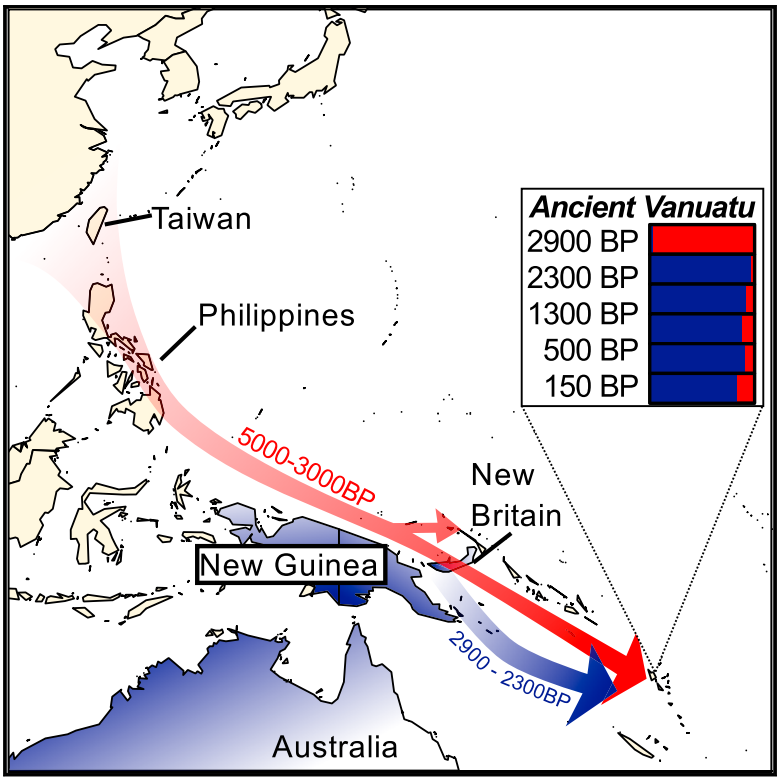
Report

Current Biology

Population Turnover in Remote Oceania Shortly after Initial Settlement

Graphical Abstract



Highlights

- The population of Vanuatu in the Pacific was largely replaced 2,900–2,300 years ago
- This second wave of migrants came from New Britain, east of New Guinea
- A third wave spread different ancestry to the far-flung islands of Polynesia

Authors

Mark Lipson, Pontus Skoglund, Matthew Spriggs, ..., Alexander J. Mentzer, Ron Pinhasi, David Reich

Correspondence

mlipson@genetics.med.harvard.edu (M.L.), reich@genetics.med.harvard.edu (D.R.)

In Brief

Lipson, Skoglund, et al. analyze ancient DNA from the Pacific island chain of Vanuatu over its entire span of occupation. After humans first arrived around 3,000 years ago, there was a nearly complete replacement of the original inhabitants by 2,300 years ago, and this second wave forms the primary ancestry of people in Vanuatu today.

Lipson et al., 2018, Current Biology *28*, 1–9
April 2, 2018
https://doi.org/10.1016/j.cub.2018.02.051

CellPress

Please cite this article in press as: Lipson et al., Population Turnover in Remote Oceania Shortly after Initial Settlement, Current Biology (2018), https://doi.org/10.1016/j.cub.2018.02.051

Current Biology
Report

Population Turnover in Remote Oceania Shortly after Initial Settlement

Mark Lipson,[1,19,*] Pontus Skoglund,[1,2,19] Matthew Spriggs,[3,4] Frederique Valentin,[5] Stuart Bedford,[4,6] Richard Shing,[4] Hallie Buckley,[7] Iarawai Phillip,[4] Graeme K. Ward,[6] Swapan Mallick,[1,8,9] Nadin Rohland,[1] Nasreen Broomandkhoshbacht,[1,8] Olivia Cheronet,[10,11] Matthew Ferry,[1,8] Thomas K. Harper,[12] Megan Michel,[1,8] Jonas Oppenheimer,[1,8] Kendra Sirak,[11,13] Kristin Stewardson,[1,8] Kathryn Auckland,[14] Adrian V.S. Hill,[14] Kathryn Maitland,[15] Stephen J. Oppenheimer,[16] Tom Parks,[14] Kathryn Robson,[17] Thomas N. Williams,[15] Douglas J. Kennett,[12] Alexander J. Mentzer,[14] Ron Pinhasi,[10,11,20] and David Reich[1,8,9,18,20,21,*]

[1]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA
[2]The Francis Crick Institute, London NW1 1AT, UK
[3]School of Archaeology and Anthropology, College of Arts and Social Sciences, The Australian National University, Canberra, ACT 2601, Australia
[4]Vanuatu National Museum, Vanuatu Cultural Centre, P.O. Box 184, Port Vila, Vanuatu
[5]Maison de l'Archéologie et de l'Ethnologie, CNRS, UMR 7041, 92023 Nanterre, France
[6]Department of Archaeology and Natural History, College of Asia-Pacific, The Australian National University, Canberra, ACT 2601, Australia
[7]Department of Anatomy, Otago Global Health Institute, School of Biomedical Sciences, University of Otago, Dunedin 9054, New Zealand
[8]Howard Hughes Medical Institute, Boston, MA 02115, USA
[9]Medical and Population Genetics Program, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[10]Department of Anthropology, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria
[11]Earth Institute, University College Dublin, Dublin 4, Ireland
[12]Department of Anthropology and Institute for Energy and the Environment, The Pennsylvania State University, University Park, PA 16802, USA
[13]Department of Anthropology, Emory University, Atlanta, GA, USA
[14]Wellcome Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK
[15]Department of Paediatrics, Faculty of Medicine, Imperial College, St Mary's Hospital, Norfolk Place, Paddington, London W2 1PG, UK
[16]School of Anthropology and Museum Ethnography, University of Oxford, Oxford OX2 6PE, UK
[17]MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford OX3 9DS, UK
[18]Max Planck-Harvard Research Center for the Archaeoscience of the Ancient Mediterranean, Cambridge, MA 02138, USA
[19]These authors contributed equally
[20]Senior author
[21]Lead Contact
*Correspondence: mlipson@genetics.med.harvard.edu (M.L.), reich@genetics.med.harvard.edu (D.R.)
https://doi.org/10.1016/j.cub.2018.02.051

SUMMARY

Ancient DNA from Vanuatu and Tonga dating to about 2,900–2,600 years ago (before present, BP) has revealed that the "First Remote Oceanians" associated with the Lapita archaeological culture were directly descended from the population that, beginning around 5000 BP, spread Austronesian languages from Taiwan to the Philippines, western Melanesia, and eventually Remote Oceania. Thus, ancestors of the First Remote Oceanians must have passed by the Papuan-ancestry populations they encountered in New Guinea, the Bismarck Archipelago, and the Solomon Islands with minimal admixture [1]. However, all present-day populations in Near and Remote Oceania harbor >25% Papuan ancestry, implying that additional eastward migration must have occurred. We generated genome-wide data for 14 ancient individuals from Efate and Epi Islands in Vanuatu from 2900–150 BP, as well as 185 present-day individuals from 18 islands. We find that people of almost entirely Papuan ancestry arrived in Vanuatu by around 2300 BP, most likely reflecting migrations a few hundred years earlier at the end of the Lapita period, when there is also evidence of changes in skeletal morphology and cessation of long-distance trade between Near and Remote Oceania [2, 3]. Papuan ancestry was subsequently diluted through admixture but remains at least 80%–90% in most islands. Through a fine-grained analysis of ancestry profiles, we show that the Papuan ancestry in Vanuatu derives from the Bismarck Archipelago rather than the geographically closer Solomon Islands. However, the Papuan ancestry in Polynesia—the most remote Pacific islands—derives from different sources, documenting a third stream of migration from Near to Remote Oceania.

RESULTS AND DISCUSSION

We generated genome-wide data for 14 ancient individuals from central Vanuatu, including 11 newly reported individuals and higher-quality data for three previously reported individuals [1]

(Table 1; Data S1). We identified and selected cochlear bone sections of petrous bones and processed them into powder in dedicated clean rooms at University College Dublin [4]. We shipped the powder to Harvard Medical School, where in a second set of clean rooms we extracted DNA [5, 6] and created individually barcoded Illumina sequencing libraries, some of which we treated with the enzyme uracil-DNA glycosylase (UDG) to greatly reduce the characteristic errors associated with ancient DNA [7, 8]. We screened these libraries for evidence of authentic ancient DNA by enriching for DNA overlapping the mitochondrial genome [9], sequencing on an Illumina NextSeq500 instrument, and measuring the rates of cytosine-to-thymine damage in the terminal nucleotide and consistency with the consensus mitochondrial genome (STAR Methods) [10]. For libraries that were promising after screening, we enriched for regions targeting approximately 1.24 million single-nucleotide polymorphisms (SNPs) and sequenced the enriched products (STAR Methods). We determined sex by examining the ratio of sequences overlapping the X and Y chromosomes, and for males, we estimated nuclear contamination based on the rate of apparent polymorphism on the X chromosome (present in only one copy in males) (STAR Methods; Data S1). The data for the 14 individuals passing quality control were derived from 46 Illumina libraries (one to eight per individual; Data S1). We assembled direct accelerator mass spectrometry radiocarbon dates for all 14 individuals, including ten newly reported dates (STAR Methods; Data S1). Finally, we generated genome-wide SNP genotype data on the Affymetrix Human Origins array for 185 present-day individuals from Vanuatu who gave informed consent for studies of genetic variation (STAR Methods; Data S1).

Genome-wide Clustering Analyses

We performed automated clustering analysis with the ADMIXTURE software [11], using a dataset consisting of the ancient and present-day Vanuatu samples together with other Oceanian, East Asian, and worldwide populations genotyped on the Human Origins array [1]. At K = 8 clusters, four ancestry components were inferred to be widespread in Oceania (Figure 1A; Figure S1). Three correlate (predominantly) to Papuan ancestry and are maximized in New Guinea (black in the plot), Mamusi and Baining from New Britain (blue), and Nasioi from Bougainville in the Solomon Islands (red). The fourth component (green), correlating to First Remote Oceanian ancestry, is maximized in the ~2900–2600 BP (before present) individuals from Vanuatu and Tonga. Other Oceanian populations display variable combinations of these components, forming gradients of ancestry between New Guinea, New Britain and New Ireland in the Bismarck Archipelago, and the Solomon Islands. The great majority of present-day and ancient groups from Vanuatu show similar ratios of the three Papuan ancestry components (although their First Remote Oceanian proportions vary), showing that they are consistent with largely deriving their Papuan ancestry from the same source. Among populations in Near Oceania, the most similar to Vanuatu in terms of the Papuan ancestry component ratio (black:blue:red) are groups from New Britain in the Bismarck Archipelago, with a majority of the blue component and smaller contributions of black and red, suggesting that the Papuan ancestry in Vanuatu derives from populations in the Bismarck Archipelago (rather than the geographically closer Solomon Islands). A similar pattern was previously inferred for the Papuan ancestry in Santa Cruz, to the immediate north of Vanuatu [12], a result that we replicate here.

We also carried out a principal-component analysis (PCA; Figure S2), which corroborated the findings from ADMIXTURE, with the primary feature being a U-shaped cline from (1) western New Britain in the Bismarck archipelago to (2) eastern New Britain, (3) most of Vanuatu, (4) the atypical Vanuatu island of Tutuba along with the Tolai of New Britain, (5) New Ireland in the Bismarck archipelago, and finally (6) Bougainville in the Solomon Islands. This cline closely correlates to the gradient of decreasing blue and increasing red components in ADMIXTURE (Figure 1A; Figure S1). The position of the Vanuatu samples in the PCA again supports the hypothesis that the inhabitants of the region after the initial Lapita settlement derived ultimately not from populations closely related to those in the closer Solomon Islands but instead from populations related to those from the island of New Britain in the Bismarck Archipelago. We also replicated this result via the statistic f_4(Australian, *Vanuatu*; *Solomon Islands*, *Bismarck Archipelago*), which is significantly positive for each choice of populations in the PCA ($Z > 2$ for all 160 comparisons; median $Z > 6$; STAR Methods), implying that Vanuatu populations share more alleles with groups from the Bismarck Archipelago than the Solomon Islands.

Papuan and First Remote Oceanian Ancestry Proportions

It has been shown that the strongest driver of genetic variation in Oceania today is the widespread but highly variable admixture between Papuan and First Remote Oceanian ancestry sources, the former representing original inhabitants of Near Oceania and the latter descendants of an expansion from East and Southeast Asia [1]. From our clustering results, a dramatic turnover is apparent in Vanuatu after around 2,900 and before around 2,300 years ago, with First Remote Oceanian populations being joined or possibly completely replaced by individuals of (almost) entirely Papuan ancestry. To provide precise estimates of mixture proportions, we used f_4-ratio statistics [13], with East Asian reference populations Atayal (aboriginal Taiwanese related to the source population of the Austronesian expansion) and Kankanaey (an Austronesian-speaking population from the Philippines—on the migratory path from Taiwan to Remote Oceania—that has been shown to be descended from the same genetic sources as the First Remote Oceanians) [1] (Figure 1; Table 1; Data S1; STAR Methods). Taking advantage of our increased coverage compared to the first study of Lapita samples, we find that the ~2900 BP Lapita individuals had a nonzero proportion of Papuan-related ancestry (2.4% ± 0.9%), although it remains striking that the initial First Remote Oceanian migrants were only minimally admixed. Given the small proportion, we did not have sufficient statistical power to determine whether this Papuan-related ancestry is derived from the region surrounding New Guinea or could perhaps have been acquired elsewhere, such as in the Philippines or eastern Indonesia. Notably, the first post-Lapita sample (2300 BP from the site of Mele-Taplins) had almost entirely Papuan ancestry, but with a small fraction derived from First Remote Oceanians (4.2% ± 1.1%). The more recent ancient individuals are similar in their proportions to present-day populations: 8%–26% First Remote

Table 1. Details of Ancient Vanuatu Samples Analyzed in This Study

Sample	Code	Radiocarbon Dates	Population Label	Location	Sex	mtDNA Haplogroup	Y Haplogroup	SNPs	Papuan Ancestry (%)
I1370	B17.P3	3080–2780 calBP (3083 ± 26 BP, Wk-21026)	Vanuatu_2900BP	Teouma, Efate	F	B4a1a1	–	237,405	−0.2 ± 1.4
I1369	B10B.P3	3020–2750 calBP (3045 ± 30 BP, Poz-81126)	Vanuatu_2900BP	Teouma, Efate	F	B4a1a1	–	271,048	3.0 ± 1.4
I1368	B30A.P3	2990–2750 calBP (2983 ± 32 BP, Wk-22657)	Vanuatu_2900BP	Teouma, Efate	F	B4a1a1	–	185,282	1.6 ± 1.4
I5951	TeoQE	2920–2720 calBP (2955 ± 20 BP, PSUAMS-2411)	Vanuatu_2900BP	Teouma Quarry Edge	M	B4a1a1	*CT*	23,107	3.9 ± 3.5
I4451	TAP1	2360–2160 calBP (2348 ± 32 BP, Wk-20390)	Vanuatu_2300BP	Mele-Taplins, Efate	M	M28a7	K2b1	340,152	95.8 ± 1.1
I4096	BURU5B	1380–1240 calBP (1380–1270 calBP [1490 ± 15 BP, PSUAMS-2460], 1350–1180 calBP [1464 ± 30 BP, Wk-25769])	Vanuatu_1300BP	Burumba, Epi Island	M	B4a1a1k	K2b1	888,003	92.2 ± 1.2
I3921	BURU5D	1340–1180 calBP (1410–1280 calBP [1530 ± 20 BP, PSUAMS-1841], 1300–1170 calBP [1395 ± 15 BP, PSUAMS-2428])	Vanuatu_1300BP	Burumba, Epi Island	M	P1d1	K2b1	855,305	91.6 ± 1.1
I5259	Mang1	630–330 calBP (559 ± 30 BP, Wk-20030)	Vanuatu_500BP	Mangaliliu, Efate	F	P1f	–	799,098	88.4 ± 1.2
I4105	WAMB1	300–0 calBP (255 ± 20 BP, PSUAMS-1922)	Vanuatu_Epi_150BP	Wambi Bay, Epi Island	M	M28a+204	O1a2	1,012,081	92.6 ± 1.2
I4106	WAMB2	280–0 calBP (225 ± 20 BP, PSUAMS-1923)	Vanuatu_Epi_150BP	Wambi Bay, Epi Island	M	B4a1a1a11	O1a2	1,020,436	90.6 ± 1.1
I4450	SEPU1	430–0 calBP (305 ± 15 BP, UCIAMS-188793)	Vanuatu_Efate_150BP	Pangpang, Efate	F	P1d2	–	735,460	90.8 ± 1.2
I4425	EF3_2_E	270–0 calBP (200 ± 20 BP, UCIAMS-188795)	Vanuatu_Efate_150BP	Ifira, Efate	F	P2	–	700,783	73.6 ± 1.4
I4424	EF_Pango1	280–0 calBP (190 ± 15 BP, UCIAMS-188794)	Vanuatu_Efate_150BP	Pango Village, Efate	M	B4a1a1	M1b	780,469	78.6 ± 1.4
I4419	BB1	260–20 calBP (135 ± 15 BP, UCIAMS-188792)	Vanuatu_Efate_150BP	Banana Bay, Efate	M	B4a1a1	K2b1	763,556	86.2 ± 1.2

The first three samples listed are previously published individuals [1] but with new libraries now added to increase coverage; the other 11 are newly published individuals. Radiocarbon date calibrations are given as 95% confidence intervals, after applying a correction for marine reservoir effect (STAR Methods; Data S1). Mitochondrial DNA haplogroups were called after merging data from all libraries. For the mtDNA and Y chromosome columns, underlining indicates typical East Asian (First Remote Oceanian) haplogroups, whereas lack of underlining indicates typical Australo-Papuan haplogroups (the italicized Y haplogroup *CT* is unclassified). BP, before present; calBP, calibrated years BP. See also Data S1.

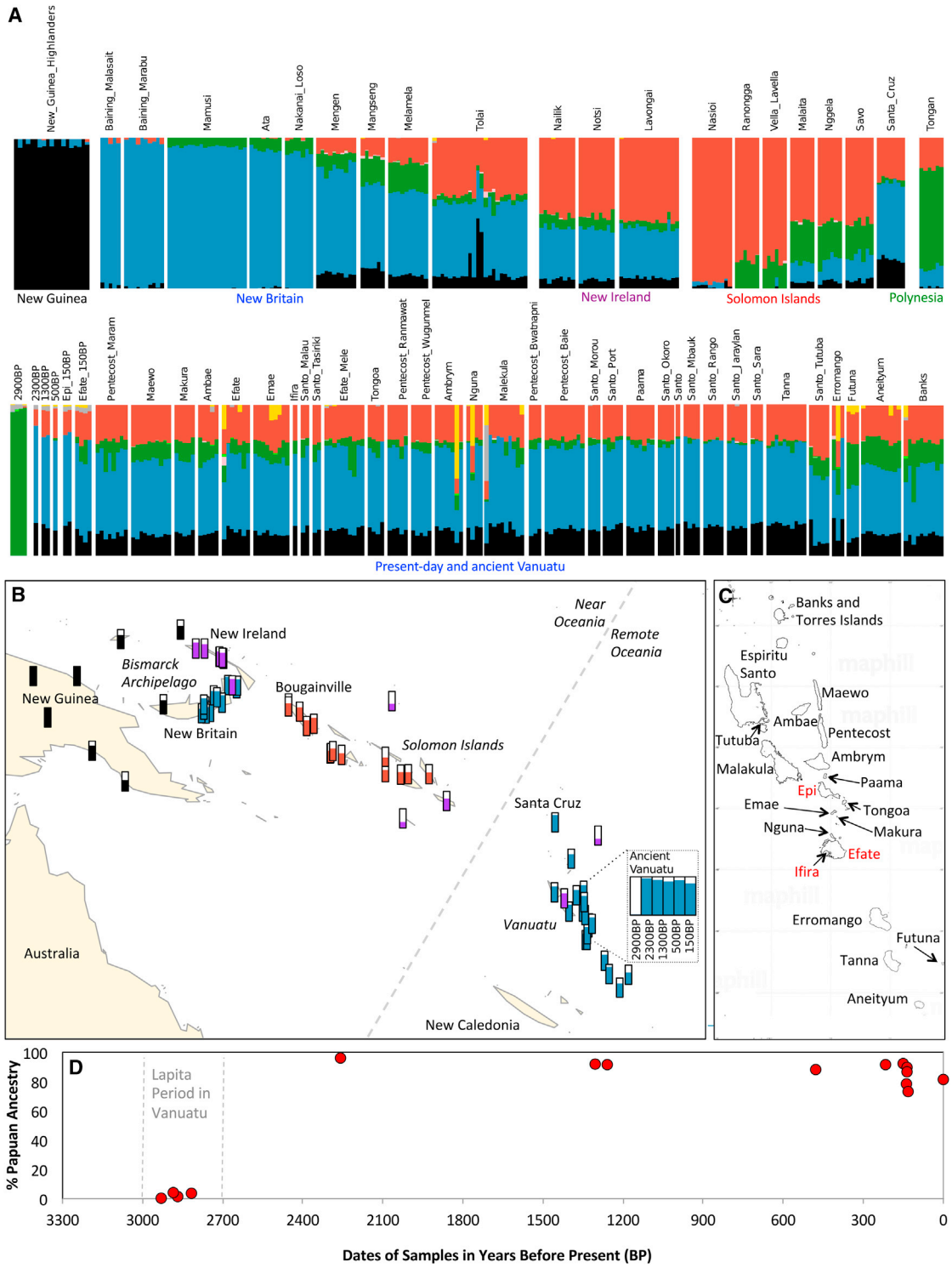


Figure 1. Locations and Broad-Scale Genetic Structure of Analyzed Populations

(A) ADMIXTURE results with K = 8 clusters for selected populations (full results are in Figure S1). The analysis suggests three primary Papuan components (black, maximized in New Guinea; blue, maximized in New Britain in the Bismarck Archipelago; red, maximized in the Solomon Islands) and a component maximized in First Remote Oceanians (green).

(B) Population locations with colored clusters assigned based on the ratio of Papuan ancestry components in ADMIXTURE. We loosely adopt the color scheme from ADMIXTURE, with black indicating a New Guinea-like profile, blue a New Britain-like profile, red a Solomon Islands-like profile, and purple a profile mixed

(legend continued on next page)

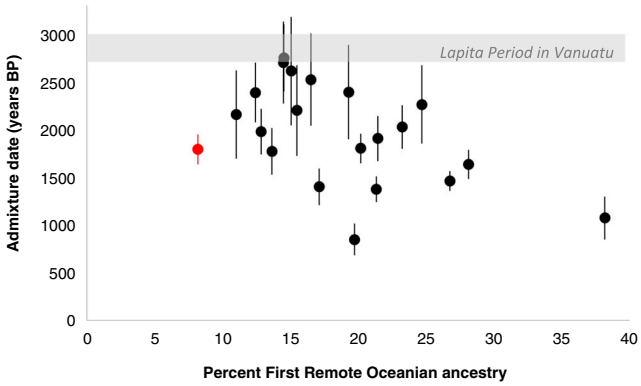


Figure 2. Ancestry Proportions and Dates of Admixture in Vanuatu

Black points represent the 20 present-day populations with the most confident admixture date estimates from ALDER (as measured by *Z* score for difference from zero), assuming 28 years per generation and showing 1 SE in each direction. The red point represents a pair of ancient individuals with a mean calibrated date of 1283 BP (18 ± 6 generations, or 1,800 ± 160 years BP). Gray shading indicates the Lapita period in Vanuatu (~3000–2700 BP). See also Data S1.

Oceanian ancestry, as compared to a range of 9%–38% today (mostly 12%–20%, and highest on the island of Futuna, which harbors a "Polynesian Outlier" population, that is, one that speaks a Polynesian language, potentially due to east-to-west back migration from Polynesia [14]). For time points with multiple samples, the individuals' mixture proportions are statistically indistinguishable, except for 150 BP Efate (point estimates of 9%, 14%, 21%, and 26% First Remote Oceanian). The post-Lapita ancestry turnover is also evident in uniparental markers, as the majority of mtDNA and Y chromosome haplogroups observed from 2300–150 BP are typical of Papuan populations, albeit with the presence of some East Asian-derived haplogroups in both mtDNA and Y, showing that members of both sexes in both ancestral populations participated in the post-2300 BP Vanuatu admixture process (Table 1).

Dates of Admixture

We estimated dates of admixture based on weighted admixture linkage disequilibrium (LD) [15] using ALDER [16], with Ami (aboriginal Taiwanese) and New Guinea Highlanders as references (Figure 2; Data S1). We obtain significant evidence for admixture LD in almost all present-day populations and three ancient population groupings (noting that power is highly dependent on sample size). The date estimates are mostly 40–100 generations before present, or 1,100–2,800 years ago assuming 28 years per generation [17], consistent with admixture having occurred soon after the early settlement of Vanuatu and continuing through time (in cases of multiple pulses of admixture, ALDER produces a single average date). We observe a significant negative correlation between admixture date and First Remote Oceanian ancestry proportion ($R^2 = 0.33$ at $p < 0.01$ for populations in Figure 2; $R^2 = 0.21$ at $p < 0.01$ for all present-day populations; STAR Methods), as expected if a subset of populations (e.g., Efate, Emae, Futuna, and Makura) received more recent pulses of gene flow from groups with high proportions of First Remote Oceanian ancestry. This scenario is plausible in light of Polynesian (Samoan) cultural influences and language replacement and the establishment of Polynesian Outlier populations on islands such as Ifira and Emae in central Vanuatu and Futuna in southern Vanuatu within the last several hundred years [14, 18].

We also obtain a direct ALDER date of 18 ± 6 generations in the past (500 ± 160 years) for a pair of ancient samples from Vanuatu radiocarbon dated to ~1,300 years ago, coinciding with the typical range of admixture dates in present-day groups (Figure 2). Together with the present-day results, this observation is relevant to the ongoing debate about the timing of admixture between people of East Asian and Papuan ancestry in Remote Oceania. Methods based on wavelet transformations have suggested mixing at a date older than 3000 BP, prior to the Lapita expansion to Remote Oceania [12, 19], whereas methods based on admixture LD have suggested more recent dates, implying that mixture occurred after later streams of gene flow [20]. It has been argued that the differences may reflect systematic biases of the methods for dates older than a couple of thousand years [12]. Thus, our finding of a definitively post-Lapita date in samples that are closer in time to the admixture provides compelling evidence for the hypothesis of more recent mixture. A plausible scenario is that the initial migration of Papuan populations occurred during the late Lapita period (before ~2700 BP), at which time archaeological evidence such as the transport of New Britain obsidian to Vanuatu documents links between the New Britain region and Remote Oceania (Santa Cruz and Vanuatu) bypassing much of the Solomon Islands [3], a pattern very similar to the population affinities seen in the genetic data. The near-complete population turnover attested to by genetic data may thus correspond to the evidence of transformation at the end of the Lapita period to more localized cultures, initiating a period of hundreds of years when inter-archipelago contacts appear to have nearly ceased [3].

Phylogeny of First Remote Oceanian Ancestry

To test whether the First Remote Oceanian ancestry in ancient and present-day groups is more closely related to Lapita samples from Tonga or Vanuatu, we compared the values of the statistics f_4(*Test*, Han; Atayal, Tonga_2600BP) and f_4(*Test*, Han; Atayal, Vanuatu_2900BP) for Oceanian populations as *Test* (STAR Methods). We found a trend toward greater allele sharing with Tonga, with significant results in Polynesian and to a lesser degree Polynesian Outlier populations (Data S1). These results show that the First Remote Oceanian ancestry in Polynesians today is derived from a source that was closer to the sampled

between New Britain and the Solomon Islands. The level of fill in the bars indicates Papuan ancestry proportion. The map was plotted in R using the "maps" package with data from http://www.naturalearthdata.com/.

(C) Close-up view of Vanuatu labeling islands for which new data are reported. Islands with ancient samples are indicated in red (for Efate and Ifira, both present-day and ancient individuals). The map was downloaded from http://www.maphill.com/vanuatu/simple-maps/blank-map/no-labels/.

(D) Papuan ancestry proportions for ancient samples over time; Efate Island is used to represent the present.

See also Figure S2 and Data S1.

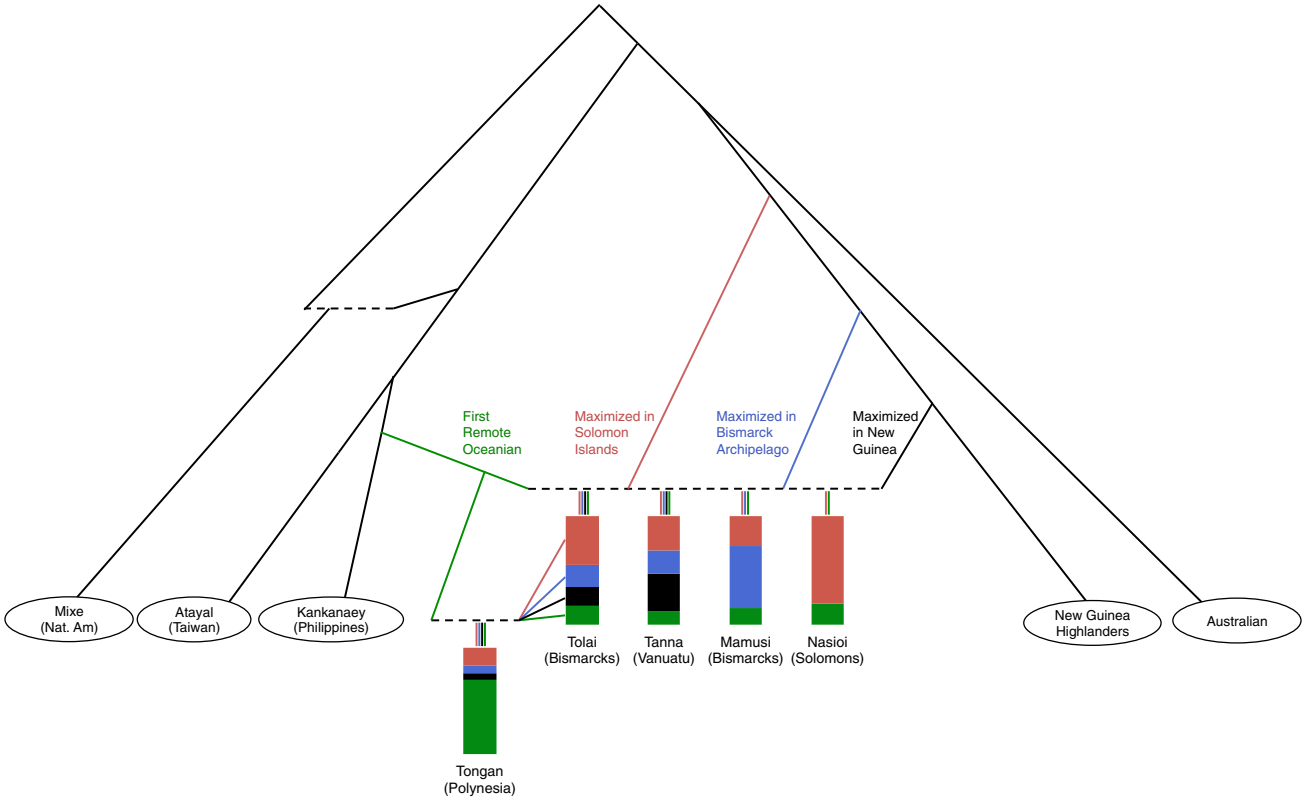


Figure 3. Inferred Admixture Graph Model with Diverse Present-Day Oceanian Populations

Dotted lines denote admixture events. For five populations, the proportions of four fitted ancestry sources maximized in First Remote Oceanians (green), the Solomon Islands (red), the Bismarck Archipelago (blue), and New Guinea (black) are shown. Papuan ancestry is inferred to be highly simlar in the Tolai and in the Tongan pouplation, allowing Tongan to be fit as a mixture of a group with ancestry similar to Tolai and additional ancestry from First Remote Oceanians. Colors were chosen to be correlated to the components inferred from ADMIXTURE (Figure S1), but the ADMIXTURE components represent combinations of the sources given here, and hence the ratios differ between the methods. Full model parameters for the admixture graph are shown in Figure S3.

Lapita-period population from Tonga than to the Vanuatu Lapita population. For post-Lapita populations (ancient and present-day) from Vanuatu, however, we do not have sufficient statistical power to determine which potential First Remote Oceanian source is closer.

Phylogeny of Papuan Ancestry

We built admixture graphs to explore in more detail the different streams of Papuan ancestry present in Oceania. We used as reference populations Australia, Kankanaey, Atayal, and Mixe together with representatives of major poles of Papuan genetic variation inferred from the ADMIXTURE analysis: Vanuatu_Tanna, Mamusi (New Britain), Nasioi (Solomon Islands), New Guinea Highlanders, and Tolai (New Britain and/or New Ireland). To avoid overfitting, we adopted a restricted framework in which the ancestry in each population was modeled as a combination of the same set of source lineages, with the exception of the unadmixed New Guinea Highlander population. We found that three Papuan source lineages were necessary in order to obtain a good fit for the model—one maximized in Mamusi, one maximized in Nasioi, and one closest to New Guinea Highlanders—showing that the implied ancestry components from ADMIXTURE (Figure 1A; Figure S1) are all well supported in formal models based on allele-sharing statistics. Additionally, the admixture graph analysis suggests that the blue (Bismarck Archipelago majority) and red (Solomon Islands majority) ADMIXTURE components represent admixed ancestry. In particular, both include First Remote Oceanian ancestry (~20% for red and ~5% for blue), and the two are additionally admixed with each other, as we could not fit a Solomon Islands population (e.g., Nasioi) and a Bismarck Archipelago population (e.g., Mamusi or Baining) simultaneously without admixture from one to the other. In our model, we included Solomon Islands-type ancestry in Mamusi (inferred as approximately one-third of its total Papuan ancestry), although we were unable to distinguish the direction(s) of gene flow. Vanuatu was confidently inferred to have ancestry from all three Papuan sources ($|Z| > 8$ for omitting any source).

We next asked whether we could add Polynesians (Tongan) as a mixture of a component related to one of the other Oceanian populations along with additional First Remote Oceanian ancestry. Such a model was successful only in one configuration, with Tongan as a mixture of a population related to the Tolai of New Britain plus additional First Remote Oceanian ancestry (all f statistics fit to within 2.0 SEs of their observed values except for one residual, f_4(Kankanaey, Tongan; Australian, Vanuatu_Tanna), at $Z = 2.7$; Figure 3; Figure S3). Our choice to include Tolai in the model was guided by the ADMIXTURE analysis, in

which the Papuan ancestry profile in Polynesians appears to match that in Tolai (and Tutuba, from near Santo island in Vanuatu) more closely than other populations. The Tolai are known to be descended from relatively recent mixture between groups from New Ireland and New Britain (resulting from displacement caused by the eruption of the Rabaul caldera ~1400 BP [18]), so their ancestors cannot represent the true source population of the Papuan ancestry in Polynesians. However, the similarity of Tolai Papuan ancestry to Polynesians suggests that the Papuan component in Polynesians could similarly be from a mixture of multiple Near Oceanian sources. Given that Tolai are genetically intermediate between populations from New Britain and New Ireland (the latter with higher Solomon Islands-related ancestry), Polynesians could plausibly have acquired New Britain-related ancestry from Vanuatu or Santa Cruz, along with ancestry more closely related to that in New Ireland or the Solomon Islands via a distinct stream of migration.

As suggested by similar mixtures of components in ADMIXTURE, the ancient Vanuatu individuals are broadly consistent with descent from the same common ancestral population as present-day groups from Vanuatu. In the admixture graphs, we could fit most of the ancient sample groups as sister populations to Vanuatu_Tanna, albeit with different proportions of First Remote Oceanian ancestry. The one exception was the 150 BP grouping of individuals from Efate (with ~18% First Remote Oceanian ancestry), which showed significant un-modeled allele sharing with Tongan (maximum residual $Z = 3.7$, after accounting for excess First Remote Oceanian ancestry). Some present-day Vanuatu populations, such as Efate, Makura, and Polynesian Outliers, show a similar pattern when tested in the model, most likely reflecting migration of Polynesians to Vanuatu in the last thousand years or less.

Conclusions

By analyzing a time transect of central Vanuatu from initial settlement through the present, combined with dense geographical sampling of present-day populations from 18 islands in Vanuatu and dozens of populations outside Vanuatu, we document a series of dramatic genetic shifts associated with consistently high human mobility through a total of at least four distinct streams of migration and admixture. First, the initial human migration to central Vanuatu involved First Remote Oceanians associated with the Lapita culture. Second, by ~2300 BP, these groups were almost completely displaced by Papuan-ancestry populations originally from the Bismarck Archipelago, who remain the source for most of the ancestry of people in Vanuatu today. Third, in Polynesia, we find evidence for a different Papuan ancestry type that reflects a distinct migration. Finally, these streams of ancestry reconnected in parts of Vanuatu, influenced by back migration from Polynesia. These results highlight a history of multiple episodes of migration and mixture in shaping the human diversity of Oceania.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Archaeological Context on Ancient Individuals with New Genome-Wide Data
  - Genotyping Data from Present-Day Vanuatu
- METHOD DETAILS
  - Ancient DNA laboratory work
  - Bioinformatic processing
  - Mitochondrial DNA haplogroup determination
  - Direct Accelerator Mass Spectrometry (AMS) Radiocarbon Dates
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Analysis dataset
  - Clustering analyses
  - Allele-sharing statistics
  - Dates of admixture
  - Admixture graph fitting
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures and one data file and can be found with this article online at https://doi.org/10.1016/j.cub.2018.02.051.

ACKNOWLEDGMENTS

We are grateful to Fiona Petchey and Tomasz Goslar for sharing unpublished information on previously reported radiocarbon dates generated at the University of Waikato and the Poznan Accelerator Mass Spectrometry laboratories. The Teouma research was supported by the Australian Research Council (Discovery Grants DP0880789 and DP0556874, M.S. and S.B.), the National Geographic Society (Scientific Research Grant 8038-06, M.S. and S.B.), the Pacific Biological Foundation (grant PBF-04-1, M.S. and S.B.), the Royal Society of New Zealand Marsden Fund (UOO0917, H.B. and S.B.), and a University of Otago Research Grant (UORG917, H.B.). F.V. was supported by CNRS-UMR 7041 (PICS 3346). We are grateful to the late Richard Shutler, Jr. for access to his original field notes and to David Burley for contributing further Shutler archives to the Vanuatu Cultural Centre that aided in interpretation. Ralph Regenvanu, former Director of the Vanuatu Cultural Centre, gave ethical guidance on the use of the present-day samples for this project. A.J.M. was supported by a Wellcome Trust Clinical Research Training Fellowship (grant reference 106289/Z/14/Z). We thank Professors John Clegg, David Weatherall, Donald Bowden, and their colleagues for their work establishing the Oceanic sample collection at the University of Oxford in the UK, with support from the Wellcome Trust and Medical Research Council. P.S. was supported by the Swedish Research Council (VR grant 2014-453). Accelerator mass spectrometry radiocarbon dating work at Pennsylvania State University (D.J.K) was supported by the NSF Archaeometry program (BCS-1460369). D.R. was supported by NIH grant GM100233, by NSF HOMINID grant BCS-1032255, and by an Allen Discovery Center of the Paul Allen Foundation and is a Howard Hughes Medical Institute investigator.

AUTHOR CONTRIBUTIONS

R.P. and D.R. supervised the study. M.S., F.V., S.B., R.S., H.B., I.P., G.W., and R.P. provided ancient samples and assembled archaeological and anthropological information. N.R., N.B., O.C., M.F., M.M., J.O., K. Sirak, and K. Stewardson performed ancient DNA laboratory work. T.K.H. and D.J.K. carried out radiocarbon dating and analyzed data. K.A., A.H., K.M., S.J.O., T.P., K.R., T.N.W., and A.J.M. provided data from present-day populations. M.L., P.S., S.M., and D.R. analyzed genetic data. M.L., P.S., M.S., and D.R. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 12, 2018
Revised: February 19, 2018
Accepted: February 19, 2018
Published: February 28, 2018

REFERENCES

1. Skoglund, P., Posth, C., Sirak, K., Spriggs, M., Valentin, F., Bedford, S., Clark, G.R., Reepmeyer, C., Petchey, F., Fernandes, D., et al. (2016). Genomic insights into the peopling of the Southwest Pacific. Nature *538*, 510–513.
2. Valentin, F., Détroit, F., Spriggs, M.J., and Bedford, S. (2016). Early Lapita skeletons from Vanuatu show Polynesian craniofacial shape: implications for Remote Oceanic settlement and Lapita origins. Proc. Natl. Acad. Sci. USA *113*, 292–297.
3. Valentin, F., Herrscher, E., Bedford, S., Spriggs, M., and Buckley, H. (2014). Evidence for social and cultural change in Central Vanuatu between 3000 and 2000 BP: comparing funerary and dietary patterns of the first and later generations at Teouma, Efate. J. Island Coast. Archaeol. *9*, 381–399.
4. Pinhasi, R., Fernandes, D., Sirak, K., Novak, M., Connell, S., Alpaslan-Roodenberg, S., Gerritsen, F., Moiseyev, V., Gromov, A., Raczky, P., et al. (2015). Optimal ancient DNA yields from the inner ear part of the human petrous bone. PLoS ONE *10*, e0129102.
5. Dabney, J., Knapp, M., Glocke, I., Gansauge, M.T., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J.L., and Meyer, M. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proc. Natl. Acad. Sci. USA *110*, 15758–15763.
6. Korlević, P., Gerber, T., Gansauge, M.T., Hajdinjak, M., Nagel, S., Aximu-Petri, A., and Meyer, M. (2015). Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. Biotechniques *59*, 87–93.
7. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., and Reich, D. (2015). Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. Philos. Trans. R. Soc. Lond. B Biol. Sci. *370*, 20130624.
8. Briggs, A.W., Stenzel, U., Meyer, M., Krause, J., Kircher, M., and Pääbo, S. (2010). Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. Nucleic Acids Res. *38*, e87.
9. Maricic, T., Whitten, M., and Pääbo, S. (2010). Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. PLoS ONE *5*, e14004.
10. Fu, Q., Meyer, M., Gao, X., Stenzel, U., Burbano, H.A., Kelso, J., and Pääbo, S. (2013). DNA analysis of an early modern human from Tianyuan Cave, China. Proc. Natl. Acad. Sci. USA *110*, 2223–2227.
11. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. *19*, 1655–1664.
12. Pugach, I., Duggan, A.T., Merriwether, D.A., Friedlaender, F.R., Friedlaender, J.S., and Stoneking, M. (2018). The gateway from Near into Remote Oceania: new insights from genome-wide data. Mol. Biol. Evol. Published online January 2, 2018. https://doi.org/10.1093/molbev/msx333.
13. Reich, D., Thangaraj, K., Patterson, N., Price, A.L., and Singh, L. (2009). Reconstructing Indian population history. Nature *461*, 489–494.
14. Garanger, J. (1972). Archéologie des Nouvelles-Hébrides (ORSTOM).
15. Moorjani, P., Patterson, N., Hirschhorn, J.N., Keinan, A., Hao, L., Atzmon, G., Burns, E., Ostrer, H., Price, A.L., and Reich, D. (2011). The history of African gene flow into Southern Europeans, Levantines, and Jews. PLoS Genet. *7*, e1001373.
16. Loh, P.R., Lipson, M., Patterson, N., Moorjani, P., Pickrell, J.K., Reich, D., and Berger, B. (2013). Inferring admixture histories of human populations using linkage disequilibrium. Genetics *193*, 1233–1254.
17. Moorjani, P., Sankararaman, S., Fu, Q., Przeworski, M., Patterson, N., and Reich, D. (2016). A genetic method for dating ancient genomes provides a direct estimate of human generation interval in the last 45,000 years. Proc. Natl. Acad. Sci. USA *113*, 5652–5657.
18. Spriggs, M. (1997). The Island Melanesians (Blackwell).
19. Xu, S., Pugach, I., Stoneking, M., Kayser, M., and Jin, L.; HUGO Pan-Asian SNP Consortium (2012). Genetic dating indicates that the Asian-Papuan admixture through Eastern Indonesia corresponds to the Austronesian expansion. Proc. Natl. Acad. Sci. USA *109*, 4574–4579.
20. Lipson, M., Loh, P.R., Patterson, N., Moorjani, P., Ko, Y.C., Stoneking, M., Berger, B., and Reich, D. (2014). Reconstructing Austronesian population history in Island Southeast Asia. Nat. Commun. *5*, 4689.
21. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079.
22. Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589–595.
23. Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. Genetics *192*, 1065–1093.
24. Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLoS Genet. *2*, e190.
25. Skoglund, P., Northoff, B.H., Shunkov, M.V., Derevianko, A.P., Pääbo, S., Krause, J., and Jakobsson, M. (2014). Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. Proc. Natl. Acad. Sci. USA *111*, 2229–2234.
26. Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.J., Kronenberg, F., Salas, A., and Schönherr, S. (2016). HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. Nucleic Acids Res. *44* (W1), W58–W63.
27. Jostins, L., Xu, Y., McCarthy, S., Ayub, Q., Durbin, R., Barrett, J., and Tyler-Smith, C. (2014). YFitter: maximum likelihood assignment of Y chromosome haplogroups from low-coverage sequence data. arXiv, arXiv:1407.7988, http://arxiv.org/abs/1407.7988.
28. Valentin, F., Spriggs, M., Bedford, S., and Buckley, H. (2011). Vanuatu mortuary practices over three millennia: Lapita to the early contact period. Journal of Pacific Archaeology *2*, 49–65.
29. Valentin, F., Shing, R., and Spriggs, M. (2005). Des restes humains datés du début de la période de Mangaasi (2400-1800 BP) découverts à Mangaliliu (Efate, Vanuatu). C. R. Palevol *4*, 420–427.
30. Shutler, M.E., and Shutler, R. (1966). A preliminary report of archaeological explorations in the Southern New Hebrides. Asian Perspect. (Honolulu) *9*, 157–166.
31. Fu, Q., Hajdinjak, M., Moldovan, O.T., Constantin, S., Mallick, S., Skoglund, P., Patterson, N., Rohland, N., Lazaridis, I., Nickel, B., et al. (2015). An early modern human from Romania with a recent Neanderthal ancestor. Nature *524*, 216–219.
32. Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., et al. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. Nature *522*, 207–211.
33. Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S.A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., et al. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. Nature *528*, 499–503.
34. Behar, D.M., van Oven, M., Rosset, S., Metspalu, M., Loogväli, E.L., Silva, N.M., Kivisild, T., Torroni, A., and Villems, R. (2012). A "Copernican" reassessment of the human mitochondrial DNA tree from its root. Am. J. Hum. Genet. *90*, 675–684.
35. Petchey, F., Spriggs, M., Bedford, S., Valentin, F., and Buckley, H. (2014). Radiocarbon dating of burials from the Teouma Lapita cemetery, Efate, Vanuatu. J. Archaeol. Sci. *50*, 227–242.
36. Longin, R. (1971). New method of collagen extraction for radiocarbon dating. Nature *230*, 241–242.
37. Lohse, J.C., Madsen, D.B., Culleton, B.J., and Kennett, D.J. (2014). Isotope paleoecology of episodic mid-to-late Holocene bison

population expansions in the Southern Plains, U.S.A. Quat. Sci. Rev. *102*, 14–26.

38. Van Klinken, G.J. (1999). Bone collagen quality indicators for palaeodietary and radiocarbon measurements. J. Archaeol. Sci. *26*, 687–695.
39. Kennett, D.J., Plog, S., George, R.J., Culleton, B.J., Watson, A.S., Skoglund, P., Rohland, N., Mallick, S., Stewardson, K., Kistler, L., et al. (2017). Archaeogenomic evidence reveals prehistoric matrilineal dynasty. Nat. Commun. *8*, 14115.
40. Stuiver, M., and Polach, H.A. (1977). Reporting of C-14 data-discussion. Radiocarbon *19*, 355–363.
41. Bronk Ramsay, C. (2009). Bayesian analysis of radiocarbon dates. Radiocarbon *51*, 337–360.
42. Reimer, P.J., Bard, E., Bayliss, A., Beck, J.W., Blackwell, P.G., Ramsey, C.B., Buck, C.E., Cheng, H., Edwards, R.L., Friedrich, M., et al. (2013). Intcal13 and Marine13 radiocarbon age calibration curves 0-50,000 years cal BP. Radiocarbon *55*, 1869–1887.
43. Galinsky, K.J., Bhatia, G., Loh, P.R., Georgiev, S., Mukherjee, S., Patterson, N.J., and Price, A.L. (2016). Fast principal-component analysis reveals convergent evolution of ADH1B in Europe and East Asia. Am. J. Hum. Genet. *98*, 456–472.
44. Lipson, M., Szécsényi-Nagy, A., Mallick, S., Pósa, A., Stégmár, B., Keerl, V., Rohland, N., Stewardson, K., Ferry, M., Michel, M., et al. (2017). Parallel palaeogenomic transects reveal complex genetic history of early European farmers. Nature *551*, 368–372.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
Ancient skeletal element	This study	S1368 / TB30A.P3
Ancient skeletal element	This study	S1369 / B10B.P3
Ancient skeletal element	This study	S1370 / B17.P3
Ancient skeletal element	This study	S3921 / BURU5D
Ancient skeletal element	This study	S4096 / BURU5B
Ancient skeletal element	This study	S4105 / WAMB1
Ancient skeletal element	This study	S4106 / WAMB2
Ancient skeletal element	This study	S4419 / BB1
Ancient skeletal element	This study	S4424 / Pango1
Ancient skeletal element	This study	S4425 / Pit 2, Loc E
Ancient skeletal element	This study	S4450 / Sepulture 1
Ancient skeletal element	This study	S4451 / TAP1
Ancient skeletal element	This study	S5259 / Mang1
Ancient skeletal element	This study	S5951 / TeoQE
Chemicals, Peptides, and Recombinant Proteins		
Pfu Turbo Cx Hotstart DNA Polymerase	Agilent Technologies	600412
Herculase II Fusion DNA Polymerase	Agilent Technologies	600679
2x HI-RPM hybridization buffer	Agilent Technologies	5190-0403
0.5 M EDTA pH 8.0	BioExpress	E177
Sera-Mag Magnetic Speed-beads Carboxylate-Modified (1 μm, 3EDAC/PA5)	GE LifeScience	65152105050250
USER enzyme	New England Biolabs	M5505
UGI	New England Biolabs	M0281
Bst DNA Polymerase2.0, large frag.	New England Biolabs	M0537
PE buffer concentrate	QIAGEN	19065
Proteinase K	Sigma Aldrich	P6556
Guanidine hydrochloride	Sigma Aldrich	G3272
3M Sodium Acetate (pH 5.2)	Sigma Aldrich	S7899
Water	Sigma Aldrich	W4502
Tween-20	Sigma Aldrich	P9416
Isopropanol	Sigma Aldrich	650447
Ethanol	Sigma Aldrich	E7023
5M NaCl	Sigma Aldrich	S5150
1M NaOH	Sigma Aldrich	71463
20% SDS	Sigma Aldrich	5030
PEG-8000	Sigma Aldrich	89510
1 M Tris-HCl pH 8.0	Sigma Aldrich	AM9856
dNTP Mix	Thermo Fisher Scientific	R1121
ATP	Thermo Fisher Scientific	R0441
10x Buffer Tango	Thermo Fisher Scientific	BY5
T4 Polynucleotide Kinase	Thermo Fisher Scientific	EK0032
T4 DNA Polymerase	Thermo Fisher Scientific	EP0062
T4 DNA Ligase	Thermo Fisher Scientific	EL0011
Maxima SYBR Green kit	Thermo Fisher Scientific	K0251

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
50x Denhardt's solution	Thermo Fisher Scientific	750018
SSC Buffer (20x)	Thermo Fisher Scientific	AM9770
GeneAmp 10x PCR Gold Buffer	Thermo Fisher Scientific	4379874
Dynabeads MyOne Streptavidin T1	Thermo Fisher Scientific	65602
Salmon sperm DNA	Thermo Fisher Scientific	15632-011
Human Cot-I DNA	Thermo Fisher Scientific	15279011
DyNAmo HS SYBR Green qPCR Kit	Thermo Fisher Scientific	F410L
Methanol, certified ACS	VWR	EM-MX0485-3
Acetone, certified ACS	VWR	BDH1101-4LP
Dichloromethane, certified ACS	VWR	EMD-DX0835-3
Hydrochloric acid, 6N, 0.5N & 0.01N	VWR	EMD-HX0603-3
Critical Commercial Assays		
High Pure Extender from Viral Nucleic Acid Large Volume Kit	Roche	5114403001
MinElute PCR Purification Kit	QIAGEN	28006
NextSeq 500/550 High Output Kit v2 (150 cycles)	Illumina	FC-404-2002
Deposited Data		
Raw and analyzed data	This paper	ENA: PRJEB24938
Software and Algorithms		
Samtools	[21]	http://samtools.sourceforge.net/
BWA	[22]	http://bio-bwa.sourceforge.net/
ADMIXTOOLS	[23]	https://github.com/DReichLab/AdmixTools
SeqPrep	https://github.com/jstjohn/SeqPrep	https://github.com/jstjohn/SeqPrep
bamrmdup	https://bitbucket.org/ustenzel/biohazard	https://bitbucket.org/ustenzel/biohazard
smartpca	[24]	https://www.hsph.harvard.edu/alkes-price/software/
ADMIXTURE	[11]	https://www.genetics.ucla.edu/software/admixture/download.html
PMDtools	[25]	https://github.com/pontussk/PMDtools
Haplogrep 2	[26]	http://haplogrep.uibk.ac.at/
Yfitter	[27]	https://sourceforge.net/projects/yfitter/
ALDER	[16]	http://cb.csail.mit.edu/cb/alder/

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, David Reich (reich@genetics.med.harvard.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Archaeological Context on Ancient Individuals with New Genome-Wide Data

We newly report data from 14 ancient skeletons. For 3 of these skeletons we are reporting new ancient DNA data increasing the quality of the dataset beyond the data reported from the same samples in a previous study [1]. For 11 samples the data are entirely new:

Teouma, Efate Island (~2900 BP) – Lapita Culture (n = 4 samples)

The Teouma Lapita culture cemetery and settlement site is discussed in detail in the Supplementary Information of Skoglund et al., 2016 and references [1]. The additional sample I5951 was displaced during quarrying activities before controlled archaeological excavations began at the site in 2004. Given its radiocarbon date it is highly likely to have been from a disturbed burial context of Lapita age and can be legitimately considered with the other Lapita-age skeletons from the site.

I5951 (TeoQE), Vanuatu_2900BP
Newly reported sample
Genetic Sex: Male

Radiocarbon Date: 2920–2720 calBP (2955 ± 20 BP, PSUAMS-2411, marine corrected)
I1370 (B17.P3), Vanuatu_2900BP
Previously reported in [1]; here we report higher coverage data
Genetic Sex: Female
Radiocarbon Date: 3080–2780 calBP (3083 ± 26 BP, Wk-21026, marine corrected)
I1369 (B10B.P3), Vanuatu_2900BP
Previously reported in [1]; here we report higher coverage data
Genetic Sex: Female
Radiocarbon Date: 3020–2750 calBP (3045 ± 30 BP, Poz-81126, marine corrected)
I1368 (TB30A.P3), Vanuatu_2900BP
Previously reported in [1]; here we report higher coverage data
Genetic Sex: Female
Radiocarbon Date: 2990–2750 calBP (2983 ± 32 BP, Wk-22657, marine corrected)

Mele-Taplins, Efate Island (~2300 BP) (n = 1 sample)

The Mele-Taplins site is described by Valentin and colleagues [28]. The skeleton comes from a subsurface grave in a rockshelter (Taplins 1) at the base of a large boulder, excavated by Graeme Ward of The Australian National University in 1973-4 and curated at Otago University, Dunedin, New Zealand. Other burials from the Taplins 2 shelter were of broadly similar age.

I4451 (TAP1), Vanuatu_2300BP
Newly reported sample
Genetic Sex: Male
Radiocarbon Date: 2360–2160 calBP (2348 ± 32 BP, Wk-20390, marine corrected)

Burumba, Epi Island (~1300 BP) (n = 2 samples)

The Burumba site is described by Valentin and colleagues [28] and excavated in 2006 by Frederique Valentin and Jacques Bole. The graves of nine adults were excavated from an open site at Kalala Plantation 200 m from the current beach, dug into sterile sand. Burial 5 was an assemblage of cranial remains of five individuals placed on a pile of coral slabs and blocks.

I3921 (BURU5D), Vanuatu_1300BP
Newly reported sample
Genetic Sex: Male
Radiocarbon Date: 1340–1180 calBP [1410–1280 calBP (1530 ± 20 BP, PSUAMS-1841), 1300–1170 calBP (1395 ± 15 BP, PSUAMS-2428), marine corrected]
I4096 (BURU5B), Vanuatu_1300BP
Newly reported sample
Genetic Sex: Male
Radiocarbon Date: 1380–1240 calBP [1380–1270 calBP (1490 ± 15 BP, PSUAMS-2460), 1350–1180 calBP (1464 ± 30 BP, Wk-25769), marine corrected]

Mangaliliu, Efate Island (~500 BP) (n = 1 sample)

The burial was excavated from a test pit in Mangaliliu village by Richard Shing in 2002 and described in detail by Valentin and colleagues [29]. The originally reported age of the burial was reassessed after direct dating of the skeleton [28].

I5259 (burial 1, Mang1), Vanuatu_500BP, Mangaliliu (Efate Island)
Newly reported sample
Genetic Sex: Female
Radiocarbon Date: 630–330 calBP (559 ± 30 BP, Wk-20030, marine corrected)

Pangpang, Efate Island (~150 BP) (n = 1 sample)

This burial, in a flexed position, was excavated by Richard Shing and Iarawai Phillip during archaeological impact assessment related to the Efate Ring Road construction between the villages of Pangpang and Forari. The body was adorned with ornaments composed of numerous tiny Conus shell and shark vertebrae beads and a large pearl shell pendant. This range of ornaments has been recorded in burial contexts of the last 400 years, prior to and during the initial phases of European contact (R.S. and I.P., unpublished data; field notes at Vanuatu National Museum).

I4450 (SEPU1, Sepulture 1), Vanuatu_Efate_150BP
Newly reported sample
Genetic Sex: Female
Radiocarbon Date: 430–0 calBP (305 ± 15 BP, UCIAMS-188793, marine corrected)

Wam Bay, Epi Island (~150 BP) (n = 2 samples)

The site appears to have been a largely Mission period cemetery of the late 19th to early 20th centuries. Three burials were exposed in proximity to a combustion feature associated with the making of lime-plaster for construction, a European introduced practice. The site was excavated by Frederique Valentin and Matthew Spriggs in 2006 (M.S. and F.V., unpublished data; field notes at Vanuatu National Museum).

I4105 (WAMB1), Vanuatu_Epi_150BP
Newly reported sample
Genetic Sex: Male
Radiocarbon Date: 300–0 calBP (255 ± 20 BP, PSUAMS-1922, marine corrected)
I4106 (WAMB2), Vanuatu_Epi_150BP
Newly reported sample
Genetic Sex: Male
Radiocarbon Date: 280–0 calBP (225 ± 20 BP, PSUAMS-1923, marine corrected)

Ifira, Efate Island (~150 BP) (n = 1 sample)

This tightly flexed burial from a feature containing skeletal remains of two individuals was excavated by Mary Elizabeth and Richard Shutler, Jr, in June 1964 on the small island of Ifira in Vila Harbor, Port Vila, during a test pit survey of the island. It is briefly mentioned in Shutler and Shutler [30]. Unpublished field notes relating to the excavation are held in the files of the Vanuatu National Museum. Ifira is notable as one of the Vanuatu Polynesian Outlier islands and this burial would date to the period of Polynesian cultural influence.

I4425 (EF3_2_E, Pit 2; Loc E), Vanuatu_Efate_150BP
Newly reported sample
Genetic Sex: Female
Radiocarbon Date: 270–0 calBP (200 ± 20 BP, UCIAMS-188795, marine corrected)

Pango Village, Efate Island (~150 BP) (n = 1 sample)

This is one of two individuals excavated by Mary Elizabeth and Richard Shutler, Jr. on the Pango Peninsula opposite the small island of Ifira in Vila Harbour, Port Vila. Unpublished field notes relating to the burial are held in the files of the Vanuatu National Museum, but the notes provide limited detail.

I4424 (EF_Pango1), Vanuatu_Efate_150BP
Newly reported sample
Genetic Sex: Male
Radiocarbon Date: 280–0 calBP (190 ± 15 BP, UCIAMS-188794, marine corrected)

Banana Bay, Efate Island (~150 BP) (n = 1 sample)

The burial was excavated by Richard Shing and Iarawai Phillip during archaeological impact assessment related to the Efate Ring Road construction in the Banana Bay area, southeast Efate. The body, lying on the back, was adorned with ornaments including numerous tiny Conus shell beads and a few European glass beads (R.S. and I.P., unpublished data; field notes at Vanuatu National Museum).

I4419 (BB1, Burial 1), Vanuatu_Efate_150BP, Banana Bay (Efate Island)
Newly reported sample
Genetic Sex: Male
Radiocarbon Date: 260–20 calBP (135 ± 15 BP, UCIAMS-188792, marine corrected)

Genotyping Data from Present-Day Vanuatu

We genotyped 185 present-day individuals from 32 populations from Vanuatu spanning 18 islands. All individuals gave informed verbal consent for studies of population history and human health, especially as they may shed light on anemia, consistent with the standards prevailing at the time the data were collected. Samples of whole blood were collected as part of a range of research projects undertaken from the late 1970s in collaborations between multiple sites and institutions in Vanuatu and the University of Oxford investigating population differences at the genetic level. In accordance with participant consent, DNA was extracted, anonymized, and stored in batches analyzable only by geographic location of participant origin. Use of the samples for genome-wide analyses including studies of population history was reviewed by the Oxford Tropical Research Ethics Community at the University of Oxford and formally approved in a letter dated July 2, 2014 (OXTREC Reference: 537-14). The use of the samples for genetic analysis was also approved by the Vanuatu Cultural Centre in a formal letter dated May 30, 2017.

METHOD DETAILS

Ancient DNA laboratory work

In dedicated clean rooms at University College Dublin, we used a dental sandblaster to separate cochlear sections from petrous bones. We milled these samples into fine powder, and shipped them to Harvard Medical School.

In dedicated clean rooms at Harvard Medical School, we extracted DNA following a previously published protocol [5], with two modifications. First, we replaced the combination of a funnel and a MinElute column with Roche columns [6]. Second, we eluted two times in 45 μl, obtaining 90 μL of extract for each sample (Data S1).

We prepared libraries from the extracts using a double-stranded protocol, affixing 7-base-pair sequences to either end to allow multiplexing of the libraries and to prevent contamination from affecting the samples after barcodes were added. We prepared some of the libraries in the presence of the enzyme UDG to remove characteristic damage associated with ancient DNA (Data S1) [7].

We enriched the libraries in solution for sequences overlapping the mitochondrial genome [9] as well as for 3000 nuclear positions, and sequenced on an Illumina NextSeq500 instrument for 2x76cycles + 2x7 cycles after adding a pair of unique 7-base-pair indices. For libraries that were promising after screening, we enriched for sequences overlapping approximately 1.24 million SNPs on the nuclear genome [10, 31–33]. We added two unique 7-base-pair indices to each enriched library and sequenced a multiplexed pool of samples with an Illumina NextSeq500 instrument for 2x76cycles + 2x7cycles. We iteratively sequenced more from each sample until the number of new SNPs covered per additional sequences generated was less than about 1 in 100.

For samples for which we wished to obtain more coverage, we prepared additional libraries from existing extract or new extract, leading to a total of up to 8 libraries for some samples. We pooled data from all libraries for further analysis. We also prepared versions of the sample data using only UDG-treated libraries. We use the suffix "_all" to refer to the versions of each sample with all libraries in Data S1 and Figure S2. We use the "_all" versions for our primary analyses, but also perform some analyses on the entirely UDG-treated versions to assess if there is evidence that any results are influenced by ancient DNA artifacts (all appear to be robust).

Bioinformatic processing

We demultiplexed reads from the NextSeq500 lanes into individual libraries based on the sequences of their two indices and two barcodes. To assign a read pair to a library we required no more than one mismatch to the total of four expected 7 base pair sequences. We merged sequences using *SeqPrep* (https://github.com/jstjohn/SeqPrep), requiring at least 15 base pairs of overlap. At positions of overlap, we used the allele and quality score of the read of higher quality.

We aligned merged sequences to the mitochondrial RSRS genome [34] (for mitochondrial DNA analyses) and to the hg19 reference (for whole genome analyses) using the command "samse" from BWA with default parameters (version 0.6.1) [22]. For non-UDG treated libraries, which are expected to have higher mismatch rates compared to the reference genome, we used more relaxed alignment parameters, "-n 0.01 -o 2 -l 16500." This setting disables seeding, allowing for less conservative alignments and helping to align damaged sequences. We chose one allele at random per site ("pseudo-haploid" genotypes) from aligned sequences to use in analyses.

Mitochondrial DNA haplogroup determination

We determined mitochondrial DNA haplogroups for each library separately as well as for pools of libraries for each sample using Haplogrep2, which provides a ranking score measuring the reliability of the haplogroup assignments [26] (Data S1). The procedure used here is designed to extract the maximally informative data from the sample during haplogroup assignment by building a mitochondrial consensus sequence in multiple ways and then using the rank score to select the most confident call. First, we restricted to sequences with characteristic ancient DNA damage in their terminal nucleotides (a procedure that removes potentially contaminating sequences at the cost of greatly reduced coverage). To restrict to damaged sequences, we used the PMDtools software [25], requiring a minimum score of pmdscore = 3, and then trimmed the sequences by 5 base pairs on either side to remove nucleotides likely to be deaminated before calling a haplogroup with Haplogrep2. As a second approach, we trimmed sequences by 0-7 base pairs on either side to eliminate characteristic ancient DNA damage and fed these sequences to Haplogrep2 without damage restriction (hence retaining more data), calling a haplogroup at each trim level.

Direct Accelerator Mass Spectrometry (AMS) Radiocarbon Dates

We prepared 10 new bone samples for AMS radiocarbon dating at the Human Paleoecology and Isotope Geochemistry Laboratory at the Pennsylvania State University. After preparation, we AMS dated the samples either at the W.M. Keck Carbon Cycle Accelerator Mass Spectrometry Laboratory at the University of California, Irvine (lab code: UCIAMS) or at the Accelerator Mass Spectrometer Laboratory at the Pennsylvania State University (lab code: PSUAMS). We co-analyzed our 10 newly generated dates with 6 previously published dates generated by other laboratories: five by the Radiocarbon Dating Laboratory at the University of Waikato (lab code: Wk) and one by the Poznan Radiocarbon Laboratory (lab code: Poz) [1, 28, 35].

Bone samples for the newly and previously reported direct AMS ^{14}C dates were manually cleaned and demineralized in weak HCl and soaked in an alkali bath (NaOH) at room temperature to remove contaminating soil humates. Samples were then rinsed to neutrality in Nanopure H_2O and gelatinized in HCL [36]. The resulting gelatin was lyophilized and weighed to determine percent yield as a measure of collagen preservation (% crude gelatin yield). Collagen was either directly AMS ^{14}C dated (Wk-20390, Wk-20030) or further purified using ultrafiltration prior to analysis. At PSUAMS we hydrolyzed two bone samples with low collagen yields and purified the resulting amino acids using XAD chromotography [37].

We use stable carbon and nitrogen isotopic analysis of bone collagen (or amino acids) as an additional quality control measure. For all samples, we examined the %C, %N and C:N ratios. C:N ratios for well-preserved samples fall between 2.9 and 3.6 [38], and all our samples met this criterion. We also used stable carbon and nitrogen isotope measurements to determine the marine

reservoir correction for each individual as described below. The detailed bone preparation and quality control methods we used for the newly reported dates from PSUAMS and UCIAMS are reported elsewhere [37, 39].

To calibrate the dates, we began with an adjustment for the marine reservoir effect, applying a correction (ΔR) of 40 ± 44 BP based on marine shell measurements to adjust for local oceanic variation in ^{14}C levels around Vanuatu as previously described by Petchey and colleagues [35]. We then corrected for mass-dependent fractionation with measured ^{13}C values [40] and calibrated in OxCal 4.3 [41] via a mixture of the IntCal13 Southern Hemisphere and Marine13 calibration curves [42] using the Mix_Curves function in OxCal according to the marine component of each individual's diet (estimated by $\delta^{13}C$ values). This method uses $\delta^{13}C$ endpoints of -21 and $-12‰$, the former representing a highly terrestrial diet and the latter indicating a diet rich in marine foods, with the percentage of marine dietary contribution estimated via linear interpolation [35]. In two cases, individuals were sampled twice for radiocarbon dating and stable isotope analysis: I3921 (PSUAMS-1841, PSUAMS-2428) and I4096 (PSUAMS-2460, Wk-25769). In these instances, we combined radiocarbon dates using the R Combine function in OxCal 4.3 and estimated the marine contribution according to the mean of two $\delta^{13}C$ measurements. For calBP dates (calibrated years before present), we use the standard definition of 1950 CE as 0 calBP.

QUANTIFICATION AND STATISTICAL ANALYSIS

Analysis dataset

All analyses are based on 593,124 SNPs on chromosomes 1-22 genotyped on the Affymetrix Human Origins array [23], with the newly reported data from ancient and present-day Vanuatu individuals merged with published Human Origins data [1]. Based on the ADMIXTURE results, we excluded 26 present-day individuals from analyses as outliers: 14 from Vanuatu (2 Efate, 3 Emae, 2 Malekula, 1 Ambrym, 1 Nguna, 1 Erromango, 3 Aneityum, and 1 Banks) with evidence of European admixture, 4 from Samoa also with evidence of European admixture, and 7 Tolai and 1 Tutuba with evidence of recent admixture or otherwise non-representative ancestry profiles.

Clustering analyses

We performed ADMIXTURE [11] clustering analysis using default parameters, with the cluster components (K) ranging from K = 2 to K = 8. We carried out principal component analysis (PCA) using the "lsqproject" and "autoshrink" options in smartpca [24, 43], computing axes using the present-day populations and projecting ancient samples. For PCA, we restricted to populations within a narrow range of Papuan ancestry proportions in order to minimize the variance due to Papuan versus First Remote Oceanian ancestry and capture components related to variation in Papuan ancestry sources. The specific range (~80% Papuan ancestry) was chosen as the highest possible that included groups from all of the major Oceanian island chains and genetic clusters. We did not project the Vanuatu_2900BP individuals because of their near-zero Papuan ancestry.

For the Papuan ancestry clusters defined in Figure 1B, we manually assigned populations based on their majority Papuan component in ADMIXTURE (out of red, blue, and black). For borderline populations with large red and blue components and small black components, we created a mixed cluster containing Kuot_Kabil, Kuot_Lamalaua, Lavongai, Madak, Nailik, Notsi, and Tigak (New Ireland); Ontong Java, Rennell and Bellona, and Tikopia (Polynesian Outliers); Makira (Solomon Islands); Tolai (New Britain); and Tutuba (Vanuatu). We assigned two borderline populations with large black components to the black cluster (Kove and Mussau).

Allele-sharing statistics

We computed allele-sharing statistics (*f*-statistics) in ADMIXTOOLS [23], with standard errors obtained by block jackknife. To replicate the findings from ADMIXTURE and PCA and test for differential allele sharing of populations from Vanuatu with groups from the Solomon Islands versus the Bismarck Archipelago, we computed the statistic f_4(Australian, *Vanuatu*; *Solomon Islands*, *Bismarck Archipelago*). We used the same populations as in the PCA (10 from Vanuatu, 8 from the Bismarck Archipelago, and 2 from the Solomon Islands; Figure S2) to avoid confounding from differential proportions of First Remote Oceanian ancestry.

We computed Papuan ancestry proportions using the f_4-ratio statistic f_4(Atayal, Australian; Kankanaey, *Test*)/f_4(Atayal, Australian; Kankanaey, New_Guinea_Highlander). The form of this statistic assumes a topology of (Atayal, (Kankanaey, First Remote Oceanian)) for the First Remote Oceanian ancestry in the *Test* population.

We measured differential affinity to Lapita samples from Vanuatu and Tonga by computing the difference between the statistics f_4(*Test*, Han; Atayal, Tonga_2600BP) and f_4(*Test*, Han; Atayal, Vanuatu_2900BP), using the qp4diff program ("allsnps" mode). In expectation, this difference is equal to the single statistic f_4(*Test*, Han; Vanuatu_2900BP, Tonga_2600BP), but our formulation allows us to increase power by utilizing the union rather than the intersection of the SNPs covered by the relatively low-coverage Vanuatu_2900BP and Tonga_2600BP samples.

Dates of admixture

We estimated dates of admixture using ALDER [14]. As reference populations we used published Human Origins data for Ami (aboriginal Taiwanese; n = 10 individuals) and New Guinea Highlanders (n = 19). In computing the correlation between ancestry proportions and dates of admixture, we performed a weighted linear regression of present-day Vanuatu population groups with ALDER date estimates (treated independently; either all 31 such groups or only the 20 with most confident estimates, as shown in Figure 2), weighting by the Z-score for the difference of the date estimate from zero [44].

Admixture graph fitting

We constructed admixture graphs using the qpGraph utility in ADMIXTOOLS [43]. The position of Mixe (a Native American population from present-day Mexico) as an outgroup relative to the other populations (in an unrooted sense) means that its eastern and western Eurasian ancestry components can be collapsed into a single lineage with no change in the model. Similarly, we can omit explicit inclusion of Denisovan admixture because of the symmetry of such ancestry in the right-hand clade of the model (as displayed in Figure S3).

DATA AND SOFTWARE AVAILABILITY

Raw sequences from the 14 individuals are available from the European Nucleotide Archive at accession number PRJEB24938. Genotype files are available at https://reich.hms.harvard.edu/datasets. To access data for the newly genotyped present-day individuals from Vanuatu, researchers should send a signed letter to D.R. containing the following text: "(a) I will not distribute the data outside my collaboration; (b) I will not post the data publicly; (c) I will make no attempt to connect the genetic data to personal identifiers for the samples; (d) I will use the data only for studies of population history; (e) I will not use the data for any selection studies; (f) I will not use the data for medical or disease-related analyses; (g) I will not use the data for commercial purposes."