Ancient DNA from Guam and the peopling of the Pacific

Irina Pugacha1, Alexander Hübner2,3,4,5,6, Hsiao-chun Hung6,7, Matthias Meyer8, Mike T. Carson6,1,9, and Mark Stoneking8,10

1Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, D-04103 Leipzig, Germany; 2Department of Archaeogenetics, Max Planck Institute for Evolutionary Anthropology, D-04103 Leipzig, Germany; 3Department of Archaeology and Natural History, Australian National University, Canberra, ACT 2601, Australia; and 4Micronesian Area Research Center, University of Guam, 96923 Mangilao, Guam

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected in 2020.

Contributed by Mark Stoneking, November 12, 2020 (sent for review October 26, 2020; reviewed by Geoffrey K. Chambers, Elizabeth A. Matisoo-Smith, and Glenn R. Summerhayes)

Humans reached the Mariana Islands in the western Pacific by ∼3,500 y ago, contemporaneous with or even earlier than the initial peopling of Polynesia. They crossed more than 2,000 km of open ocean to get there, whereas voyages of similar length did not occur anywhere else until more than 2,000 y later. Yet, the settlement of Polynesia has received far more attention than the settlement of the Marianas. There is uncertainty over both the origin of the first colonizers of the Marianas (with different lines of evidence suggesting variously the Philippines, Indonesia, New Guinea, or the Bismarck Archipelago) as well as what, if any, relationship they might have had with the first colonizers of Polynesia. To address these questions, we obtained ancient DNA data from two skeletons from the Ritidian Beach Cave Site in northern Guam, dating to ∼2,200 y ago. Analyses of complete mitochondrial DNA genome sequences and genome-wide SNP data strongly support ancestry from the Philippines, in agreement with some interpretations of the linguistic and archaeological evidence, but in contradistinction to results based on computer simulations of sea voyaging. We also find a close link between the ancient Guam skeletons and early Lapita individuals from Vanuatu and Tonga, suggesting that the Marianas and Polynesia were colonized from the same source population, and raising the possibility that the Marianas played a role in the eventual settlement of Polynesia.

Significance

We know more about the settlement of Polynesia than we do about the settlement of the Mariana Islands in the western Pacific. There is debate over where people came from to get to the Marianas, with various lines of evidence pointing to the Philippines, Indonesia, New Guinea, or the Bismarck Archipelago, and over how the ancestors of the present Mariana Islanders, the Chamorro, might be related to Polynesians. We analyzed ancient DNA from Guam from two skeletons dating to ∼2,200 y ago and found that their ancestry is linked to the Philippines. Moreover, they are closely related to early Lapita skeletons from Vanuatu and Tonga, suggesting that the early Mariana Islanders may have been involved in the colonization of Polynesia.

Many books have been written about where the Polynesians came from but nobody cares a straw about where the Guamanians came from. And yet it is probable that they can tell at least as much about the peopling of the Pacific as can the Polynesians.

——William Howells, The Pacific Islanders (1)

The human settlement of the Mariana Islands, in western Micronesia, was in some respects more remarkable than the settlement of Polynesia. And yet, as noted in the quote above and by others (2), the settlement of Polynesia has received far more attention than that of the Mariana Islands. Consisting of 15 islands (of which Guam is the largest and southernmost) stretching across some 750 km of sea, the Marianas archipelago is located ∼2,500 km east of the Philippines and ∼2,200 km north of New Guinea (Fig. 1). The earliest archaeological sites date to around 3.5 thousand y ago (3), and paleoenvironmental evidence suggests even older occupation, starting around 4.3 kya (4). Thus, the first human presence in the Marianas was at least contemporaneous with, and possibly even earlier than, the earliest Lapita sites in Island Melanesia and western Polynesia that date to after 3.3 kya (5) and are associated with the ancestors of Polynesians. However, reaching the Marianas necessitated crossing more than 2,000 km of open ocean, whereas voyages of similar length were not accomplished by Polynesian ancestors until they ventured into eastern Polynesia within the past 1,000 y (2, 6).

Where these intrepid voyagers originated, and how they relate to Polynesians, are open questions. Mariana Islanders are unusual in many respects when compared with other Micronesians and Polynesians. Chamorro, the indigenous language of Guam, is classified as a Western Malayo-Polynesian language within the Austronesian language family, along with the languages of western Indonesia (the islands west of Wallace’s Line) (Fig. 1), Sulawesi, and the Philippines. Palauan, another indigenous language of western Micronesia, is also a Western Malayo-Polynesian language, whereas all other Micronesians and all Polynesian languages belong to the Oceanic subgroup of Eastern Malayo-Polynesian (7). The most definitive features of Lapita pottery, associated with the earliest presence of Austronesians in Island Melanesia and western Polynesia (8), are absent in the Marianas, as are the domestic animals, such as pigs, dogs, and chickens typically associated with Lapita sites and Polynesian settlement (9). Moreover, rice cultivation seems to have been present as an indigenous tradition in the Marianas (10), but so far no such evidence has been found elsewhere in Remote Oceania.

Author contributions: M.T.C. and M.S. designed research; M.M. performed research; H.-C.H. and M.T.C. contributed new reagents/analytic tools; I.P. and A.H. analyzed data; and I.P. and M.S. wrote the paper with contributions and comments from all authors.

Reviewers: G.K.C., Victoria University of Wellington; E.A.M.-S., University of Otago; and G.R.S., University of Otago.

The authors declare no competing interest.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

1To whom correspondence may be addressed. Email: mtcarson@triton.uog.edu or stoneking@eva.mpg.de.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2022112118/-/DCSupplemental.

Published December 21, 2020.
These linguistic and cultural differences have led most scholars to conclude that the settlement of western Micronesia and Polynesia had little to do with one another. To be sure, indications have been noted of morphological (1), cranial (11), and genetic (12–16) affinities between Micronesians and Polynesians [see also Addison and Matsuo-Smith (17)], and stylistic links between the pottery of the Philippines, the Marianas, and the Lapita region have also been illustrated (18). Nonetheless, the standard narrative for Polynesian origins (Fig. 1) is that they reflect a movement of Austronesian-speaking people from Taiwan around 3.5 kya into western Polynesia, with subsequent short, intervisible water crossings, while the islands of Remote Oceania required substantial water crossings that were not intervisible and that were not achieved until ∼3.5 kya or later. Red dots indicate the locations of the early Lapita samples from Vanuatu and Tonga; the blue arrow indicates the conventional route for the Austronesian expansion to the Bismarck Archipelago, which was then the source of initial voyages to Remote Oceania; the solid red arrow indicates the route for the settlement of the Marianas supported by this study; and the dashed red arrow indicates the potential additional contribution of Mariana Islanders to further settlement of the Pacific, suggested by this study.

In principle, to address this issue, the ancestry of the modern inhabitants of the Marianas could be analyzed for Papuan-related ancestry. However, a common finding of ancient DNA studies is that the ancestry of people in a region today may not reflect the ancestry of people living in that region thousands of years ago (42). In particular for the Marianas, the archaeological evidence indicates substantial cultural change around ∼1 kya (28, 43), coinciding with the construction of stone-pillar houses in formal village arrangements (latte) at a time when nearly all of

Fig. 1. Map of the western Pacific, showing locations and areas mentioned in the text. The inset shows the location of the Ritidian Site on Guam. Location names in red have been suggested as potential sources for the settlement of the Marianas Islands. Wallace’s Line divides biogeographic regions and lies at the boundary of the prehistoric continental landmasses of Sunda and Sahul. The dashed blue line indicates the boundary between Near and Remote Oceania; the islands of Near Oceania were colonized beginning 45 to 50 kya and involved relatively short, intervisible water crossings, while the islands of Remote Oceania required substantial water crossings that were not intervisible and that were not achieved until ∼3.5 kya or later. Red dots indicate the locations of the early Lapita samples from Vanuatu and Tonga; the blue arrow indicates the conventional route for the Austronesian expansion to the Bismarck Archipelago, which was then the source of initial voyages to Remote Oceania; the solid red arrow indicates the route for the settlement of the Marianas supported by this study; and the dashed red arrow indicates the potential additional contribution of Mariana Islanders to further settlement of the Pacific, suggested by this study.
the Pacific Islands were populated and connected by long-distance sea voyaging (8). The presence of mtDNA haplogroup B4 sequences in modern Chamorro has been attributed to contact during the latte period (28).

In addition, population contacts and movements became more complicated during the European colonial period, starting with the arrival of Magellan in 1521 in the Marianas and continuing with the Manila-Acapulco galleons (and slave trade) from 1565.
to 1815; Guam was a regular stopover on these voyages. European colonialism also involved multiple relocations and reductions in population size across the archipelago. These events undoubtedly had an impact on the genetic ancestry of the modern Chamorros, making it more difficult to assess their origins and potential relationships with Polynesians. It would therefore be preferable to address these issues with ancient DNA from the Marianas.

At the Ritidian Site in northern Guam (SI Appendix, Fig. S1), two skeletons clearly predating the late period were found outside a ritual cave site (44). These individuals, RBC1 and RBC2, were buried side-by-side in extended positions, with heads and torsos removed (SI Appendix, Fig. S2). Direct radiocarbon dating of a bone from RBC2 produced a result of 2,180 ± 30 calibrated years BP (44), which is thus some 1,000 y after the initial settlement of Guam, but also some 1,000 y before the late period. Here we report the analysis of ancient DNA retrieved from these remains; our results contribute to the debate over the starting point for the first voyages that led to human settlement of the Marianas, and we provide additional insights into the role of the Marianas in the larger view of the peopling of the Pacific.

Results

Shotgun sequencing of libraries constructed from DNA extracted from the ancient Guamanian skeletons revealed elevated C→T substitutions at the ends of fragments, as expected for ancient DNA (SI Appendix, Fig. S3 and Dataset S1). The percent endogenous DNA was too low for further shotgun sequencing (Dataset S1); we therefore proceeded by capture enrichment for the mtDNA genome, and for a panel of 1.2 million SNPs used in previous ancient DNA studies (24–26, 45, 46), prior to sequencing.

mtDNA and Y Chromosome. After merging the sequence data from libraries enriched for mtDNA while excluding those that were highly contaminated (Dataset S2), we were able to obtain mtDNA genome sequences at an average coverage of 95.2-fold for RBC1 and 261.3-fold for RBC2. Estimated contamination in the mtDNA sequences, using a likelihood-based approach (47) subsequently referred to as contamMix (48), was 17.9% for RBC1 and 6.6% for RBC2 (Dataset S3). The sequences are identical where they overlap, and even with this relatively high level of contamination, both sequences are confidently assigned to haplogroup E2a (Dataset S3). In addition to the diagnostic mutations for haplogroup E2a, both sequences carry a novel acid substitution (Gln → Arg) in the ATP6 gene.

Haplogroup E2a is the most common haplogroup in the modern Chamorro population of Guam (28), with a frequency of 65%. Elsewhere it is reported to occur sporadically in populations from the Philippines and Indonesia (49–52), and in a single individual from the Solomon Islands (53); otherwise, it is absent from Oceania and has not been reported from Mainland Southeast Asia. The finding of this haplogroup in the ancient Guamanian skeletons thus suggests links to the Philippines and Indonesia, rather than New Guinea or the Bismarck Archipelago. Of additional importance, the high frequency of this haplogroup in modern Chamorros suggests a degree of genetic continuity with the population represented by the ancient skeletons, persisting through modern Chamorros suggests a degree of genetic continuity with the interceding cross-population contacts since the latte period after ~1,000 y and later European colonial events.

Based on the ratio of the average coverage of X chromosome vs. Y chromosome + autosomal reads in the shotgun sequencing data, RBC1 is male and RBC2 is female (SI Appendix, Fig. S4). The Y chromosome of RBC1 is assigned to haplogroup O2a2 (formerly haplogroup O3a3), based on having the derived allele for the diagnostic marker P201 (54); genotypes at all other informative Y-chromosome SNPs for which there are data from RBC1 are consistent with this haplogroup (Dataset S4). Haplogroup O2a2-P201 is widespread across Mainland and Island Southeast Asia and Oceania, and has been associated with the Austronesian expansion (55, 56).

Genome-Wide SNP Data: Ancient Guam Origins. We enriched 13 sequencing libraries from RBC1 and 11 from RBC2 for ~1.2 million SNPs (Dataset S5) and obtained data (Dataset S6) for 128,772 SNPs (39,760 in deaminated reads) for RBC1 and 361,982 SNPs (143,451 in deaminated reads) for RBC2. Given the relatively high contamination estimates for some of the libraries (Dataset S5), we either redid analyses using only deaminated reads (if there were enough deaminated reads), or included data from Europeans in the analysis, to ensure that contamination with modern European DNA was not influencing the results. The results reported below are based on all reads, as we did not find any indication of contamination influencing the results.

We first checked if RBC1 and RBC2 might be related by calculating the fraction of pairwise differences for the 33,400 overlapping SNPs between them, and comparing this to mean pairwise distances for first-, second-, and third-degree relatives in the 1000 Genomes Project dataset (57), using sites on the Human Origins Array (SI Appendix, Fig. S5A). While the pairwise distance between RBC1 and RBC2 is similar to that for first-degree relatives in the 1000 Genomes dataset, suggesting that they might be first-degree relatives, we obtained similar mean pairwise distances for other ancient samples from Southeast Asia and Oceania (SI Appendix, Fig. S5B). We therefore conclude that the limited amount of data and low overall genetic diversity characteristic of the ancient samples preclude accurate assessment of relatedness.

We then projected RBC1 and RBC2 onto principal components (PCs) constructed with modern samples genotyped on the Affymetrix 6.0 platform and with data from the Simons Genome Diversity Project (SGDP); the overlap with this array is provided in Dataset S6, and details on the modern samples are in Dataset S7. This dataset has good coverage of populations from Island Southeast Asia, in particular from eastern Indonesia, which exhibit both Asian-related and Papuan-related ancestry and hence are a potential proxy for the ancestry in the ancient Guam samples if they in fact have Papuan-related ancestry. The results for the first two PCs (Fig. 2) show three axes of variation, with Europe/South Asia, New Guinea, and Southeast Asia at the vertices. The two ancient Guam samples overlap samples from Taiwan and the Philippines. There is no indication of any Papuan-related ancestry in the ancient Guam samples, particularly when compared with eastern Indonesian samples, all of which have some Papuan-related ancestry and hence are clearly separated from other Southeast Asian samples.

We next carried out ADMIXTURE analysis of the same dataset; while the results for K = 3 are associated with the lowest cross-validation error (SI Appendix, Fig. S6A), the results for K = 6 distinguish different ancestry components for Mainland vs. Island Southeast Asia, so we show these results in Fig. 2B and the results for K = 2 to K = 8 in SI Appendix, Fig. S7. Notably, the yellow ancestry component, which is characteristic of New Guinea and is also present in eastern Indonesia, is completely lacking in the ancient Guam samples for all analyzed values of K (Fig. 2B and SI Appendix, Fig. S7). Moreover, at K = 6 the two ancient Guam samples have the dark blue ancestry component, which is at highest frequency in individuals from the Philippines and Taiwan (Fig. 2B). RBC1 also has a purple component, which likely reflects recent European DNA contamination.

Thus, these PC and ADMIXTURE analyses suggest that there is no Papuan-related ancestry in the ancient Guam samples, and moreover indicate that they are most similar to modern samples from the Philippines and Taiwan. However, the number of SNPs in the Affymetrix 6.0 dataset that overlap the ancient Guam
samples is too small for more formal tests of population relationships (Dataset S6), and moreover this dataset has limited coverage of modern Oceanian populations. We therefore carried out all further analyses with the Human Origins dataset, which includes more modern samples from Near and Remote Oceania (Dataset S7), more overlap with the ancient Guamanian data (Dataset S6), and also includes data from ancient samples from Asia and the Pacific (Dataset S8), including early Lapita samples from Vanuatu and Tonga.

A principle components analysis (PCA) of these samples with the ancient samples projected (Fig. 3A) places the early Lapita samples at one vertex, East Asia at another, and New Guinea at the third vertex; the ancient Guamanian samples are now projected away from modern Taiwan and Philippine samples, in the direction of the early Lapita samples. An ADMIXTURE analysis of these data for $K = 9$ (Fig. 3B; results for $K = 5$ to $K = 12$ in SI Appendix, Fig. S8), which has the lowest cross-validation error (SI Appendix, Fig. S6B), now reveals two primary ancestry components in the ancient Guamanian samples: A dark blue component as before that is at highest frequency in Indonesia and the Philippines, and an orange component that is at highest frequency in Polynesia; the additional minor purple component likely reflects recent European contamination. As before, there is no indication from either the PCA or the ADMIXTURE analysis of any Papuan-related ancestry in the ancient Guamanian samples.

While the presence of these two ancestry components in the ancient Guamanian samples could indicate admixture between a source population related to Indonesia/Philippines and another related to Polynesians, other explanations for the presence of multiple ancestry components are possible (37, 38). In particular, it could be that the ancient Guamanian samples are ancestrally related to both Indonesia/Philippines and to Polynesians, and that subsequent divergence and genetic drift has facilitated the identification of separate Indonesia/Philippines and Polynesia-related ancestry components in the ADMIXTURE analysis, both of which are present in the ancient Guamanian samples. To investigate the relationships of the ancient Guamanian samples in more detail, we analyzed outgroup-$f_3$ and -$f_4$ statistics. The outgroup-$f_3$ analysis, which compares the amount of drift (i.e., ancestry) shared by the ancient Guamanian samples with other populations relative to an outgroup (Mbuti), shows that the ancient Guamanian samples share the most drift with the Lapita Vanuatu and Lapita Tonga samples, followed by an ancient sample from the Philippines and then by modern samples from the Philippines and Taiwan and late Neolithic samples from the Taiwan Strait Islands (Fig. 4A). Notably, the drift shared with New Guinea, and with the French, is less than that with any other population, indicating that the ancient Guamanian samples share the least relatedness with these two populations. These results further support the lack of any Papuan-related ancestry in the ancient Guamanian samples, and moreover also indicate that recent European contamination is not influencing these results.

We then constructed an $f_4$ statistic of the form $f_4(\text{test}, \text{Kankanae}; \text{New Guinea highlands, Mbuti})$; values of this statistic that are equal to zero indicate that the test population forms a clade with Kankanae relative to New Guinea; values less than zero indicate that Kankanae shares more ancestry with New Guinea than does the test population; and values greater than zero indicate that the test population shares more ancestry with New Guinea than does Kankanae. We used the ancient Guamanian samples and all other populations from Oceania as the test population; the results (Fig. 4B) indicate that all populations from Oceania tested share ancestry with New Guinea in comparison to Kankanae, except for the ancient Guamanian samples. These form a clade with Kankanae, as the $Z$-statistic is not significantly different from zero ($Z = -1.93$) Genome-Wide SNP Data: Relationships with Early Lapita Samples.

The PCA, ADMIXTURE, and outgroup-$f_3$ analyses not only indicate affinities between the ancient Guamanian samples and Philippine/Taiwan populations, but additionally suggest strong affinities between the ancient Guamanian and early Lapita samples. To investigate in more detail the relationships among the ancient Guamanian and early Lapita samples with samples from Asia and Oceania, we conducted $f_4$ analyses of the form (ancient Guamanian, early Lapita; Asia/Oceania, Mbuti), separately for the early Lapita Vanuatu and Tonga samples and for all modern and ancient Asian and Oceanian samples in the dataset. Values of this $f_4$ statistic that are consistent with zero imply that the ancient Guamanian and early Lapita samples form a clade; negative values indicate excess shared ancestry between the early Lapita samples and the Asia/Oceania population; and positive values indicate excess shared ancestry between the ancient Guamanian samples and the Asia/Oceania population. The results (SI Appendix, Fig. S9) show that the ancient Guamanian and early Lapita samples always form a clade with one another when compared with any Asian population. However, both of the early Lapita samples share more ancestry with ancient and modern Polynesian samples (but not with any other samples from Oceania) than do the ancient Guamanian samples. This is further supported by outgroup-$f_3$ comparisons of the ancient Guamanian and early Lapita samples with other populations (SI Appendix, Fig. S10): Both early Lapita samples share more drift with the modern and ancient Remote Oceanians sampled than do the ancient Guamanian samples. Nonetheless, $f_4$ statistics of the form (Oceania, early Lapita; ancient Guamanian, Mbuti) are always significantly negative for both early Lapita samples, regardless of which Oceanian population is included in the test (SI Appendix, Fig. S11). These $f_4$ results indicate that there is shared drift between the early Lapita and ancient Guamanian samples when compared with any other Oceanian sample, in keeping with the outgroup-$f_3$ results (Fig. 4). Overall, the $f_3$ and $f_4$ results imply that while the early Lapita and ancient Guamanian samples are closely related to each other, the early Lapita samples are a better proxy for the Polynesian-related ancestry in modern and ancient Oceanian samples than are the ancient Guamanian samples.

We next used admixture graphs (i.e., trees that allow for admixture or migration) to further investigate the relationships among the ancient Guamanian, early Lapita, and other Asian and Oceanian samples. Included in these analyses were: New Guinea Highlanders as a source of Papuan ancestry; Han Chinese as a source of Asian ancestry; Kankanae as a source of Austronesian ancestry; Tolai (mixed Papuan/Austronesian ancestry) and Bainning Marabu (Papuan ancestry only) from New Britain to investigate relationships with the Bismarck Archipelago; modern Vanuatu with mixed Papuan/Austronesian ancestry; and the ancient Guamanian, Lapita Vanuatu, and Lapita Tonga samples. We also included Mbuti as an outgroup. We first constructed a maximum-likelihood tree and added migration edges, using the software TreeMix (59); a tree with two migration edges (Fig. 5A) has all residuals within 3 SE (SI Appendix, Fig. S12) and thus provides a reasonable fit. This tree indicates shared drift between the ancient Guamanian and early Lapita samples, with the migration edges bringing Lapita-related ancestry into the modern Vanuatu and Tolai samples.

We additionally investigated admixture graphs using a Markov chain Monte Carlo method, implemented in the software AdmixtreeBayes (60), to sample the space of possible admixture graphs. The graph with the highest posterior probability (17.6%) supports shared drift between the ancient Guamanian and early Lapita samples (SI Appendix, Fig. S13); moreover, a consensus graph that depicts the nodes present in at least 50% of the posterior sample of 1,000 admixture graphs (Fig. 5B) indicates that the shared drift between the ancient Guamanian and early Lapita samples (node n3 in Fig. 5B) appears in 99% of the topologies. We
further examined this topology, inferred in an unsupervised manner by both TreeMix and AdmixtureBayes, with a combination of f statistics using the qpGraph software. This topology has a worst-fitting Z-score of 4.56 (Fig. 5C), which is above the conventional threshold of the worst-fitting |Z-score| < 3 for an "acceptable" graph. In general, deviations between the fitted and observed data can be explained either by an incorrect topology (which, in the case of qpGraph, is specified by the user and not inferred from the data) or by unmodeled admixture. The worst-fitting f statistics tend to involve Han Chinese; when they are
excluded the worst-fitting Z-score is reduced to \(-3.72\). This graph has five \(f^3\) statistics with \(|Z\text{-score}| > 3\), all of which involve Mbuti and New Guinea Highlanders, so this graph probably provides a reasonable depiction of the relationships of the Oceanian samples, in particular the shared drift between the ancient Guam and early Lapita samples. For the two populations with mixed ancestry, the modern Vanuatu sample is inferred to have 65% Papuan-related and 35% Austronesian-related ancestry, while the Tolai sample has 85% Papuan-related and 15% Austronesian/Lapita-related ancestry; these estimates are in close agreement with those from AdmixtureBayes (Vanuatu: 66% Papuan-related and 34% Austronesian-related ancestry; Tolai: 87% Papuan-related and 13% Austronesian-related ancestry).

We further investigated the shared drift between the ancient Guam and early Lapita samples by including ancient samples from Liangdao that share ancestry with aboriginal Taiwanese (61) in the admixture graph analyses. While the results suggest that Liangdao is a better proxy than modern samples for the Austronesian-related ancestry in the ancient Guam and early Lapita samples (SI Appendix, Fig. S14), there is still shared drift between the ancient Guam and early Lapita samples.

**Discussion**

Some caution is warranted in interpreting the results of this study of ancient DNA from Guam, as they are based on two skeletons that may be related and that date from \(\sim 1,400\) y after the first human settlement of Guam. Previous studies of ancient DNA from early Lapita sites in Remote Oceania have found that initial results based on a limited number of samples (26) did not capture the full complexity revealed when additional samples were analyzed (24, 25). Nonetheless, the relationships that these ancient Guam samples exhibit with other ancient samples, as well as with modern samples from the region, provide some interesting insights into the peopling of Guam and the further settlement of Remote Oceania that should be the basis for further investigations.

**Origins of the Ancient Guam Samples.** The mtDNA and Y chromosome haplogroups of the ancient Guam samples suggest links with Southeast Asia rather than New Guinea or the Bismarck Archipelago. Moreover, none of the analyses of the genome-wide data found any trace of Papuan-related ancestry in the ancient Guam samples. Our results thus rule out any source for the ancestry of these individuals that is east of Wallace’s Line, as substantial amounts of Papuan-related ancestry are present in eastern Indonesia, New Guinea, and the Bismarck Archipelago. The most likely source is the Philippines, although western Indonesia is also possible; further sampling of Philippine and Indonesian populations—and ancient DNA from these regions—would help pinpoint the source. Moreover, in considering the archaeological evidence, finer-scale sampling is needed to contend with a rapid geographic spread of the red-slipped pottery horizon around 3.5 kya, reflecting population dispersal from the Philippines both eastward into the Marians and southward into Sulawesi, as well as eventually farther.

A Philippine source for the foundational population of Guam is consistent with the findings of modern DNA sampling (28), the linguistic evidence (2, 21), and the archaeological signature at the time of first Marianas settlement about 3.5 kya (30, 43). However, computer simulations of sea voyaging instead have indicated New Guinea or the Bismarck Archipelago as probable origin points of voyages reaching the Marians (33, 34). One potential scenario to reconcile these two lines of evidence is that people traveled from the Philippines to New Guinea or the Bismarcks, without mixing with any populations along the way, and then voyaged from New Guinea/the Bismarcks to Guam, again without first mixing with any resident populations. However, the TreeMix and AdmixtureBayes results (Fig. 5) do not support this scenario, nor does the linguistic and archaeological evidence. In particular, the earliest pottery in the Marians, dating to around 3.5 kya (3, 43), likely predates the oldest Lapita sites to the east of New Guinea, dated to not more than 3.3 kya (5). Yet the pottery, fine
shell ornaments, and other cultural objects in the Marianas dating to 3.5 kya are quite distinct from the Lapita tradition, and instead are connected to material markers in the Philippines that date to 3.8 to 3.5 kya (19, 30, 62), thus supporting movement from the Philippines to the Marianas. Moreover, the computer simulations of sea voyaging do not adequately consider the ability of ancient voyagers to travel against strong ocean currents and prevailing winds; in particular, the single outrigger canoes of the Chamorros—the “flying proas”—impressed early visitors with their greater speed and maneuverability, compared with Spanish ships (63). There is even at

Fig. 5. Tree and graph depictions of the relationships of ancient Guam, early Lapita, and select Asian and Oceanian populations. (A) Maximum-likelihood tree with two migration edges. All residuals (SI Appendix, Fig. S12) are within 3 SE. (B) Consensus graph with nodes present in at least 50% of the topology sets recovered with AdmixtureBayes. (C) Admixture graph obtained with qpGraph for the topology found by AdmixtureBayes with the highest posterior probability (SI Appendix, Fig. S13). The colored arrows in B and C indicate drift (ancestry) shared between the ancient Guam and early Lapita samples.
least one historically documented event of a Chinese trader drifting in a “sampun” from Manila to Guam during the 1600s (64). Ancient DNA from early Lapita skeletons in the Bismarcks would provide a further test of the hypothesis that people moved from the Bismarcks to Guam. It is, of course, possible that later periods of cross-contact voyaging brought additional groups of people to the Marianas from elsewhere (including perhaps the Bismarcks, among other places).

**Relationships between Ancient Guam and Early Lapita Samples.**

What about a Micronesian route [for the colonization of Polynesia]? It is not in favor with the anthropologists, though after all it was not anthropologists who settled Polynesia.

—William Howells, *The Pacific Islanders* (1)

All analyses consistently point to a surprisingly close relationship between the ancient Guam and early Lapita samples. This closeness is particularly evident in the outgroup–relationship between the ancient Guam and early Lapita samples. Moreover, admixture graphs indicate that the ancient Guam samples diverged first, and do not support movement of people from the Bismarcks to Guam (Fig. 5 and SI Appendix, Figs. S13 and S14). However, the admixture graph results should be viewed with caution, as they may be influenced by including a mix of ancient and modern DNA samples in the analyses (usually with fewer ancient than modern samples for each population), with possible attractions between ancient samples due to similar patterns of contamination or sequencing errors due to damage. Nonetheless, it appears that people either moved from the Marianas to the Bismarcks (or elsewhere in Island Melanesia) and then to other parts of Remote Oceania, or that the ancestors of the ancient Guam and early Lapita samples migrated separately, and by different routes, from the same source population.

Our results do not allow us to distinguish between these two possibilities. Arguing against a direct role for the Marianas in the later colonization of Polynesia is the lack of a linguistic connection, definitive Lapita pottery, or domesticated animals characteristic of Polynesia. However, languages spoken by people today may reflect subsequent developments, domestic animals may have been introduced via other routes, and the pottery of the Marianas predates Lapita pottery by a few centuries and is considered by some to be a related variety of the finely decorated pottery that subsequently became elaborated in Lapita pottery (18, 43). Moreover, we point out that a direct movement of people from the Philippines (or nearby areas) to the Bismarcks, either via the Marianas or by some other path that bypassed eastern Indonesia and the rest of New Guinea, would account for one peculiar observation, and that is the lack of Papuan-related ancestry in the early Lapita samples from Vanuatu and Tonga (24–26). If the ancestors of Polynesians migrated from Taiwan or the Philippines to the Bismarcks by island-hopping through eastern Indonesia and along the coast of New Guinea (Fig. 1), in a process that took a few hundred years (perhaps 10 to 15 generations), then they would have encountered people with Papuan-related ancestry along the way, and there would have been ample opportunity for them to have picked up some Papuan-related ancestry. Perhaps the ancestors of Polynesia did move via this route, but did not immediately mix with the people along the way, because of social or other perceived differences. However, any such barrier to mixing did not last long, as Papuan-related ancestry shows up in Vanuatu almost at the same time as the early Lapita samples (25), and there is evidence for substantial later Papuan-related contact in Vanuatu, Santa Cruz, and Fiji that then spread throughout Polynesia (24, 25, 27, 37). An alternative explanation that is worth considering is that the early ancestors of Polynesians lack Papuan-related ancestry because they did not encounter people with Papuan-related ancestry until they reached the Bismarcks, perhaps because they voyaged via the Marianas or otherwise bypassed eastern Indonesia and coastal New Guinea.

As the quotation from Howells (1) at the beginning of this section indicates, the settlement of Polynesia via Micronesia has generally not been considered by researchers. However, this possibility has been suggested based on pottery evidence (18), and the genetic evidence presented here provides further insights into the connections between Micronesians and Polynesians noted previously (11–17). Howell’s (1) suggestion of a role for Micronesia (specifically, the Marianas) in the settlement of Polynesia merits further consideration.

**Methods**

**Site Description and Samples.** The two skeletons, RBC1 and RBC2, were uncovered outside Ritidian Beach Cave (also called Ritidian First Cave), within the larger Ritidian Site of northern Guam (SI Appendix, Figs. S1 and S2). The two individuals had been buried side by side, in extended position inside distinctive pits. The heads and torsos had been removed slightly later. Details of these findings have been reported elsewhere and situated within the larger site chronology and context (44). The two skeletons from Ritidian offer a rare view of ancient burial practice in the Marianas region, as similar burial practices have been observed in the Philippines (65, 66) and Indonesia (67). While the site and indeed this specific cave revealed multiple cultural occupation layers dating back to the first regional settlement about 3.5 kya, these two burials of RBC1 and RBC2 were found within the uppermost layer, confirmed by direct radiocarbon date from a bone of RBC2 of 2,180 ± 30 yr BP (44). A tarsal bone was provided from each skeleton for ancient DNA analysis.

**DNA Extraction, Library Preparation, and Whole-Genome Sequencing.** In an ancient DNA clean room, ∼1 mm of material was removed from the surface of each specimen and ∼50 mg bone powder obtained by drilling into the bone with a dentistry drill at low speed. DNA was extracted following a protocol provided elsewhere, using spin columns and binding buffer option “D” (68). DNA libraries were prepared from 10-μl aliquots of each DNA extract using an automated protocol for single-stranded library preparation (69) with a Bravo NGS workstation. Negative controls were included both during DNA extraction and library preparation; these contained water instead of sample powder or DNA extract, respectively. The number of library molecules obtained from each sample DNA extract was more than 10 times higher than in the extraction and library negative controls (Dataset S1). All libraries, including the negative controls, were then amplified and double-indexed (70) as described elsewhere (69).

Whole-genome sequencing data were generated on the Illumina HiSeq 2500 platform (2× 76-bp paired-end sequencing). After de-multiplexing (requiring a perfect index), overlapping paired-end sequences were merged into full-length molecule sequences (71), and subsequently aligned to the human reference genome hg19 with decovoy sequences (ftp://ftp-trace.ncbi.nlm.nih.gov/ refseq/primate/hg19) using bwa-mem (72) using parameters optimized for direct reads (requiring a perfect index), overlapping paired-end sequences were merged into full-length molecule sequences (71), and subsequently aligned to the human reference genome hg19 with decovoy sequences (ftp://ftp-trace.ncbi.nlm.nih.gov/ refseq/primate/hg19) using bwa-mem (72) using parameters optimized for direct reads (requiring a perfect index). Sequences were assigned to their respective source libraries, and the genetic evidence presented here provides further insights into the connections between Micronesians and Polynesians noted previously (11–17). Howells’s (1) suggestion of a role for Micronesia (specifically, the Marianas) in the settlement of Polynesia merits further consideration.

**MDDNA Enrichment and Sequencing.** Libraries were enriched for human mitochondrial DNA using a synthetic probe set (75) encompassing the reference Cambridge sequence reference (rCRS) (76) in 1-bp tiling. Hybridization capture was performed in two successive rounds, following an on-bead capture protocol of (77) implemented in the Bravo NGS workstation (26). The enriched libraries were pooled with libraries from other projects and sequenced on an Illumina MiSeq in paired-end mode (2× 76 cycles).

The sequencing data were processed as described above for the whole-genome sequencing data, but mapped to the rCRS using bwa aln with the same settings. Sequences were assigned to their respective source libraries, requiring perfect matching of both indices. Sequences that were shorter than 35 bp or that did not produce alignments with a map quality of at least 25 were discarded. PCR duplicates were removed using DeDup (74). After discarding libraries with contamination >25% (Dataset S2), estimated by a likelihood-based method (47), we obtained 32,386 unique reads for RBC1 and 94,116 unique reads for RBC2 (Dataset S3). Enhanced frequencies of C→T

Pugach et al.

Ancient DNA from Guam and the peopling of the Pacific

PNAS | 9 of 12

https://doi.org/10.1073/pnas.2022112118
substitutions at the beginning and end of sequence alignments, which result from cytosine deamination in ancient DNA (79, 80), were detected in the mtDNA reads (Dataset S2).

An in-house pipeline (https://github.com/alexhbnr/mitoBench-ancientMT) was used to call the mtDNA consensus sequence, which required a minimum of three reads and used snpAD (81) to infer the consensus allele while taking into account ancient DNA damage. Contamination was estimated by a likelihood-based method (47), and HaploGrep2 (82) was used to call mitochondrial haplogroups.

**Genome-Wide SNP Capture Enrichment and Sequencing.** Enrichment of the libraries for a panel of ~1.2 million SNPs was performed using a set of DNA capture probes [“1240k”, composed of SNP panels 1 and 2, as described elsewhere (45)] and two successive rounds of in-solution hybridization capture (75). Sequencing of the enriched libraries and raw data processing were performed as described for the whole-genome sequencing above. Genotypes were inferred by randomly sampling an allele observed at each site after masking Ts at the five terminal bases at each read end by replacing them with Ns. For determining the Y chromosome haplogroup of the male sample, we subset the genotypes to the sites located on the human Y chromosome and analyzed them using yhtaplo (83) with the nondefault option “–ancStopThrash 1ed”.

The sequencing data have been made available at the European Nucleotide Archive under accession no. PRJEB40707.

**Genome-Wide SNP Data Analysis.**

**Comparative datasets.** Newly generated data from Guam were merged with published data from modern and ancient samples (Datasets S7 and S8) as follows. First, for comparisons to populations from Island Southeast Asia, the Guamanian samples were merged with previously-curated data from 25 modern populations genotyped on the Affymetrix 6.0 array (27, 84, 85); to provide worldwide context these data were further merged with a subset of the whole genome sequences from the SGDP (86). Related individuals were identified based on kinship coefficients, estimated using the software KING (87), with subsequent removal of one individual from each pair. Pruning of SNPs in linkage disequilibrium (LD) was done using the PLINK tool (87) with the following settings: –indep-pairwise 200 25.0 0.4 (88). After these quality-filtering steps, there were 136,162 SNPs and 303 individuals from 72 populations from Eurasia and Oceania remaining for the analyses. This dataset was used only for PCA and ADMIXTURE analyses.

Second, to better resolve relationships with populations from Near and Remote Oceania, as well as with other ancient samples from Asia and the Pacific, we used data from 53 modern populations from Oceania and 39 populations from East Asia genotyped on the Affymetrix Human Origins array (24, 26, 89–91), as well as previously published shotgun and capture-enrichment sequencing data from 82 ancient samples (24–26, 45, 61, 92). After removing related individuals as described above for the Affymetrix 6.0 data, this dataset consisted of 1,194 individuals and 593,124 SNPs. Not all samples were used for all of the analyses. For PCA and ADMIXTURE analyses, we used an LD-pruned dataset of 216,996 SNPs. In addition, ancient samples with more than 15,000 missing sites were excluded from the ADMIXTURE analysis.

**Data analyses.** We attempted to estimate relatedness between RBC1 and RBC2 by calculating the fraction of pairwise differences at 33,040 overlapping sites that are included on the Human Origins array. For comparison, we also calculated this fraction for modern samples from the 1000 Genomes Project dataset (57), which includes individuals with known degrees of relatedness, and for ancient samples from Southeast Asia and Oceania; the ancient DNA data were obtained from the Reich laboratory website (https://reich.hms.harvard.edu/downloadable-genotypes-present-day-and-ancient-dna-data-copiled-published-papers-v24-4). PCA was performed as described previously (93) with one modification, namely for the analyses which included ancient samples, the principle axes were calculated based on modern samples, and the ancient samples were projected using least-squares projection (which is more appropriate than orthogonal projection for samples with high amounts of missing data), as described in the documentation to the smartpca software (94).

To infer individual ancestry components and analyze population structure, we used the ADMIXTURE software (95) in the unsupervised mode. For each dataset, we first removed SNPs in strong LD ($r^2 > 0.4$) using the PLINK tool (96), and for the Human Origins dataset we further excluded ancient samples that had fewer than 15,000 SNPs remaining. We varied the number of ancestral populations ($K$ value) from $K = 2$ to $K = 8$ for the Affymetrix 6.0 dataset, and from $K = 5$ to $K = 12$ for the Human Origins Array dataset. We performed 100 independent runs for each value of $K$, and used the cross-validation procedure implemented in the ADMIXTURE software to assess the best value of $K$.

To formally test population relationships suggested by PCA and ADMIXTURE analyses, we used outgroup F3- and R2 statistics, implemented in the ADMIXTOOLS software suite (90). All data processing and analyses were carried out using the admixr R package (97).

To model the relationships between modern and ancient samples, we first used the following TreeMix (59) and AdmixtureBayes methods (60) to infer topologies, that were then tested using the qpGraph software implemented in ADMIXTOOLS (90). We performed 10 independent runs of TreeMix with zero to five migration events, and report the tree with the highest likelihood. For the AdmixtureBayes analyses we increased the default number of Markov chain Monte Carlo steps to 1,000,000, as recommended by the developers to avoid convergence problems for a model with 10 populations. We used the 10 topologies with the highest posterior probabilities estimated by AdmixtureBayes as input graphs for qpGraph, which we ran with parameters: bigsize: 0.05; forceomode: YES; lsqmode: YES; diag: 0.0001; bigger: 6; hires: YES, r-scale: 1. All three methods were applied to the exact same dataset, using all samples available for each population (Dataset S7), with the exception of the admixed modern Vanuatu. The amount of Polynesian ancestry in this population is highly variable, with a range of 9 to 38% today, so for our analyses we took all individuals from the island of Futuna, where Polynesian-related ancestry is highest (24). TreeMix and AdmixtureBayes do not allow sites with missing data, so for each SNP each population is required to have at least one genotype call. Since our model included three ancient populations, the number of sites available for these analyses was reduced to 76,284. For qpGraph it is possible to use an option which would maximize the number of sites for each computed statistic, but since we have modern and ancient data in the same dataset this would result in dramatically different sets for different comparisons. As this could bias the results, we therefore chose not to use this option. Mbti was used as an outgroup in all of the admixture graph analyses.

Statistical programming was done using the statistical program R v4.0.1 (https://www.R-project.org/). We used the tidyverse (98), data.table (https://CRAN.R-project.org/package=data.table), Hmisc (https://CRAN.R-project.org/package=Hmisc), and pheatmap (https://CRAN.R-project.org/package=pheatmap) packages.

**Data Availability.** All data used in this paper are in the main text or in the SI Appendix. The new data reported in this paper have been deposited in the European Nucleotide Archive, https://www.ebi.ac.uk/ena/browser/home (accession no. PRJEB40707).

**Acknowledgments.** We thank S. Nagel, B. Nickel, B. Schellbach, A. Schmidt, A. Weihmann, and M. Wunsch for performing the laboratory work; the Bioinformatics Group in the Department of Evolutionary Genetics at the Max Planck Institute for Evolutionary Anthropology for the initial processing of the sequencing data; P. Bellwood and R. Blust for comments on an earlier draft; and M. Hajdinjak for helpful discussion. Research at the Ritidian Site in Guam was performed under Special Use Permit 12518-1601 and Archaeological Resources Protection Act Permit 15GUA001, in cooperation with the Guam National Wildlife Refuge and the US Fish and Wildlife Service, and was funded by the Chiang Ching-kuo Foundation for International Scholarly Exchange (Grant RG021-P-10) and by the Australian Research Council (Grant DP150104458). Research was funded by the Max Planck Society. Disclaimer: PNAS policy is that reviewers of inaugural Articles should not have authored a paper with any author within the past 4 years. Although reviewer G.R.S. and author H.-c.H. were coauthors of an invited festschrift chapter that appeared in *Terra Australis* in early 2017, their active collaboration for this work occurred in 2014, and G.R.S. and H.-c.H. have not collaborated since then. G.R.S. was therefore allowed to review the manuscript.

19. A. Wollstein
31. M. G. Vilar
27. A. Wollstein
28. M. G. Vilar et al.
30. H. C. Hung et al.
30. H. C. Hung et al.
30. H. C. Hung et al.