

Genomic insights into the peopling of the Southwest Pacific

Pontus Skoglund^{1,2,3}, Cosimo Posth^{4,5}, Kendra Sirak^{6,7}, Matthew Spriggs^{8,9}, Frederique Valentin¹⁰, Stuart Bedford^{9,11}, Geoffrey R. Clark¹¹, Christian Reepmeyer¹², Fiona Petchey¹³, Daniel Fernandes^{6,14}, Qiaomei Fu^{1,15,16}, Eadaoin Harney^{1,2}, Mark Lipson¹, Swapan Mallick^{1,2}, Mario Novak^{6,17}, Nadin Rohland¹, Kristin Stewardson^{1,2,18}, Syafiq Abdullah¹⁹, Murray P. Cox²⁰, Françoise R. Friedlaender²¹, Jonathan S. Friedlaender²², Toomas Kivisild^{23,24}, George Kokij²⁵, Pradiptajati Kusuma²⁶, D. Andrew Merriwether²⁷, Francois-X. Ricaut²⁸, Joseph T. S. Wee²⁹, Nick Patterson², Johannes Krause⁵, Ron Pinhasi⁶ & David Reich^{1,2,18}§

The appearance of people associated with the Lapita culture in the South Pacific around 3,000 years ago¹ marked the beginning of the last major human dispersal to unpopulated lands. However, the relationship of these pioneers to the long-established Papuan people of the New Guinea region is unclear. Here we present genome-wide ancient DNA data from three individuals from Vanuatu (about 3,100–2,700 years before present) and one from Tonga (about 2,700–2,300 years before present), and analyse them with data from 778 present-day East Asians and Oceanians. Today, indigenous people of the South Pacific harbour a mixture of ancestry from Papuans and a population of East Asian origin that no longer exists in unmixed form, but is a match to the ancient individuals. Most analyses have interpreted the minimum of twenty-five per cent Papuan ancestry in the region today as evidence that the first humans to reach Remote Oceania, including Polynesia, were derived from population mixtures near New Guinea, before their further expansion into Remote Oceania^{2–5}. However, our finding that the ancient individuals had little to no Papuan ancestry implies that later human population movements spread Papuan ancestry through the South Pacific after the first peopling of the islands.

Pacific islanders today derive from a mixture of two highly divergent ancestral populations³. The first ancestral modern human population arrived in island southeast Asia more than 40,000 years before present (BP), and contributed to the ancestry of both indigenous Australians and Papuans, and hence to other Pacific islanders⁴. The second ancestral population is more closely related to mainland East Asians⁴, and is not found in unadmixed form today. The first humans to reach Remote Oceania—a term we use to refer to the region unoccupied before approximately 3,000 BP beyond the main Solomon Islands and, in this case, excluding Micronesia—were associated with the Lapita culture, which existed between 3,450–3,250 and 2,700–2,500 BP. These people spread into Remote Oceania using the first boats capable of long-distance sea travel and introduced new domesticated animals and plants, and their successors reached the most isolated islands of the eastern and southern Pacific by 1,000–700 BP⁶. Several hypotheses have

been proposed to explain why present-day indigenous people of Near Oceania (New Guinea, the Bismarck Islands, and the Solomon Islands area) and Remote Oceania have ancestry both from Papuans and from populations of ultimate East Asian origin. In one set of models that has been favoured by recent genetic studies^{3–5,7}, the mixture occurred at around 3,000 BP, during the expansion of populations of East Asian origin through the New Guinea region⁸. In the other set of models, the population of ultimate East Asian origin initially mixed little with Papuans⁹, and thus later gene exchanges account for the ubiquitous Papuan ancestry today^{2,10}.

We obtained genome-wide ancient DNA data from three individuals from Teouma, an archaeological site on Efate island, Vanuatu (Supplementary Information section 1), which were all directly radiocarbon dated to between 3,110 and 2,740 BP, an interval that is chronologically part of the Lapita period (Extended Data Table 1). We also obtained genome-wide ancient DNA data from an individual from the Talasiu site on Tongatapu island, Tonga, directly radiocarbon dated to 2,680–2,340 BP, a period spanning the late Lapita and immediately post-Lapita period (Supplementary Information section 2 and Extended Data Table 1). In dedicated clean rooms, we prepared powder from petrous bones¹¹, extracted DNA¹², and prepared up to four double-stranded libraries from each extract¹³. We enriched the libraries for 1.24 million targeted single nucleotide polymorphisms (SNPs)¹⁴, sequenced the products, and represented each individual by a single randomly drawn sequence for each SNP. This procedure resulted in 139,461–231,944 SNPs that were covered at least once in each of the individuals. The low ratio of sequences aligning to Y-chromosome targets compared to targets on other chromosomes¹⁵ reveals that all four individuals are females (Extended Data Table 1). We obtained three mitochondrial DNA sequences from Vanuatu and all were haplogroup B4a1a1a, the classic ‘Polynesian motif’¹⁶.

Multiple features of the data suggest that the DNA was authentic and minimally contaminated. First, in all individuals, around 40% of all sites that are cytosines in the human reference sequence appear as thymines in the terminal nucleotide, as expected for genuine ancient

¹Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. ²Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. ³Archaeological Research Laboratory, Department of Archaeology and Classical Studies, Stockholm University, 10691 Stockholm, Sweden. ⁴Institute for Archaeological Sciences, Archaeology and Palaeogenetics, University of Tübingen, Tübingen 72070, Germany. ⁵Max Planck Institute for the Science of Human History, 07745 Jena, Germany. ⁶School of Archaeology and Earth Institute, Belfield, University College Dublin, Dublin 4, Dublin, Ireland. ⁷Department of Anthropology, Emory University, Atlanta, Georgia 30322, USA. ⁸School of Archaeology and Anthropology, College of Arts and Social Sciences, The Australian National University, Canberra, Australian Capital Territory 2601, Australia. ⁹Vanuatu National Museum, Vanuatu Cultural Centre, Port Vila, Vanuatu. ¹⁰Maison de l'Archéologie et de l'Ethnologie, CNRS, UMR 7041, 92023 Nanterre, France. ¹¹Department of Archaeology and Natural History, College of Asia and the Pacific, The Australian National University, Canberra, Australian Capital Territory 2601, Australia. ¹²College of Arts, Society and Education, James Cook University, Queensland 4870, Australia. ¹³Radiocarbon Dating Laboratory, University of Waikato, Hamilton 3240, New Zealand. ¹⁴CIAS, Department of Life Sciences, University of Coimbra, 3000-456 Coimbra, Portugal. ¹⁵Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, IVPP, CAS, Beijing 100044, China. ¹⁶Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig 04103, Germany. ¹⁷Institute for Anthropological Research, 10000 Zagreb, Croatia. ¹⁸Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts 02115, USA. ¹⁹RIPAS Hospital, Bandar Seri Begawan, Brunei Darussalam. ²⁰Institute of Fundamental Sciences, Massey University, Palmerston North 4442, New Zealand. ²¹Independent Scientist, Sharon, Connecticut 06069, USA. ²²Department of Anthropology, Temple University, Gladfelter Hall, Philadelphia, Pennsylvania 19122, USA. ²³Estonian Biocentre, Evolutionary Biology group, Tartu, 51010, Estonia. ²⁴Division of Archaeology, University of Cambridge, Fitzwilliam Street, Cambridge CB2 1QH, UK. ²⁵Papua New Guinea Institute of Medical Research, Goroka, Eastern Highlands Province 441, Papua New Guinea. ²⁶Eijkman Institute for Molecular Biology, Jakarta 10430, Indonesia. ²⁷Department of Anthropology, Binghamton University, Binghamton, New York 13902, USA. ²⁸Evolutionary Medicine Group, Laboratoire d'Anthropologie Moléculaire et Imagerie de Synthèse UMR 5288 CNRS, Université de Toulouse, Toulouse 31073, France. ²⁹National Cancer Centre Singapore, Singapore 169610, Singapore.

§These authors jointly supervised this work.

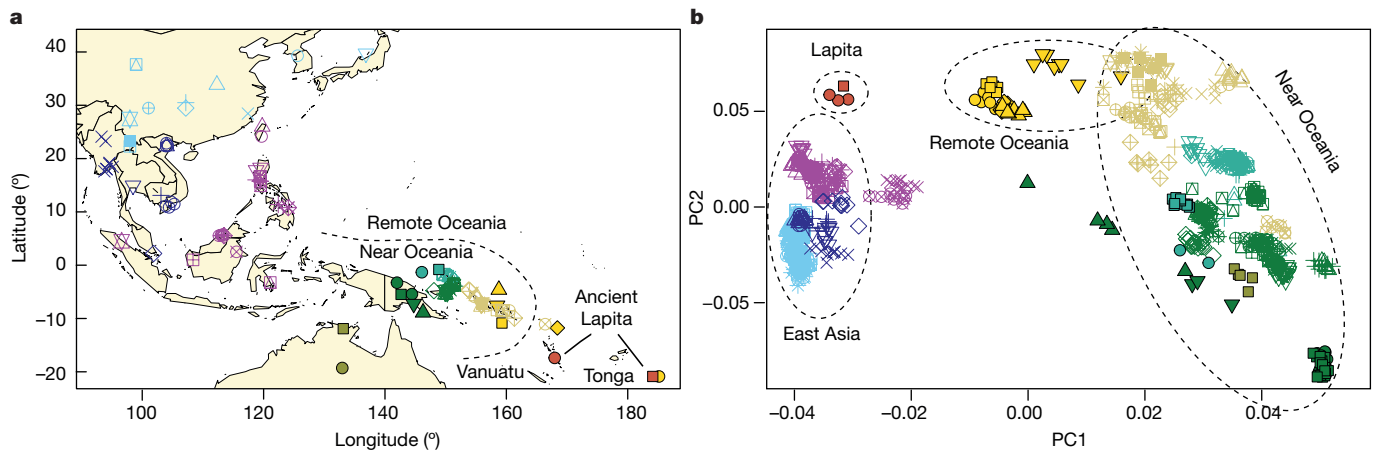


Figure 1 | Data from ancient and present-day populations. **a**, Locations of 778 present-day individuals genotyped on the Affymetrix Human Origins Array and 4 ancient individuals (red symbols). **b**, Ancient individuals projected onto principal components (PC) 1 and 2 computed using only present-day samples. Individual population labels are given in Extended Data Fig. 2.

DNA (Extended Data Fig. 1a). Second, when we carried out principal component analysis (PCA; Fig. 1) of 778 present-day people from 83 East Asian and Oceanian populations genotyped at 621,799 SNPs (of which 356 individuals from 38 groups were newly genotyped for this study; Extended Data Table 2) and projected the ancient individuals, we found that all clustered tightly with each other and with data from the same individuals restricting to sequences with cytosine-to-thymine changes at the terminal nucleotide (these sequences are unlikely to be contaminants^{17,18}) (Extended Data Fig. 1b). Third, the cluster of ancient individuals does not overlap with present-day populations, indicating that the data are from a population that is not present in unmixed form today (Fig. 1). The distinctiveness of the ancient individuals is also highlighted by their high differentiation from all present-day groups ($0.05 < F_{ST} < 0.26$; between all modern individuals and the ancient Vanuatu individuals, using the statistic F_{ST} , which compares within- and between-group squared allele frequency differences) (Extended Data Table 3).

The ancient Vanuatu and Tongan individuals are not shifted in the PCA in the direction of Papuan ancestry, in contrast to all present-day

Remote Oceanians. In this respect, they are similar to indigenous Taiwanese populations such as the Ami and Atayal as well as to populations from the Philippines such as the Kankanaey, who have no detectable Papuan ancestry (Fig. 1). To test whether the ancient individuals had any evidence of Papuan ancestry, we used the *qpWave*/*qpAdm* software (Methods) to analyse allele frequency correlation statistics¹⁹. The results were consistent with the ancient individuals and the Taiwanese Ami having descended from a common ancestral population to the exclusion of 14 worldwide outgroups ($P > 0.05$ for the ancient individuals from both Vanuatu and Tonga). We estimate the possible range of Papuan ancestry in the Vanuatu individuals to be 0–11% and in the Tongan individual to be 0–17% (99% confidence intervals truncated at zero), which does not overlap the point estimates of at least 25% Papuan ancestry in all present-day Oceanians (Fig. 2a). To test the hypothesis that the ancient Remote Oceanian individuals might be from the source population of the non-Papuan ancestry in Oceanians today, we computed the statistic f_4 (Africa, *Test*; Australian, Polynesian), which evaluates the degree of allele sharing of a candidate *Test* population with Polynesians (at sites where Polynesians differ

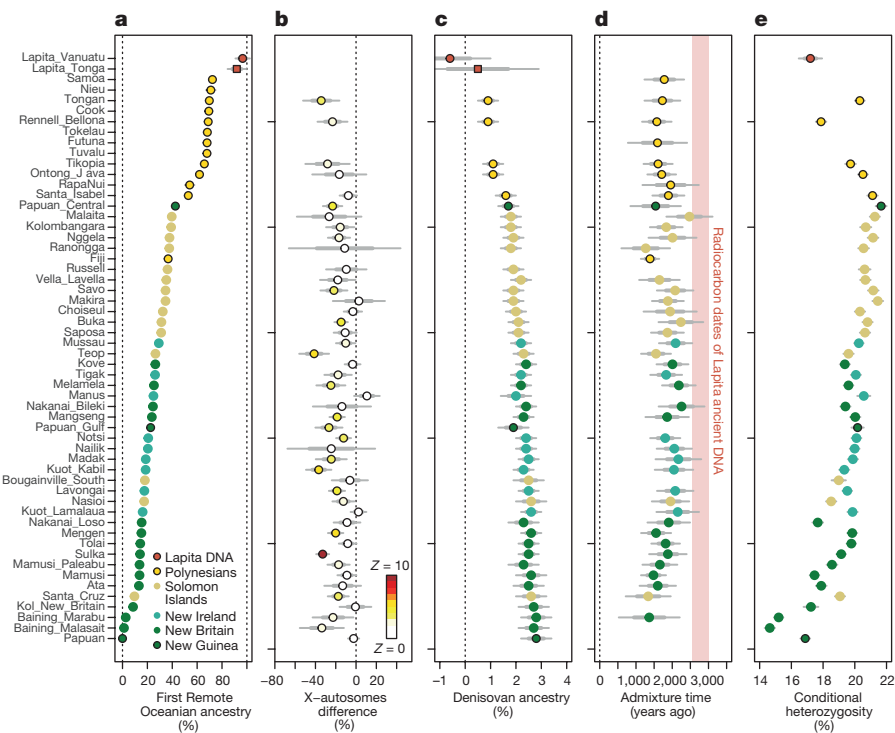


Figure 2 | Genetic characteristics of the Oceanian ancestry cline.

a, Estimated proportion of First Remote Oceanian ancestry. The Papuan ancestry can be estimated as 100% minus the estimate of First Remote Oceanian ancestry. **b**, Difference between First Remote Oceanian ancestry estimates on chromosome X and the autosomes. **c**, Denisovan ancestry estimates are inversely related to First Remote Oceanian ancestry estimates. **d**, Estimated date of admixture in all populations with at least four individuals and significant evidence of decay of weighted admixture linkage disequilibrium as measured in ALDER. We used Han and New Guinean Highlanders as surrogates for the ancestral populations. We assumed a generation interval of 28.1 years, and show 95% confidence intervals (thin whiskers) incorporating uncertainty both in the ALDER date and the value of the human generation interval. We show the range of radiocarbon dates for the ancient individuals. **e**, Conditional heterozygosity (genetic diversity) estimated by drawing two random chromosomes from different individuals at each locus, using only SNPs ascertained in a single Yoruba, and restricting to transversion SNPs to avoid any concerns about inflated heterozygosity owing to ancient DNA degradation. Thick and thin error bars in all five panels correspond to 1 and 1.96 standard error of the estimate, respectively.

from Australians), and found that it was maximized when *Test* was the ancient Vanuatu or Tonga individuals (Extended Data Fig. 2b), as expected if a population related to them was the true source. We conclude that the non-Papuan ancestry that is ubiquitous in Oceania is derived from a population related to the ancient individuals we analysed, and that this ancestry reached uninhabited islands in Remote Oceania with little or possibly no mixture with Papuans. We call the population of which both the ancient Vanuatu and Tongan individuals were a part the 'First Remote Oceanians' and find that the ancestry fraction from this population is the single most important factor shaping genetic variation among Pacific islanders, accounting for most variation in measurements including genetic diversity (Pearson's $R = 0.86$, $P = 2 \times 10^{-12}$ for 42 non-Polynesian groups; Extended Data Fig. 2) and the proportion of archaic Denisovan ancestry ($R = -0.96$, $P < 10^{-16}$ for all 56 Oceanian groups; Fig. 2).

Our evidence that early and geographically diverse Remote Oceanian individuals had little or no Papuan ancestry contradicts models in which there were significant Papuan contributions to Lapita people before their dispersal into Remote Oceania^{3–5}. Instead, our results show that the Papuan genetic signature appeared in many Remote Oceanian populations only subsequent to initial settlement. To gain further insight into when the Papuan ancestry may have become ubiquitous in Remote Oceanians, we leveraged the fact that chromosome segments from ancestral populations break up at a known rate owing to recombination and that the length distribution of these segments translates to a date of mixture²⁰. We estimate dates of approximately 50–80 generations ago using ALDER²¹, or 1,500–2,300 BP assuming 28.1 years (see Methods) per generation²² (Fig. 2d and Extended Data Fig. 3). We combined the statistical error of the genetic estimate and the uncertainty about the generation interval (Methods), and obtained a 95% confidence interval of 1,239–1,927 BP for a pool of Polynesians, all of whom have similar Papuan ancestry proportions. This finding that Papuan–First Remote Oceanian mixture occurred long after the end of the Lapita period implies that the Polynesian ancestral population was not fully formed at that time, although we note that alternative methods for dating Papuan admixture in Remote Oceanians arrived at older dates^{4,23–25}. However, our ALDER dates are supported by direct ancient DNA evidence, as the Tongan individual at 2,680–2,340 BP carried little or no Papuan ancestry, providing unambiguous confirmation that the ancestral population of Polynesians was not fully formed or widespread by the end of the Lapita period.

We used *qpGraph* to explore models of population separation and mixture that might accommodate the ancient DNA data²⁶ (Supplementary Information section 3). We obtained fits using models in which Polynesians today are mixtures of First Remote Oceanians and a Papuan population related to Highland New Guineans (Fig. 3a). We also obtained consistent findings using *TreeMix*²⁷ (Extended Data Fig. 4). In Fig. 3 we show the best fitting model, which suggests that the ancient individuals from Vanuatu and Tonga descended from an ancestral (presumably Lapita) population that separated earlier from the population that is the primary component in present-day Polynesians. This implies that not just Papuan ancestry but also deeply branching First Remote Oceanian ancestry was introduced to Remote Oceania through movement of people after the time of the ancient individuals. Thus, the minimum 25% Papuan ancestry seen in present-day Remote Oceanians is a conservative underestimate of the later population displacement. It is unlikely that there was 100% replacement, however, as we observed weak excess affinity of present-day Tongans to the ancient Tongan individuals in symmetry tests (see Methods). More deeply in time, our modelling indicates that Philippine populations (Kankanaey) are the closest outgroup to the First Remote Oceanians, indigenous Taiwanese (Atayal) second closest, and mainland south-east Asians such as the Dai most remote, consistent with models of population movement along a route from Taiwan to the Philippines to Near Oceania to Remote Oceania²⁸. We were surprised that we could not fit Australians as outgroups to New Guinean Highlanders and the

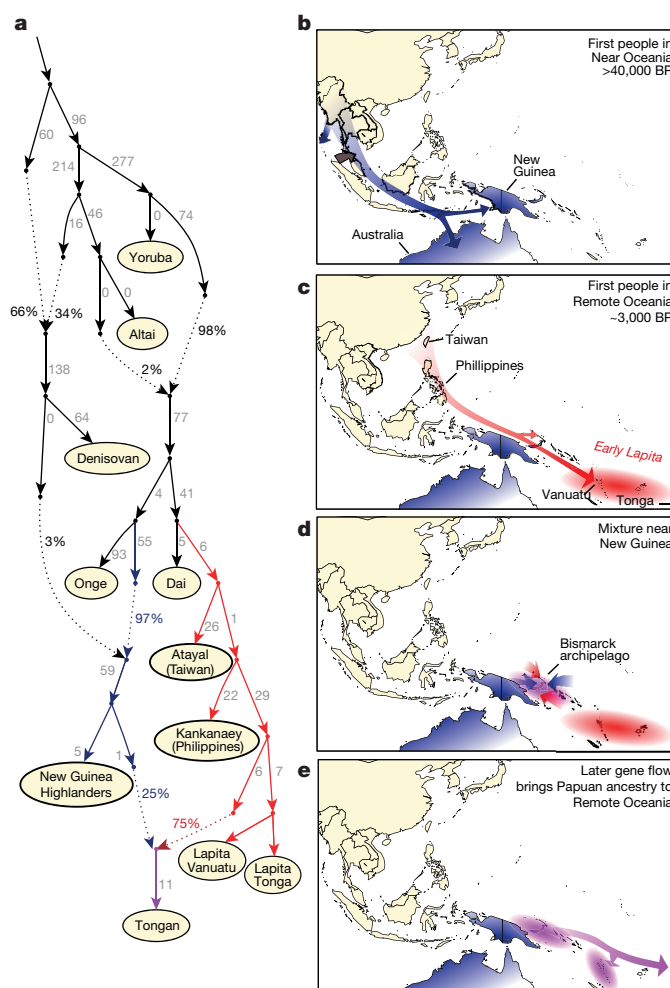


Figure 3 | A model of population history. **a**, A model of population relationships that fits allele frequency patterns (all empirical f -statistics within 3 standard errors of expectation). Branch lengths are shown in units of $F_{ST} \times 1,000$. Admixture edges show mixture proportions. Altai, the Altai Neanderthal genome. **b**, A model of population movements more than 40,000 years ago in which modern humans arrived in the Australia–New Guinea region (blue shading) and mixed with archaic Denisovans (brown arrow). **c**, A model of events before 3,000 years ago, in which the First Remote Oceanian population formed by spread of a population of ultimate East Asian origin to a region including Vanuatu and Tonga, and experienced little or no mixture with the Papuans they encountered along the journey (red shading). Note that geographic routes are speculative. **d**, A model of populations of mixed Papuan–First Remote Oceanian ancestry in Near Oceania less than 3,000 years ago in a patchwork of islands with different proportions of First Remote Oceanian ancestry (pink shading). **e**, A model of secondary expansion of admixed populations bringing Papuan ancestry into Remote Oceania, which was still not complete in Tonga by the date of the Talasiu individual at 2,680–2,340 BP.

Papuan ancestry in Polynesians (Extended Data Fig. 5). However, we could fit Australians as deriving from a mixture of an ancient Australian lineage and a Papuan lineage from the same group that expanded into Polynesia. This is plausible if there was continuing gene flow between New Guinea and Australia. Another parsimonious model is that the ancestry in present-day Polynesians is not all Papuan, but a Papuan–Australian mix.

Previous studies of mitochondrial DNA and Y chromosomes suggested that present-day people of the South Pacific harbour more East Asian ancestry from female than from male ancestors³. Our genome-wide analyses confirm a significant excess of First Remote Oceanian ancestry on the X chromosome compared to the autosomes (Z scores up to 10) (Fig. 2b). Females carry two-thirds of the X chromosomes in

a population but only half of the autosomes (Extended Data Fig. 6), and we compared the ancestry estimates in these two parts of the genome to obtain the most accurate estimates of sex-biased admixture in diverse Oceanians to date (Extended Data Fig. 6 and Extended Data Table 4). It has been suggested that matrilineal social structure in the primarily First Remote Oceanian ancestry populations of the region is one likely factor to explain these patterns^{29,30}. However, it is also possible that some of these patterns reflect a scenario in which the later movement of Papuan ancestry into Remote Oceania was largely mediated by males who then mixed with resident females.

Our study has shown that many of the first humans in Remote Oceania had little, if any, Papuan ancestry, in stark contrast to the situation today. While our findings cannot rule out the possibility that multiple groups—some of which carried substantial amounts of Papuan ancestry—settled Remote Oceania early on, the lack of such ancestry in both Vanuatu and Tonga can be more parsimoniously explained by later population movements bringing the Papuan ancestry. The scenario emerging from ancient DNA analysis is thus radically different from that suggested by previous genetic studies, which have generally posited that the first people in Remote Oceania and Polynesia^{2–5} had substantial Papuan ancestry. Our finding of major post-Lapita flow of Papuan ancestry into Remote Oceania also cannot be related to the later arrival of Papuan ancestry that has been suggested for Fiji, which is estimated to have occurred at least a millennium later at 500 BP⁴ or 1,100 BP²⁴ (Fig. 2). Systematic study of ancient DNA from throughout Remote Oceania should make it possible to provide a detailed chronicle of the population movements and sex-biased population mixtures that shaped the ancestry of present-day Oceanians.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 April; accepted 13 September 2016.

Published online 3 October 2016.

- Sheppard, P. J., Chiu, S. & Walter, R. Re-dating Lapita movement into Remote Oceania. *J. Pacific Archaeol.* **6**, 26–36 (2015).
- Kayser, M. *et al.* Genome-wide analysis indicates more Asian than Melanesian ancestry of Polynesians. *Am. J. Hum. Genet.* **82**, 194–198 (2008).
- Kayser, M. The human genetic history of Oceania: near and remote views of dispersal. *Curr. Biol.* **20**, R194–R201 (2010).
- Wollstein, A. *et al.* Demographic history of Oceania inferred from genome-wide data. *Curr. Biol.* **20**, 1983–1992 (2010).
- Matisoo-Smith, E. Ancient DNA and the human settlement of the Pacific: a review. *J. Hum. Evol.* **79**, 93–104 (2015).
- Bellwood, P. S. *First Farmers: the Origins of Agricultural Societies* (Blackwell Publishing, 2005).
- Duggan, A. T. *et al.* Maternal history of Oceania from complete mtDNA genomes: contrasting ancient diversity with recent homogenization due to the Austronesian expansion. *Am. J. Hum. Genet.* **94**, 721–733 (2014).
- Kayser, M. *et al.* Melanesian origin of Polynesian Y chromosomes. *Curr. Biol.* **10**, 1237–1246 (2000).
- Blust, R. Remote Melanesia: one history or two? An addendum to Donohue and Denham. *Oceanic Linguistics* **47**, 445–459 (2008).
- Friedlaender, J. S. *et al.* The genetic structure of Pacific Islanders. *PLoS Genet.* **4**, e19 (2008).
- Pinhasi, R. *et al.* Optimal ancient DNA yields from the inner ear part of the human petrous bone. *PLoS One* **10**, e0129102 (2015).
- Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl Acad. Sci. USA* **110**, 15758–15763 (2013).
- Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Phil. Trans. R. Soc. Lond. B* **370**, 20130624 (2015).
- Fu, Q. *et al.* An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216–219 (2015).
- Skoglund, P., Storå, J., Götherström, A. & Jakobsson, M. Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J. Archaeol. Sci.* **40**, 4477–4482 (2013).
- Melton, T. *et al.* Polynesian genetic affinities with Southeast Asian populations as identified by mtDNA analysis. *Am. J. Hum. Genet.* **57**, 403–414 (1995).
- Skoglund, P. *et al.* Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science* **336**, 466–469 (2012).
- Skoglund, P. *et al.* Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc. Natl Acad. Sci. USA* **111**, 2229–2234 (2014).
- Reich, D. *et al.* Reconstructing Native American population history. *Nature* **488**, 370–374 (2012).
- Moorjani, P. *et al.* The history of African gene flow into Southern Europeans, Levantines, and Jews. *PLoS Genet.* **7**, e1001373 (2011).
- Loh, P.-R. *et al.* Inferring admixture histories of human populations using linkage disequilibrium. *Genetics* **193**, 1233–1254 (2013).
- Fenner, J. N. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am. J. Phys. Anthropol.* **128**, 415–423 (2005).
- Xu, S., Pugach, I., Stoneking, M., Kayser, M. & Jin, L. Genetic dating indicates that the Asian-Papuan admixture through Eastern Indonesia corresponds to the Austronesian expansion. *Proc. Natl Acad. Sci. USA* **109**, 4574–4579 (2012).
- Pugach, I., Matveyev, R., Wollstein, A., Kayser, M. & Stoneking, M. Dating the age of admixture via wavelet transform analysis of genome-wide data. *Genome Biol.* **12**, R19 (2011).
- Lipson, M. *et al.* Reconstructing Austronesian population history in Island Southeast Asia. *Nat. Commun.* **5**, 4689 (2014).
- Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
- Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967 (2012).
- Bellwood, P. Holocene population history in the Pacific region as a model for worldwide food producer dispersals. *Curr. Anthropol.* **52**, S363–S378 (2011).
- Jordan, F. M., Gray, R. D., Greenhill, S. J. & Mace, R. Matrilineal residence is ancestral in Austronesian societies. *Proc. R. Soc. Lond. B* **276**, 1957–1964 (2009).
- Stephen Lansing, J. *et al.* An ongoing Austronesian expansion in Island Southeast Asia. *J. Anthropol. Archaeol.* **30**, 262–272 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the 356 volunteers who donated samples for genome-wide analysis; M. Stoneking for co-funding genotyping of the Bismarck samples; M. Brilliant, H. Norton, and L. Scheinfeldt, for help in the preparation of the Bismarck samples and establishment of a repository for them at the Marshfield Foundation; A. Wissgott for help in data generation from the ancient Tongan individual; A. Kim, I. Pugach, and M. Stoneking for comments, and I. Mathieson for critiques and advice on estimating sex-specific ancestral contributions. The maps in Figs 1a and 3b–e maps were plotted in R using the *worldO* map of the ‘fields’ and ‘maps’ packages (using public domain data from the CIA World Data Bank II). P.S. was supported by the Wenner-Gren foundation, SciLifeLab, and the Swedish Research Council (VR grant 2014-453). The Teouma research by M.S. and S.B. was supported by the Australian Research Council (Discovery Grants DP0880789 and DP110101415), the National Geographic Society, and the Australia-Pacific Science Foundation. F.V. was supported by CNRS-UMR 7041. M.N. was supported by an Irish Research Council grant (GOIPD/2013/1). D.F. was supported by an Irish Research Council grant (GOIPG/2013/36). Q.F. was funded by the Key Research Program of Frontier Sciences of CAS (QYDZB-SS W-DQC003), the National Natural Science Foundation of China (L1524016) and the Chinese Academy of Sciences Discipline Development Strategy Project (2015-DX-C-03). T.K. was supported by ERC starting grant FP7-261213. C.P. and J.K. were supported by the Baden Württemberg Foundation. J.K. was supported by the DFG grant KR 4015/1-1 and the Max Planck Society. R.P. was supported by ERC starting grant ADNABIOARC (263441). D.R. was supported by NIH grant GM100233, by NSF HOMINID BCS-1032255, and is a Howard Hughes Medical Institute investigator.

Author Contributions N.P., J.K., R.P. and D.R. supervised the study. M.S., F.V., S.B., G.R.C., and C.R. assembled archaeological material and information. P.S., C.P., Q.F., M.L., S.M., N.R. and D.R. analysed genetic data. C.P., K.Si., F.P., D.F., E.H., M.N., N.R. and K.St. performed laboratory work. S.A., M.P.C., F.R.F., J.S.F., T.K., G.K., P.K., D.A.M., F-X.R., and J.T.S.W. assembled the sample collection from present-day populations. P.S. and D.R. wrote the manuscript with major input from C.P., M.S., F.V., G.R.C., M.P.C., J.S.F., J.K. and R.P. and additional input from all other co-authors.

Author Information The aligned sequences are available through the European Nucleotide Archive under accession number PRJEB14728. The newly reported SNP genotyping data for the subset of individuals who provided informed consent consistent with fully public distribution are available at http://genetics.med.harvard.edu/reichlab/Reich_Lab/Datasets.html. To access data for the remaining samples, 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.” Extended Data Table 2 specifies which samples are consistent with which type of data distribution. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.S. (skoglund@genetics.med.harvard.edu), R.P. (ron.pinhasi@ucd.ie) or D.R. (reich@genetics.med.harvard.edu).

Reviewer Information *Nature* thanks P. Bellwood, C. Capelli and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Ancient DNA sampling, extraction, library preparation, enrichment and sequencing. The Vanuatu skeletal samples B30A, B10B, B17 were analysed with permission from the Vanuatu National Museum and the excavators of the Teouma site. The Tonga skeletal sample SK10 was analysed with permission from the excavators of the Talasiu site.

All preparation of skeletal samples, DNA extraction, and library preparation was carried out in dedicated ancient DNA laboratories at University College Dublin, Ireland (sample preparation of the three Vanuatu individuals), at Harvard Medical School in Boston, USA (DNA extraction and library preparation of the three Vanuatu individuals), and at the Max Planck Institute for the Science of Human History in Jena, Germany (sample preparation, DNA extraction and library preparation of the Tonga individual). Each of these facilities is spatially separated from other molecular biology laboratories, and measures are taken to protect ancient individuals from contamination including HEPA filtered air, head-to-toe suits, face masks with visors, multiple layers of gloves, bleaching of all surfaces, ultraviolet light (UVC) decontamination of (non-sensitive) consumables and chemicals, and UVC decontamination of the facility when researchers are not in the room³¹. The final step of the library preparation (amplification) was performed outside the ancient DNA laboratory.

We prepared powder from the cochlea of petrous bones, extracted DNA¹², and prepared libraries with standard protocols (ref. 13 for the Vanuatu individuals and ref. 32 for the Tonga individual). For the three Vanuatu individuals, the first library was prepared in the presence of uracil DNA glycosylase (UDG) to cut out errors due to ancient DNA damage, whereas the remaining three libraries as well as the Tonga library were prepared without UDG as this preserves more DNA for any given sample. We performed in-solution enrichment using previously reported protocols^{13,14,33,34} for a targeted set of 1,237,207 SNPs that comprises two previously separately reported sets of 394,577 SNPs³⁴ and 842,630 SNPs¹⁴. We sequenced the product on Illumina NextSeq500 instruments for 2 × 75 cycles. Following demultiplexing, and, for the Vanuatu samples, removal of both oligonucleotide barcodes that were used to identify the libraries and trailing adaptor sequences, we merged the forward and reverse reads of each read pair requiring a 15-base pair overlap (allowing one mismatch). We then aligned merged sequences to the human genome *hg19* using *BWA 0.6.1* (ref. 35). We removed sequences aligned to identical outer coordinates, choosing the highest quality sequence for each duplication cluster. We merged the data from the four libraries for each Vanuatu individual.

Genomic analysis. We determined sex by comparing the number of X and Y chromosome alignments¹⁵. We estimated damage patterns using *PMDtools* v0.60¹⁸, separating damage patterns observed inside and outside a CpG context. Since all four individuals were female, we could not estimate contamination using X chromosome data. We investigated whether there was evidence of excess relatedness between any pair of individuals among the Vanuatu individuals, but found that the pairwise mismatch rate using panel 5 of the Affymetrix Human Origins array (see below) was 19.8% ± 0.4% for I1368/I1369, 19.7% ± 0.6% for I1368/I1370, and 20.5% ± 0.4% for I1369/I1370. This suggests no atypical pair of individuals and a similar within-population mismatch rate (heterozygosity) as some present-day Polynesian populations (Fig. 2).

Genotyping of present-day humans. We genotyped 356 individuals from 38 southeast Asian and Oceanian populations on the Affymetrix Human Origins array (Extended Data Table 2). The individuals all contributed DNA samples voluntarily and provided informed consent consistent with studies of human genetic variation and history. Ethical approval of the component studies was provided by the Singapore Health IRB, the Research Ethics Committee at the Facultés de Médecine de Toulouse, the Brunei Medical and Health Research Ethics Committee, the University of Cambridge Biology Research Ethics Committee, the Government of Papua New Guinea Medical Research Advisory Committee, and the Temple University IRB. The collection of genome-wide variation data on de-identified samples was approved by the Harvard Human Research Protection Program (Protocol 11681), re-reviewed on 12 July 2016.

We restricted analysis to samples that had >95% genotyping completeness and that were not visual outliers in PCA with respect to the main cluster of samples in the group. We merged with previously reported Affymetrix Human Origins SNP array data^{26,36–39}. We also co-analysed our data with samples genotyped on the Affymetrix 6.0 platform where we removed three previously published³⁹ Rapa Nui individuals (*5s5j*, *XB3B*, and *3p3p*), and two previously published⁴⁰ Samoan individuals (*PLY_07* and *PLY_11*), all of which appeared to have recent European ancestry based on clustering analyses. We finally compared our data to high-coverage genomes from an archaic Neanderthal and an archaic Denisovan, both from Denisova Cave in the Altai Mountains of Siberia^{41–43}.

Population genetic analysis. When overlapping with the Affymetrix Human Origins SNP array data set of present-day human populations, we have between 74,000 and 126,000 SNPs covered at least once for each of the four individuals

(Extended Data Table 1). This is more than the minimum coverage required for high-resolution analysis using allele frequency correlation statistics, e.g. 10,000 SNPs per individual according to Supplementary Information section 6.2 of ref. 44, a study that had the same median coverage (0.19 ×) as ours (the range in the present study is 0.14–0.26 ×). For all analyses, we called genotypes by randomly sampling a single non-duplicate sequence read at each position⁴⁵. This procedure is standard for analysis of low-coverage ancient DNA data and is also often used for higher-coverage data to minimize reference genome biases that can be introduced when determining diploid genotypes^{14,17,34,36,41,44–50}. For the *qpAdm*, *qpWave* and *qpGraph* analyses we excluded transition SNPs to avoid potential biases from post-mortem damage (see below).

We performed PCA using *smartpca*⁵¹, with the option *inbreed: YES* in order to sample a single genotype from each individual randomly to match the pseudo-haploid nature of the ancient DNA genotypes from the ancient individuals⁵². We computed f_3 , f_4 and D -statistics as in ref. 26, and F_{ST} using the Hudson estimator and randomly sampled a single haploid sequence to represent each individual at each SNP position, using *popstats*³⁸. We estimated the date of admixture using *ALDER*²¹. We tested the consistency of a matrix of f_4 -statistics with one or more sources of ancestry with respect to a set of outgroups (New_Guinea, Denisova, Sardinian, English, Yakut, Chukchi, Mala, Japanese, Ju_hoan_North, Mixe, Onge, Yoruba, and Mbuti) using *qpWave*^{19,34}.

For the ancient individuals and all present-day populations genotyped on the Human Origins array, we used *qpAdm*³⁴, which estimates ancestry proportions from two or more proxy source populations assuming that the proxies are more closely related to the real source populations than they are to a set of outgroups (*qpAdm* also provides a formal statistical test for whether this is the case, which passes in the context that we use it here). We estimated First Remote Oceanian and Papuan ancestry using Denisova, Sardinian, English, Yakut, Chukchi, Mala, Japanese, Ju_hoan_North, Mixe, Onge, Yoruba, and Mbuti as outgroups and New_Guinea and Ami as proxies for the Papuan and First Remote Oceanian source populations, respectively. For the ancient individuals, we excluded all transition SNPs to avoid possible biases due to post-mortem damage, resulting in 35,194 transversion SNPs for Vanuatu (covered by at least one of the individuals) and 22,030 for Tonga. For estimating *qpAdm* ancestry proportions in the Affymetrix 6.0 Polynesian data, we used whole-genome sequences from the same populations as outgroups⁵³. We estimated Denisovan ancestry using the Denisovan genome and Japanese as the two sources, and chimpanzee (*Pan troglodytes*), Ju_hoan_North, Mbuti, Yoruba, Dinka and the Altai Neanderthal genome as outgroups.

We computed conditional heterozygosity using panel 5 of the Affymetrix Human Origins array, which contains SNPs ascertained as heterozygous in a single West African Yoruba individual. This provides an unbiased estimate of relative heterozygosity since the Yoruba individual is approximately symmetrically related to all Oceanians (Denisovan ancestry violates this assumption but is not expected to change the ranking of populations). We estimated heterozygosity as the average pairwise mismatch rate when sampling 2 chromosomes from two different individuals using *popstats*³⁸, restricting to transversion SNPs for all populations and computing standard errors using a weighted block jack-knife.

For authentication, we used *PMDtools*¹⁸ to extract sequences with clear evidence of post-mortem damage patterns (PMD score of at least 3), disregarding individual bases with phred-scaled base quality <30. We randomly sampled new pseudo-haploid genotypes from the resulting set of sequences and projected the ancient individuals onto the principal components inferred from the present-day populations as above. After this filtering, we retained 68,450 SNPs for I1368; 98,722 SNPs for I1369; 83,024 SNPs for I1370; and 117,023 SNPs for CP30. The ninety-nine per cent confidence intervals for *qpAdm* estimates of Papuan ancestry (see above) using the PMD score-restricted data were 0–21% for the ancient Vanuatu individuals and 0–24% for the ancient Tonga individual, consistent with the confidence intervals obtained from the full data.

To test whether the ancient Vanuatu and the ancient Tonga individuals form a clade, we used *qpWave* to test whether a model of Dai, Ami, Kankanaey and a fourth population were consistent with being outgroups to the two ancient sample groups (we used Dai, Ami and Kankanaey as these span present-day Mainland East Asia, Taiwan, and the Philippines, and lack Papuan ancestry to the limits of our resolution). The analysis used the ~12,000 SNPs that remained after excluding transition SNPs and SNPs missing in one of the two ancient sample groups. We found that the model was consistent with the data for all tested Oceanian and Asian populations shown in Fig. 1, but that the lowest P value was observed for present-day Tongans ($P = 0.09$). We also found that $f_4(\text{Ami, Present day Tongan; Lapita_Vanuatu, Lapita_Tonga}) = 0.006$, $Z = 3.2$, when using all SNPs. This suggests a possible affinity between present-day Tongans and the ancient Tongan individual, consistent with the hypothesis that the ancient population of Tonga with little or no Papuan ancestry may have contributed some of the ancestry of present-day Tongans.

Admixture date estimation. To estimate the date of historical admixture between First Remote Oceanians and Papuans, we used ALDER^{21,25} on the full Human Origins array data, with New Guinean Highlanders and Han Chinese as the two sources. We use Han Chinese for this analysis owing to their substantial sample size compared to populations more closely related to the ancestral First Remote Oceanian population such as the ancient individuals we analysed, indigenous Taiwanese, and indigenous Philippine groups. ALDER estimates are known to be robust even when using imperfect surrogates for the ancestral populations in this way²⁶. We estimate an admixture date for a pool of Polynesian populations by combining data from Tongan, Tikopia, Russell and Bellona populations, all genotyped on the Affymetrix Human Origins SNP array.

ALDER and other methods based on admixture linkage disequilibrium estimate dates in units of generations, which need to be converted to years. For this purpose, we require an estimate of the generation interval—the average age of a parent at the time their gametes were formed—weighted by the fraction of recombination events that occur in each sex (62.3% of all autosomal crossovers are estimated to occur in females, based on table 1 of ref. 54.). Using estimates from the anthropological literature, this quantity is 27.8 years for hunter–gathering societies, 28.6 years for developed nation states, and 29.6 years for less developed nation states²². These numbers are in the range of the point estimate we use of 28.1 years based on breakdown of admixture linkage disequilibrium in radiocarbon-dated ancient genomes⁵⁵. To account for the substantial variability in generation intervals across human societies, we use the sample standard error of 2.15 years measured across eleven diverse hunter-gatherer groups based on Table 4 of ref. 22. The date estimates in Fig. 2 and Extended Data Fig. 4 thus use a generation interval of 28.1 years, and combine the standard error from ALDER (a) with the uncertainty in generation time, that is, $\sqrt{a^2 \times 2.15^2 + A^2 \times 2.15^2 + 28.1^2 \times a^2}$, where A is the ALDER point estimate in number of generations.

We do not subtract 66 years from the dates produced by ALDER to obtain BP dates (conventionally the date before 1950 AD, 66 years ago), because what ALDER is estimating is a number that is close to the BP date. To see this, note that ALDER estimates the date between when chromosomes of the two ancestries began crossing over (one generation after mixing began), and the date of the last crossover (when the germ cells that mixed to produce the present-day samples in our study were formed, likely one or two generations before 2016 AD). Accounting for these corrections means that ALDER is estimating a date of mixture that is likely to be within a generation of the true BP date.

Fitting models of population history. We used *qpGraph*^{26,56} to assess the fit of admixture graph models to allele frequency correlation patterns as measured by f_2 , f_3 , and f_4 statistics. We started with a skeleton phylogenetic tree consisting of Yoruba, New Guinea, Dai, Atayal, Kankanaey and the pool of ancient Vanuatu individuals. We added Tongan, Mamanwa (a Philippine Negrito group), Nasioi and Kolombangara, respectively, to all possible edges in the tree, and retained only the graph solutions that provided no individual f_4 statistics with $|Z| > 3$ between empirical and predicted statistics. For the extended version of the admixture graph, we also added Australians to all possible edges of the graph that included these populations. Finally, we modelled the previously documented admixture history relating Denisovans and the Altai Neanderthal genome to the outgroup chimpanzee and the anatomically modern human populations, to which we added the Andamanese Onge and the ancient Tongan individual. The final graph visualized in Fig. 3 used 10,893 SNPs after restricting to transversion SNPs to avoid complications due to ancient DNA damage and also SNPs with coverage in all groups. For more information on the admixture graph inference procedure, see Supplementary Information section 3.

As an alternative inference method, we used *Treemix* v1.12 (ref. 27) to test models for Yoruba, Dai, Atayal, Kankanaey, Tongan, New Guinean Highlanders, the ancient Vanuatu individual and the ancient Tongan individual. The total number of SNPs after excluding transitions, SNPs with minor allele count of less than 4 in the selected data, and SNPs where one population had missing data, was 10,119, which we divided into 337 blocks of 30 consecutive SNPs each to estimate the covariance matrix. We first fitted a maximum likelihood tree of all populations, but found that several of the fitted allele frequency covariances deviated from those empirically observed by up to 16.4 standard errors. We then used the automated heuristic optimization in *Treemix* to infer a graph model with one admixture event using the same populations, and found that the optimal fit was for a model with an admixture event in the history of Tongans, where one portion of their ancestry diverged before the split of the ancestors of the ancient Vanuatu and Tonga individuals, and the other (25% \pm 3%) derived from the New Guinean lineage. This maximum deviation between empirical and model covariances observed for the graph with one admixture edge was 1.6, indicating a good fit, consistent with our investigation of models using *qpGraph*.

Female and male ancestral contributions. To estimate the proportion of female ancestors (F) and male ancestors (M) for a given population, we used two different methods both based on the estimates of ancestry for the X chromosome and autosomes. Both used the same underlying model, in which the observed admixture proportion estimates that \hat{H}_{auto} and \hat{H}_X for the autosomes and X chromosome, respectively, depend on M and F such that:

$$\hat{H}_{auto} = \left(\frac{M}{2} + \frac{F}{2} \right) \quad (1)$$

$$\hat{H}_X = \left(\frac{M}{3} + \frac{2F}{3} \right) \quad (2)$$

The first approach obtains unbounded point estimates of M and F by rearranging equations:

$$M = 4 \times \hat{H}_{auto} - 3 \times \hat{H}_X \quad (3)$$

$$F = 3 \times \hat{H}_X - 2 \times \hat{H}_{auto} \quad (4)$$

Similarly, we obtained standard errors for M and F using the weighted block jackknife standard errors for \hat{H}_{auto} and \hat{H}_X , SE_{auto} and SE_X , as

$$SE_M = \sqrt{(9 \times SE_X^2) + (16 \times SE_{auto}^2)} \quad (5)$$

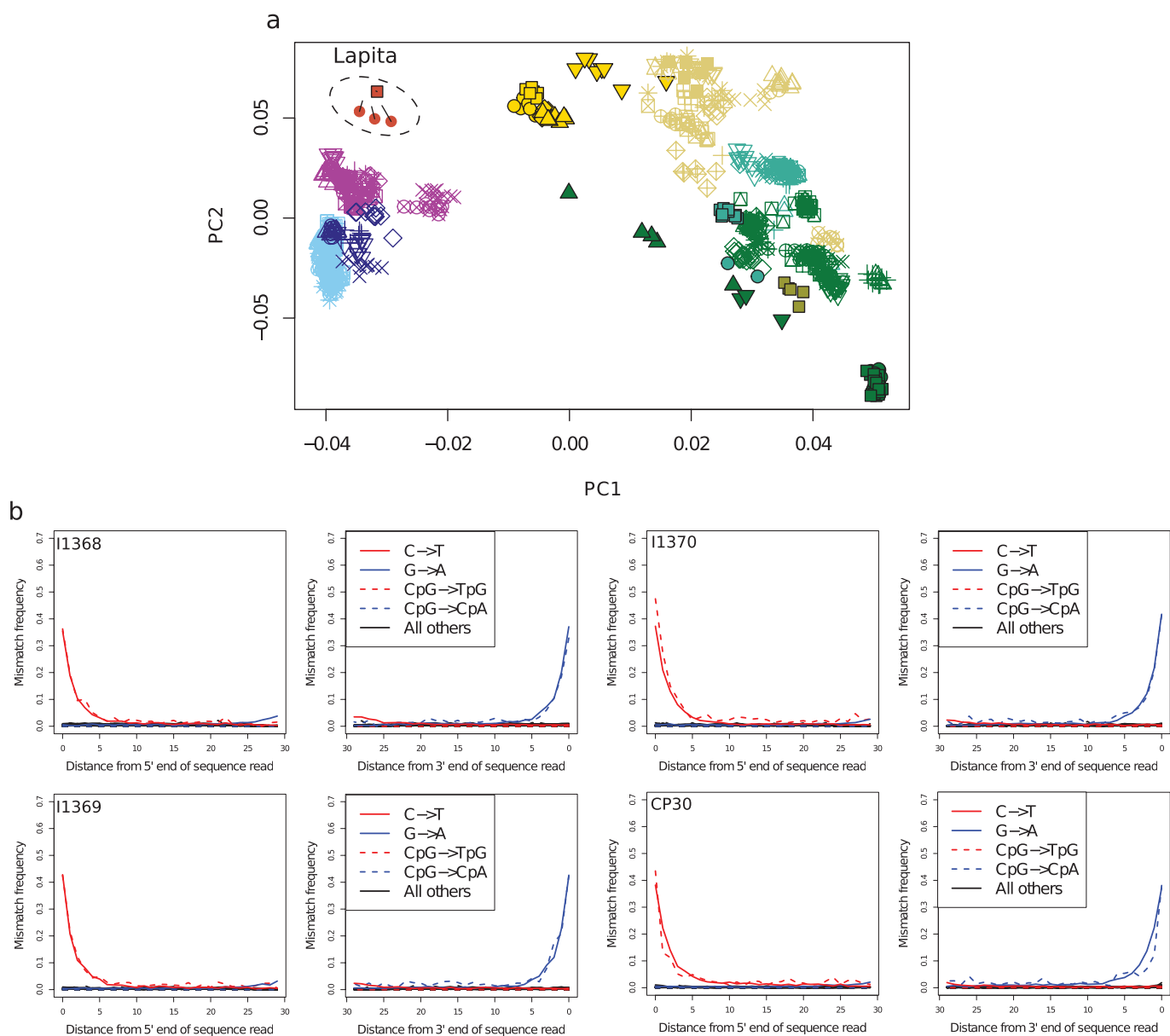
$$SE_F = \sqrt{(9 \times SE_X^2) + (4 \times SE_{auto}^2)} \quad (6)$$

As an alternative to estimating M and F , we took an approximate Bayesian approach by performing 1 million simulations in which M and F were sampled from a uniform prior distribution (0, 1). We then simulated ancestry estimates specifying normal distributions with means and standard errors matching the empirical values (equations 1 and 2). We used the *abc* R package⁵⁷ to run a rejection algorithm retaining the 1% of all simulation replicates with the closest Euclidean distances to the empirical \hat{H}_{auto} and \hat{H}_X , and performed local linear regression on log-transformed summary statistics to obtain a posterior distribution. The results of the two methods are qualitatively similar. In Extended Data Fig. 6, we plot the posterior intervals of these distributions for selected populations.

Sample size. No statistical methods were used to predetermine sample size.

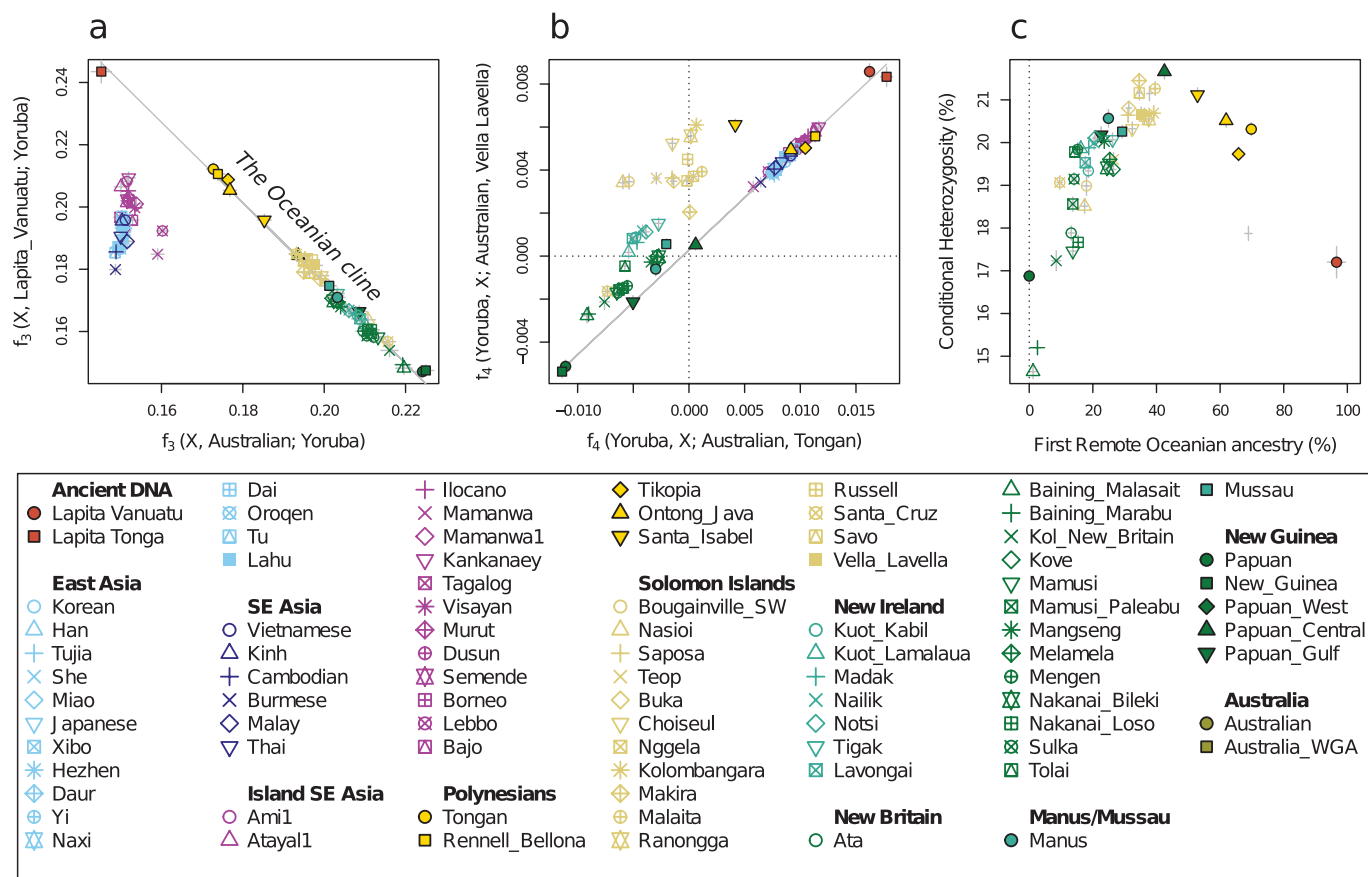
- Knapp, M., Clarke, A. C., Horsburgh, K. A. & Matisoo-Smith, E. A. Setting the stage – Building and working in an ancient DNA laboratory. *Ann. Anat. Anat. Anz.* **194**, 3–6 (2012).
- Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Prot.* <http://dx.doi.org/10.1101/pdb.prot5448> (2010).
- Fu, Q. *et al.* DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl Acad. Sci. USA* **110**, 2223–2227 (2013).
- Haak, W. *et al.* Massive migration from the steppe is a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
- Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
- Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014).
- Qin, P. & Stoneking, M. Denisovan ancestry in East Eurasian and Native American populations. *Mol. Biol. Evol.* **32**, 2665–2674 (2015).
- Skoglund, P. *et al.* Genetic evidence for two founding populations of the Americas. *Nature* **525**, 104–108 (2015).
- Moreno-Mayar, J. V. *et al.* Genome-wide ancestry patterns in Rapanui suggest pre-European admixture with Native Americans. *Curr. Biol.* **24**, 2518–2525 (2014).
- Reich, D. *et al.* Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. *Am. J. Hum. Genet.* **89**, 516–528 (2011).
- Reich, D. *et al.* Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053–1060 (2010).
- Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222–226 (2012).
- Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014).
- Allentoft, M. E. *et al.* Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172 (2015).
- Green, R. E. *et al.* A draft sequence of the Neanderthal genome. *Science* **328**, 710–722 (2010).
- Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87–91 (2014).
- Raghavan, M. *et al.* The genetic prehistory of the New World Arctic. *Science* **345**, 1255832 (2014).

48. Rasmussen, M. *et al.* The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* **506**, 225–229 (2014).
49. Rasmussen, M. *et al.* The ancestry and affiliations of Kennewick Man. *Nature* **523**, 455–458 (2015).
50. Skoglund, P. *et al.* Genomic diversity and admixture differs for Stone-Age Scandinavian foragers and farmers. *Science* **344**, 747–750 (2014).
51. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
52. Skoglund, P. & Jakobsson, M. Archaic human ancestry in East Asia. *Proc. Natl Acad. Sci. USA* **45**, 18301–18306 (2011).
53. Sudmant, P. H. *et al.* Global diversity, population stratification, and selection of human copy-number variation. *Science* **349**, 1174–1181 (2015).
54. Kong, A. *et al.* A high-resolution recombination map of the human genome. *Nat. Genet.* **31**, 241–247 (2002).
55. Moorjani, P. *et al.* 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 (2016).
56. Reich, D., Thangaraj, K., Patterson, N., Price, A. L. & Singh, L. Reconstructing Indian population history. *Nature* **461**, 489–494 (2009).
57. Csilléry, K., Blum, M. G. B., Gaggiotti, O. E. & François, O. Approximate Bayesian Computation (ABC) in practice. *Trends Ecol. Evol.* **25**, 410–418 (2010).
58. Bronk Ramsey, C. *OxCal Program v4.2.4*. (Radiocarbon Accelerator Unit, Univ. Oxford, 2016).
59. Reimer, P. J. *et al.* IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 Years cal BP. *Radiocarbon* http://dx.doi.org/10.2458/azu_js_rc.55.16947 (2013).
60. Ambrose, S. H. Isotopic analysis of paleodiets: methodological and interpretive considerations in *Food and Nutrition in History and Anthropology (USA)* <http://agris.fao.org/aos/records/US9514697?output=xml> (1993).
61. Petchey, F., Spriggs, M., Bedford, S., Valentin, F. & Buckley, H. Radiocarbon dating of burials from the Teouma Lapita cemetery, Efate, Vanuatu. *J. Archaeol. Sci.* **50**, 227–242 (2014).
62. Petchey, F., Anderson, A., Zondervan, A., Ulm, S. & Hogg, A. New marine ΔR values for the South Pacific subtropical gyre region. *Radiocarbon* **50**, 373–397 (2008).



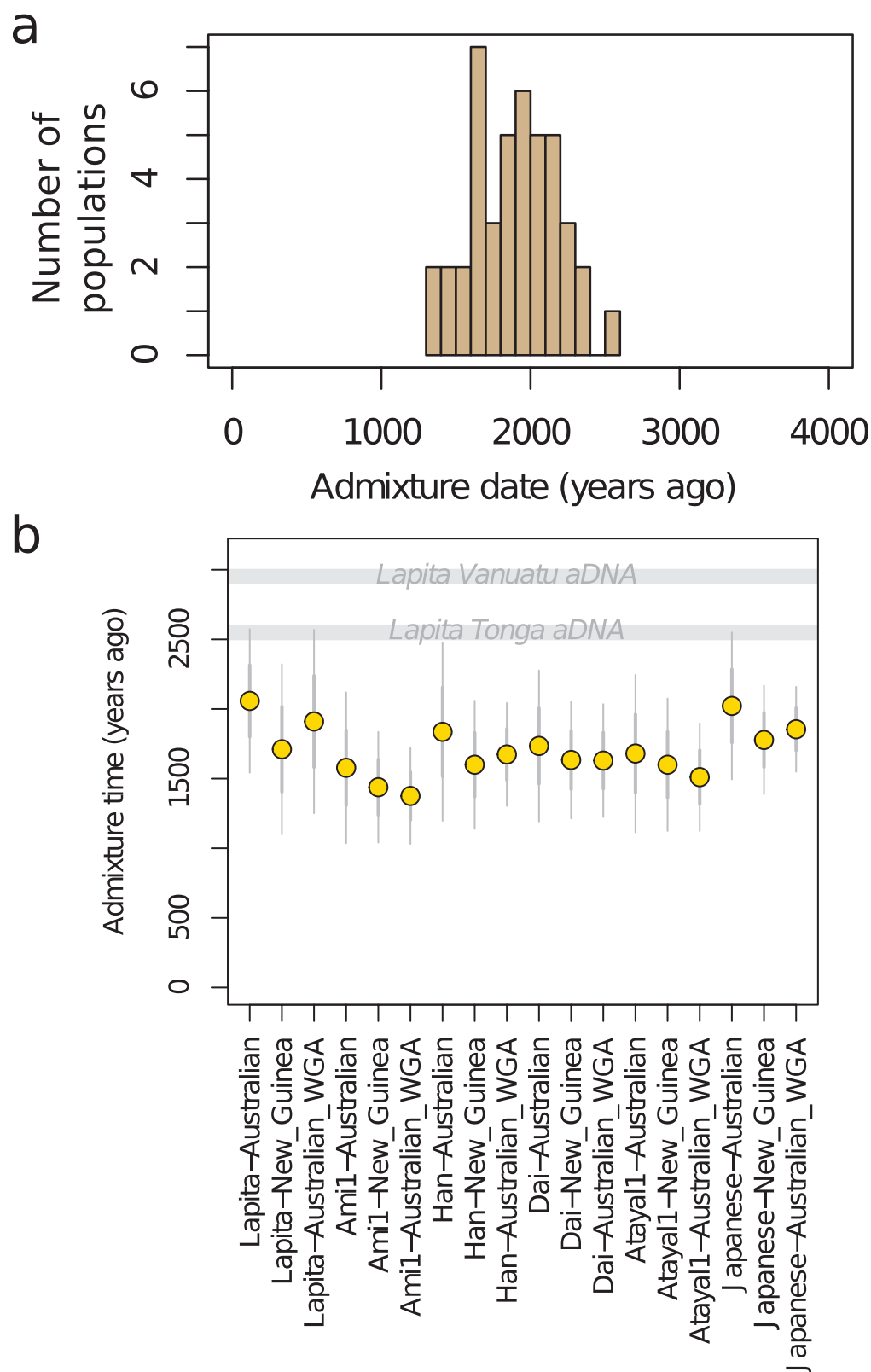
Extended Data Figure 1 | Ancient DNA authenticity. **a**, PCA performed as for Fig. 1, but with the four ancient individuals represented only by sequences that show clear evidence of post-mortem damage (PMD score of at least 3) to remove contaminating sequences that might be present^{17,18}. The numbers of SNPs remaining after restriction to damaged sequences is 68,450 SNPs for I1368; 98,722 SNPs for I1369; 83,024 SNPs for I1370; and 117,023 SNPs for CP30. The lines indicate the projection of the samples when no damage-restriction is performed. The large number

of SNPs retained, and the fact that the ancient individuals cluster tightly and have the same qualitative positioning in the plot as Fig. 1, indicates that contamination did not contribute to our findings. We also find that estimates of Papuan ancestry using PMD score restricted data are consistent with those obtained using the full data (see Methods). **b**, Post-mortem damage patterns for genome-wide in-solution enrichment data from four ancient individuals.



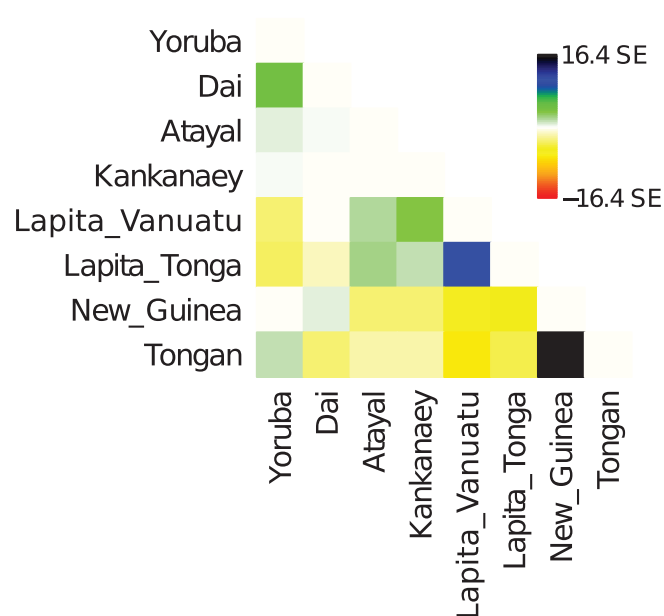
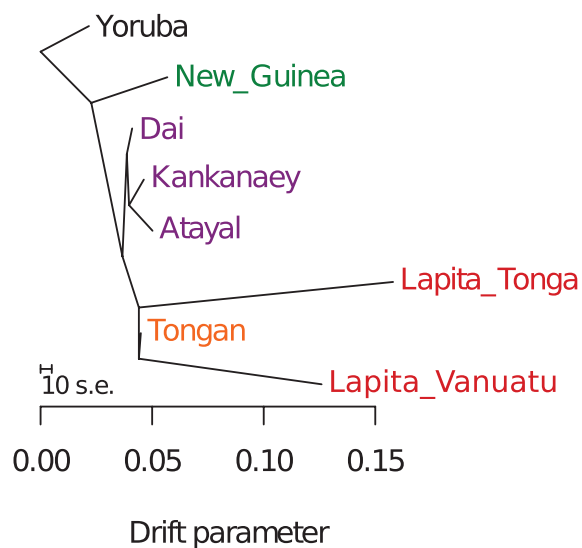
Extended Data Figure 2 | f -statistics document the Oceanian ancestry cline. **a**, Shared genetic drift with the ancient Vanuatu individuals is negatively related to shared drift with Australians. Except for the ancient Tongan individual, populations from Taiwan, the Philippines and Polynesia share the most genetic drift with the ancient Vanuatu individuals, who are not shown in the plot because they are used as reference in the computation. The trend line was fitted without the East Asian populations in the off-cline cluster. The absence of off-cline Oceanian individuals suggests the possibility that present-day Oceanians may largely be derived from a mixture of two source populations.

b, The ancient Vanuatu individuals and the ancient Tongan individual maximize statistics of the form $f_4(\text{Yoruba}, \text{Test}; \text{Australian}, \text{Oceanian})$, suggesting that they are the most closely related to the East Asian ancestry in Oceanians of any sampled population. The trend line was fitted using populations >0.005 on the x -axis, together with the two populations with the lowest values on the x -axis (Papuan and New_Guinea). **c**, Biplot of First Remote Oceanian ancestry proportions against conditional heterozygosity. Populations with intermediate admixture proportion show the greatest genetic diversity. Thick and thin error bars in all panels are 1 and 1.96 standard errors of the estimate, respectively.

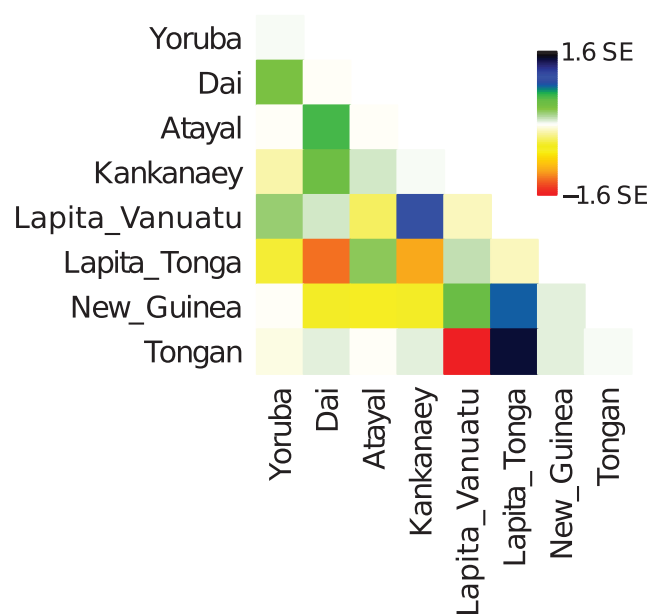
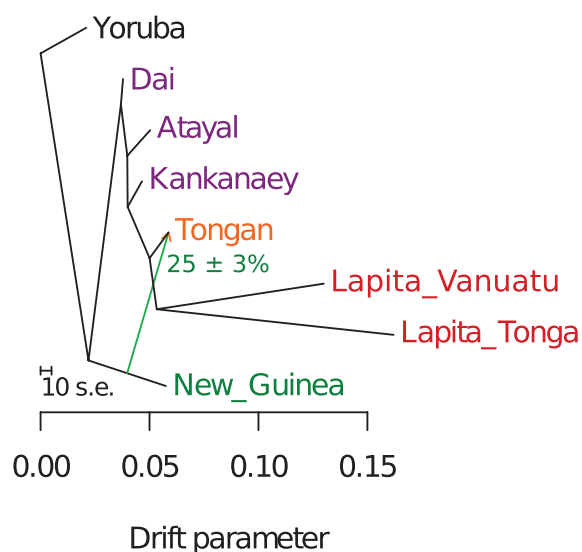


Extended Data Figure 3 | Admixture date estimates. **a**, Histogram of the point estimate dates in Fig. 2d. **b**, Admixture date estimates for Tongans using different pairs of source populations ('Lapita' in this figure refers to the pool of ancient Vanuatu individuals). Error bars show 1 (thick whiskers) and 1.96 (thin whiskers) standard errors. WGA, whole-genome amplified DNA.

a) rejected tree model



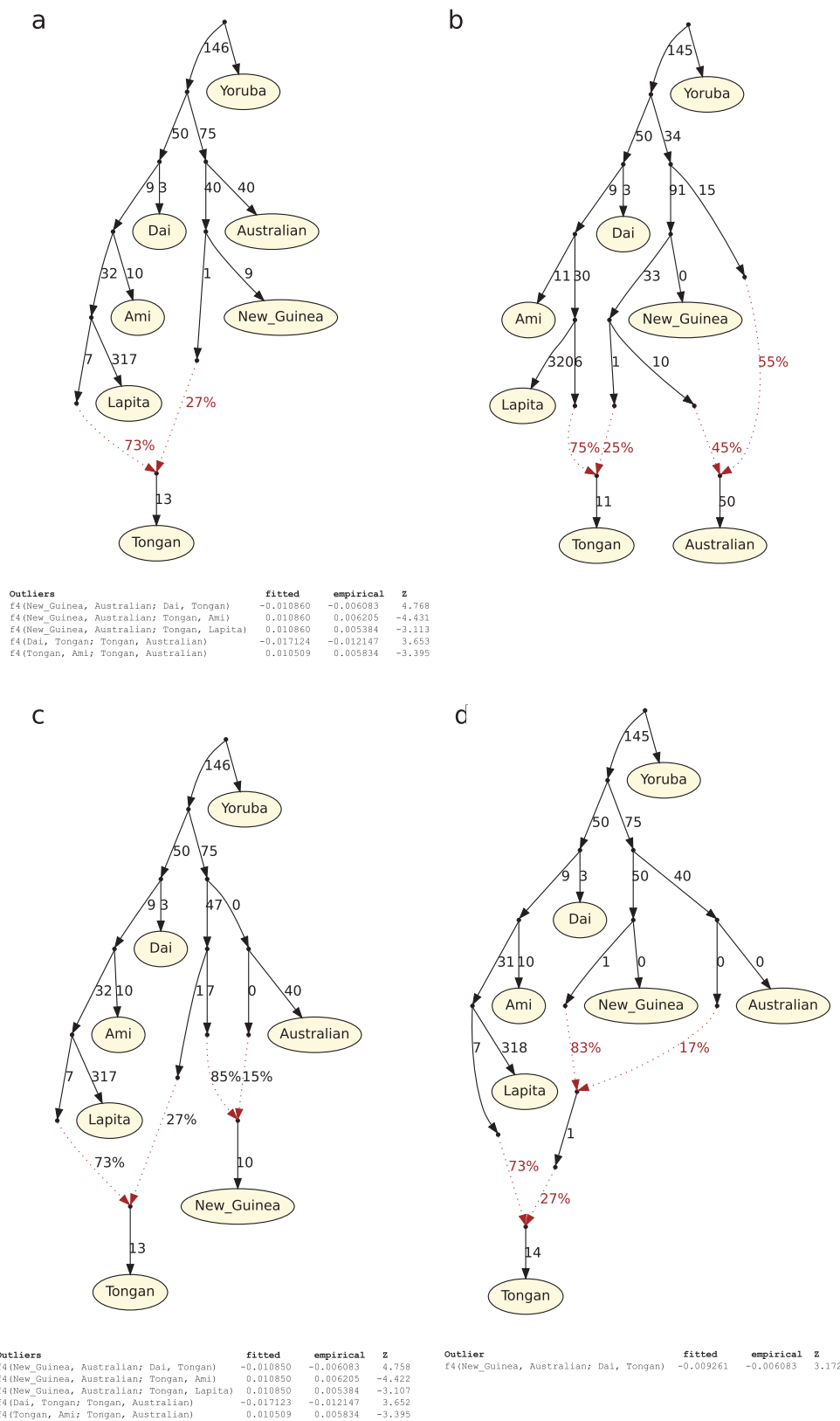
b) graph with 1 migration edge



Extended Data Figure 4 | Admixture graph inferred using *Treemix*.

a. A simple tree-like model without admixture fits the data poorly, as can be seen from the matrix of residuals between empirical and modelled allele frequency covariance on the right. **b.** The optimal placement of a single 25% admixture event is from the lineage related to New Guinean

Highlanders into the lineage leading to Tongans. Tongans derive the other portion of their ancestry from the lineage leading to the two ancient groups of individuals. This graph has no significant deviations between empirical and modelled allele frequency covariances.

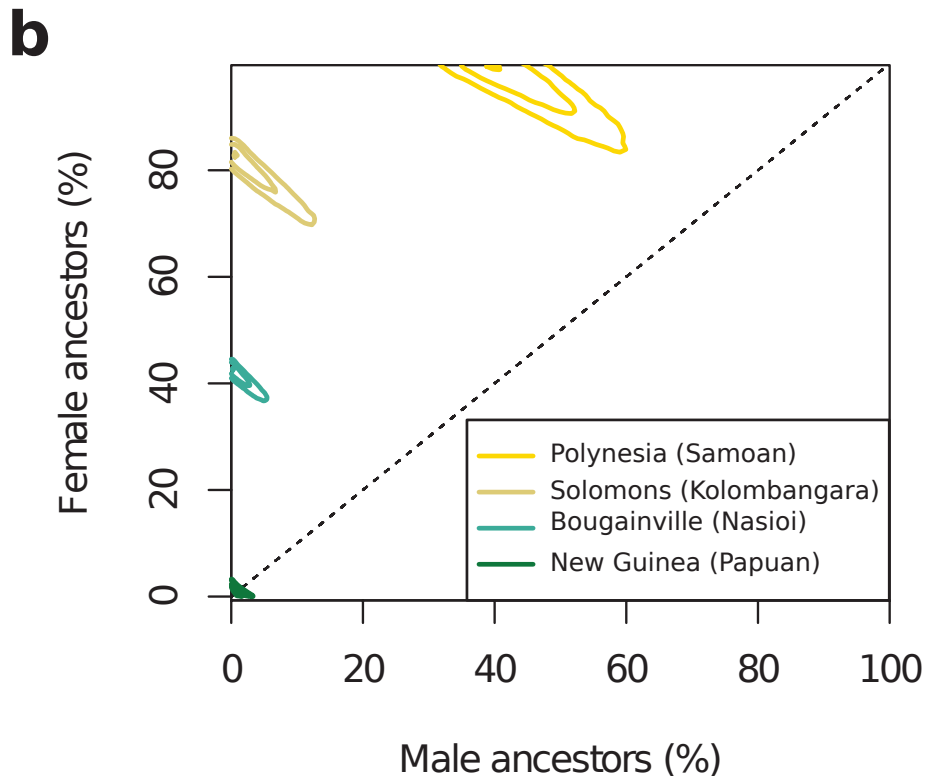
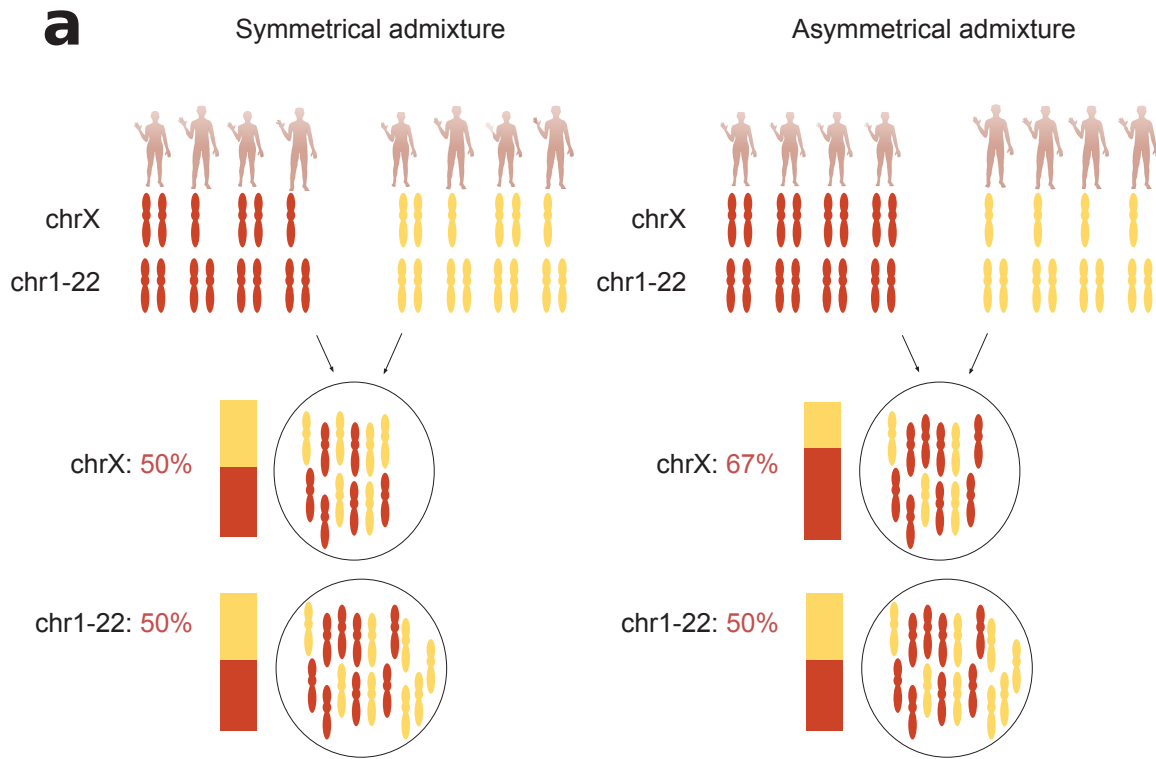


Extended Data Figure 5 | Admixture graphs modelling the population history of Australians. Outlier f_4 -statistics are shown ($|Z| > 3$).

a, A model with a single admixture edge positing that Australians are an outgroup to the Papuan ancestry in Tongans does not fit the data (5 outlier statistics). **b**, An alternative model with 2 admixture edges in which the Papuan ancestry in Tongans also contributed to Australians fits the data (no outliers). **c**, A model with 2 admixture edges in which New Guinean

Highlanders are admixed from an Australian source after the divergence of the Papuan source in Tongans does not fit the data (5 outliers).

d, A model with 2 admixture edges in which the Papuan ancestry in Tongans is intermediate between the New Guinean Highlander lineage and the Australian lineage. Branch lengths are in units of $F_{ST} \times 1,000$. Lapita in this figure refers only to Vanuatu, which is the only group for which we have multiple individuals.



Extended Data Figure 6 | First Remote Oceanian ancestry today comes primarily from females. **a**, Illustration of the rationale for using the X chromosome to study asymmetrical admixture between males and females. The example on the left illustrates admixture with equal proportion of males and females in both the red and the yellow ancestral population. The example on the right illustrates an extreme case of asymmetrical admixture where the red ancestral population only contributes females and the yellow ancestral population only contributes

males to the admixed generation, demonstrating the disproportional contribution of X chromosomes by females to the admixed population. **b**, Female and male ancestral contributions based on an admixture model fitted to estimated ancestry proportions on the autosomes and X chromosome. We show the 95%, 70%, and 5% highest posterior intervals for four selected populations from Polynesia (Samoans), the Solomon Islands (Kolombangara), Bougainville (Nasioi), and mainland New Guinea (Papuan).

Extended Data Table 1 | In-solution DNA enrichment and sequencing of ancient individuals

Location	Sample information					Coverage on chromosomes 1-22			Sex determination		
	ID1	ID2	Bone for aDNA	Bone for dating	¹⁴ C Date: Calibrated 95.4% Conf. Int. (Uncalibrated date, Lab number)	Mean depth	All SNPs	SNPs over-lapping array	Y SNPs	X SNPs	Sex
Vanuatu	I1368	B30A	Petrous	Skull	2990-2740 BP (2983±32 BP, Wk-22657)	0.26	139,461	74,631	321	18,231	F
Vanuatu	I1369	B10B	Petrous	Petrous	3000-2750 BP (3045±30 BP, Poz-81126)	0.14	199,500	107,523	341	24,255	F
Vanuatu	I1370	B17	Petrous	Skull	3110-2780 BP (3083±26 BP, Wk-21026)	0.21	167,311	90,402	231	19,303	F
Tonga	CP30	SK10	Petrous	Fibula	2680-2340 BP (2594±20 BP, Wk-41883)	0.16	231,994	125,908	75	25,943	F

All dates are calibrated using OxCal v4.2.4⁵⁸ with a mixture of the Marine13 and Intcal13 curves⁵⁹ as determined by linear interpolation between dietary terrestrial and marine $\delta^{13}\text{C}$ isotopic endpoints (-21‰ / -12‰) with an uncertainty of $\pm 10\%$ on the per cent marine carbon result following previous recommendations⁶⁰. Two of the dates have been previously reported (for I1368/B30A and I1370/B17)⁶¹, and in this study we add two new dates: for I1369/B10B from Vanuatu (on the same petrous bone used for ancient DNA analysis) and on CP30/SK10 from Tonga (on a fibula). Measured ^{13}C and ^{15}N values for I1369/B10B are -14.5‰ and 13.7‰ respectively, and for SK10 -16.44‰ and 10.48‰ . As justified in ref. 61, we also applied a location-specific reservoir correction (ΔR) of 40 ± 44 ^{14}C years to the marine curve to adjust for regional oceanic variation in ^{14}C around Vanuatu, and 11 ± 83 ^{14}C years for Tongatapu⁶².

Extended Data Table 2 | 356 individuals newly genotyped on the Human Origins Array

Population	N	Country of origin	Land mass	Language	Lat.	Long.	Co-authors for samples	Protocol Numbers for informed consent	Data distribution
Ata	8	Papua New Guinea	New Britain	Papuan	-5.7	150.9	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Baining_Malasait	5	Papua New Guinea	New Britain	Papuan	-4.47	151.9	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Baining_Marabu	10	Papua New Guinea	New Britain	Papuan	-4.63	152.3	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Bajo	10	Indonesia	Sulawesi	Austronesian	-3.97	122.59	M.P.C., P.K., F.-X.R.	4.13.2013 approval AMIS-UPS Ethics Committee	Signed Letter
Buka	8	Papua New Guinea	Bougainville	Austronesian	-5.42	154.67	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Burmese	10	Myanmar	Asia	Sino-Tibetan	16.41	95.89	T.K., J.T.S.W.	2010/820/A and HBREC.2011.01	Fully public
Dusun	10	Brunei	Borneo	Austronesian	4.71	114.67	T.K., S.A.	MHREC/EDU/2012/3(1) and HBREC.2011.01	Fully public
Ilocano	2	Philippines	Luzon	Austronesian	14.6	120.98	T.K., J.T.S.W.	2010/820/A and HBREC.2011.01	Fully public
Kankanaey	10	Philippines	Luzon	Austronesian	17.07	121.03	T.K., J.T.S.W.	2010/820/A and HBREC.2011.01	Fully public
Kol_New_Britain	2	Papua New Guinea	New Britain	Papuan	-5.38	151.63	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Kove	18	Papua New Guinea	New Britain	Austronesian	-5.47	148.95	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Kuot_Kabil	9	Papua New Guinea	New Ireland	Papuan	-3.07	151.7	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Kuot_Lamalaua	8	Papua New Guinea	New Ireland	Papuan	-3	151.5	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Lavongai	15	Papua New Guinea	New	Austronesian	-2.57	150.43	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Lebbo	8	Indonesia	Borneo	Austronesian	1.66	117.16	M.P.C., P.K., F.-X.R.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Madak	10	Papua New Guinea	New Ireland	Austronesian	-3.1	151.7	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Malay	9	Singapore	Asia	Austronesian	1.35	103.82	T.K., J.T.S.W.	2010/820/A and HBREC.2011.01	Fully public
Mamusi	20	Papua New Guinea	New Britain	Austronesian	-6	151	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Mamusi_Paleabu	6	Papua New Guinea	New Britain	Austronesian	-5.95	150.9	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Mangseng	6	Papua New Guinea	New Britain	Austronesian	-5.93	150.7	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Manus	2	Papua New Guinea	Manus	Austronesian	-2.08	147	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Melamela	10	Papua New Guinea	New Britain	Austronesian	-5	151.25	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Mengen	10	Papua New Guinea	New Britain	Austronesian	-5.1	151.4	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Murut	10	Brunei	Borneo	Austronesian	4.62	115.14	T.K., S.A.	MHREC/EDU/2012/3(1) and HBREC.2011.01	Fully public
Mussau	10	Papua New Guinea	St. Matthias	Austronesian	-1.58	149.73	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Nakanai_Bileki	10	Papua New Guinea	New Britain	Austronesian	-5.75	150.8	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Nakanai_Loso	7	Papua New Guinea	New Britain	Austronesian	-5.48	150.8	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Nailik	9	Papua New Guinea	New Ireland	Austronesian	-2.98	151.52	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Notsi	9	Papua New Guinea	New Ireland	Austronesian	-3.05	151.65	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Saposa	10	Papua New Guinea	Bougainville	Austronesian	-5.58	154.67	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Southwest_Bougainville	2	Papua New Guinea	Bougainville	Papuan	-6.6	155.5	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Sulka	20	Papua New Guinea	New Britain	Papuan	-4.5	152.3	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Tagalog	5	Philippines	Luzon	Austronesian	14.6	120.98	T.K., J.T.S.W.	2010/820/A and HBREC.2011.01	Fully public
Teop	10	Papua New Guinea	Bougainville	Austronesian	-5.85	155.18	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Tigak	10	Papua New Guinea	New Ireland	Austronesian	-2.57	150.83	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Tolai	24	Papua New Guinea	New Britain	Papuan	-4.31	152.14	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Vietnamese	10	Vietnam	Asia	Austroasiatic	10.82	106.64	T.K., J.T.S.W.	2010/820/A and HBREC.2011.01	Fully public
Visayan	4	Philippines	Mindanao	Austronesian	9.76	125.51	T.K., J.T.S.W.	2010/820/A and HBREC.2011.01	Fully public

Extended Data Table 3 | f -statistics for populations on the Oceanian cline

	$F_{ST}(\text{Lapita_Vanuatu, Test})$		$f_3(\text{Lapita_Vanuatu, Australian; Test})$	
Test population	Estimate	SE	Estimate	Z-score
Baining_Malasait	0.263	0.005	0.143	38.4
Baining_Marabu	0.249	0.004	0.115	40.1
Papuan	0.225	0.004	0.066	33.3
Kol_New_Britain	0.216	0.006	0.086	16.7
Mamusi	0.204	0.004	0.059	26.5
Ata	0.197	0.004	0.050	21.6
Nakanai_Loso	0.194	0.004	0.055	21.9
Mamusi_Paleabu	0.187	0.004	0.040	17.8
Santa_Cruz	0.185	0.004	0.021	10.7
Nasioi	0.178	0.004	0.028	14.7
Bougainville_South	0.176	0.006	0.034	8.1
Sulka	0.174	0.004	0.022	13.1
Mengen	0.168	0.004	0.017	9.1
Tolai	0.163	0.004	0.009	5.6
Kuot_Kabil	0.162	0.004	0.014	7.7
Lavongai	0.160	0.004	0.009	5.2
Kuot_Lamalaua	0.160	0.004	0.008	4.3
Nakanai_Bileki	0.153	0.004	0.013	7.0
Melamela	0.152	0.004	0.010	5.1
Madak	0.151	0.004	0.003	1.8
Papuan_Gulf	0.151	0.005	0.002	0.7
Kove	0.150	0.004	0.013	7.4
Mangseng	0.145	0.004	0.003	1.7
Nailik	0.145	0.004	0.000	-0.1
Teop	0.144	0.004	0.005	2.8
Notsi	0.144	0.004	-0.002	-1.4
Manus	0.141	0.006	-0.005	-1.4
Tigak	0.141	0.004	0.002	1.2
Mussau	0.132	0.004	-0.005	-3.3
Choiseul	0.123	0.004	-0.007	-3.7
Saposa	0.122	0.004	-0.011	-6.5
Buka	0.118	0.004	-0.016	-9.9
Vella_Lavella	0.112	0.004	-0.016	-8.6
Ranongga	0.110	0.004	-0.015	-8.2
Savo	0.109	0.004	-0.020	-12.1
Russell	0.108	0.005	-0.017	-6.8
Kolombangara	0.108	0.004	-0.015	-8.0
RenBel	0.106	0.004	0.035	14.0
Gela	0.103	0.004	-0.024	-14.3
Makira	0.101	0.004	-0.024	-16.1
Malaita	0.096	0.004	-0.025	-15.5
Papuan_Central	0.094	0.004	-0.031	-18.2
Bajo	0.082	0.004	0.022	12.8
Isabel	0.079	0.004	-0.024	-15.9
Tikopia	0.077	0.004	-0.003	-1.3
Ontong_Java	0.069	0.004	-0.018	-10.4
Tongan	0.053	0.004	-0.018	-9.9

Standard error (SE) is shown for F_{ST} between each *Test* population and the pool of ancient Vanuatu individuals. The Z score is given for the statistic $f_3(\text{Lapita_Vanuatu, Australian; Test})$, where $Z < 3$ provides significant evidence that the *Test* is admixed between sources related to the ancient Vanuatu and Australians.

Extended Data Table 4 | Ancestry estimates for populations on the Oceanian cline

Test	First Remote Oceanian ancestry estimate							Anc. contrib. (method of moments)				Two source model
	Auto.	SE	chrX	SE	Diff.	SE	Z	Male	SE	Female	SE	P-value
Tongan	69.8%	1.0%	104.1%	8.9%	-34.3%	9.0%	-3.8	-33.1%	27.0%	172.7%	22.0%	0.22
Rennel & Bellona	68.9%	1.2%	92.1%	7.3%	-23.2%	7.4%	-3.1	-0.7%	22.4%	138.5%	32.8%	0.89
Tikopia	65.8%	1.0%	93.9%	10.9%	-28.1%	10.9%	-2.6	-18.5%	32.9%	150.1%	39.9%	0.58
Ontong_Java	61.9%	0.9%	78.4%	13.3%	-16.5%	13.3%	-1.2	12.4%	40.1%	111.4%	12.7%	0.83
Santa Isabel	52.9%	0.9%	60.5%	4.2%	-7.6%	4.3%	-1.8	30.1%	13.1%	75.7%	14.2%	0.87
Papuan_Central	42.5%	0.9%	65.6%	4.7%	-23.1%	4.8%	-4.8	-26.8%	14.6%	111.8%	47.7%	0.14
Malaita	39.6%	0.9%	66.1%	15.9%	-26.5%	15.9%	-1.7	-39.9%	47.8%	119.1%	20.8%	0.25
Kolombangara	39.1%	1.0%	54.7%	6.9%	-15.6%	7.0%	-2.2	-7.7%	21.1%	85.9%	16.3%	0.87
Nggela	37.8%	0.9%	54.6%	5.4%	-16.8%	5.5%	-3.1	-12.6%	16.6%	88.2%	82.5%	0.59
Ranongga	37.6%	1.0%	48.8%	27.5%	-11.2%	27.5%	-0.4	4.0%	82.6%	71.2%	29.8%	0.15
Russell	36.2%	1.2%	45.8%	9.9%	-9.6%	10.0%	-1.0	7.4%	30.1%	65.0%	26.2%	0.19
Vella_Lavella	35.2%	1.0%	53.2%	8.7%	-18.0%	8.8%	-2.1	-18.8%	26.4%	89.2%	19.9%	0.10
Savo	34.6%	0.9%	56.4%	6.6%	-21.8%	6.7%	-3.3	-30.8%	20.1%	100.0%	38.4%	0.18
Makira	34.6%	0.9%	31.8%	12.8%	2.8%	12.8%	0.2	43.0%	38.6%	26.2%	13.4%	0.08
Choiseul	32.4%	1.0%	35.5%	4.4%	-3.1%	4.5%	-0.7	23.1%	13.8%	41.7%	9.8%	0.20
Buka	31.3%	0.9%	46.0%	3.2%	-14.7%	3.3%	-4.4	-12.8%	10.3%	75.4%	16.9%	0.52
Saposa	31.1%	0.9%	41.8%	5.6%	-10.7%	5.7%	-1.9	-1.0%	17.2%	63.2%	14.2%	0.25
Mussau	29.2%	0.9%	39.4%	4.7%	-10.2%	4.8%	-2.1	-1.4%	14.6%	59.8%	22.0%	0.54
Teop	26.5%	0.9%	67.8%	7.3%	-41.3%	7.4%	-5.6	-97.4%	22.2%	150.4%	11.5%	0.86
Kove	26.4%	0.9%	29.8%	3.8%	-3.4%	3.9%	-0.9	16.2%	12.0%	36.6%	20.2%	0.01
Tigak	26.2%	0.9%	43.9%	6.7%	-17.7%	6.8%	-2.6	-26.9%	20.4%	79.3%	21.4%	0.87
Melamela	25.3%	0.9%	50.1%	7.1%	-24.8%	7.2%	-3.5	-49.1%	21.6%	99.7%	18.8%	0.94
Manus	24.9%	1.2%	14.2%	6.2%	10.7%	6.3%	1.7	57.0%	19.2%	-7.2%	42.9%	0.87
Nakanai_Bileki	24.5%	0.9%	38.4%	14.3%	-13.9%	14.3%	-1.0	-17.2%	43.1%	66.2%	11.2%	0.28
Mangseng	23.6%	0.9%	42.2%	3.7%	-18.6%	3.8%	-4.9	-32.2%	11.7%	79.4%	19.9%	0.34
Papuan_Gulf	22.6%	1.1%	49.3%	6.6%	-26.7%	6.7%	-4.0	-57.5%	20.3%	102.7%	10.9%	0.07
Notsi	20.7%	0.8%	33.0%	3.6%	-12.3%	3.7%	-3.3	-16.2%	11.3%	57.6%	64.8%	0.13
Nailik	20.4%	0.8%	44.8%	21.6%	-24.4%	21.6%	-1.1	-52.8%	64.9%	93.6%	22.9%	0.08
Madak	18.7%	0.8%	43.1%	7.6%	-24.4%	7.6%	-3.2	-54.5%	23.0%	91.9%	18.7%	0.11
Kuot_Kabil	18.6%	0.9%	55.4%	6.2%	-36.8%	6.3%	-5.9	-91.8%	18.9%	129.0%	26.5%	0.04
Bougainville_Sout	18.0%	1.4%	24.1%	8.8%	-6.1%	8.9%	-0.7	-0.3%	27.0%	36.3%	11.8%	0.26
Lavongai	17.6%	0.8%	36.4%	3.9%	-18.8%	4.0%	-4.7	-38.8%	12.1%	74.0%	16.0%	0.07
Nasioi	17.4%	1.0%	29.9%	5.3%	-12.5%	5.4%	-2.3	-20.1%	16.4%	54.9%	11.5%	0.40
Kuot_Lamalaua	16.2%	0.8%	13.8%	3.8%	2.4%	3.9%	0.6	23.4%	11.8%	9.0%	19.6%	0.07
Nakanai_Loso	15.4%	1.1%	24.3%	6.5%	-8.9%	6.6%	-1.4	-11.3%	20.0%	42.1%	11.2%	0.37
Mengen	15.2%	0.8%	35.3%	3.7%	-20.1%	3.8%	-5.3	-45.1%	11.6%	75.5%	13.0%	0.41
Tolai	14.3%	0.8%	22.5%	4.3%	-8.2%	4.4%	-1.9	-10.3%	13.3%	38.9%	9.4%	0.01
Sulka	14.1%	0.8%	47.0%	3.1%	-32.9%	3.2%	-	-84.6%	9.8%	112.8%	16.3%	0.48
Mamusi_Paleabu	13.7%	1.0%	30.9%	5.4%	-17.2%	5.5%	-3.1	-37.9%	16.7%	65.3%	13.9%	0.21
Mamusi	13.7%	0.9%	22.7%	4.6%	-9.0%	4.7%	-1.9	-13.3%	14.3%	40.7%	27.1%	0.35
Ata	13.2%	1.0%	26.4%	9.0%	-13.2%	9.1%	-1.5	-26.4%	27.3%	52.8%	15.4%	0.24
Santa_Cruz	9.6%	0.9%	27.1%	5.1%	-17.5%	5.2%	-3.4	-42.9%	15.7%	62.1%	23.3%	0.17
Kol_New_Britain	8.5%	1.4%	9.1%	7.7%	-0.6%	7.8%	-0.1	6.7%	23.8%	10.3%	30.1%	0.66
Baining_Marabu	2.6%	1.0%	25.3%	10.0%	-22.7%	10.0%	-2.3	-65.5%	30.3%	70.7%	32.5%	0.24
Baining_Malasait	1.2%	1.2%	34.9%	10.8%	-33.7%	10.9%	-3.1	-99.9%	32.8%	102.3%	8.2%	0.12
Papuan	0.0%	0.5%	2.4%	2.7%	-2.4%	2.7%	-0.9	-7.2%	8.3%	7.2%	0.0%	0.58

Auto., estimate on the autosomes (chromosomes 1–22). Diff., difference between the autosome and X chromosome estimates.

Supplementary Note 1: The Teouma site

The Teouma Lapita site is located on the south coast of Efate Island, Vanuatu, and is currently 800 meters from Teouma Bay. Tectonic uplift, volcanic ashfalls prior to and during the period of human utilisation of the site, and alluvial deposition from an adjacent stream have altered its immediately coastal location during Lapita times. Excavations by a joint ANU-Vanuatu National Museum team took place between 2004 and 2010¹⁻³, during which 68 burial features, including remains of just over 100 individuals were found, concentrated in a band running northeast to southwest, parallel to the former beach and some 10-15m wide. Teouma has numerous indicators of being an initial colonisation site for Efate, including extinct faunal remains, early ceramic forms and decoration, and New Britain obsidian. This combined with the extensive and previously very rare Lapita skeletal remains underline its significance to those investigating colonization in this region.

There are two key zones of the site; the Cemetery Zone (CZ) and the Midden Zone (MZ) located to the east of the Lapita cemetery, that both comprise three layers. The Lapita midden contained a number of domesticated species (i.e., *Sus scrofa*, *Gallus gallus* and the commensal *Rattus exulans*), as well as a large number of bones from indigenous birds and animals, many now extinct, including land tortoise (*Meiolania damelipi*), crocodile (*Mekosuchus kalpokasi*) and a range of bird species^{2,4,5}. Layer 3 of the Midden Zone (MZ3) deposit is considered to be contemporary with Layer 3 of the cemetery (CZ3) and both are emplaced onto a pre-human tephra deposit⁶. A later midden deposit (Layer 2 Cemetery Zone [CZ2]), up to 50 cm thick, covered the cemetery and extended downslope over the former beach and within alluvial deposits from the adjacent stream, representing a shift to purely habitation use at the main site. This is associated with Post-Lapita Arapus and Early Erueti ceramics currently dated to c. 2800-2500 BP². Continued tectonic uplift and alluvial deposition led to abandonment of the settlement as immediate access to the sea became problematic because of shoreline progradation. Layer 1 from both zones has not been dated but represents natural post-occupation accumulation, rich in tephra from the 2300 BP eruption of Nguna Island to the north. In Area 7C some 70m south of the main Teouma site further burials were located, dating to within the Erueti phase at about 2400 BP⁷.

An extensive dating program has been undertaken at the site on a range of materials including charcoal and marine shell, as well as human, terrestrial native fauna and domesticate bones^{8,9}. That analysis places the start date of the Lapita deposits at 2920-2870 cal BP with a possible use as early as 3000 cal BP, and an end date of 2870-2750 cal BP. The results are in keeping with evaluation of the burials (36 direct dating on human cranial and infracranial elements⁸), which places earliest use of the cemetery at c. 2970 cal BP with regular use underway by c. 2940-2880 cal BP and the last internment occurring c. 2770-2710 cal BP. Continued use of the area is evident by the later burials nearby dating to the end of the Erueti phase and by components of the

CZ2 midden deposits of post-Lapita age.

The exceptionally well-preserved and extensive Lapita skeletal remains have provided an opportunity to start to establish a population profile, gaining insights into the health, morphology and diet of a group of Lapita settlers who were among the first to have ventured into Remote Oceania. The 68 burials at Teouma display a wide range of mortuary practices and indicate that ritual was a multi-faceted and on-going process, rather than a one-off event. The burials were generally placed in shallow graves dug into the underlying tephra deposits amongst gaps in the uplifted reef and coral boulders on the upper part of the beach. Evidence indicates manipulation of the corpse prior to burial or at least during the early stages of decomposition and of the skeletal remains after body decomposition. All the adult burials had their skulls, and often many other bones, removed during this process and only a few of these bones were secondarily re-deposited at the site^{1,3,10,11}. Nine skulls were found out of a total of 52 infracranial skeletons in articulation. Although disconnected from infracranial skeletons, the cranial elements appears to have belonged to the same group as shown by similarities in isotopic values measured in bone collagen and dental enamel. Seven of them, belonging to female and male adults, were preserved well enough for morphological examination. Five of them were useable for metrical analysis that demonstrated that each displays a lack of Australian or Papuan affinities and an alignment with ancient and more recent Asiatic populations¹². Ancient DNA was successfully obtained from three skulls from striking mortuary contexts: a jar burial containing a single skull (B17), an alignment of three skulls lying on the chest of a skeleton without a skull (B10B), and a triangular bone arrangement, with skulls at each vertex, lying on the legs of a skeleton without a skull (B30A).

References

- 1 - Bedford, S., Spriggs, M. & Regenvanu, R. 2006. The Teouma Lapita Site and the Early Human Settlement of the Pacific Islands. *Antiquity* 80:812-828.
- 2 - Bedford, S., Spriggs, M., Buckley, H., Valentin, F. & Regenvanu, R. 2009. The Teouma Lapita Site, South Efate, Vanuatu: A Summary of Three Field Seasons (2004-2006). In P. Sheppard, T. Thomas and G. Summerhayes (eds) *Lapita: Ancestors and Descendants. New Zealand Archaeological Association Monograph* 28, pp. 215-234.
- 3 - Bedford, S., Spriggs, M., Bukley, H., Valentin, F., Regenvanu, R. & Abong, M. 2010. A cemetery of first settlement: Teouma, South Efate, Vanuatu/ Un cimetiere de premier peuplement: le site de Teouma, sud d'Efate, Vanuatu. In C. Sand and S. Bedford (eds) *Lapita: Oceanic Ancestors/Lapita: Ancetres Oceaniens*, pp. 140-161. Paris: Musee de Quai Branly/Somogy.
- 4 - White, A., Worthy, T., Hawkins, S., Bedford, S. & Spriggs, M. 2010. Megafaunal meiolaniid horned turtles survived until early human settlement in Vanuatu, Southwest Pacific. *Proceedings of the National Academy of Sciences* 107:15512-15516.

- 5 - Worthy, T., Hawkins, S., Bedford, S. & Spriggs, M. 2015. The avifauna from the Teouma Lapita site, Efate Island, Vanuatu, including a new genus and species of megapode. *Pacific Science* 69(2):205-254.
- 6 - Spriggs, M. & Bedford, S. 2013. Is there an incised Lapita phase after dentate-stamped pottery ends? Data from Teouma, Efate Island, Vanuatu. In G. Summerhayes and H. Buckley (eds) *Pacific Archaeology: Documenting the Past 50,000 Years*, pp. 148-156. University of Otago Studies in Archaeology 25. Dunedin: Department of Anthropology and Archaeology, University of Otago.
- 7 - Valentin F., Herrscher E., Bedford S., Spriggs M., 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. *Journal of Island and Coastal Archaeology* 9:3, 381-399.
- 8 - Petchey F., Spriggs M., Bedford S., Valentin F., Buckley H., 2014. Radiocarbon dating of burials from the Teouma Lapita cemetery, Efate, Vanuatu. *Journal of Archaeological Science*, 50: 227-242.
- 9 - Petchey F., Spriggs M., Bedford S., Valentin F., 2015. The chronology of occupation at Teouma, Vanuatu: Use of a modified chronometric hygiene protocol and Bayesian modeling to evaluate midden remains. *Journal of Archaeological Science: Reports* 4 (2015), p. 95-105.
- 10 - Valentin, F., Bedford, S., Buckley H., Spriggs, M., 2010. Lapita Burial Practices: Evidence for Complex Body and Bone Treatment at the Teouma Cemetery, Vanuatu, Southwest Pacific. *Journal of Island and Coastal Archaeology*, 5 (2): 212-235.
- 11 - Valentin F., Choi J.-I., Lin H., Bedford S., Spriggs M. 2015. Three thousand year old jar burials at the Teouma cemetery (Vanuatu): a Southeast Asian-Lapita connection?, in C. Sand, S. Chiu and N. Hogg (eds) *The Lapita Cultural Complex in Time and Space: Expansion Routes, Chronologies and Typologies*, Nouméa and Taipei, IANCP and Center for Archaeological Studies, Research Center for Humanities and Social Sciences, Academia Sinica. (Archaeologia Pasifika 4), p. 81-101.
- 12 - Valentin, F., Déroit, F., Spriggs, M. J. T. & Bedford, S. Early Lapita skeletons from Vanuatu show Polynesian craniofacial shape: Implications for Remote Oceanic settlement and Lapita origins. *Proceedings of the National Academy of Sciences*, 113 (2): 292-297 (2016).

Supplementary Note 2: The Talasiu site

The Talasiu site (TO-Mu-2) is located on the palaeoreef-limestone shoreline of the Fanga 'Uta Lagoon around 10 kilometers from Nuku'alofa and immediately north of Lapaha Village. The site consists of a dense shell midden deposit ~90 cm thick covering some 450 square meters and includes fire features and burials. The palaeoshoreline at Talasiu is fronted by reclaimed land that was deposited in the lagoon during the expansion of the Tu'i Tonga chiefdom 800-200 cal. BP. During the Lapita period in Tonga (2900-2650 cal. BP) (95.4% prob.)¹, the Talasiu site would have been an attractive location for human settlement as it overlooked an embayment holding large numbers of sessile and gregarious shellfish close to a fresh water solution channel, with gardening soils immediately inland. The site was first investigated in 1957 by Jack Golson (ANU) and in the 1980s by Dirk Spennemann (ANU) who surface collected pottery (including sherds decorated with dentate-stamping) lithics and human bone^{2,3}. Site monitoring after road grading in 2008 identified a concentration of burned and partially burned human bone which was excavated and found to be a cremation mortuary context containing the incomplete remains of four individuals⁴. In 2011, new burials were found eroding from the road cut and as the area was about to be intensively gardened a rescue archaeology project to recover human remains was directed by Frederique Valentin (CNRS) and Geoffrey Clark (ANU) in 2013-2014 with the support of the Ministry of Internal Affairs (Kingdom of Tonga) and funded by the French Government (MAEDI, Commission des fouilles à l'étranger).

The site stratigraphy consists of a series of six layers distinguished by varying quantities of shellfish, ceramics, charcoal and bone resting on sterile basal clay (Layer 6). The 95% probability range of radiocarbon determinations obtained on coconut endocarp (n=5), unidentified charcoal (n=2), worked shell grave goods (n=3) and human bone sample on articulated burials (n=6) fall between 2870 and 2340 cal. BP. Calibrated charcoal and bone ¹⁴C ages between 2600 and 2300 BP are influenced by curve flattening resulting in multiple curve intercepts and a substantial widening in the calibrated age range of determinations. However, a high-resolution chronology based on U-Th dating of coral files and AMS determinations on short-lived material demonstrates the Lapita period on Tongatapu spanned 2900-2650 cal. BP¹ and it is highly likely that the midden and burials at Talasiu date to ~2700-2600 cal. BP² and are of late Lapita/immediately post-Lapita age.

Analysis of the midden using archaeozoological and microbotanical techniques provided evidence for a broad spectrum mixed economy in Tonga². Pottery was predominantly plain with dentate-stamped vessels bearing simple open designs typical of late Lapita ceramics throughout the deposit. Shell artefacts included short and long shell units, broad *Conus* spp. rings, and narrow rings made in *Conus* spp. and *Tridacna* spp. that are characteristic Lapita ornaments. Lithics included adzes, flakes, grind stones and oven stones. Thin-section and compositional study of the lithics with pXRF and LA-ICP-MS identified obsidian flakes from Tafahi in

northern Tonga⁵, adzes made in volcanic material chemically sourced to Central Tonga, eastern Fiji and Samoa, and manuports and grindstones/oven stones from nearby 'Eua Island and Central Tonga^{2,5,6}.

The 19 late-Lapita/immediately post-Lapita burial contexts at Talasiu were placed at the base of the midden or interstratified within the midden. Intact lenses of ash and marine shell overlying several burials conclusively demonstrate that interments were made during midden formation. The burials were single or multiple interments containing the skeletal remains of one to six individuals (male and female adults and children of various ages). Field evidence indicates a wide range of mortuary practices involving primary burial in various positions, secondary deposits of burnt⁴ or unburnt bone, including placement of isolated skulls, and post-decomposition removal of skulls and long bones. This diversified mortuary pattern recalls that of late Lapita and immediately post-Lapita burials from sites in Vanuatu⁷, while the body treatment of some individuals is the same as at the Lapita site of Teouma⁸⁻⁹. These 19 burial contexts have yielded the skeletal remains of 62 individuals, providing an opportunity to obtain information on health, diet and morphology of a group of late Lapita/immediately post-Lapita people from Tongatapu, using macroscopic and microscopic observations and biogeochemical and palaeomolecular data.

Preliminary metrical analysis performed on one skull (BG3) shows that the individual lacks Australian or Papuan affinities and instead has affinities to mainland Asian populations, as do the Teouma Lapita skulls¹⁰. Ancient DNA was successfully obtained from the right petrous bone of burial SK10, a single primary interment of an adult female with a marine-adjusted calibrated age of 2680-2340 cal. BP (95.4% confidence interval; see Extended Data Table 1). The SK10 ¹⁴C calibration is influenced by the radiocarbon plateau around 2600-2300 BP, which results in a wide spread in the calibrated age range. It is important to note that this burial was sealed by a layer of intact midden which is inferred from essentially identical radiocarbon results on carbonized coconut shell from the upper/middle and basal layers to have been deposited around 2700-2600 cal. BP².

1. Burley, D.V., Edinborough, K., Weisler, M., Zhao, J.-x., 2015. Bayesian modeling and chronological precision for Polynesian settlement of Tonga. PLoS One 10 (3): e0120795. <http://dx.doi.org/10.1371/journal.pone.0120795>.
2. Clark, G., Grono, E., Ussher, E. and Reepmeyer, C. 2015. Early settlement and subsistence on Tongatapu, Kingdom of Tonga: Insights from a 2700-2650 cal BP midden deposit. *Journal of Archaeological Science: Reports* DOI: 10.1016/j.jasrep.2015.08.005.
3. Spennemann, D.H.R. 1986. *Archaeological fieldwork in Tonga 1985-1986. Preliminary report on the archaeological activities during the 1985/86 field season of the Tongan Dark Ages Research Programme*. Department of Prehistory, The Australian National University, Canberra.

4. Valentin, F. and Clark, G. 2013. Early Polynesian mortuary behaviour at the Talasiu site, Kingdom of Tonga. *Journal of Pacific Archaeology* 4(1): 1-14.
5. Reepmeyer, C., Clark, G. and Sheppard, P. 2012. Obsidian source use in Tongan prehistory: New results and implications. *Journal of Island and Coastal Archaeology* 7(2): 255-271.
6. Clark, G., Reepmeyer, C., Melekiola, N., Woodhead, J., Dickinson, W.R. and Martinsson-Wallin, H. 2014. Stone tools from the ancient Tongan state reveal prehistoric interaction centres in the Central Pacific. *Proceedings of the National Academy of Sciences U.S.A.* 111(29): 10491-10496.
7. Bedford, S., Buckley, H., Valentin, F., Tayles, N., Longga, F. 2011. Lapita burials, a new Lapita cemetery and post-Lapita burials from Malakula, northern Vanuatu, Southwest Pacific. *Journal of Pacific Archaeology*, 2(2): 26-48.
8. Valentin, F., Bedford, S., Buckley H., Spriggs, M., 2010. Lapita Burial Practices: Evidence for Complex Body and Bone Treatment at the Teouma Cemetery, Vanuatu, Southwest Pacific. *Journal of Island and Coastal Archaeology*, 5 (2): 212-235.
9. Valentin F., Choi J.-I., Lin H., Bedford S., Spriggs M. 2015. Three thousand year old jar burials at the Teouma cemetery (Vanuatu): a Southeast Asian-Lapita connection?, in C. Sand, S. Chiu and N. Hogg (eds) *The Lapita Cultural Complex in Time and Space: Expansion Routes, Chronologies and Typologies*, Nouméa and Taipei, IANCP and Center for Archaeological Studies, Research Center for Humanities and Social Sciences, Academia Sinica. (Archaeologia Pasifika 4), p. 81-101.
10. Valentin, F., Déroît, F., Spriggs, M. J. T. & Bedford, S. 2016. Early Lapita skeletons from Vanuatu show Polynesian craniofacial shape: Implications for Remote Oceanic settlement and Lapita origins. *Proceedings of the National Academy of Sciences*, 113(2): 292-297.

Supplementary Note 3: Population history graph models

We used admixture graph models to model the population history of Oceania and Polynesia. For our main analyses, we used *qpGraph*^{1,2} to test systematically a variety of historical models, and to assess their fit to the data, but see also our complementary analysis with *TreeMix* documented in the Methods section and Extended Data Figure 4. For *qpGraph*, we focused on assessing whether graphs fulfill the criterion that all or nearly all f_4 -statistics predicted by the model are consistent with the empirical statistics. We refer to individual f_4 -statistics with $|Z| > 3$ between empirical and predicted statistics as 'outliers', providing evidence for unmodeled shared genetic drift between populations. We also assess the relative fit of models by considering a sum of squared differences-statistic (RMSE) summarizing the overall fit between predicted and empirical f -statistics, but note that the distribution of this statistic under arbitrary admixture graph models is not known.

We focused on a set of populations representing broad genetic and geographic diversity:

1. Outgroup: African Yoruba
2. Unadmixed Papuans: New Guinean Highlanders
3. Mainland East Asia: Dai Chinese
4. Indigenous Taiwanese: Atayal
5. Philippines population without Papuan ancestry: Kankanaey
6. Philippines population with Papuan ancestry: Mamanwa
7. Northern Solomons Papuan-speakers who are a mixture of Papuan and First Remote Oceanian ancestry: Nasioi (HGDP Bougainville)
8. Solomon Islander Austronesian-speakers who are a mixture of Papuan and First Remote Oceanian ancestry: Kolombangara
9. Polynesians: Tongan
10. Australians

A 'skeleton graph' of a tree-like population history that fits the ancient DNA data. We first added Lapita_Vanuatu to all 8 possible edges of a skeleton phylogeny of (Yoruba, (New_Guinea, (Dai, (Atayal, Kankanaey))))), which we found fitted the data without outliers either using all SNPs or only transversions. Two principal graph topologies fitted with no outliers, one of which modeled the Lapita_Vanuatu as being most closely related to the Kankanaey of the Philippines, and the other which modeled the Lapita_Vanuatu, Atayal, and Kankanaey as a trifurcation. Since the trifurcation model also had a slightly greater chi-squared statistic (RMSE=1.71) than the model in which the Lapita_Vanuatu share on the order of $F_{ST} = 0.001$ drift with the Kankanaey lineage (RMSE=1.58), we continued to use the latter graph, but note that the shared drift between the Kankanaey and Lapita_Vanuatu is small.

Adding populations to the skeleton graph

We proceeded by adding four genetically differentiated populations as mixtures of all 45 possible pairs of edges of the graph: the Polynesian Tongans, the Solomon Islander Kolombangara, the Northern Solomons Nasioi (HGDP Bougainville), and the Philippine Mamanwa³.

Tongans, Kolombangara, and Nasioi

We found that Tongans, Kolombangara and Nasioi can only be fit as having derived their ancestry from the First Remote Oceanian lineage on the one hand, and the New Guinean Highlander lineage on the other hand, with no outlier statistics and $\chi^2 < 1.9$ for all three populations. All other graphs for Tongans and Kolombangara have 32 or more outlier statistics, and $\chi^2 > 15$. The second best fitting graph for Tongans posits that Tongans received their Austronesian ancestry from the Atayal lineage, but this graph has 50 outlier f_4 -statistics; for example f_4 (Yoruba, Tongan; Atayal, ancient Oceanians) which is predicted to be ~ 0 in this graph is empirically 0.021 ($Z=9.6$). In addition, this graph has a chi-squared statistic of $\chi^2 = 41$ compared to 1.6 for the best fitting graph. For Nasioi the second best fitting graph has 10 outlier statistics, but a slightly smaller χ^2 -statistic (5.3). We conclude that there is overwhelming evidence that all three populations received their non-Papuan ancestry from a source most closely related to the Lapita_Vanuatu, which in this paper we call the First Remote Oceanian lineage.

Mamanwa

We found that there were three phylogenies that fit the Mamanwa without outliers. One portion of Mamanwa ancestry was in all cases derived from the lineage leading to New Guinean Highlanders, but the second source was either the First Remote Oceanian lineage, the Kankanaey lineage, or the lineage ancestral to Kankanaey and the Lapita_Vanuatu. In all three cases the inferred drift length separating this second source of the Mamanwa ancestry to the (Lapita_Vanuatu, Kankanaey) ancestral node was $F_{ST} = 0.0001$. We are thus not able to uniquely place the Mamanwa with respect to the First Remote Oceanians and Kankanaey, but use the placement along the Kankanaey lineage due to a smaller chi-squared statistic.

Modeling Tongans, Solomon Islanders and Bougainville populations simultaneously

We next proceeded to testing all possible models that included two populations out of the set (Tongan, Kolombangara, Nasioi). To do this we used the unique fitting graph for each population from our initial additions to the skeleton graph, and added the two others as mixture between all possible edges of that graph. We find that for all these combinations of populations there are at least 23 graph topologies without outliers, and we thus do not have enough resolution in the data to distinguish between different graphs that include these populations simultaneously. We therefore focus on graphs where only one of these populations is included at the same time.

Adding populations to a skeleton graph with Australians

We also consider a graph that models Australians as a clade with New Guineans (Yoruba, ((New_Guinea, Australian), (Dai, (Atayal, (Kankanaey, Lapita_Vanuatu)))). The Lapita_Vanuatu fit in this graph as a clade with Kankanaey without outliers, but when we add Tongan, Kolombangara, Nasioi and Mamanwa as admixed between all possible 66 pairs of graph edges, there are no solutions without outlier statistics. For the Tongan, Kolombangara, and Nasioi, the best fitting graph posits that they are mixed between the Lapita_Vanuatu and New Guinea Highlander lineages, although both the Nasioi and Tongan graph show outliers (2 and 7, respectively) suggestive of unmodeled affinity to Australians. For Tongans:

$f_4(\text{New_Guinea, Australia; Dai, Tongan})$: model=-0.011118, empirical=-0.006075, ($Z = 4.9$)

$f_4(\text{New_Guinea, Australia; Atayal, Tongan})$: model=-0.011118, empirical=-0.005472, ($Z = 4.9$)

The Mamanwa are optimally fitted as a lineage basal to Australians and New_Guineans, but a single $Z = 3$ outlier ($f_4[\text{Yoruba, Mamanwa; New_Guinean, Australian}]$) suggests that the portion of Mamanwa ancestry related to Papuans and Australians is slightly closer to Australians. The difference between this statistic and the statistic in Table S1 (see below) is that the graph statistic is based on transversion SNPs where we have at least one Lapita_Vanuatu genotype.

Graphs that can account for the excess affinity between Oceanians and Australians

We have shown that when we model the non-First Remote Oceanian ancestry of Tongans as entirely Papuan—with the lineage leading to Australians basal to both lineages—we observe more allele sharing between Tongans and Australians than is predicted by this model. In contrast, differences between Australians and Papuans affinity are not detectable in non-Oceanians, who are approximately symmetrically related to Australians and Papuans (Table S1).

Table S1. Australians and New Guinean Highlanders are approximately symmetrically related to non-Oceanians based on the statistic $f_4(\text{Yoruba, } X, \text{Australian, New_Guinea})$.

X	f_4	Z
Primate_Chimp	0.000044	0.1
Dinka	0.000078	0.5
Sardinian	-0.000558	-1.6
Onge	0.000287	0.7
Dai	0.000214	0.6
Atayal	0.000261	0.6
Karitiana	0.000289	0.6
Mixe	0.000076	0.2
Mamanwa	0.000819	2.3
Tongan	0.002998	7.7
Kolombangara	0.002590	6.6

We tested different permutations of a graph where either Papuans, Australians, or Tongans are admixed (and excluded the Kankanaey for simplicity), and found that the only permutation tested that resulted in no outliers posited that the Papuan population that contributed ~25% of the ancestry of Tongans also contributed ~45% of the ancestry of Australians (Extended Data Figure 3). An alternative graph that only has a single $Z=3.17$ outlier posits that the non-First Remote Oceanian portion of Tongan ancestry is itself admixed between Papuan- and Australian-related sources. We cannot statistically distinguish between these two models, but conclude that there is evidence for complexity in the history of Papuans, Australians and Polynesians.

Graph model with Tongans and Mamanwa

Combining the observations from the systematic tests of all possible placements of different populations on skeleton graphs above, we find that:

1. Tongans, Nasioi and Kolombangara are generally fitted as mixed between the same two lineages: First Remote Oceanian represented by the ancient samples, and Papuan.
2. The Philippine Mamanwa can be fitted as being mixed from a lineage related to the Philippine Kankanaey, and a lineage that split off prior to the separation of the ancestors of Papuans and Australians.
3. Australians cannot be successfully fit as an outgroup to the non-First Remote Oceanian ancestry in Tongans, and will be treated separately.

We found that a graph with Mamanwa and Tongans results in an excellent fit with no outliers. We also confirmed that Tongans could be successfully replaced with Nasioi or Kolombangara without resulting in a poor fit.

Extended graph model with archaic human genomes and Onge Andaman Islanders

In the above, we do not model the Neanderthal ancestry that separates non-Africans from Africans, or the Denisovan ancestry that separates Australians and Papuans from other populations. While this ancestry is important when considering the histories of these populations, the reason that we can successfully fit Australians and Papuans without modeling their Denisovan ancestry in our graphs is that the difference in archaic ancestry between the Australians and Papuans and mainland non-Africans can be accounted for by shifting the bifurcation point with Yoruba Africans. If multiple African populations with different divergences from non-Africans had been included the Denisovan component of Australian and Papuan ancestry would be expected to share different amounts of drift with the African populations as they have different degrees of relatedness to non-Africans. This would make it impossible for a model that does not model the archaic admixtures to accommodate the real data.

To fit an extended model incorporating archaic admixture inferred by previous studies, we used chimpanzee as an outgroup, and both the archaic Denisova and Altai Neanderthal genomes^{4,5}. Following previous findings, we modeled the Denisovans as being admixed between the Neanderthal lineage and a more basal ('unknown archaic') lineage⁴. We also modeled

Neanderthal admixture in the ancestral non-African population⁶, and Denisovan admixture in the ancestral population of Australians and Papuans⁷. As a final expansion of our population history model, we added the Andamanese Onge on the lineage ancestral to Australians and Papuans but diverging prior to the Denisovan admixture event³. We found that the resulting graph (Figure 3A) fits well given the large number of populations included. There are only two outliers that deviate by slightly more than 3 standard errors from empirical statistics:

$f_4(\text{Yoruba, Tongan; Denisova, Altai})$: model = 0.001284, empirical = 0.006947, ($Z = 3.1$)

$f_4(\text{Atayal, Mamanwa; Onge, Tongan})$: model = -0.010912, empirical = -0.007444, ($Z = 3.2$)

The first outlier could be interpreted as suggesting unmodeled affinity between Tongans and Neanderthals, but the second one is more difficult to interpret, as it is consistent for example with unmodeled affinity between Mamanwa and Tongan or Onge and Atayal.

Adding the ancient Tonga sample

For the graph we display in Figure 3, we do not include the Mamanwa but added the Lapita_Tonga sample which despite being represented by only a single individual can be adequately fitted as being a clade with the Lapita_Vanuatu sample. The three minor outliers with Z-scores between 3 and 3.5 were all for population configurations that have little to do with the history of admixture and population splits in Oceania that we are most interested in:

$f_4(\text{Yoruba, New_Guinea; Yoruba, Dai})$: model = 0.143671, empirical = 0.133643, ($Z = 3.3$)

$f_4(\text{Yoruba, Dai; Yoruba, Dai})$: model = 0.193597, empirical = 0.183619, ($Z = 3.0$)

$f_4(\text{Yoruba, Dai; Dai, Atayal})$: model = 0.004789, empirical = -0.000185, ($Z = 3.4$)

We note that we can also fit modified models with the Lapita_Tonga as being a clade with the First Remote Oceanian lineage contributing to present-day Tongans with approximately the same support, but we chose to conservatively display the graph in Figure 3.

References:

- 1 Reich, D., Thangaraj, K., Patterson, N., Price, A. L. & Singh, L. Reconstructing Indian population history. *Nature* 461, 489-494 (2009).
- 2 Patterson, N. *et al.* Ancient admixture in human history. *Genetics* 192, 1065-1093 (2012).
- 3 Reich, D. *et al.* Denisova Admixture and the First Modern Human Dispersals into Southeast Asia and Oceania. *The American Journal of Human Genetics* 89, 516-528, (2011).
- 4 Prufer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505, 43-49, doi:10.1038/nature12886 (2014).
- 5 Meyer, M. *et al.* A High-Coverage Genome Sequence from an Archaic Denisovan Individual. *Science* 338, 222-226, doi:10.1126/science.1224344 (2012).
- 6 Green, R. E. *et al.* A Draft Sequence of the Neanderthal Genome. *Science* 328, 710-722 (2010).
- 7 Reich, D. *et al.* Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468, 1053-1060 (2010).