S. Pääbo, J.-L. Arsuaga, M. Meyer, Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 15758–15763 (2013). doi:10.1073/pnas.1314445110 Medline

43. P. Korlević, T. Gerber, M.-T. Gansauge, M. Hajdinjak, S. Nagel, A. Aximu-Petri, M. Meyer, Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* **59**, 87–93 (2015). doi:10.2144/000114320 Medline

44. N. Rohland, I. Glocke, A. Aximu-Petri, M. Meyer, Extraction of highly degraded DNA from ancient bones, teeth and sediments for high-throughput sequencing. *Nat. Protoc.* **13**, 2447–2461 (2018). doi:10.1038/s41596-018-0050-5 Medline

45. M. Meyer, M. Kircher, Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, prot5448 (2010). doi:10.1101/pdb.prot5448 Medline

46. M.-T. Gansauge, M. Meyer, Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat. Protoc.* **8**, 737–748 (2013). doi:10.1038/nprot.2013.038 Medline

47. M. Kircher, S. Sawyer, M. Meyer, Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3 (2012). doi:10.1093/nar/gkr771 Medline

48. A. W. Briggs, U. Stenzel, M. Meyer, J. Krause, M. Kircher, S. Pääbo, Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res.* **38**, e87 (2010). doi:10.1093/nar/gkp1163 Medline

49. N. Rohland, E. Harney, S. Mallick, S. Nordenfelt, D. Reich, Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20130624 (2015). doi:10.1098/rstb.2013.0624 Medline

50. Q. Fu, A. Mittnik, P. L. F. Johnson, K. Bos, M. Lari, R. Bollongino, C. Sun, L. Giemsch, R. Schmitz, J. Burger, A. M. Ronchitelli, F. Martini, R. G. Cremonesi, J. Svoboda, P. Bauer, D. Caramelli, S. Castellano, D. Reich, S. Pääbo, J. Krause, A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* **23**, 553–559 (2013). doi:10.1016/j.cub.2013.02.044 Medline

51. Q. Fu, M. Hajdinjak, O. T. Moldovan, S. Constantin, S. Mallick, P. Skoglund, N. Patterson,

N. Rohland, I. Lazaridis, B. Nickel, B. Viola, K. Prüfer, M. Meyer, J. Kelso, D. Reich, S. Pääbo, An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216–219 (2015). doi:10.1038/nature14558 Medline

52. W. Haak, I. Lazaridis, N. Patterson, N. Rohland, S. Mallick, B. Llamas, G. Brandt, S. Nordenfelt, E. Harney, K. Stewardson, Q. Fu, A. Mittnik, E. Bánffy, C. Economou, M. Francken, S. Friederich, R. G. Pena, F. Hallgren, V. Khartanovich, A. Khokhlov, M. Kunst, P. Kuznetsov, H. Meller, O. Mochalov, V. Moiseyev, N. Nicklisch, S. L. Pichler, R. Risch, M. A. Rojo Guerra, C. Roth, A. Szécsényi-Nagy, J. Wahl, M. Meyer, J. Krause, D. Brown, D. Anthony, A. Cooper, K. W. Alt, D. Reich, Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015). doi:10.1038/nature14317 Medline

53. I. Mathieson, I. Lazaridis, N. Rohland, S. Mallick, N. Patterson, S. A. Roodenberg, E. Harney, K. Stewardson, D. Fernandes, M. Novak, K. Sirak, C. Gamba, E. R. Jones, B. Llamas, S. Dryomov, J. Pickrell, J. L. Arsuaga, J. M. B. de Castro, E. Carbonell, F. Gerritsen, A. Khokhlov, P. Kuznetsov, M. Lozano, H. Meller, O. Mochalov, V. Moiseyev, M. A. R. Guerra, J. Roodenberg, J. M. Vergès, J. Krause, A. Cooper, K. W. Alt, D. Brown, D. Anthony, C. Lalueza-Fox, W. Haak, R. Pinhasi, D. Reich, Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015). doi:10.1038/nature16152 Medline

54. N. Rohland, S. Mallick, M. Mah, R. Maier, N. Patterson, D. Reich, Three reagents for in-solution enrichment of ancient human DNA at more than a million SNPs. bioRxiv 2022.01.13.476259 [Preprint] (2022); https://doi.org/10.1101/2022.01.13.476259.

55. E. Elhaik, E. Greenspan, S. Staats, T. Krahn, C. Tyler-Smith, Y. Xue, S. Tofanelli, P. Francalacci, F. Cucca, L. Pagani, L. Jin, H. Li, T. G. Schurr, B. Greenspan, R. Spencer Wells, Genographic Consortium, The GenoChip: A new tool for genetic anthropology. *Genome Biol. Evol.* **5**, 1021–1031 (2013). doi:10.1093/gbe/evt066 Medline

56. D. F. O'Shaughnessy, A. V. Hill, D. K. Bowden, D. J. Weatherall, J. B. Clegg, Globin genes in Micronesia: Origins and affinities of Pacific Island peoples. *Am. J. Hum. Genet.* **46**, 144–155 (1990). Medline

57. D. M. Behar, M. van Oven, S. Rosset, M. Metspalu, E.-L. Loogväli, N. M. Silva, T. Kivisild, A. Torroni, R. Villems, A "Copernican" reassessment of the human mitochondrial DNA tree from its root. *Am. J. Hum. Genet.* **90**, 675–684 (2012). doi:10.1016/j.ajhg.2012.03.002 Medline

58. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009). doi:10.1093/bioinformatics/btp324 Medline

59. P. Skoglund, J. Storå, A. Götherström, M. Jakobsson, Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J. Archaeol. Sci.* **40**, 4477–4482 (2013). doi:10.1016/j.jas.2013.07.004

60. A. Furtwängler, E. Reiter, G. U. Neumann, I. Siebke, N. Steuri, A. Hafner, S. Lösch, N. Anthes, V. J. Schuenemann, J. Krause, Ratio of mitochondrial to nuclear DNA affects contamination estimates in ancient DNA analysis. *Sci. Rep.* **8**, 14075 (2018). doi:10.1038/s41598-018-32083-0 Medline

61. G. Renaud, V. Slon, A. T. Duggan, J. Kelso, Schmutzi: Estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* **16**, 224 (2015). doi:10.1186/s13059-015-0776-0 Medline

62. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics* **15**, 356 (2014). doi:10.1186/s12859-014-0356-4 Medline

63. H. Weissensteiner, D. Pacher, A. Kloss-Brandstätter, L. Forer, G. Specht, H.-J. Bandelt, F. Kronenberg, A. Salas, S. Schönherr, HaploGrep 2: Mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* **44**, W58–W63 (2016). doi:10.1093/nar/gkw233 Medline

64. V. M. Narasimhan, N. Patterson, P. Moorjani, N. Rohland, R. Bernardos, S. Mallick, I. Lazaridis, N. Nakatsuka, I. Olalde, M. Lipson, A. M. Kim, L. M. Olivieri, A. Coppa, M. Vidale, J. Mallory, V. Moiseyev, E. Kitov, J. Monge, N. Adamski, N. Alex, N. Broomandkhoshbacht, F. Candilio, K. Callan, O. Cheronet, B. J. Culleton, M. Ferry, D. Fernandes, S. Freilich, B. Gamarra, D. Gaudio, M. Hajdinjak, É. Harney, T. K. Harper, D. Keating, A. M. Lawson, M. Mah, K. Mandl, M. Michel, M. Novak, J. Oppenheimer, N. Rai, K. Sirak, V. Slon, K. Stewardson, F. Zalzala, Z. Zhang, G. Akhatov, A. N. Bagashev, A. Bagnera, B. Baitanayev, J. Bendezu-Sarmiento, A. A. Bissembaev, G. L. Bonora, T. T. Chargynov, T. Chikisheva, P. K. Dashkovskiy, A. Derevianko, M. Dobeš, K. Douka, N. Dubova, M. N. Duisengali, D. Enshin, A. Epimakhov, A. V. Fribus, D. Fuller, A. Goryachev, A. Gromov, S. P. Grushin, B. Hanks, M. Judd, E. Kazizov, A. Khokhlov, A. P. Krygin, E. Kupriyanova, P. Kuznetsov, D. Luiselli, F. Maksudov, A. M. Mamedov, T. B. Mamirov, C. Meiklejohn, D. C. Merrett, R. Micheli, O. Mochalov, S. Mustafokulov, A. Nayak, D. Pettener, R. Potts, D. Razhev, M. Rykun, S. Sarno, T. M. Savenkova, K. Sikhymbaeva, S. M. Slepchenko, O. A. Soltobaev, N. Stepanova, S. Svyatko, K. Tabaldiev, M. Teschler-Nicola, A. A. Tishkin, V. V. Tkachev, S. Vasilyev, P. Velemínský, D. Voyakin, A. Yermolayeva, M. Zahir, V. S. Zubkov, A. Zubova, V. S. Shinde, C. Lalueza-Fox, M. Meyer, D. Anthony, N. Boivin, K. Thangaraj, D. J. Kennett, M. Frachetti, R. Pinhasi, D. Reich, The formation of human populations in South and Central Asia. *Science* **365**, eaat7487 (2019). doi:10.1126/science.aat7487 Medline

65. G. J. van Klinken, Bone collagen quality indicators for palaeodietary and radiocarbon measurements. *J. Archaeol. Sci.* **26**, 687–695 (1999). doi:10.1006/jasc.1998.0385

66. C. B. Ramsey, Radiocarbon calibration and analysis of stratigraphy: The OxCal program. *Radiocarbon* **37**, 425–430 (1995). doi:10.1017/S0033822200030903

67. T. J. Heaton, P. Köhler, M. Butzin, E. Bard, R. W. Reimer, W. E. N. Austin, C. Bronk Ramsey, P. M. Grootes, K. A. Hughen, B. Kromer, P. J. Reimer, J. Adkins, A. Burke, M. S. Cook, J. Olsen, L. C. Skinner, Marine20—The marine radiocarbon age calibration curve (0–55,000 cal BP). *Radiocarbon* **62**, 779–820 (2020). doi:10.1017/RDC.2020.68

68. K. A. Hughen, M. G. L. Baillie, E. Bard, J. Warren Beck, C. J. H. Bertrand, P. G. Blackwell, C. E. Buck, G. S. Burr, K. B. Cutler, P. E. Damon, R. L. Edwards, R. G. Fairbanks, M. Friedrich, T. P. Guilderson, B. Kromer, G. McCormac, S. Manning, C. Bronk Ramsey, P. J. Reimer, R. W. Reimer, S. Remmele, J. R. Southon, M. Stuiver, S. Talamo, F. W. Taylor, J. VAN der Plicht, C. E. Weyhenmeyer, Marine04 marine radiocarbon age

calibration, 0–26 cal kyr BP. *Radiocarbon* **46**, 1059–1086 (2004). doi:10.1017/S0033822200033002

69. P. J. Reimer, M. G. L. Baillie, E. Bard, A. Bayliss, J. W. Beck, P. G. Blackwell, C. Bronk Ramsey, C. E. Buck, G. S. Burr, R. L. Edwards, M. Friedrich, P. M. Grootes, T. P. Guilderson, I. Hajdas, T. J. Heaton, A. G. Hogg, K. A. Hughen, K. F. Kaiser, B. Kromer, F. G. McCormac, S. W. Manning, R. W. Reimer, D. A. Richards, J. R. Southon, S. Talamo, C. S. M. Turney, J. van der Plicht, C. E. Weyhenmeyer, IntCal09 and Marine09 radiocarbon age calibration curves, 0—50,000 years cal BP. *Radiocarbon* **51**, 1111–1150 (2009). doi:10.1017/S0033822200034202

70. P. J. Reimer, E. Bard, A. Bayliss, J. W. Beck, P. G. Blackwell, C. B. Ramsey, C. E. Buck, H. Cheng, R. L. Edwards, M. Friedrich, P. M. Grootes, T. P. Guilderson, H. Haflidason, I. Hajdas, C. Hatté, T. J. Heaton, D. L. Hoffmann, A. G. Hogg, K. A. Hughen, K. F. Kaiser, B. Kromer, S. W. Manning, M. Niu, R. W. Reimer, D. A. Richards, E. M. Scott, J. R. Southon, R. A. Staff, C. S. M. Turney, J. van der Plicht, IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. *Radiocarbon* **55**, 1869–1887 (2013). doi:10.2458/azu_js_rc.55.16947

71. M. Stuiver, G. W. Pearson, T. Braziunas, Radiocarbon age calibration of marine samples back to 9000 cal yr BP. *Radiocarbon* **28**, 980–1021 (1986). doi:10.1017/S0033822200060264

72. E. Alves, K. Macario, R. Souza, O. Aguilera, A. Carolina Goulart, R. Scheel-Ybert, C. Bachelet, C. Carvalho, F. Oliveira, K. Douka, Marine reservoir corrections on the southeastern coast of Brazil: Paired samples from the Saquarema shellmound. *Radiocarbon* **57**, 517–525 (2015). doi:10.2458/azu_rc.57.18404

73. R. W. Reimer, P. J. Reimer, An online application for ΔR calculation. *Radiocarbon* **59**, 1623–1627 (2017). doi:10.1017/RDC.2016.117

74. M. T. Carson, Peopling of Oceania: Clarifying an initial settlement horizon in the Mariana Islands at 1500 BC. *Radiocarbon* **62**, 1733–1754 (2020). doi:10.1017/RDC.2020.89

75. G. K. Ward, S. R. Wilson, Procedures for comparing and combining radiocarbon age determinations: A critique. *Archaeometry* **20**, 19–31 (1978). doi:10.1111/j.1475-4754.1978.tb00208.x

76. F. Petchey, M. Spriggs, S. Bedford, F. Valentin, H. Buckley, Radiocarbon dating of burials from the Teouma Lapita cemetery, Efate, Vanuatu. *J. Archaeol. Sci.* **50**, 227–242 (2014). doi:10.1016/j.jas.2014.07.002

77. K. Prüfer, F. Racimo, N. Patterson, F. Jay, S. Sankararaman, S. Sawyer, A. Heinze, G. Renaud, P. H. Sudmant, C. de Filippo, H. Li, S. Mallick, M. Dannemann, Q. Fu, M. Kircher, M. Kuhlwilm, M. Lachmann, M. Meyer, M. Ongyerth, M. Siebauer, C. Theunert, A. Tandon, P. Moorjani, J. Pickrell, J. C. Mullikin, S. H. Vohr, R. E. Green, I. Hellmann, P. L. F. Johnson, H. Blanche, H. Cann, J. O. Kitzman, J. Shendure, E. E. Eichler, E. S. Lein, T. E. Bakken, L. V. Golovanova, V. B. Doronichev, M. V. Shunkov, A. P. Derevianko, B. Viola, M. Slatkin, D. Reich, J. Kelso, S. Pääbo, The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014). doi:10.1038/nature12886 Medline

78. P. Skoglund, S. Mallick, M. C. Bortolini, N. Chennagiri, T. Hünemeier, M. L. Petzl-Erler, F. M. Salzano, N. Patterson, D. Reich, Genetic evidence for two founding populations of the Americas. *Nature* **525**, 104–108 (2015). doi:10.1038/nature14895 Medline

79. S. Mallick, H. Li, M. Lipson, I. Mathieson, M. Gymrek, F. Racimo, M. Zhao, N. Chennagiri, S. Nordenfelt, A. Tandon, P. Skoglund, I. Lazaridis, S. Sankararaman, Q. Fu, N. Rohland, G. Renaud, Y. Erlich, T. Willems, C. Gallo, J. P. Spence, Y. S. Song, G. Poletti, F. Balloux, G. van Driem, P. de Knijff, I. G. Romero, A. R. Jha, D. M. Behar, C. M. Bravi, C. Capelli, T. Hervig, A. Moreno-Estrada, O. L. Posukh, E. Balanovska, O. Balanovsky, S. Karachanak-Yankova, H. Sahakyan, D. Toncheva, L. Yepiskoposyan, C. Tyler-Smith, Y. Xue, M. S. Abdullah, A. Ruiz-Linares, C. M. Beall, A. Di Rienzo, C. Jeong, E. B. Starikovskaya, E. Metspalu, J. Parik, R. Villems, B. M. Henn, U. Hodoglugil, R. Mahley, A. Sajantila, G. Stamatoyannopoulos, J. T. S. Wee, R. Khusainova, E. Khusnutdinova, S. Litvinov, G. Ayodo, D. Comas, M. F. Hammer, T. Kivisild, W. Klitz, C. A. Winkler, D. Labuda, M. Bamshad, L. B. Jorde, S. A. Tishkoff, W. S. Watkins, M. Metspalu, S. Dryomov, R. Sukernik, L. Singh, K. Thangaraj, S. Pääbo, J. Kelso, N. Patterson, D. Reich, The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016). doi:10.1038/nature18964 Medline

80. A. Bergström, S. A. McCarthy, R. Hui, M. A. Almarri, Q. Ayub, P. Danecek, Y. Chen, S. Felkel, P. Hallast, J. Kamm, H. Blanché, J.-F. Deleuze, H. Cann, S. Mallick, D. Reich, M. S. Sandhu, P. Skoglund, A. Scally, Y. Xue, R. Durbin, C. Tyler-Smith, Insights into human genetic variation and population history from 929 diverse genomes. *Science* **367**, eaay5012 (2020). doi:10.1126/science.aay5012 Medline

81. G. S. Jacobs, G. Hudjashov, L. Saag, P. Kusuma, C. C. Darusallam, D. J. Lawson, M. Mondal, L. Pagani, F.-X. Ricaut, M. Stoneking, M. Metspalu, H. Sudoyo, J. S. Lansing, M. P. Cox, Multiple deeply divergent Denisovan ancestries in Papuans. *Cell* **177**, 1010–1021.e32 (2019). doi:10.1016/j.cell.2019.02.035 Medline

82. A.-S. Malaspinas, O. Lao, H. Schroeder, M. Rasmussen, M. Raghavan, I. Moltke, P. F. Campos, F. S. Sagredo, S. Rasmussen, V. F. Gonçalves, A. Albrechtsen, M. E. Allentoft, P. L. F. Johnson, M. Li, S. Reis, D. V. Bernardo, M. DeGiorgio, A. T. Duggan, M. Bastos, Y. Wang, J. Stenderup, J. V. Moreno-Mayar, S. Brunak, T. Sicheritz-Ponten, E. Hodges, G. J. Hannon, L. Orlando, T. D. Price, J. D. Jensen, R. Nielsen, J. Heinemeier, J. Olsen, C. Rodrigues-Carvalho, M. M. Lahr, W. A. Neves, M. Kayser, T. Higham, M. Stoneking, S. D. J. Pena, E. Willerslev, Two ancient human genomes reveal Polynesian ancestry among the indigenous Botocudos of Brazil. *Curr. Biol.* **24**, R1035–R1037 (2014). doi:10.1016/j.cub.2014.09.078 Medline

83. The 1000 Genomes Project Consortium, A global reference for human genetic variation. *Nature* **526**, 68–74 (2015). doi:10.1038/nature15393 Medline

84. P. H. Sudmant, T. Rausch, E. J. Gardner, R. E. Handsaker, A. Abyzov, J. Huddleston, Y. Zhang, K. Ye, G. Jun, M. H. Fritz, M. K. Konkel, A. Malhotra, A. M. Stütz, X. Shi, F. P. Casale, J. Chen, F. Hormozdiari, G. Dayama, K. Chen, M. Malig, M. J. P. Chaisson, K. Walter, S. Meiers, S. Kashin, E. Garrison, A. Auton, H. Y. K. Lam, X. J. Mu, C. Alkan, D. Antaki, T. Bae, E. Cerveira, P. Chines, Z. Chong, L. Clarke, E. Dal, L. Ding, S. Emery, X. Fan, M. Gujral, F. Kahveci, J. M. Kidd, Y. Kong, E.-W. Lameijer, S.

McCarthy, P. Flicek, R. A. Gibbs, G. Marth, C. E. Mason, A. Menelaou, D. M. Muzny, B. J. Nelson, A. Noor, N. F. Parrish, M. Pendleton, A. Quitadamo, B. Raeder, E. E. Schadt, M. Romanovitch, A. Schlattl, R. Sebra, A. A. Shabalin, A. Untergasser, J. A. Walker, M. Wang, F. Yu, C. Zhang, J. Zhang, X. Zheng-Bradley, W. Zhou, T. Zichner, J. Sebat, M. A. Batzer, S. A. McCarroll, R. E. Mills, M. B. Gerstein, A. Bashir, O. Stegle, S. E. Devine, C. Lee, E. E. Eichler, J. O. Korbel, The 1000 Genomes Project Consortium, An integrated map of structural variation in 2,504 human genomes. *Nature* **526**, 75–81 (2015). doi:10.1038/nature15394 Medline

85. H. McColl, F. Racimo, L. Vinner, F. Demeter, T. Gakuhari, J. V. Moreno-Mayar, G. van Driem, U. Gram Wilken, A. Seguin-Orlando, C. de la Fuente Castro, S. Wasef, R. Shoocongdej, V. Souksavatdy, T. Sayavongkhamdy, M. M. Saidin, M. E. Allentoft, T. Sato, A.-S. Malaspinas, F. A. Aghakhanian, T. Korneliussen, A. Prohaska, A. Margaryan, P. de Barros Damgaard, S. Kaewsutthi, P. Lertrit, T. M. H. Nguyen, H. C. Hung, T. Minh Tran, H. Nghia Truong, G. H. Nguyen, S. Shahidan, K. Wiradnyana, H. Matsumae, N. Shigehara, M. Yoneda, H. Ishida, T. Masuyama, Y. Yamada, A. Tajima, H. Shibata, A. Toyoda, T. Hanihara, S. Nakagome, T. Deviese, A.-M. Bacon, P. Duringer, J.-L. Ponche, L. Shackelford, E. Patole-Edoumba, A. T. Nguyen, B. Bellina-Pryce, J.-C. Galipaud, R. Kinaston, H. Buckley, C. Pottier, S. Rasmussen, T. Higham, R. A. Foley, M. M. Lahr, L. Orlando, M. Sikora, M. E. Phipps, H. Oota, C. Higham, D. M. Lambert, E. Willerslev, The prehistoric peopling of Southeast Asia. *Science* **361**, 88–92 (2018). doi:10.1126/science.aat3628 Medline

86. A. Bergström, S. J. Oppenheimer, A. J. Mentzer, K. Auckland, K. Robson, R. Attenborough, M. P. Alpers, G. Koki, W. Pomat, P. Siba, Y. Xue, M. S. Sandhu, C. Tyler-Smith, A Neolithic expansion, but strong genetic structure, in the independent history of New Guinea. *Science* **357**, 1160–1163 (2017). doi:10.1126/science.aan3842 Medline

87. I. Lazaridis, N. Patterson, A. Mittnik, G. Renaud, S. Mallick, K. Kirsanow, P. H. Sudmant, J. G. Schraiber, S. Castellano, M. Lipson, B. Berger, C. Economou, R. Bollongino, Q. Fu, K. I. Bos, S. Nordenfelt, H. Li, C. de Filippo, K. Prüfer, S. Sawyer, C. Posth, W. Haak, F. Hallgren, E. Fornander, N. Rohland, D. Delsate, M. Francken, J.-M. Guinet, J. Wahl, G. Ayodo, H. A. Babiker, G. Bailliet, E. Balanovska, O. Balanovsky, R. Barrantes, G. Bedoya, H. Ben-Ami, J. Bene, F. Berrada, C. M. Bravi, F. Brisighelli, G. B. J. Busby, F. Cali, M. Churnosov, D. E. C. Cole, D. Corach, L. Damba, G. van Driem, S. Dryomov, J.-M. Dugoujon, S. A. Fedorova, I. Gallego Romero, M. Gubina, M. Hammer, B. M. Henn, T. Hervig, U. Hodoglugil, A. R. Jha, S. Karachanak-Yankova, R. Khusainova, E. Khusnutdinova, R. Kittles, T. Kivisild, W. Klitz, V. Kučinskas, A. Kushniarevich, L. Laredj, S. Litvinov, T. Loukidis, R. W. Mahley, B. Melegh, E. Metspalu, J. Molina, J. Mountain, K. Näkkäläjärvi, D. Nesheva, T. Nyambo, L. Osipova, J. Parik, F. Platonov, O. Posukh, V. Romano, F. Rothhammer, I. Rudan, R. Ruizbakiev, H. Sahakyan, A. Sajantila, A. Salas, E. B. Starikovskaya, A. Tarekegn, D. Toncheva, S. Turdikulova, I. Uktveryte, O. Utevska, R. Vasquez, M. Villena, M. Voevoda, C. A. Winkler, L. Yepiskoposyan, P. Zalloua, T. Zemunik, A. Cooper, C. Capelli, M. G. Thomas, A. Ruiz-Linares, S. A. Tishkoff, L. Singh, K. Thangaraj, R. Villems, D. Comas, R. Sukernik, M. Metspalu, M. Meyer, E. E. Eichler, J. Burger, M. Slatkin, S. Pääbo, J. Kelso, D. Reich, J. Krause, Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014). doi:10.1038/nature13673 Medline

88. P. Qin, M. Stoneking, Denisovan ancestry in East Eurasian and Native American populations. *Mol. Biol. Evol.* **32**, 2665–2674 (2015). doi:10.1093/molbev/msv141 Medline

89. I. Olalde, S. Mallick, N. Patterson, N. Rohland, V. Villalba-Mouco, M. Silva, K. Dulias, C. J. Edwards, F. Gandini, M. Pala, P. Soares, M. Ferrando-Bernal, N. Adamski, N. Broomandkhoshbacht, O. Cheronet, B. J. Culleton, D. Fernandes, A. M. Lawson, M. Mah, J. Oppenheimer, K. Stewardson, Z. Zhang, J. M. Jiménez Arenas, I. J. Toro Moyano, D. C. Salazar-García, P. Castanyer, M. Santos, J. Tremoleda, M. Lozano, P. García Borja, J. Fernández-Eraso, J. A. Mujika-Alustiza, C. Barroso, F. J. Bermúdez, E. Viguera Mínguez, J. Burch, N. Coromina, D. Vivó, A. Cebrià, J. M. Fullola, O. García-Puchol, J. I. Morales, F. X. Oms, T. Majó, J. M. Vergès, A. Díaz-Carvajal, I. Ollich-Castanyer, F. J. López-Cachero, A. M. Silva, C. Alonso-Fernández, G. Delibes de Castro, J. Jiménez Echevarría, A. Moreno-Márquez, G. Pascual Berlanga, P. Ramos-García, J. Ramos-Muñoz, E. Vijande Vila, G. Aguilella Arzo, Á. Esparza Arroyo, K. T. Lillios, J. Mack, J. Velasco-Vázquez, A. Waterman, L. Benítez de Lugo Enrich, M. Benito Sánchez, B. Agustí, F. Codina, G. de Prado, A. Estalrrich, Á. Fernández Flores, C. Finlayson, G. Finlayson, S. Finlayson, F. Giles-Guzmán, A. Rosas, V. Barciela González, G. García Atiénzar, M. S. Hernández Pérez, A. Llanos, Y. Carrión Marco, I. Collado Beneyto, D. López-Serrano, M. Sanz Tormo, A. C. Valera, C. Blasco, C. Liesau, P. Ríos, J. Daura, M. J. de Pedro Michó, A. A. Diez-Castillo, R. Flores Fernández, J. Francès Farré, R. Garrido-Pena, V. S. Gonçalves, E. Guerra-Doce, A. M. Herrero-Corral, J. Juan-Cabanilles, D. López-Reyes, S. B. McClure, M. Merino Pérez, A. Oliver Foix, M. Sanz Borràs, A. C. Sousa, J. M. Vidal Encinas, D. J. Kennett, M. B. Richards, K. Werner Alt, W. Haak, R. Pinhasi, C. Lalueza-Fox, D. Reich, The genomic history of the Iberian Peninsula over the past 8000 years. *Science* **363**, 1230–1234 (2019). doi:10.1126/science.aav4040 Medline

90. N. Patterson, A. L. Price, D. Reich, Population structure and eigenanalysis. *PLOS Genet.* **2**, e190 (2006). doi:10.1371/journal.pgen.0020190 Medline

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

92. N. Patterson, P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, T. Genschoreck, T. Webster, D. Reich, Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012). doi:10.1534/genetics.112.145037 Medline

93. D. Reich, N. Patterson, D. Campbell, A. Tandon, S. Mazieres, N. Ray, M. V. Parra, W. Rojas, C. Duque, N. Mesa, L. F. García, O. Triana, S. Blair, A. Maestre, J. C. Dib, C. M. Bravi, G. Bailliet, D. Corach, T. Hünemeier, M. C. Bortolini, F. M. Salzano, M. L. Petzl-Erler, V. Acuña-Alonzo, C. Aguilar-Salinas, S. Canizales-Quinteros, T. Tusié-Luna, L. Riba, M. Rodríguez-Cruz, M. Lopez-Alarcón, R. Coral-Vazquez, T. Canto-Cetina, I. Silva-Zolezzi, J. C. Fernandez-Lopez, A. V. Contreras, G. Jimenez-Sanchez, M. J. Gómez-Vázquez, J. Molina, A. Carracedo, A. Salas, C. Gallo, G. Poletti, D. B. Witonsky, G. Alkorta-Aranburu, R. I. Sukernik, L. Osipova, S. A. Fedorova, R. Vasquez, M. Villena, C. Moreau, R. Barrantes, D. Pauls, L. Excoffier, G. Bedoya, F. Rothhammer, J.-M. Dugoujon, G. Larrouy, W. Klitz, D. Labuda, J. Kidd, K. Kidd, A. Di Rienzo, N. B.

Freimer, A. L. Price, A. Ruiz-Linares, Reconstructing Native American population history. *Nature* **488**, 370–374 (2012). doi:10.1038/nature11258 Medline

94. H. Ringbauer, J. Novembre, M. Steinrücken, Human parental relatedness through time-detecting runs of homozygosity in ancient DNA. bioRxiv 2020.05.31.126912 [Preprint] (2020); https://doi.org/10.1101/2020.05.31.126912.

95. J. S. Friedlaender, F. R. Friedlaender, F. A. Reed, K. K. Kidd, J. R. Kidd, G. K. Chambers, R. A. Lea, J.-H. Loo, G. Koki, J. A. Hodgson, D. A. Merriwether, J. L. Weber, The genetic structure of Pacific Islanders. *PLOS Genet.* **4**, e19 (2008). doi:10.1371/journal.pgen.0040019 Medline

96. M. T. Carson, Palaeohabitat of first settlement sites 1500–1000 BC in Guam, Mariana Islands, western Pacific. *J. Archaeol. Sci.* **38**, 2207–2221 (2011). doi:10.1016/j.jas.2011.03.021

97. J. R. Amesbury, D. R. Moore, R. L. Hunter-Anderson, Cultural adaptations and late Holocene sea level change in the Marianas: Recent excavations at Chalan Piao, Saipan, Micronesia. *Bull. Indo-Pacific Prehistory Assoc.* **15**, 53–69 (1996).

98. J. L. Craib, Early occupation at Unai Chulu, Tinian, Commonwealth of the Northern Mariana Islands. *Bull. Indo-Pacific Prehistory Assoc.* **13**, 116–134 (1993).

99. P. Bellwood, R. Wood, G. Irwin, A. Waluyo, "Excavations in the Uattamdi rockshelters, Kayoa Island" in *The Spice Islands in Prehistory Archaeology in the Northern Moluccas, Indonesia*, vol. 50 of *Terra Australis*, P. Bellwood, Ed. (ANU Press, 2019), pp. 67–76.

100. H. Hung, M. T. Carson, P. Bellwood, F. Z. Campos, P. J. Piper, E. Dizon, M. J. L. A. Bolunia, M. Oxenham, Z. Chi, The first settlement of Remote Oceania: The Philippines to the Marianas. *Antiquity* **85**, 909–926 (2011). doi:10.1017/S0003598X00068393

101. W. R. Dickinson, Hydro-isostatic and tectonic influences on emergent Holocene paleoshorelines in the Mariana Islands, western Pacific Ocean. *J. Coast. Res.* **16**, 735–746 (2000).

102. W. R. Dickinson, Impact of mid-Holocene hydro-isostatic highstand in regional sea level on habitability of islands in Pacific Oceania. *J. Coast. Res.* **19**, 489–502 (2003).

103. C. Higham, T. Higham, A. Kijngam, Cutting a Gordian Knot: the Bronze Age of Southeast Asia: origins, timing and impact. *Antiquity* **85**, 583–598 (2011). doi:10.1017/S0003598X00067971

104. E. E. Cochrane, T. M. Rieth, D. Filimoehala, The first quantitative assessment of radiocarbon chronologies for initial pottery in Island Southeast Asia supports multi-directional Neolithic dispersal. *PLOS ONE* **16**, e0251407 (2021). doi:10.1371/journal.pone.0251407 Medline

105. R. L. Hunter-Anderson, Cultural responses to a Late Holocene climatic oscillation in the Mariana Islands, Micronesia: Lessons from the past. *Hum. Ecol. Rev.* **17**, 148–159 (2010).

106. D. R. Moore, What's new and what's cooking in the *latte* period pots. *Micronesica* **42**, 121–147 (2012).

107. U. Büntgen, V. S. Myglan, F. C. Ljungqvist, M. McCormick, N. Di Cosmo, M. Sigl, J. Jungclaus, S. Wagner, P. J. Krusic, J. Esper, J. O. Kaplan, M. A. C. de Vaan, J. Luterbacher, L. Wacker, W. Tegel, A. V. Kirdyanov, Cooling and societal change during the Late Antique Little Ice Age from 536 to around 660 AD. *Nat. Geosci.* **9**, 231–236 (2016). doi:10.1038/ngeo2652

108. S. Helama, P. D. Jones, K. R. Briffa, Dark ages cold period: A literature review and directions for future research. *Holocene* **27**, 1600–1606 (2017). doi:10.1177/0959683617693898

109. R. L. Hunter-Anderson, M. Yousuf, M. W. Rowe, Pictographs from Mahlac Cave, Guam: Radiocarbon dating and chemical studies. *Am. Indian Rock Art* **40**, 995–1016 (2013).

110. M. T. Carson, An overview of *latte* period archaeology. *Micronesica* **42**, 1–79 (2012).

111. R. L. Hunter-Anderson, G. B. Thompson, D. R. Moore, Rice as a prehistoric valuable in the Mariana Islands, Micronesia. *Asian Perspect.* **34**, 69–89 (1995).

112. J. Chappell, Sea levels and sediments: Some features of the context of coastal archaeological sites in the tropics. *Archaeol. Ocean.* **17**, 69–78 (1982). doi:10.1002/j.1834-4453.1982.tb00040.x

113. W. H. Easton, T. L. Ku, R. H. Randall, Recent reefs and shorelines of Guam. *Micronesica* **14**, 1–11 (1978).

114. M. T. Carson, Paleo-terrain research: Finding the first settlement sites of Remote Oceania. *Geoarchaeology* **29**, 268–275 (2014). doi:10.1002/gea.21457

115. W. H. Alkire, *An Introduction to the Peoples and Cultures of Micronesia* (Cummings Publishing Company, 1977).

116. J. M. Davidson, Preliminary archaeological investigations on Ponape and other eastern Caroline Islands. *Micronesica* **3**, 81–95 (1967).

117. M. Pietrusewsky, L. M. Willacker, The search for Father Bachelot: First Catholic missionary to the Hawaiian Islands (1827–1837). *J. Forensic Sci.* **42**, 208–212 (1997). doi:10.1520/JFS14099J Medline

118. M. D. McCoy, J. S. Athens, Sourcing the megalithic stones of Nan Madol: An XRF study of architectural basalt stone from Pohnpei, Federated States of Micronesia. *J. Pacific Archaeol.* **3**, 105–114 (2012).

119. M. D. McCoy, H. A. Alderson, R. Hemi, H. Cheng, R. L. Edwards, Earliest direct evidence of monument building at the archaeological site of Nan Madol (Pohnpei, Micronesia) identified using 230 Th/U coral dating and geochemical sourcing of megalithic architectural stone. *Quat. Res.* **86**, 295–303 (2016). doi:10.1016/j.yqres.2016.08.002

120. J. S. Athens, "Archaeology of the Eastern Caroline Islands, Micronesia" in *The Oxford Handbook of Prehistoric Oceania*, E. E. Cochrane, T. L. Hunt, Eds. (Oxford Univ. Press, 2018), pp. 271–301.

121. K. Seikel, Mortuary contexts and social structure at Nan Madol, Pohnpei. *J. Island Coast. Archaeol.* **6**, 442–460 (2011). doi:10.1080/15564894.2011.559615

122. J. S. Athens, "The rise of the Saudeleur: dating the Nan Madol chiefdom, Pohnpei." in

Vastly Ingenious: The Archaeology of Pacific Material Culture in Honour of Janet M. Davidson, A. Anderson, K. Green, F. Leach, Eds. (Otago Univ. Press, 2007), pp. 191–208.

123. J. E. Bath, J. S. Athens, Prehistoric social complexity on Pohnpei: The Saudeleur to Nahnmwarki transformation. *Micronesica* **2**, 275–290 (1990).

124. D. L. Hanlon, *Upon a Stone Altar: A History of the Island of Pohnpei to 1890* (Pacific Islands Monograph Series, Univ. Hawaii Press, 1988), vol. 5.

125. M. Lipson, Applying f_4-statistics and admixture graphs: Theory and examples. *Mol. Ecol. Resour.* **20**, 1658–1667 (2020). doi:10.1111/1755-0998.13230 Medline

126. M. P. Cox, T. M. Karafet, J. S. Lansing, H. Sudoyo, M. F. Hammer, Autosomal and X-linked single nucleotide polymorphisms reveal a steep Asian-Melanesian ancestry cline in eastern Indonesia and a sex bias in admixture rates. *Proc. Biol. Sci.* **277**, 1589–1596 (2010). doi:10.1098/rspb.2009.2041 Medline

127. R. Blust, *The Austronesian Languages* (revised edition) (ANU Press, 2013).