Bridging Near and Remote Oceania: mtDNA and NRY Variation in the Solomon Islands

F. Delfin, S. Myles, Y. Choi, D. Hughes, R. Illek, M. van Oven, Brigitte Pakendorf, M. Kayser, M. Stoneking

Bridging Near and Remote Oceania: mtDNA and NRY Variation in the Solomon Islands

Frederick Delfin,§,[1] Sean Myles,†§,[1] Ying Choi,[2] David Hughes,‡,[1] Robert Illek,[3] Mannis van Oven,[2] Brigitte Pakendorf,[4] Manfred Kayser,*§,[2] and Mark Stoneking*§,[1]

[1]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
[2]Department of Forensic Molecular Biology, Erasmus MC University Medical Center Rotterdam, Rotterdam, the Netherlands
[3]Zoology Department, Naturhistorisches Museum, Vienna, Austria
[4]Max Planck Research Group on Comparative Population Linguistics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
†Present address: Department of Plant and Animal Sciences, Nova Scotia Agricultural College, Truro, Nova Scotia, Canada.
‡Present address: Group of Comparative Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai, China.
§These authors, respectively, contributed equally to this study.
*Corresponding author: E-mail: m.kayser@erasmusmc.nl; stoneking@eva.mpg.de.
Associate editor: Anne Stone

Abstract

Although genetic studies have contributed greatly to our understanding of the colonization of Near and Remote Oceania, important gaps still exist. One such gap is the Solomon Islands, which extend between Bougainville and Vanuatu, thereby bridging Near and Remote Oceania, and include both Austronesian-speaking and Papuan-speaking groups. Here, we describe patterns of mitochondrial DNA (mtDNA) and nonrecombining Y chromosome (NRY) variation in over 700 individuals from 18 populations in the Solomons, including 11 Austronesian-speaking groups, 3 Papuan-speaking groups, and 4 Polynesian Outliers (descended via back migration from Polynesia). We find evidence for ancient (pre-Lapita) colonization of the Solomons in old NRY paragroups as well as from M2-M353, which probably arose in the Solomons ~9,200 years ago and is the most frequent NRY haplogroup there. There are no consistent genetic differences between Austronesian-speaking and Papuan-speaking groups, suggesting extensive genetic contact between them. Santa Cruz, which is located in Remote Oceania, shows unusually low frequencies of mtDNA and NRY haplogroups of recent Asian ancestry. This is in apparent contradiction with expectations based on archaeological and linguistic evidence for an early (~3,200 years ago), direct colonization of Santa Cruz by Lapita people from the Bismarck Archipelago, via a migration that "leapfrogged" over the rest of the Solomons. Polynesian Outliers show dramatic island-specific founder events involving various NRY haplogroups. We also find that NRY, but not mtDNA, genetic distance is correlated with the geographic distance between Solomons groups and that historically attested spheres of cultural interaction are associated with the recent genetic structure of Solomons groups, as revealed by mtDNA HV1 sequence and Y-STR haplotype diversity. Our results fill an important lacuna in human genetic studies of Oceania and aid in understanding the colonization and genetic history of this region.

Key words: Solomon Islands, mitochondrial DNA, Y chromosome.

Introduction

Genetic, linguistic, and archaeological evidence are converging on a unified view of the peopling of Near and Remote Oceania (Kayser 2010; Kirch 2010). Near Oceania (consisting of Sahul, the combined New Guinea and Australia landmasses; the Bismarck Archipelago; and the Solomon Islands except for the Reef/Santa Cruz islands) was colonized at least 49 thousand years ago (KYA) (Groube et al. 1986; Torrence et al. 2004; Summerhayes et al. 2010). Contemporary populations in New Guinea and nearby islands that speak non-Austronesian (Papuan) languages, an extremely diverse and heterogeneous collection of some 700 languages, may be descended from this early migration. Although there was undoubtedly additional contact with Southeast Asia following the initial colonization of Near Oceania, the next major migration to Near Oceania involved Austronesian speakers (Kirch 2010), who spread from East Asia (most likely Taiwan) through the Philippines and Indonesia beginning around 4 KYA (Bellwood and Dizon 2005), arriving in the Bismarck Archipelago and along the northern coast of New Guinea by 3 KYA (Summerhayes 2001). There they came into contact with and admixed with Papuan-speaking groups before continuing to spread eastward through the Solomon Islands to Vanuatu and Fiji and from Fiji to the remainder of Remote Oceania (Kirch 1997; Green 2000). Fijians culturally and phenotypically have more in common with Near Oceanian groups than do other Remote Oceanian groups (Spriggs 1997), which may be explained by additional contact between Near Oceania and Fiji that did not extend further into Remote Oceania (Wollstein et al. 2010).

However, some researchers have questioned the extent to which Austronesian languages, archaeological evidence such as Lapita pottery, and gene flow from Asia after the initial settlement of Near Oceania all reflect the same single major wave of migration (Terrell 2009; Soares et al. 2011). Nonetheless, an admixed Asian/Near Oceanian ancestry of Remote Oceanians is strongly supported by analyses of mitochondrial DNA (mtDNA) and nonrecombining Y chromosome (NRY) data (Kayser et al. 2000, 2006; Kayser, Choi, et al. 2008) as well as genome-wide data (Friedlaender et al. 2008; Kayser, Lao, et al. 2008; Wollstein et al. 2010). The estimated date for this admixture, based on genome-wide SNP data, is about 3 KYA (Wollstein et al. 2010; Pugach et al. 2011), in excellent agreement with the archaeological and linguistic evidence (Spriggs 2003; Gray et al. 2009; Specht 2009). Moreover, mtDNA and NRY analyses indicate that this admixture was sex biased (Melton et al. 1995; Kayser et al. 2000; Su et al. 2000; Hurles et al. 2002): about 94% of Polynesian mtDNAs are of Asian ancestry, whereas about 66% of Polynesian Y chromosomes are of Near Oceania ancestry (Kayser et al. 2006). Although the mtDNA support for this sex-biased admixture hypothesis has recently been questioned (Soares et al. 2011), genome-wide SNP data do indicate significantly more Asian versus New Guinea ancestry for the X chromosome of Polynesians than for the autosomes (Wollstein et al. 2010), in agreement with the sex-biased admixture scenario. In addition, Papuan-speaking groups in New Guinea show higher frequencies of Asian mtDNA haplogroups than of Asian NRY haplogroups (Kayser, Choi, et al. 2008). After arriving in Near Oceania, Austronesian-speaking groups thus presumably incorporated more Papuan-speaking men than women, whereas Papuan-speaking groups incorporated more Austronesian-speaking women than men, possibly reflecting a matrilocal residence pattern for the Austronesian-speaking groups versus a patrilocal residence pattern for the Papuan-speaking groups (Hage and Marck 2003).

Although our knowledge concerning the colonization of Near and Remote Oceania has thus advanced considerably in recent years, there are still important lacunae with respect to genetic data, in particular the Solomon Islands, which bridge Near and Remote Oceania (fig. 1). The main Solomons chain was occupied at least 28 KYA (Wickler and Spriggs 1988) and there is no evidence that humans occupied the Santa Cruz/Reef Islands or any islands further east prior to the Austronesian expansion (Sheppard and Walter 2006). Although most groups in the Solomons speak Austronesian languages, there are a few Papuan-speaking groups, who thus represent the most eastward extension of Papuan languages. It is generally considered that the few Papuan-speaking groups in the Solomons are descended from an early migration there (Dunn et al. 2002, 2005), although it is also possible that some Papuan-speaking groups migrated to the Solomons at the same time as Austronesian speakers. In addition, some groups in the Solomons are closely related culturally and linguistically to Polynesians, and these "Polynesian Outliers" thus reflect recent back migration from Polynesia (Kirch 1984; Green 1995).

There are additional major cultural differences across the Solomons, as revealed in the archaeological and linguistic record, which might indicate different colonization events. Ceramics are known only from the northern and western main Solomons; they are largely absent from the central/southeast main Solomons archaeological record (Sheppard and Walter 2006). This situation is mirrored by a sharp linguistic division, known as the Tryon–Hackman line, that divides all Austronesian languages spoken on Isabel (except Bughotu, on the easternmost tip) and all islands to the north and west (including Bougainville) from those spoken on Guadalcanal and all islands to the south and east (Tryon and Hackman 1983; Ross 1989). Various explanations have been proposed for this boundary, involving multiple migrations of different Austronesian-speaking groups from either Near or Remote Oceania that interacted to different degrees with existing Papuan groups (Ross 1989; Sheppard and Walter 2006).

The Santa Cruz/Reef Islands provide evidence for further interactions between the Solomons and western parts of Near Oceania, specifically the Bismarck Archipelago north of Papua New Guinea. Separated from the main Solomons by about 400 km, the Santa Cruz/Reef Islands were first colonized by people with Lapita pottery about 3.2 KYA, and there is an extensive archaeological record of obsidian coming from the Bismarcks (over 2,000 km away) over a period of some 500 years (Sheppard and Walter 2006; Walter and Sheppard 2009). Lapita sites and obsidian from the Bismarcks do not appear elsewhere in the Solomons until much later (<2.7 KYA); thus it appears that the Santa Cruz/Reef Islands were colonized directly by, and maintained contact with, people from the Bismarcks who "leapfrogged" over the rest of the Solomons (Sheppard and Walter 2006). Linguistic evidence supports the leapfrog hypothesis; the Santa Cruz/Reef languages were originally considered to be Papuan with perhaps some Austronesian features (Wurm 1978) but are now considered a primary branch of Oceanic languages (Næss 2006; Ross and Næss 2007; Næss and Boerger 2008) that are not closely related to other Austronesian languages in the Solomons. Moreover, it has been suggested that the Tryon–Hackman line may reflect a migration of people from the Santa Cruz/Reef Islands back to the eastern and central Solomons (Sheppard and Walter 2006)

The history of the Solomon Islands thus potentially reflects multiple migrations from both the east and the west; how have these different migrations influenced the current gene pool in the Solomons? To what extent do Austronesian-speaking and Papuan-speaking groups differ genetically? Are the Polynesian Outliers indeed more closely related genetically to Polynesians than to other groups in the Solomons? Can we identify source populations in Near Oceania, Asia, or Remote Oceania that contributed genetically to the Solomons? Although there have been a few genetic studies of populations from the Solomon Islands (Friedlaender et al. 2002, 2007; Cox and Mirazón-Lahr 2006; Ricaut et al. 2010), these are quite limited in scope both in terms of genetic markers studied and geographic coverage. To date there has been no systematic investigation of genetic variation across the Solomon Islands that
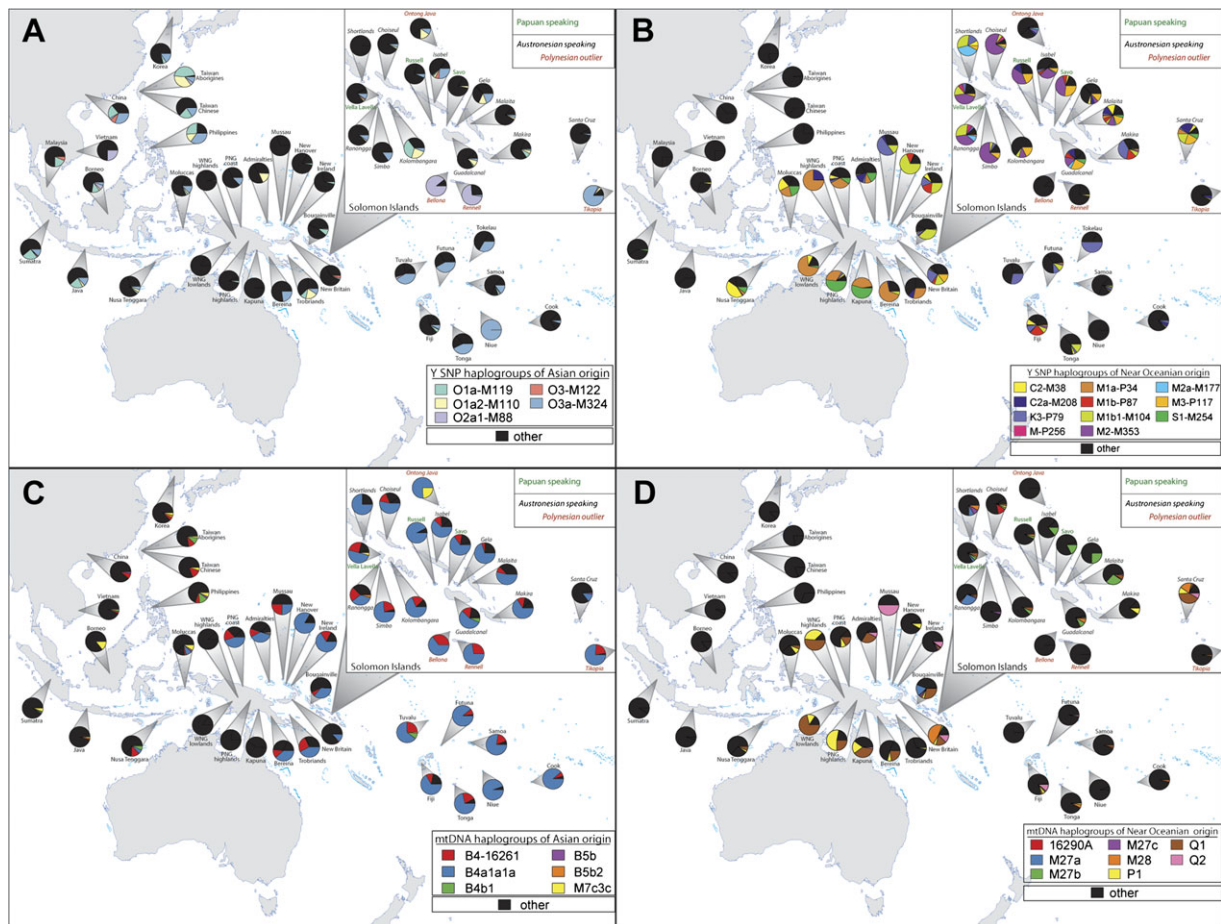
FIG. 1. Frequencies of NRY (*A,B*) and mtDNA (*C,D*) haplogroups of assumed Asian (*A,C*) or Near Oceanian (*B,D*) origin. Only those haplogroups found in the Solomon Islands are depicted.

can address the above questions concerning the genetic history of this region. To remedy this situation, we carried out extensive sampling across the Solomon Islands and report here the results of mtDNA and NRY analyses of more than 700 individuals from 18 island populations, including Papuan-speaking and Austronesian-speaking groups as well as Polynesian Outliers. We also update our previous data from nearly 1,700 individuals from East Asia, Southeast Asia, Near Oceania, and Remote Oceania (Kayser et al. 2006; Kayser, Choi, et al. 2008; Mona et al. 2009) with data for new NRY markers.

Materials and Methods

Samples

In August–September 2004, MS, BP, MK, RI, SM, and DH collected cheek swab samples from across the Solomon Islands. Self-described information on the birthplace, language, and ancestry of each donor was obtained, going back to at least the grandparental generation when known. The ancestry information was used to assign the island/language of origin of each sample separately for the mtDNA lineage (using maternal ancestry) and the NRY lineage (using paternal ancestry). After excluding known relatives, samples with maternal/paternal ancestry from outside the Solomon Islands, and

islands with sample sizes less than 10 individuals, there remained 703 samples for mtDNA analysis and 712 samples for NRY analysis. These come from 18 islands (fig. 1) and include 11 Austronesian-speaking groups (AN: from Choiseul, Gela, Guadalcanal, Isabel, Kolombangara, Makira, Malaita, Ranongga, Santa Cruz, the Shortlands, and Simbo), 3 Papuan-speaking groups (PAP: from Russell, Savo, and Vella Lavella), and 4 Polynesian Outliers (PO: from Bellona, Rennell, Tikopia, and Ontong Java). The samples from Choiseul, Malaita, the Shortlands, Ontong Java, and Tikopia were not collected from these islands but rather from individuals living in other locations who traced their ancestry to these islands. All donors gave written informed consent, and this study was approved by the Ministry for Education and Training and the Ministry of Health and Medical Services for the Solomon Islands; the Ethics Commission of the University of Leipzig Medical Faculty; and the Erasmus University Medical Center Rotterdam.

mtDNA Sequencing and NRY Genotyping

DNA was extracted from the cheek swabs as described previously (Miller et al. 1988). The first hypervariable segment (HV1) of the mtDNA control region was amplified and both strands sequenced as described previously (Nasidze et al. 2009). The sequences were aligned to the revised

Cambridge Reference Sequence (Andrews et al. 1999) and polymorphisms scored accordingly. In addition, the 9-bp deletion in the noncoding region between the COII and lysine tRNA genes was screened as described previously (Redd et al. 1995). mtDNA haplogroups were assigned based on Build 10 of the mtDNA tree available at http://www.phylotree.org (van Oven and Kayser 2009). HV1 sequences with a C to A transversion at position 16290, previously observed in Vanuatu (Sykes et al. 1995), were assigned as lineage 16290A, as in the absence of complete mtDNA genome sequence data the exact phylogenetic position of this lineage cannot be determined. The HV1 sequences have been deposited in GenBank (GenBank accession numbers JN017205–JN017907).

For the NRY, 36 Y-SNPs were genotyped by means of five separate multiplex assays using standard polymerase chain reaction and SNaPshot (Applied Biosystems, Foster, CA) single-base primer extension technologies: 1) M9, P132, P256, M214, M74 (also known as N12), M173 (also known as P241), P202 (phylogenetically equivalent to M230; Karafet et al. 2008), M254, and M226; 2) M9, P79, M353, M177 (also known as SRY9138), and P117; 3) M4, M5, P34, M83, P87, M104 (also known as P22), and M16; 4) M175, M119, M110, M268, M95, M88, M122, M324, M7, and M134; and 5) RPS4Y (also known as M130), M38, M208, P33, P54, and M217. Y-SNPs were ascertained with expectancy to be polymorphic in the Solomons based on previous work in Southeast Asia and Oceania (Su et al. 1999; Shi et al. 2005; Cox and Mirazón-Lahr 2006; Kayser et al. 2006; Scheinfeldt et al. 2006; Sengupta et al. 2006; Xue et al. 2006; Cox et al. 2007; Kayser, Choi, et al. 2008). Genotyping details will be provided elsewhere (van Oven, Choi and Kayser, in preparation) and are available from the authors upon request. NRY haplogroups were assigned according to the revised Y Chromosome Consortium tree (Karafet et al. 2008). Seventeen Y-STR loci (DYS19, DYS385a/b, DYS389I, DYS389II-I, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, and YGATAH4) were genotyped using the AmpFlSTR Yfiler kit (Applied Biosystems) following the manufacturer's recommendations.

Comparative NRY and mtDNA data were taken from our previous publications for populations from East Asia, Southeast Asia, New Guinea, and nearby islands, Fiji, and Polynesia (Kayser et al. 2006; Kayser, Choi, et al. 2008; Mona et al. 2009). NRY SNPs not genotyped before in these samples, but included here for the Solomon samples, were updated and are published here for the first time (with the exception of P256, which was typed only in the Nusa Tenggara reference samples). We also included published mtDNA and NRY data from the Bismarck Archipelago (Scheinfeldt et al. 2006; Friedlaender et al. 2007) in some analyses.

Data Analysis

The mtDNA HV1 sequences (nucleotide positions 16024–16365) from the Solomons and the reference data were aligned with MAFFT version 6.708 (Katoh et al. 2005)

and the aligned sequences were edited for gaps and missing data using BioEdit version 7.0 (Hall 1999). For comparison with the reference data set, the following seven Y-STR loci were used: DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, and DYS393; genotypes for DYS389II were obtained by subtracting the DYS389I allele. Diversity statistics, analysis of molecular variance (AMOVA), and Mantel tests were carried out with Arlequin version 3.11 (Excoffier et al. 2005). Mann–Whitney U tests, multidimensional scaling (MDS), and correspondence analyses (CA) were performed using the STATISTICA software package (StatSoft 2007).

Network analyses were performed using Network version 4.6 and Network Publisher version 1.1.0.7 (http://fluxus-engineering.com). All NRY networks made use of the seven Y-STR loci enumerated above except for the M2-M353* haplogroup network. As the M2-M353* haplogroup was observed mainly in the Solomon Islands, the NRY network made use of 15 of the 17 Y-STR loci typed (excluding DYS385, as it is a duplicated locus for which unambiguous assignment of alleles is not possible). NRY networks based on either 7 or 15 Y-STR loci were generated using a network weighting scheme based on Y-STR locus-specific mutation rates (Goedbloed et al. 2009). Initial analyses were performed using the Median Joining and Maximum Parsimon (MJ-MP) algorithms; however, this generated complex networks that were difficult to visualize and interpret, hence network reduction schemes were applied. To reduce network complexity, the Reduced-Median (RM) algorithm (Bandelt et al. 1995) was used followed by the MJ and MP algorithms. The RM-MJ-MP networks were compared with the initial MJ-MP networks to ensure that the simplified network was representative of the complex network. For the HV1 sequences within particular haplogroups, MJ networks were generated (Bandelt et al. 1999) followed by MP postprocessing (Polzina and Daneshmand 2003).

The BATWING program (Wilson et al. 2003) was used to estimate the time since the most recent common ancestor (TMRCA). For a BATWING run, data for the estimation was composed of Y-SNPs and seven Y-STR loci data from the entire data set (the Solomon Islands and the reference data) all of which constituted eight geographical population groups. The Y-SNP data were used as unique event polymorphisms (UEP sites) that constituted the known Y-SNP phylogeny, which was used to constrain the gene genealogy model. Y-STR data were used under a step-wise mutation model with locus-specific mutation rate priors for the Y-STRs (Goedbloed et al. 2009). Population structure was modeled to be a substructured population (of eight groups) and population size was modeled to be of an initial constant size followed by exponential growth. Three independent BATWING (Markov Chain Monte Carlo [MCMC]) runs were performed. Each run had a different random number seed, a total of 10^9 MCMC chains and a 10% burn-in period. Built-in BATWING functions were used to evaluate the MCMC run. The 95% posterior density was computed, and TMRCA point estimates (number of

generations) were converted to time in years using a generation time of 31 years per generation (Fenner 2005).

Results and Discussion

mtDNA and NRY Haplogroups in the Solomons

mtDNA HV1 sequences and 9-bp deletion status, and Y-SNP and Y-STR genotypes, were obtained for 703 and 712 samples, respectively. Based on the HV1 sequences and the 9-bp deletion information, the mtDNA haplogroup could be confidently assigned to 682 samples. All mtDNA analyses were carried out at both haplogroup level and haplotype (sequence) levels. A total of 14 mtDNA and 19 NRY haplogroups were observed (fig. 1; supplementary tables 1, 2, and 7, Supplementary Material online); the markers used to assign mtDNA and NRY haplogroups and their phylogenetic relationships are provided in supplementary figures 1 and 2 (Supplementary Material online). Each haplogroup was assigned a probable origin in Asia, Near Oceania (NO), or Remote Oceania (RO), based on its frequency distribution and associated HV1 sequence or Y-STR variation. It should be stressed that haplogroups of presumed Asian origin in Near or Remote Oceania were not necessarily brought there via the Austronesian expansion but rather reflect haplogroups that are not of indigenous NO origin and hence reflect some past migration from Asia.

Overall, the mtDNA haplogroups in the Solomons are predominantly of Asian origin, whereas the NRY haplogroups are predominantly of NO origin. The major mtDNA lineages in the Solomons belong to haplogroup B (fig. 1; supplementary table 2, Supplementary Material online); in particular, haplogroups B4-16261 and B4a1a1a (also known as the Polynesian Motif) account for 76% of the Solomons HV1 sequences. We assigned an Asian origin for all B haplogroups, in accordance with previous evidence (Trejaut et al. 2005; Kayser et al. 2006), although it should be noted that an origin in the Bismarcks of the "final" mutation that defines haplogroup B4a1a1a (at position 16247) has recently been proposed (Soares et al. 2011). The only other mtDNA haplogroup of Asian origin in the Solomons is M7c3c; previously known as M7c1c (Kivisild et al. 2002; Trejaut et al. 2005; Hill et al. 2007), at an overall frequency of 1.5% but restricted to Vella Lavella and Ontong Java. Several of the Near Oceanian mtDNA haplogroups (16290A, M27a, M27b, M27c, M28) are restricted to the Bismarcks or Bougainville and hence probably arose there (Friedlaender et al. 2007), whereas others (P1, Q1, Q2) are more widespread across Near Oceania (supplementary table 2, Supplementary Material online). Overall, the mtDNA haplogroups in the Solomons reflect the predominant haplogroups of both Asian and Near Oceanian origin in New Guinea and nearby islands and in particular indicate close connections between the Bismarcks and the Solomons.

In contrast to the mtDNA haplogroups, the NRY haplogroups in the Solomons exhibit several surprising features. Of particular note is haplogroup M2-M353*,
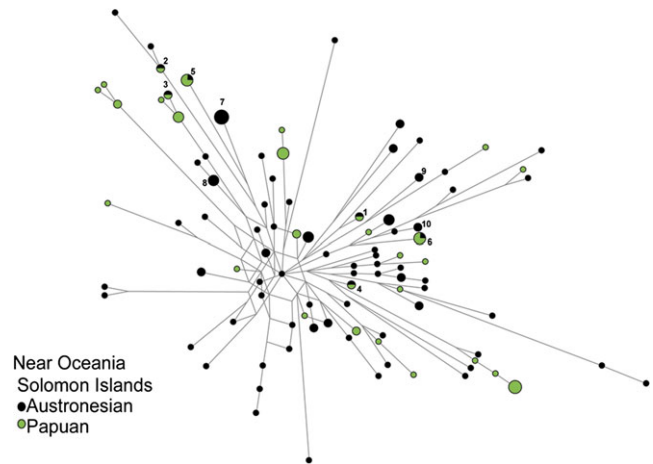

Fig. 2. Network of Y-STR haplotypes (based on 15 loci) for NRY haplogroup M2-M353*. Haplotypes shared by more than one population are numbered as in supplementary table 4 (Supplementary Material online).

previously found only in a few individuals from Vanuatu, Fiji, and Polynesia (Kayser et al. 2006; Karafet et al. 2010), whereas a sublineage, M2a-M177, was reported from a single Nasioi from Bougainville and a single individual from Malaita in the Solomons (Cox and Mirazón-Lahr 2006). Surprisingly, M2-M353* is the most frequent NRY haplogroup in the Solomons, at an overall frequency of 20% (fig. 1; supplementary table 1, Supplementary Material online). It is found in all groups from the Solomons except in the extreme west (Shortland Islands) and east (Santa Cruz) and is also absent from the Polynesian Outliers (fig. 1). The sublineage M2a-M177 occurs sporadically in the Solomons at an overall frequency of 2.4%, although it reaches a frequency of 38% in the Shortlands (supplementary table 1, Supplementary Material online). The frequency of M2-M353* is significantly higher in PAP (42%) than in AN (21%) groups ($P < 0.0001$), and the STR variation (mean number of pairwise differences between haplotypes), based on 15 loci, is slightly higher in the former (7.0 ± 3.4) than in the latter (6.5 ± 3.1) groups. Moreover, several Y-STR haplotypes (based on 15 loci) are shared between AN and PAP groups (fig. 2). The TMRCA for M2-M353* is about 9.2 KYA (supplementary table 3, Supplementary Material online), with a 95% confidence interval (CI) of 6.9–13.0 KYA, which thus predates the Austronesian arrival into the Solomons. The age of M2-M353*, along with the higher frequency and slightly higher diversity in Papuan-speaking groups, makes it likely that this haplogroup arose within a Papuan-speaking group in the Solomons and prior to the arrival of Austronesian speakers.

Another interesting feature is that three NRY haplogroups in the Solomons occur as paragroups (i.e., derived for a mutation defining the haplogroup but ancestral for all mutations tested that define lineages within the haplogroup): K-M9*, C-RPS4Y*, and M-P256*. K-M9* is widespread throughout Southeast Asia and Near and Remote Oceania (supplementary table 1, Supplementary Material online) and has a TMRCA of about 30.3 KYA (supplementary

549

table 3, Supplementary Material online). K-M9* STR haplotypes from the Solomons occur in several branches of the K-M9* network (supplementary fig. 3, Supplementary Material online). Disentangling the origin(s) of K-M9* lineages in the Solomons will therefore require the identification of further sublineages.

Haplogroup C-RPS4Y* (TMRCA: 23.3 KYA; supplementary fig. 3, Supplementary Material online) is found at similarly low frequencies in the Solomons (fig. 1; supplementary table 1, Supplementary Material online). Elsewhere, C-RPS4Y* occurs mostly in Southeast Asia and is practically absent from Near or Remote Oceania. Haplogroup C in Near and Remote Oceania is mostly represented by C2-M38* (TMRCA: 17 KYA; supplementary table 3, Supplementary Material online) and sublineages thereof (supplementary table 1 and fig. 2, Supplementary Material online). One possible explanation for C-RPS4Y* in the Solomons is that it was brought by the Austronesian expansion via Southeast Asia. However, the network for STR haplotypes of C-RPS4Y* does not support this explanation, as Solomons haplotypes are quite diverged from Southeast Asian haplotypes (supplementary fig. 4, Supplementary Material online), compared with the extensive haplotype sharing for other haplogroups of likely Austronesian origin among the Solomons, New Guinea, and Southeast Asia (e.g., O1a2-M110 and O3a-M324*, discussed in more detail below). These results instead suggest that C-RPS4Y* was present in New Guinea and the Solomons at an early time and that the M38 (and subsequent) mutations arose in New Guinea and spread, replacing C-RPS4Y* in New Guinea, while C-RPS4Y* was maintained in the Solomons.

This scenario of an ancient presence of C-RPS4Y* in New Guinea and the Solomons, followed by subsequent sublineage-defining mutations that replaced the paragroup in New Guinea but not in the Solomons, may be mirrored by M-P256*. The M clade is restricted to eastern Indonesia and Oceania and occurs almost exclusively as haplogroups M1, M2, and M3 and subhaplogroups thereof (Karafet et al. 2008). Previously M-P256* (TMRCA: 18 KYA; supplementary table 3, Supplementary Material online) was reported in a few individuals from Eastern Indonesia and Papua New Guinea (Karafet et al. 2010), although it is absent in our sample of over 300 East Indonesians (Mona et al. 2009) (supplementary table 1, Supplementary Material online). In the Solomons, M-P256* occurs sporadically in both AN and PAP groups at low frequency (supplementary table 1, Supplementary Material online). A possible explanation for these results is that the P256 mutation originated in either Eastern Indonesia or New Guinea, spread to the Solomon Islands, and (as with C-RPS4Y*) subsequent mutations occurred on the background of P256 that ultimately mostly replaced it elsewhere but not completely in the Solomons. However, K-M9* samples from Near Oceania have not been tested for the P256 mutation; thus some of these may turn out to be haplogroup M-P256*, in which case, the presence of M-P256* in the Solomons could reflect recent gene flow from New Guinea. Still, further

support for an ancient presence of M-P256* in the Solomons comes from haplogroup M2-M353*, which as discussed above arose in the Solomons ~9.2 KYA on the background of M-P256 and has since become the most frequent NRY haplogroup in the Solomons.

Several other haplogroups (C2a1-P33, O2a1-M88, and O3a-M324*) show interesting features in the PO groups and will be discussed further below. Some NRY haplogroups in the Solomons (in particular, K3-P79, M1b1-M104*, and M3-P117) probably arose in the Bismarcks (Scheinfeldt et al. 2006); thus both mtDNA and NRY haplogroups support an important role for the Bismarcks in the colonization of the Solomons. Two additional haplogroups, O1a-M119* and (especially) O1a2-M110, are likely to be of Taiwanese origin (Kayser, Choi, et al. 2008; Karafet et al. 2010) and hence support a link between Taiwan, Near Oceania, and the Solomons that extends into Remote Oceania (fig. 1; supplementary table 1, Supplementary Material online).

Austronesian-Speaking versus Papuan-Speaking Groups from the Solomons

The distributions of Asian versus NO mtDNA or NRY haplogroups do not show any consistent patterns between AN and PAP groups from the Solomons (fig. 1 and table 1). Overall, there is a much higher frequency of Asian mtDNA haplogroups in AN and PAP groups (72.6%) than of Asian NRY haplogroups (15.8%), in keeping with previous observations of a much higher contribution of Asian mtDNAs than of Asian Y chromosomes to Oceanic populations (Kayser et al. 2000, 2006; Kayser, Choi, et al. 2008). However, contrary to expectations based on the origin of the languages, the frequency of Asian mtDNA haplogroups is significantly higher in PAP groups than in AN groups (84.5% vs. 69.5%, $P = 0.002$, chi-square contingency test). This difference is almost entirely due to Santa Cruz, which has by far the lowest frequency of Asian mtDNA haplogroups (15.2%; next lowest is Ranongga with 62.5%). When Santa Cruz is removed from the analysis, the frequency of Asian mtDNA haplogroups is still higher in PAP groups (84.5%) than in AN groups (76.1%) but not significantly so ($P > 0.05$). By contrast, the frequency of Asian NRY haplogroups is significantly higher overall in AN groups (17.5%) than in PAP groups (8.5%; $P = 0.027$). However, this difference is driven by the high frequency of Asian NRY haplogroups in Kolombangara (65% overall, with most of this due to O1a-M119* and O1a2-M110 at a combined frequency of 59% in Kolombangara vs. an average of 4.5% elsewhere; supplementary table 1, Supplementary Material online), as the difference is no longer significant when Kolombangara is removed from the analysis.

Among AN groups of the Solomons, there is significant heterogeneity in the frequency of Asian versus NO mtDNA haplogroups ($P < 0.001$); removing two potential outliers (Santa Cruz with 15.2% Asian mtDNA haplogroups and Simbo with 95.7% Asian mtDNA haplogroups) reduces but does not eliminate the significant heterogeneity among AN groups ($P = 0.026$). There is also significant heterogeneity in the frequency of Asian versus NO NRY haplogroups

Table 1. Origin of mtDNA and NRY Haplogroups in Solomon Island Populations.

Population	mtDNA	%Asian	%NO
Austronesian	*423*	*69.5*	*30.5*
Choiseul	29	75.9	24.1
Gela	40	75.0	25.0
Guadalcanal	56	80.4	19.6
Isabel	44	84.1	15.9
Kolombangara	16	87.5	12.5
Makira	36	88.9	11.1
Malaita	73	63.0	37.0
Ranongga	48	62.5	37.5
Santa Cruz	46	15.2	84.8
Shortlands	12	75.0	25.0
Simbo	23	95.7	4.3
Papuan	*110*	*84.5*	*15.5*
Russell	26	92.3	7.7
Savo	38	81.6	18.4
Vella Lavella	46	82.6	17.4
Polynesian Outliers	*149*	*99.3*	*0.7*
Bellona	34	100.0	0.0
Ontong Java	33	100.0	0.0
Rennell	44	100.0	0.0
Tikopia	38	97.4	2.6
Total	**682**	**78.4**	**21.6**

Population	NRY	%Asian	%NO	%RO	%Unknown
Austronesian	*456*	*17.5*	*70.4*	*3.1*	*9.0*
Choiseul	36	8.3	91.7	0.0	0.0
Gela	47	31.9	29.8	2.1	36.2
Guadalcanal	53	15.1	73.6	0.0	11.3
Isabel	53	37.7	47.2	3.8	11.3
Kolombangara	17	64.7	35.3	0.0	0.0
Makira	39	15.4	66.7	0.0	17.9
Malaita	98	8.2	78.6	8.2	5.1
Ranongga	44	11.4	88.6	0.0	0.0
Santa Cruz	36	2.8	88.9	8.3	0.0
Shortlands	13	0.0	100.0	0.0	0.0
Simbo	20	15.0	85.0	0.0	0.0
Papuan	*106*	*8.5*	*76.4*	*0.9*	*14.2*
Russell	32	6.3	71.9	0.0	21.9
Savo	30	3.3	80.0	0.0	16.7
Vella Lavella	44	13.6	77.3	2.3	6.8
Polynesian Outliers	*150*	*71.3*	*6.0*	*21.3*	*1.3*
Bellona	28	89.3	0.0	10.7	0.0
Ontong Java	32	25.0	12.5	56.3	6.3
Rennell	38	73.7	2.6	23.7	0.0
Tikopia	52	88.5	7.7	3.8	0.0
Total	**712**	**27.5**	**57.7**	**6.6**	**8.1**

NOTE.—The origin of each haplogroup was assigned as Asian or Near Oceanian (NO) based on the frequency distribution and associated diversity in either HV1 sequences for mtDNA haplogroups or Y-STR haplotypes for NRY haplogroups (assigned origins are indicated in supplementary tables 1 and 2, Supplementary Material online). NRY haplogroup C2a1-P33 is of Remote Oceanian (RO) origin, while the specific origin of haplogroups C-RPS4Y* and K-M9* cannot be determined with certainty and hence are designated as "Unknown" haplogroups. Group labels and computed values for the three groupings of Solomon Island populations are bold and italicized.

($P < 0.001$), with the frequency of Asian NRY haplogroups ranging from 0% (in the Shortlands) to 64.7% (in Kolombangara). By contrast, the frequencies of Asian versus NO haplogroups do not differ significantly among the Solomon PAP groups with respect to either mtDNA or the NRY ($P > 0.05$ for both), but this difference between AN and PAP groups may simply reflect the smaller number of PAP groups.

None of the diversity measures (table 2: haplogroup diversity, HV1 sequence diversity or Y-STR haplotype diver-

Table 2. Diversity Statistics for Solomon Island Populations for mtDNA and the NRY.

Population	Sample Size	nHg	HgD	SE	nHT	HtD	SE	MPD	SE
mtDNA									
Austronesian	*423*	*13*	*0.66*	*0.02*	*81*	*0.78*	*0.02*	*4.71*	*2.31*
Choiseul	29	6	0.63	0.09	9	0.57	0.11	3.19	1.70
Gela	40	3	0.46	0.07	9	0.70	0.06	3.35	1.76
Guadalcanal	56	7	0.64	0.06	22	0.73	0.07	3.44	1.78
Isabel	44	3	0.47	0.08	11	0.64	0.08	2.44	1.35
Kolombangara	16	4	0.52	0.13	6	0.84	0.05	3.28	1.78
Makira	36	3	0.42	0.09	4	0.46	0.09	1.52	0.94
Malaita	73	6	0.63	0.04	19	0.83	0.03	5.04	2.48
Ranongga	48	8	0.75	0.03	14	0.82	0.05	5.41	2.65
Santa Cruz	46	8	0.77	0.02	17	0.91	0.02	7.56	3.59
Shortlands	12	4	0.45	0.17	6	0.76	0.12	3.77	2.04
Simbo	23	3	0.42	0.10	5	0.56	0.11	1.15	0.77
Papuan	*110*	*8*	*0.52*	*0.05*	*21*	*0.63*	*0.05*	*3.03*	*1.59*
Russell	26	3	0.15	0.09	7	0.72	0.08	2.27	1.29
Savo	38	3	0.49	0.08	6	0.44	0.09	2.37	1.32
Vella Lavella	46	7	0.66	0.06	12	0.70	0.07	3.85	1.97
Polynesian Outliers	*149*	*4*	*0.43*	*0.04*	*12*	*0.43*	*0.05*	*1.11*	*0.73*
Bellona	34	2	0.47	0.05	5	0.55	0.08	0.62	0.50
Ontong Java	33	2	0.38	0.08	4	0.58	0.07	2.55	1.41
Rennell	44	2	0.41	0.06	6	0.33	0.09	0.35	0.35
Tikopia	38	3	0.38	0.08	5	0.20	0.09	0.73	0.56
NRY									
Austronesian	*456*	*18*	*0.91*	*0.01*	*361*	*0.999*	*0.000*	*9.76*	*4.48*
Choiseul	36	7	0.48	0.10	29	0.983	0.013	7.93	3.77
Gela	47	9	0.87	0.02	43	0.995	0.006	9.36	4.37
Guadalcanal	53	12	0.89	0.02	48	0.996	0.005	10.06	4.67
Isabel	53	13	0.84	0.03	43	0.991	0.006	9.18	4.29
Kolombangara	17	6	0.82	0.06	12	0.956	0.033	8.39	4.09
Makira	39	9	0.80	0.05	31	0.987	0.009	8.99	4.23
Malaita	98	16	0.92	0.01	81	0.996	0.002	9.95	4.59
Ranongga	44	8	0.77	0.04	37	0.992	0.007	8.86	4.17
Santa Cruz	36	9	0.86	0.03	27	0.975	0.016	9.13	4.30
Shortlands	13	5	0.78	0.08	11	0.962	0.050	8.88	4.38
Simbo	20	5	0.57	0.12	18	0.990	0.019	8.16	3.95
Papuan	*106*	*14*	*0.78*	*0.03*	*75*	*0.992*	*0.003*	*9.40*	*4.35*
Russell	32	6	0.78	0.05	17	0.952	0.019	9.46	4.46
Savo	30	7	0.74	0.06	22	0.979	0.013	9.00	4.26
Vella Lavella	44	10	0.78	0.05	37	0.987	0.009	9.05	4.25
Polynesian Outliers	*150*	*9*	*0.73*	*0.02*	*57*	*0.917*	*0.017*	*7.65*	*3.59*
Bellona	28	2	0.20	0.09	6	0.489	0.112	2.48	1.38
Ontong Java	32	7	0.66	0.08	23	0.974	0.015	8.00	3.82
Rennell	38	4	0.45	0.08	16	0.774	0.071	5.39	2.66
Tikopia	52	6	0.32	0.08	16	0.856	0.035	5.03	2.48

NOTE.—nHG, number of haplogroups; HgD, haplogroup diversity; nHt, number of different HV1 sequences (for mtDNA) or Y-STR haplotypes (for the NRY); HtD, haplotype diversity; MPD, mean number of pairwise differences; SE, standard error. Group labels and computed values for the three groupings of Solomon Island populations are bold and italicized.

sity, and MPD) differ significantly between AN and PAP groups for either the mtDNA or the NRY (Mann–Whitney U tests, $P > 0.05$ for all comparisons). The PAP group from Russell has an exceptionally low mtDNA haplogroup diversity value (0.15, next lowest is 0.42 from Makira and Simbo), but otherwise there are no obvious outliers among AN and PAP groups with respect to the diversity measures.

The MDS plot based on Φst distances for mtDNA HV1 sequences, and the CA plot based on NRY haplogroup frequencies, do not show any consistent groupings of AN versus PAP groups (fig. 3). For the mtDNA MDS plot (fig. 3A), Santa Cruz is a clear outlier, whereas the PAP
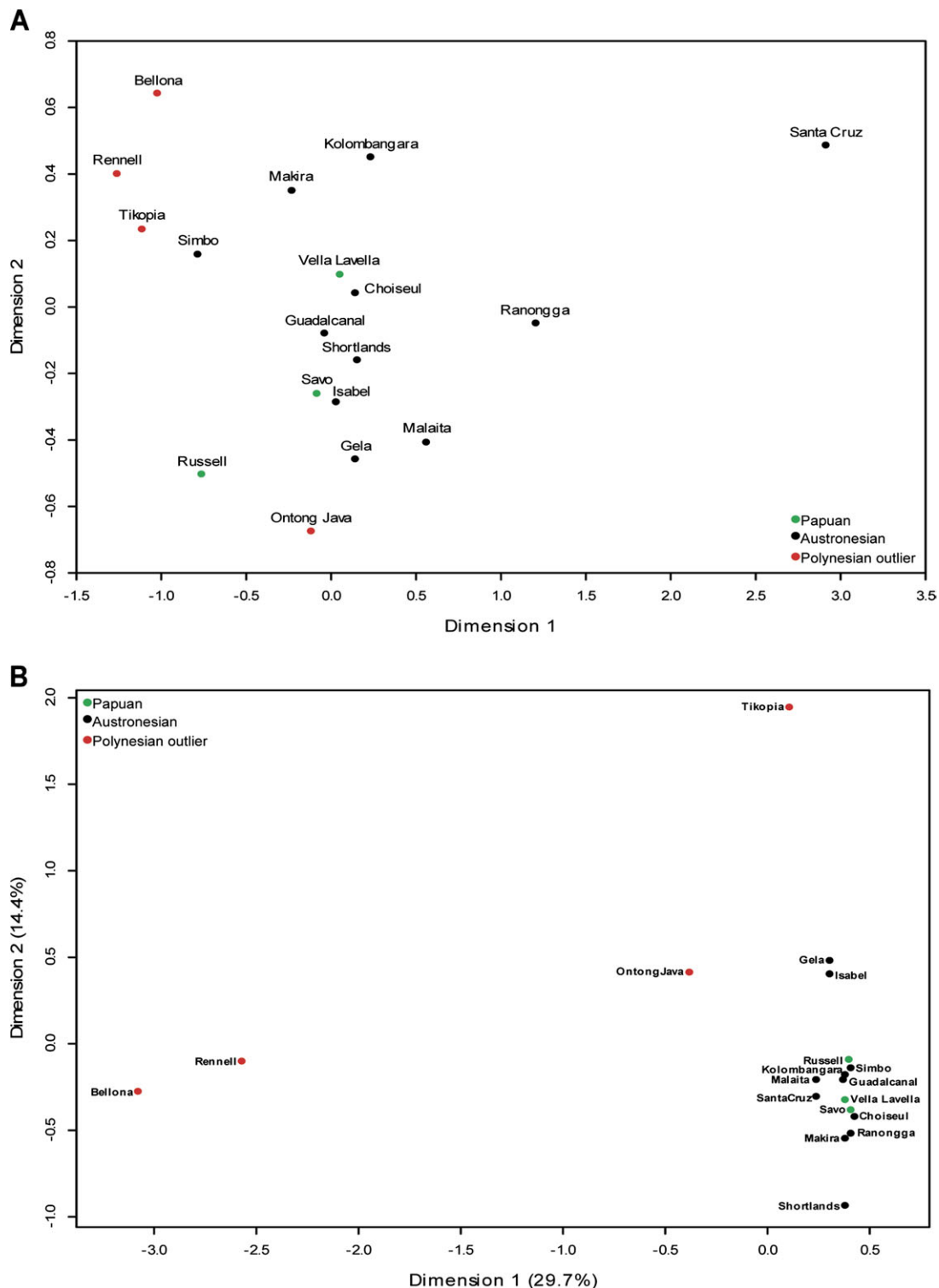
FIG. 3. (*A*) MDS plot based on Φst distances calculated from HV1 sequences from Solomons groups. The stress value is 0.088. (*B*) CA plot for Solomons groups based on NRY haplogroup frequencies.

groups fall within a cluster of the remaining AN groups plus the PO groups. Santa Cruz is an outlier because of the high frequency of NO mtDNA haplogroups (especially Q1; supplementary table 2, Supplementary Material online). A CA plot based on mtDNA haplogroup frequencies indicates Santa Cruz as well as Ranongga as outliers (supplementary fig. 5, Supplementary Material online); Ranongga is an out-

lier presumably because of the high frequency of hap-logroup M27a (29% in Ranongga vs. 0–8% elsewhere; supplementary table 2, Supplementary Material online). For the CA plot based on NRY haplogroup frequencies (fig. 3B), the PO groups are distinguished from the other groups, whereas the PAP groups fall in a cluster with all of the AN groups except Gela, Isabel, and the Shortlands.

Table 3. AMOVA Based on mtDNA and NRY Haplogroups for Various Groupings of Solomon Island Populations.

Grouping	mtDNA Haplogroups			mtDNA Haplotype		
	Among Groups	Among Populations Within Groups	Within Populations	Among Groups	Among Populations Within Groups	Within Populations
None	—	11.1*	88.9	—	16.4*	83.6
AN versus PAP versus PO	0.4	10.9*	88.7	1.4	15.5*	83.1
AN versus PAP	−1.7	11.8*	89.9	0.0	15.9*	86.1
AN versus PAP (omitting Santa Cruz)	−1.6	6.7*	94.8	0.0	6.7*	94.6
PO	—	6.0*	94.0	—	13.3*	86.7
Tryon–Hackman line (omitting Santa Cruz and PO)	0.5	5.7*	93.8	0.1	7.2*	92.7
Interaction areas—Isabel in West group	4.4*	6.2*	89.5	9.3*	4.2*	86.5
Interaction areas—Isabel in Central group	5.0*	5.5*	89.4	9.5*	3.9*	86.6

Grouping	NRY Haplogroups			NRY Haplotype		
	Among Groups	Among Populations Within Groups	Within Populations	Among Groups	Among Populations Within Groups	Within Populations
None	—	21.6*	78.4	—	15.8*	84.2
AN versus PAP versus PO	7.6*	16.7*	75.8	7.5*	11.1*	81.5
AN versus PAP	0.3	10.1*	89.6	0.7	8.9*	90.5
AN versus PAP (omitting Santa Cruz)	−0.1	10.1*	90.1	0.1	8.6*	91.4
PO	—	52.8*	47.2	—	27.2*	72.3
Tryon–Hackman line (omitting Santa Cruz and PO)	4.5*	8.6*	86.9	6.3*	5.8*	87.9
Interaction areas—Isabel in West group	5.8*	7.3*	86.9	7.3*	6.0*	86.6
Interaction areas—Isabel in Central group	6.0*	7.0*	87.0	6.9*	6.2*	86.8

NOTE.—For explanation of various groups, see text.
*$P < 0.05$.

The AMOVA (table 3) furthermore does not support any genetic distinction between AN and PAP groups. When considering all Solomon Island groups, the within-population component is higher for mtDNA (89%) than for NRY haplogroups (78%), as is commonly observed in human populations (Seielstad et al. 1998). When comparing AN versus PAP groups, for both mtDNA and NRY the among-group component is nonsignificant and negligible compared with the among-populations-within-groups component, and this is true whether or not Santa Cruz is included (table 3).

Overall, none of these analyses indicate any strong/significant genetic differences between AN and PAP groups in the Solomon Islands. Although in principal this finding could be explained by a common origin of AN and PAP groups, with little genetic contribution from the Austronesians, recent gene flow between AN and PAP groups seems more likely. For example, NRY haplogroup M2-M353 probably arose in a Solomons PAP group about 9.2 KYA (as discussed previously), and yet, there is extensive haplotype sharing between AN and PAP groups (fig. 2). A similarly fine-scaled analysis of AN and PAP groups in the Bismarcks also found evidence for significant genetic exchange between different language groups (Hunley et al. 2008); the large differences between AN and PAP languages are therefore not a barrier to genetic exchange.

However, patterns of Y-STR haplotype sharing (supplementary table 4, Supplementary Material online) and HV1 sequence sharing (supplementary table 5, Supplementary Material online) among and between AN and PAP groups do indicate an interesting dichotomy (fig. 4). Of 32 haplotypes (based on 15 Y-STR loci and on the background of specific haplogroups) shared by two or more Solomons groups, 18 occur in both AN and PAP groups (fig. 4), and the overall frequency of shared haplotypes exclusively within AN or PAP groups versus between AN and PAP groups does not differ significantly from that expected by chance ($P = 0.22$). But only 7 of 20 HV1 sequences shared by two or more Solomons groups occur in both AN and PAP groups (fig. 4), and the overall frequency of HV1 sequence sharing exclusively within AN or PAP groups is significantly higher than expected by chance ($P < 0.001$).

Thus, at the level of recent migration that is revealed by shared haplotypes/sequences, there is a significant tendency for females, but not males, to move to groups speaking languages belonging to the same classification. This need not mean that language itself is a barrier to female migration, as PAP and AN groups may differ in other cultural characteristics that are impacting migration. We come back to this issue below, when we consider the

NRY Haplogroup	Ht #	N	AN	PAP
C2a-M208*	1	4	■	
	2	4	■	
K-M9*	1	4	■	
	2	2	■	
K3-P79	1	2	■	
M-P256*	1	3	■	
	2	2	■	
M1b1-M104*	1	2		■
	2	3		■
M2-M353*	1	2	■	
	2	2	■	
	3	2	■	
	4	2	■	
	5	4	■	
	6	4	■	
	7	6	■	
	8	3	■	
	9	2	■	
	10	2	■	
M2a-M177	1	2		■
M3-P117	1	6	■	
	2	3	■	
	3	2	■	
	4	2	■	
	5	2	■	
	6	3	■	
	7	2		■
O1a2-M110	1	2	■	
O3a-M324*	1	3	■	
	2	2	■	
	3	3	■	
	4	5	■	

MtDNA Haplogroup	Ht #	N	AN	PAP
B4-16261	1	48	■	
	7	2	■	
B4a1a1a	1	207	■	■
	2	6	■	
	4	4	■	
	5	4	■	
	7	9	■	
	8	16	■	
	16	2	■	
	17	2	■	
	18	2	■	
	19	2	■	
	20	3	■	
M27b	1	45		■
	2	2		■
	3	3		■
M28	1	2	■	■
P1	1	2		■
	2	2	■	■
	4	3		■

Fig. 4. Patterns of Y-STR haplotype (15 Y-STR loci) and mtDNA HV1 sequence sharing, on the background of specific haplogroups, among Austronesian-speaking (AN) and Papuan-speaking (PAP) groups. Ht#, haplotype number according to supplementary tables 4 (Y-STR haplotypes) and 5 (HV1 sequences) (Supplementary Material online); N, sample size. Solid blocks indicate the presence of that haplotype in AN or PAP groups, respectively.

Tryon–Hackman line and other cultural groupings of Solomon populations.

Santa Cruz

Santa Cruz deserves special mention, both because of its geographic location in Remote Oceania (all of the other Solomon groups in this study come from Near Oceania) and because based on archaeological and linguistic evidence it has been suggested that Santa Cruz was colonized directly from the Bismarcks, leapfrogging over the rest of the Solomons (Sheppard and Walter 2006; Næss and Boerger 2008). Santa Cruz has a very different mtDNA composition compared with other groups from the Solomons (fig. 1; supplementary table 2, Supplementary Material online): Only 15% of the mtDNAs in Santa Cruz are B4 or B5 lineages (of putative Asian origin) versus an average of 80% in the other Solomon groups. This is significantly lower (contingency test, $P < 0.03$) than observed in a previous study of mtDNA variation in Santa Cruz (Friedlaender et al. 2002), which found B4/B5 lineages at a frequency of 33%. This difference may reflect geographic variation within Santa Cruz, as the previous study sampled individuals from along the western shore of Graciosa Bay (Friedlaender et al. 2002), which probably has experienced more contact with groups from elsewhere; our study includes villages primarily from the southwestern interior and coast and northeastern coast.

Correspondingly, there is an extraordinarily high frequency of mtDNA sequences of NO origin of 85% in Santa Cruz (supplementary table 2, Supplementary Material online); the group with the next highest frequency of NO

mtDNA sequences is Ranongga, with 37.5%. The high frequency of NO mtDNA haplogroups in Santa Cruz is unlikely to represent a simple founder event or bottleneck, as several NO haplogroups occur at relatively high frequency in Santa Cruz, including M28, P1, Q1, and Q2, and moreover, there are several different HV1 sequence types within each haplogroup. With respect to NRY haplogroups, Santa Cruz is also noteworthy for having the second-lowest frequency of Asian haplogroups of all Solomon groups studied (2.8%; lowest is the Shortlands with 0% Asian NRY haplogroups, albeit with a sample size of just 13 individuals; table 1). Thus, Santa Cruz stands out by having unusually low frequencies of both mtDNA and NRY haplogroups of Asian origin.

At the time that genetic sampling was undertaken, the Reef–Santa Cruz languages were considered to be Papuan languages with some Austronesian features due to contact influence (Wurm 1978) or perhaps even mixed Austronesian–Papuan languages (Smith 1995). However, subsequent work has shown that the Reef–Santa Cruz languages are instead Austronesian languages with no evidence of any Papuan contact influence (Næss 2006; Ross and Næss 2007; Næss and Boerger 2008). The Reef–Santa Cruz languages form a distinct primary subgroup of Oceanic languages (Næss and Boerger 2008; Gray et al. 2009) that is not closely related to other Austronesian languages of the Solomons and hence are thought to represent a separate earlier migration of Austronesian speakers, probably from the Bismarcks, that bypassed the main Solomons chain (Næss and Boerger 2008). Archaeologically, Santa Cruz is the only place in the Solomons with early Lapita sites, indicating initial settlement over 3 KYA (Green 1991). Moreover, it is unique in the Solomons for the amount of obsidian from New Britain, indicating continuous long-distance contact between the Reef–Santa Cruz Islands and the Bismarcks (Sheppard and Walter 2006).

The archaeological and linguistic evidence would thus suggest that an Austronesian-speaking group with Lapita pottery arrived in Santa Cruz very soon after such people arrived in the Bismarcks from island Southeast Asia, which presumably would not leave time for much genetic admixture between these people and the Papuan-speaking groups already present in the Bismarcks. Thus, one would expect that Santa Cruz should exhibit a high frequency of the Asian mtDNA and NRY haplogroups that were presumably characteristic of the earliest Austronesian-speaking groups. And yet, we find precisely the opposite pattern namely Santa Cruz is significantly depauperate in mtDNA and NRY haplogroups of Asian origin compared with other Solomon groups.

It is not clear how to resolve this quandary. One possible explanation is that there was pre-Lapita colonization of Santa Cruz and that subsequent migrations introduced the Austronesian language and Lapita pottery without having much genetic impact. The open ocean voyage required to reach Santa Cruz from the main Solomon chain is comparable to that between New Guinea and Manus Island, which was colonized about 18 KYA (Ambrose 2002; Specht

2005), so a similarly early colonization of Santa Cruz is not inconceivable. And, as with Santa Cruz, only Austronesian languages are spoken on Manus today, while there are appreciable frequencies of NO mtDNA (~40%) and NRY (~82%) haplogroups (Kayser, Choi, et al. 2008). However, archaeological investigations have been fairly extensive on Santa Cruz (Walter and Sheppard 2009) and have yet to find any evidence of pre-Lapita colonization.

Another possible explanation would be a rapid shift in the language and culture of a Papuan-speaking group in the Bismarcks to an Austronesian language and Lapita culture and that these people then went on to colonize Santa Cruz. A rapid pace of interaction between people with the Lapita culture and other people has been suggested previously (Spriggs 1995; Kirch 1997). Moreover, as in the Solomons, AN and PAP groups in the Bismarcks do not show significant differences in either mtDNA or NRY haplogroup distributions (Scheinfeldt et al. 2006; Friedlaender et al. 2007) nor in autosomal microsatellite loci (Friedlaender et al. 2008). These observations suggest that there has been substantial contact between PAP and AN groups in the Bismarcks, and depending on how early and rapid this contact was, could support this explanation.

Alternatively, it may be that the initial colonization of Santa Cruz was indeed by Austronesian speakers with Lapita pottery and a high frequency of Asian mtDNA and NRY haplogroups, but that subsequent contact with the Bismarcks, as documented in the obsidian record (Spriggs 1997; Sheppard and Walter 2006), led to a gradual replacement of the original gene pool with NO mtDNA and NRY haplogroups. Archaeological evidence indicates that following initial colonization, there was extensive ongoing trade between Santa Cruz and the Bismarcks for at least 500 years (Spriggs 1997), raising the possibility of extensive gene flow as well. In support of this explanation, the major mtDNA and NRY haplogroups in Santa Cruz also occur in the Bismarcks (supplementary tables 1 and 2, Supplementary Material online), although this is true of the main Solomons as well.

Another potential explanation, namely extensive gene flow between Santa Cruz and the main Solomons (Spriggs 1997), is ruled out by the complete absence of NRY haplogroup M2-M353*, the most frequent NRY haplogroup in the main Solomons, on Santa Cruz. We also note that while we find the archaeological, linguistic, and genetic evidence for associating Austronesian languages and Lapita culture with a major migration from Asia to Oceania persuasive (Bellwood and Dizon 2005; Kayser 2010; Kirch 2010), others do not (Terrell 2009; Donohue and Denham 2010; Soares et al. 2011); the apparent lack of fit between the genetic data versus the archaeological/linguistic evidence for Santa Cruz could be argued to support this latter view. In sum, the colonization history of Santa Cruz remains puzzling; analyses of genome-wide data should provide further insights.

Polynesian Outliers

Our sampling in the Solomons included four groups classified as Polynesian Outliers (PO), consisting of Rennell and Bellona, Tikopia, and Ontong Java (fig. 1). The people from these islands share cultural traits with Polynesians, speak Polynesian languages, and consider in their oral tradition that they came from Polynesia (Kirch 1984; Green 1995). Hence, these groups are thought to represent back migrations to the Solomons from Polynesia.

Overall, the PO have high frequencies of three NRY haplogroups that account for 88% of their Y chromosomes: C2a1-P33; O2a1-M88; and O3a-M324* (supplementary table 1, Supplementary Material online). However, each haplogroup occurs at high frequency in a different PO group, indicating island-specific founder events in their paternal history. Haplogroup C2a1-P33 has been found before only in Polynesia and has therefore been suggested to be of Polynesian origin (Cox et al. 2007). However, we also find this haplogroup in Fiji (supplementary table 1, Supplementary Material online) and the TMRCA estimate is ~4.3 KYA (95% CI: 3.1–5.7 KYA; supplementary table 3, Supplementary Material online), which appears to be too early for an origin in Polynesia. This haplogroup therefore may have arisen elsewhere in Remote Oceania, before the colonization of Polynesia. A Remote Oceanian origin of C2a1-P33 in the PO is supported by network analysis (fig. 5A), as Y-STR haplotypes found in the PO are shared with or are closely related to those from Remote Oceania, whereas in addition there are numerous haplotypes in Remote Oceania that are not shared with the PO. C2a1-P33 is at much higher frequency in Ontong Java (56%) than in the other PO (4–24%; supplementary table 1, Supplementary Material online) and also occurs at low frequency in some AN and PAP groups in the Solomons, which is discussed in more detail below.

Haplogroup O2a1-M88 has a peculiar distribution in that it is found at low frequency in East and Southeast Asia and Remote Oceania but nowhere in Near Oceania (supplementary table 1, Supplementary Material online). Among the PO, it occurs at very high frequency in Rennell (71%) and Bellona (89%), consistent with a close shared history of these two groups (Elbert and Monberg 1965) but in only one Tikopian and not at all in Ontong Java. The network of O2a1-M88 haplotypes (fig. 5B) indicates a strong founder effect, with 85% of the individuals from Rennell/Bellona with this haplogroup sharing one haplotype. All individuals from Rennell/Bellona who are not haplogroup O2a1-M88 are haplogroup C2a1-P33, except for one individual from Rennell who is haplogroup O3a-M324*, and one who is haplogroup C2-M38* (supplementary table 1, Supplementary Material online). Linguistic and archaeological evidence, as well as oral traditions, suggest that Renell and Bellona were occupied prior to the arrival of people from Polynesia about AD 1400 (Elbert 1962; Elbert and Monberg 1965; Poulsen 1972). However, there is no indication of any significant genetic contribution of the pre-Polynesians to the current population of Rennell and Bellona, as all HV1 sequences and NRY haplotypes shared between Rennell/Bellona and other populations occur in Fiji and/or Polynesia, with the exception of one B4-16261 HV1 sequence from Rennell (supplementary table 5, Supplementary Material online).
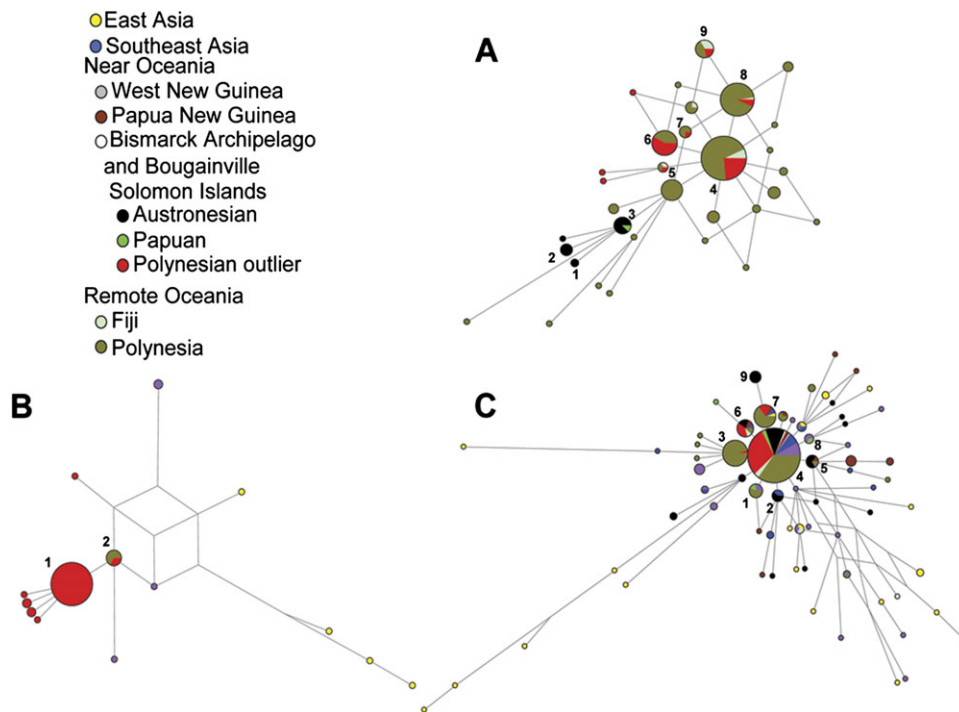
FIG. 5. Networks of Y-STR haplotypes (based on 7 loci) for NRY haplogroups found at high frequency in the Polynesian Outliers. (*A*) Haplogroup C2a1-P33. (*B*) Haplogroup O2a1-M88. (*C*) Haplogroup O3a-M324*. Haplotypes shared by more than one population are numbered as in supplementary table 6 (Supplementary Material online).

The complete absence of O2a1-M88 elsewhere in Near Oceania is puzzling, given that it likely arose in East Asia about 8 KYA (supplementary table 3, Supplementary Material online) and then spread to Remote Oceania via Near Oceania. As none of the O2 sublineage markers were typed in the largest study of Y chromosome variation from the Bismarcks and Bougainville (Scheinfeldt et al. 2006), it is possible that their O-M175* Y chromosomes are in reality haplogroup O2a1-M88, but this would only be a maximum of 10 individuals (of 641). Given the low frequency of O2a1-M88 in East and Southeast Asia (supplementary table 1, Supplementary Material online), it seems likely that it was a low-frequency haplogroup brought to Remote Oceania via the Austronesian expansion and was either subsequently lost via drift in Near Oceania or still exists at low frequency in Near Oceania in groups that have not yet been sampled. In our RO groups, O2a1-M88 is found only in Samoa and Tuvalu, at low frequency.

Haplogroup O3a-M324* occurs at much higher frequency in Tikopians (83%) than in the other PO (0–9%). This haplogroup is found throughout East and Southeast Asia and Near and Remote Oceania (supplementary table 1, Supplementary Material online) and has a TMRCA of ~11 KYA (supplementary table 3, Supplementary Material online). The network of O3a-M324* haplotypes (fig. 5C) shows a single major haplotype shared from East Asia to Remote Oceania, indicating a recent and rapid spread of this haplogroup. Although O3a-M324* does occur in other Solomon groups, all four haplotypes in PO are also found in Polynesians, but only two of these are also found in other Solomon groups (fig. 5C; supplementary table 6,

Supplementary Material online). Thus, a Polynesian source for O3a-M324* in the PO is indicated.

Of the remaining NRY haplogroups in the PO, all are also found in Remote Oceania, with the exception of a single individual from Ontong Java with haplogroup M2a-M177. In sum, with respect to NRY variation, the PO reflect a Polynesian origin with strong founder events, as different haplogroups predominate in different PO groups: C2a1-P33 on Ontong Java, O2a1-M88 on Rennell and Bellona, and O3a-M324* on Tikopia.

The mtDNA results similarly indicate a restricted number of founders for the PO, as 21% of the HV1 sequences are haplogroup B4-16261 and 72% are haplogroup B4a1a1a (supplementary table 2, Supplementary Material online). However, these values do not differ significantly from those found in Remote Oceania (14% B4-16261 and 78% B4a1a1a). And while there is significant heterogeneity in mtDNA haplogroup frequencies across the PO, this is due entirely to Ontong Java, as the other three PO do not differ significantly in mtDNA haplogroup frequencies. Ontong Java lacks haplogroup B4-16261 completely and is the only PO with haplogroup M7c3c, at a frequency of 24% (supplementary table 2, Supplementary Material online). This haplogroup is otherwise rare in Near and Remote Oceania but exists in almost all Southeast Asian groups analyzed here. The Ontong Java individuals with this haplogroup all have the same HV1 sequence, which is also shared with individuals from the Bismarcks and from Southeast Asia (supplementary fig. 6, Supplementary Material online). The only other mtDNA haplogroup in the PO is Q1, which is of NO origin, found in a single Tikopian

NRY Haplogroup	Ht #	N	TH-West	TH-East
C-M208*	2	4		■
K3-P79	1	2	■	
M-P256*	1	2	■	
M1b1-M104*	2	3	■	
M2-M353*	7	6	■	
	8	3	■	
	9	2	■	
	10	2	■	
M2a-M177	1	2	■	
M3-P117	1	4	■	■
	4	2		■
	5	2		■
	6	3	■	■
O3a-M324*	3	3	■	■
	4	5	■	

MtDNA Haplogroup	Ht #	N	TH-West	TH-East
B4-16261	1	38	■	■
B4a1a1a	1	151	■	■
	2	6	■	
	4	4	■	
	5	4	■	
	7	9	■	
	8	16	■	
	16	2		■
	17	2	■	
	18	2		■
	19	2	■	
	20	3	■	
M27b	1	36	■	■
	2	2	■	
	3	3		■
P1	2	2	■	■

Fig. 6. Patterns of Y-STR haplotype (15 Y-STR loci) and mtDNA HV1 sequence sharing, on the background of specific haplogroups, among AN groups to the west (TH-West, consisting of Choiseul, Isabel, Kolombangara, Ranongga, Shortlands, and Simbo) versus east (TH-East, consisting of Gela, Guadalcanal, Makira, and Malaita) of the Tryon–Hackman line. Ht#, haplotype number according to supplementary tables 4 (Y-STR haplotypes) and 5 (HV1 sequences) (Supplementary Material Online); N, sample size. Solid blocks indicate the presence of that haplotype in TH-East or TH-West groups, respectively.

(supplementary table 2, Supplementary Material online). Of 14 HV1 sequences shared between the PO and other groups, 9 are shared with Remote Oceania, more so than any other group (supplementary table 5, Supplementary Material online). Thus, the mtDNA evidence also strongly supports a Remote Oceanian origin of the PO.

Overall, there appears to be a stronger founder effect in the NRY lineages in the PO than in the mtDNA lineages. This is indicated by the AMOVA (table 3), in which 53% of the variation is between groups for NRY haplogroups, but only 6% for mtDNA haplogroups, and 13% for HV1 sequences. However, this could partly reflect the reduced mtDNA variation in the Polynesian source population, relative to NRY variation (Kayser et al. 2006), and there are clear indications of reduced mtDNA variability as well in the PO (table 2). In sum, these results suggest that Rennell/Bellona, Ontong Java, and Tikopia were each colonized by a small number of individuals and that there has been little (if any) genetic contact between them and other groups in the Solomons. In the case of Tikopia, these genetic results contrast sharply with archaeological evidence for extensive contact between Tikopia and other groups in the Solomons (Kirch 1984). However, the mtDNA HV1 sequences shared between the Tikopians and Remote Oceanians are also widespread in the Solomons (supplementary table 5, Supplementary Material online); complete mtDNA genome sequencing may help distinguish further between a Remote Oceanian versus Solomons origin for these shared HV1 sequences.

Tryon–Hackman Line and Other Culture Areas
The Tryon–Hackman line, running roughly between Isabel and Guadalcanal, is considered a major linguistic boundary within the Oceanic languages that reflects different migra-

tion histories and/or spheres of contact influence in northwestern versus southeastern Solomons (Tryon and Hackman 1983; Ross 1989; Sheppard and Walter 2006). To investigate the effect of this boundary on genetic structure within the Solomons, we carried out an AMOVA (table 3) for the AN groups, omitting Santa Cruz and the Polynesian Outliers. For both mtDNA and NRY haplogroups, the geographic grouping implied by the Tryon–Hackman line has a higher among-groups component than does the linguistic grouping of AN versus PAP, and the among-groups component is significantly different from zero for NRY haplogroups. Still, the among-populations-within-groups component exceeds the among-groups component, indicating that the Tryon–Hackman line is a relatively poor fit to the genetic data (since there is more variation within the geographic groupings of populations than between them).

However, the analysis of shared HV1 sequences and Y-STR haplotypes reveals a different picture. With respect to HV1 sequences (fig. 6; supplementary table 5, Supplementary Material online), 10 of 16 shared HV1 sequences occur in groups from different sides of the Tryon–Hackman line and the sharing of identical sequences among AN groups from the same side of the Tryon–Hackman line versus different sides of the Tryon–Hackman line does not differ from random expectations ($P = 0.08$). With respect to Y-STR haplotypes (fig. 6; supplementary table 4, Supplementary Material online), there is significantly more sharing of identical haplotypes among AN groups from the same side of the Tryon–Hackman line ($P < 0.001$); only 2 of 15 shared Y-STR haplotypes occur in AN groups from different sides. Thus, for recent migration as reflected in sharing of identical HV1 sequences or Y-STR haplotypes, males are moving significantly more often among groups

from within the northwestern or within the southeastern Solomons, respectively, but recent female migration shows no such geographic constraint.

To further investigate the influence of geography on the genetic structure of Solomon groups, we carried out Mantel tests of the significance of correlations between the geographic distance and the genetic distances (Φst distances based on mtDNA HV1 sequences, Fst distances based on NRY haplogroup frequencies, and Rst distances based on Y-STR haplotypes) between Solomon groups (excluding the Polynesian Outliers). We observed significant correlations between geographic distances and genetic distances based on both mtDNA ($r = 0.62$, $P = 0.05$) and NRY (Fst: $r = 0.29$, $P = 0.04$; Rst: $r = 0.54$, $P < 0.001$) genetic distances. However, for the mtDNA Φst distances the significant correlation with geographic distances is caused by Santa Cruz, which exhibits both large genetic and geographic distances from the other Solomon groups, as when Santa Cruz is removed from the analysis the correlation is no longer significant ($r = 0.00$, $P = 0.47$). By contrast, the correlation between NRY Fst distances and geographic distances remains significant even when Santa Cruz is removed from the analysis (Fst: $r = 0.33$, $P = 0.01$; Rst: $r = 0.44$, $P = 0.0015$). Thus, genetic differences among Solomon groups are related to geographic distances between groups for the NRY but not for mtDNA, in keeping with the Y-STR haplotype and HV1 sequence sharing patterns described above. Overall, the higher rate of female migration without geographic constraints may reflect the importance of traditional headhunting alliances, as such alliances dictated both marriage partners within alliances as well as capture and integration of females (but not males) from opposing alliances (Dureau 2000; Thomas 2009).

Another factor that could influence the genetic relationships of the Solomon groups is spheres of cultural interaction, reflected in trade, headhunting, and/or slaving alliances. Based on historical accounts of such cultural interactions (Rivers 1914; Aswani and Sheppard 2003), the Solomons can be divided into four areas of interaction that cut across language groups (and including groups for which we have data): Far West (Bougainville, Shortlands); West (Choiseul, Kolombangara, Ranongga, Russells, Simbo); Central (Gela, Guadalcanal, Savo); and East (Malaita, Makira); Santa Cruz is not considered part of these interaction areas. To investigate how these interaction areas correspond to the genetic structure of Solomon groups, we carried out an AMOVA; because some groups on Isabel are in the Western interaction group and some are in the Central interaction group, and we do not know to which of these groups our Isabel samples belong, we carried out two analyses, one with Isabel in the West group and one with Isabel in the Central group. The results (table 3) indicate that for both mtDNA and NRY haplogroups, the interaction areas do not correspond to the genetic structure of Solomon groups because the among-groups component is smaller than the among-populations-within-groups component. However, for both Y-STR haplotypes and HV1 sequences, the interaction areas do fit the genetic structure in that the

among-groups component is larger than the among-populations-within-groups component (table 3) regardless of how Isabel is classified. Analyses based on haplotypes/sequences instead of haplogroups are expected to be more influenced by recent events that impact haplotype/sequence sharing but not haplogroup sharing. Therefore, the AMOVA results suggest that the interaction areas may have influenced recent gene flow in the Solomons.

Bridging Near and Remote Oceania

To what extent do the Solomons indeed connect Near and Remote Oceania? To address this issue, we included comparable mtDNA and NRY data from over 2000 individuals from East and Southeast Asia, Near Oceania, and Remote Oceania (supplementary tables 1 and 2, Supplementary Material online). The MDS (based on HV1 sequences) and CA (based on NRY haplogroup) plots for these data are presented in figure 7. The MDS plot (fig. 7A) separates East and Southeast Asian groups from Near and Remote Oceanian groups, with Near Oceanian groups tending toward one extreme and Remote Oceanian groups toward the other. Santa Cruz is associated with New Britain, along with other Near Oceanian groups predominantly from New Guinea, whereas the other AN and PAP Solomon groups are with Fiji, the Trobriand Islands, and some Bismarck groups (New Hanover and New Ireland), close to Polynesian groups. Three of the PO groups are near Polynesian groups, whereas the fourth, Ontong Java, is with the AN and PAP Solomon groups, probably because of the high frequency of haplogroup M7c3c in Ontong Java. The CA plot for NRY haplogroups (fig. 7B) does not show any discrete clusters but rather a continuum in the first dimension with West and Papua New Guinea at one extreme, to East and Southeast Asia, with two of the Polynesian Outliers at the other extreme. The AN and PAP groups from the Solomons are separated somewhat in the second dimension, along with several groups from the Bismarcks and Remote Oceania.

Overall, the CA plots indicate a tendency for the Solomon AN and PAP groups to cluster with populations from the Bismarcks and Remote Oceania, which may suggest a primary origin for these Solomon groups from the Bismarcks and subsequent gene flow from the Solomons to Remote Oceania. Analyses of specific haplogroups, discussed previously, support this conclusion. In particular, NRY haplogroups M3-P117, K3-P79, and M1b1-M104*, which are likely to have originated in the Bismarcks (Scheinfeldt et al. 2006), are widespread in the Solomons and also occur in Remote Oceania (Kayser, Choi, et al. 2008). In fact, the only NRY haplogroup present in significant frequencies ($>5\%$) in both the Solomons and Remote Oceania but absent from the Bismarcks, and that did not originate in the Solomons (M2-M353*) or Remote Oceania (C2a1-P33), is haplogroup O3a-M324*. Haplogroup O3a-M324* has a TMRCA of 10.5 KYA (supplementary table 3, Supplementary Material online) and is widespread throughout East and Southeast Asia, Near Oceania, and Remote Oceania, with numerous shared
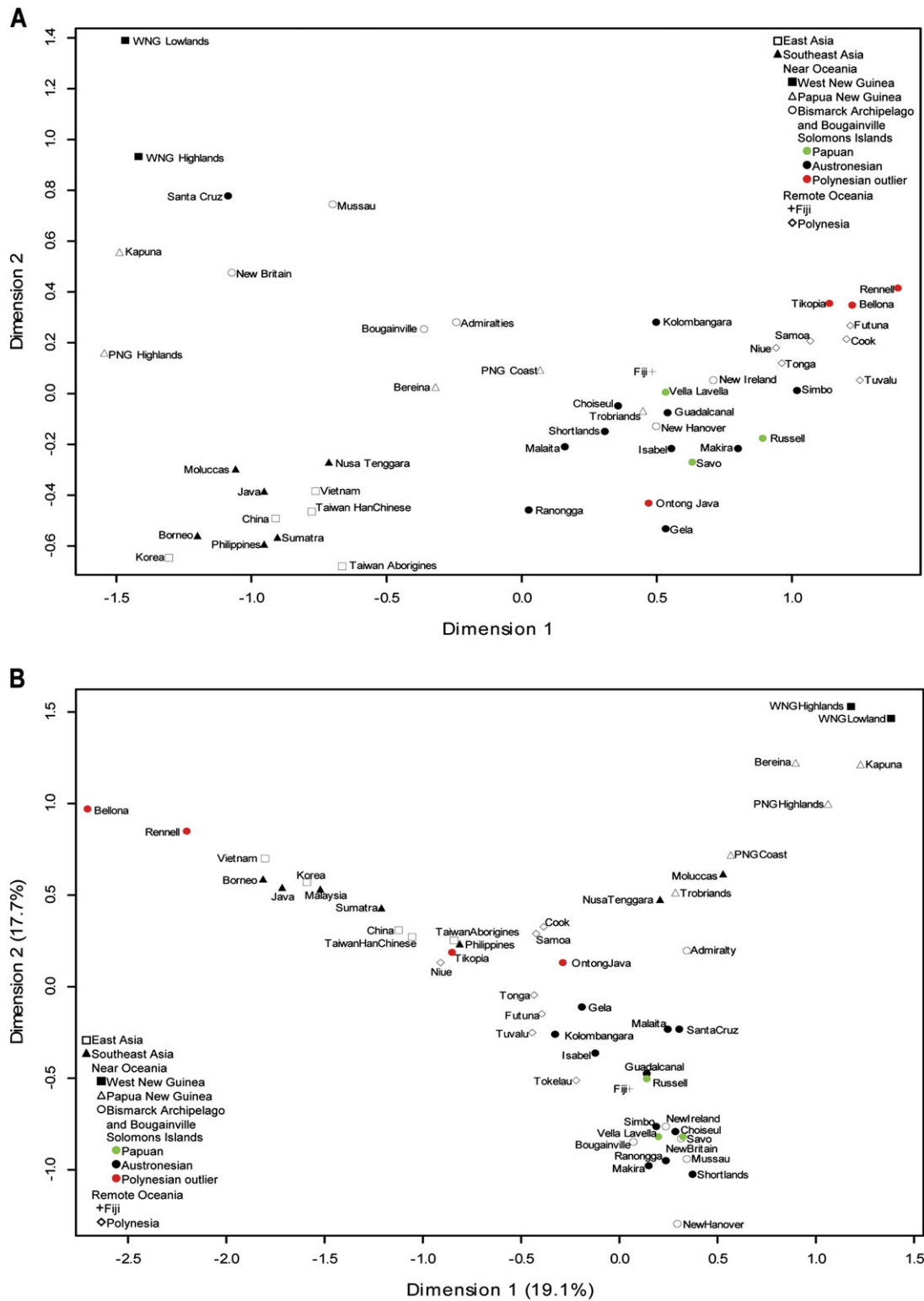
FIG. 7. (A) MDS plot based on ΦST distances calculated from HV1 sequences from Solomons and reference groups. The stress value is 0.089. (B) CA plot for Solomons and reference data, based on NRY haplogroups.

haplotypes (fig. 5C), indicating a rapid spread of this haplogroup. It occurs at low frequency (0.7%) in the Admiralty Islands (Kayser, Choi, et al. 2008) and is also expected to be rare in the nearby Bismarcks; although the M324 marker was not analyzed in the previous extensive study of NRY haplogroups from the Bismarcks,

individuals with M324 would have been classified as the O3-M122 haplogroup, which was found at a frequency of only 2.4% (Scheinfeldt). The most likely explanation for the low frequency of O3a-M324* in the Bismarcks, given the otherwise widespread distribution of this haplogroup across Asia and Oceania, would appear to be loss via drift.

Several mtDNA haplogroups also indicate a prominent association between the Bismarcks and the Solomons. This is most clearly evident in haplogroups M27a, M27b, M27c, and M28, which together account for 13.5% of Solomon mtDNAs and elsewhere are found only in the Bismarcks (and Remote Oceania for M28), where they probably originated (Scheinfeldt et al. 2006). The two most frequent mtDNA haplogroups, B4-16261 (15%) and B4a1a1a (61%), are geographically widespread across Southeast Asia and Near and Remote Oceania (supplementary table 2, Supplementary Material online) with numerous extensively shared haplotypes (supplementary figs 7 and 8, Supplementary Material online), including some shared exclusively or predominantly between the Bismarcks and Solomons. Overall, the Bismarcks were the major source of both NRY and mtDNA haplogroups in the Solomons as well as Remote Oceania, as suggested by previous studies (Scheinfeldt et al. 2006; Friedlaender et al. 2007; Kayser, Choi, et al. 2008), and in good agreement with linguistic and archaeological evidence suggesting a major role for the Bismarcks in the spread of Austronesian languages and Lapita pottery (Ross 1989; Kirch 2010).

The homeland of the Austronesian expansion has been suggested to be Taiwan, based on linguistic and archaeological evidence (Bellwood and Dizon 2005). Previous studies indicated a genetic trail for the major Remote Oceanian mtDNA haplogroup, B4a1a1a, going back to Taiwan (Redd et al. 1995; Melton and Stoneking 1996; Kayser et al. 2006), in good agreement with the "Out-of-Taiwan" hypothesis for Austronesian origins. However, a recent study concluded instead that haplogroup B4a1a1a originated in the Bismarcks prior to the arrival of the Austronesians and spread via voyaging corridors to the west and to the east (Soares et al. 2011). This interpretation is questionable, given that the confidence intervals for the age of B4a1a1a from different geographic regions overlap considerably and moreover do not take into account any uncertainty in the mutation rate estimates. An Austronesian-associated expansion for B4a1a1a is therefore not excluded—but neither can it be assumed. Other genetic data do, however, point to a genetic impact of the Austronesian expansion. In particular, NRY haplogroup O1a2-M110 has been reported previously only in aboriginal Taiwanese, island Southeast Asia, the Admiralties, and Remote Oceania (Kayser et al. 2006; Kayser, Choi, et al. 2008; Delfin et al. 2011), and it could potentially exist in the Bismarcks as well, as the major study of NRY variation in the Bismarcks did not type the M110 marker (Scheinfeldt et al. 2006) but did find a low frequency of the "parent" haplogroup O1a-M119. The presence of O1a2-M110 in many Solomon groups, at an overall frequency of ∼5% (supplementary table 1, Supplementary Material online), further attests to the importance of this haplogroup in Oceania. As previously suggested, the probable origin of O1a2-M110 is in Taiwan (Kayser, Choi, et al. 2008), with recent expansion indicated by the numerous haplotypes shared across wide geographic regions (supplementary fig. 9, Supplementary Material online).

To what extent can the Solomons be said to be an intermediate source of Remote Oceanian NRY and mtDNA lineages? There are five NRY haplogroups at a frequency of 4% or more in Remote Oceania (excluding C2a1-P33, which is of Remote Oceanian origin): K-M9*, K3-P79, M1b-P87*, M1b1-M104*, and O3a-M324*; all of these also occur in the Solomons (supplementary table 1, Supplementary Material online). Haplogroup M1b-P87* is of particular interest as in our data for Remote Oceania it is found only in Fiji, at a frequency of 23%; it has also been reported from Vanuatu at a frequency of ∼4% (Karafet et al. 2010). There are two Y-STR haplotypes on the background of M1b-P87* shared between Fiji and at least one other group, and both of these involve only Solomons groups (supplementary table 6, Supplementary Material online). Haplogroup M1b-P87* may thus reflect a more recent migration from Near Oceania to Fiji that did not extend further eastward, as inferred from genome-wide SNP data (Wollstein et al. 2010). Moreover, several other Y-STR haplotypes are shared between Remote Oceania and the Solomons (supplementary table 6, Supplementary Material online). Regarding mtDNA, practically all of the mtDNA haplogroups found in Remote Oceania are also found in the Solomons (supplementary table 2, Supplementary Material online), and for the two major haplogroups, B4-16261 and B4a1a1a, most of the HV1 sequences shared between Remote Oceania and some other group include at least one group from the Solomons (supplementary table 5, Supplementary Material online).

Vanuatu would also be expected to show connections with the Solomons, but unfortunately existing mtDNA and NRY data are either too restricted, or have not been typed for sufficient overlapping markers, to be used in the comparative analyses with the Solomons data (Cox 2006; Friedlaender et al. 2007; Karafet et al. 2010). However, one such marker indicating a link from the Solomons via Vanuatu to Fiji is M1b-P87, which as mentioned above, is found in Vanuatu at a frequency of ∼4% (Karafet et al. 2010). Although the major NRY and mtDNA lineages found in Vanuatu to date are also present in the Solomons, more detailed genetic studies of the Vanuatu archipelago would be desirable.

Overall, both the NRY and the mtDNA data support connections between the Solomons and Remote Oceania. However, one puzzling feature is the very low frequency in Remote Oceania (4% in Fiji, otherwise only one individual from Futuna; supplementary table 1, Supplementary Material online) of the most frequent NRY haplogroup in the Solomons, M2-M353*. This might suggest a primary role for Santa Cruz, rather than the main Solomons, in the colonization of Remote Oceania, as M2-M353* is not found on Santa Cruz. However, some important Remote Oceanian haplogroups (i.e., present at an overall frequency of >5%; supplementary table 1, Supplementary Material online) are either missing or at very low frequency in Santa Cruz, including NRY haplogroups K-M9*, K3-P79, M1b-P87* (which, however, might reflect a secondary migration as discussed above), and O3a-M324* and mtDNA haplogroup B4-16261. Although the absence of these haplogroups in Santa Cruz may reflect genetic drift, and

migration from Santa Cruz may have played a role in the subsequent colonization of Remote Oceania, it seems that there must also have been at least some gene flow from the main Solomons.

Also of interest is the possibility of back migrations to the Solomons from Remote Oceania (besides the Polynesian Outliers). Given the overall close relationships between the Solomons and Remote Oceania, it is difficult to find clear indications of a back migration in the genetic data. However, NRY haplogroup C2a1-P33 does provide a possible indication. The network of Y-STR haplotypes on the background of C2a1-P33 (fig. 5A) includes a branch of four Y-STR haplotypes that are found only in several AN and PAP groups in the Solomons (plus one found in a Polynesian) but not in the Polynesian Outliers. The presumptive ancestral haplotype for this branch is also not found in the Polynesian Outliers. These results strongly suggest that this particular branch of C2a1-P33 was brought to the Solomons via a migration from Remote Oceania that did not involve the Polynesian Outliers. More detailed genetic investigations are required to substantiate this indication of a back migration from Remote Oceania and more NO populations need to be analyzed to determine if this putative back migration did not extend any further westward than the Solomons (as indicated by the absence of C2a1-P33 in all other NO groups analyzed to date).

Conclusions

The results from this study of the Solomon Islands add substantially to our understanding of the genetic history of Near and Remote Oceania. In particular, we find

1. Old NRY paragroups (C-RPS4Y* and possibly M-P256*) are present in the Solomons but lacking (or nearly so) in the rest of Near Oceania, in keeping with a relatively old colonization of the Solomons. NRY haplogroup M2-M353* supports this view, as our data suggest that it arose ~9.2 KYA in the Solomons.
2. AN and PAP groups do not differ significantly with respect to either mtDNA or NRY patterns of variation or diversity, suggesting substantial genetic contact between groups speaking very different languages. This view is reinforced by NRY haplogroup M2-M353*, which probably arose in the Solomons, and for which there is extensive sharing of Y-STR haplotypes between AN and PAP groups, supporting extensive gene flow between them.
3. Santa Cruz is a conundrum, as archaeological and linguistic evidence indicate that it was colonized relatively soon after the arrival of Austronesian speakers in Near Oceania, and yet it has unusually low frequencies of NRY and mtDNA haplogroups of Asian origin. Possible explanations include: pre-Lapita settlement of Santa Cruz followed by language shift when Austronesian speakers arrived; a rapid language and cultural shift by a PAP group in the Bismarcks after the arrival of Austronesians there that subsequently colonized Santa Cruz; or gradual genetic replacement due to the ongoing extensive contact with the Bismarcks following initial colonization. Regardless of the explanation, Santa Cruz has clearly remained genetically isolated from the rest of the Solomons.

4. Polynesian Outliers have a very distinct history, with separate and severe island-specific founder events evident in NRY variation, as well as reduced mtDNA diversity. The Polynesian Outliers show signs of isolation from other groups in the Solomons, especially Rennell and Bellona, for which the frequency of NRY haplogroups that likely came from Polynesia (C2a1-P33, O2a1-M88, and O3a-M324*) is 97.4% and 100%, respectively, and the frequency of mtDNA haplogroups that likely came from Polynesia (B4-16261 and B4a1a1a) is 100% for both islands. This genetic isolation probably reflects a combination of the geographic isolation of these islands as well as cultural barriers.
5. The Tryon–Hackman line, which marks an important division among Oceanic languages, does not find any correspondence in the genetic structure of Solomon groups. Overall, patterns of NRY, but not mtDNA, variation are correlated with geographic distances between Solomon groups, suggesting that male, but not female, migration is influenced by geographic distance. Spheres of historical cultural interaction among Solomon groups do correspond to the genetic structure of Y-STR haplotypes and mtDNA HV1 sequences, suggesting that these interaction areas have impacted recent gene flow among Solomons groups as well.
6. The major source of Solomon NRY and mtDNA types is the Bismarcks, in good agreement with linguistic and archaeological evidence as well as expectations based on previous genetic evidence. Furthermore, the Solomons appear to be the main source of Remote Oceanian NRY and mtDNA types. Overall, the Solomons therefore bridge Near and Remote Oceania in terms of the genetic history of Oceania. There is also an indication of a back migration from Remote Oceania to the Solomons in the NRY haplogroup C2a1-P33 network of Y-STR haplotypes, distinct from the migrations that settled the Polynesian Outliers.
7. Overall, we obtained more detailed insights from NRY than from mtDNA variation, which partly reflects the reduced mtDNA variation in this region and partly reflects the reduced resolution afforded by sequencing only HV1. We expect additional insights into the genetic history of the Solomons to arise from complete mtDNA genome sequencing as well as from analyses of genome-wide data.

Supplementary Material

Supplementary tables 1–7 and figures 1–9 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

References

Ambrose W. 2002. From the very old to new, obsidian artefacts in the Admiralty Islands. In: Kaufmann C, Schmid CK, Opnemus S, editors. Admiralty Islands: art from the South Seas. Zurich (Switzerland): Museum Rietberg. p. 67–72.

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet. 23:147.

Aswani S, Sheppard P. 2003. The archaeology and ethnohistory of exchange in precolonial and colonial Roviana. Gifts, commodities, and inalienable possessions. Curr Anthropol. 44:S51–S78.

Bandelt H-J, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 16:37–48.

Bandelt H-J, Forster P, Sykes BC, Richards MB. 1995. Mitochondrial portraits of human populations using median networks. Genetics 141:743–753.

Bellwood P, Dizon E. 2005. The Batanes Archaeological Project and the "Out of Taiwan" hypothesis for Austronesian dispersal. J Austronesian Studies. 1:1–31.

Cox MP. 2006. Extreme patterns of variance in small populations: placing limits on human Y-chromosome diversity through time in the Vanuatu Archipelago. Ann Hum Genet. 71: 390–406.

Cox MP, Mirazón-Lahr M. 2006. Y-chromosome diversity is inversely associated with language affiliation in paired Austronesian- and Papuan-speaking communities from Solomon Islands. Am J Hum Biol. 18:35–50.

Cox MP, Redd AJ, Karafet TM, Ponder CA, Lansing JS, Sudoyo H, Hammer MF. 2007. A Polynesian motif on the Y chromosome: population structure in Remote Oceania. Hum Biol. 79:525–535.

Delfin F, Salvador JM, Calacal GC, et al. (17 co-authors). 2011. The Y-chromosome landscape of the Philippines: extensive heterogeneity and varying genetic affinities of Negrito and non-Negrito groups. Eur J Hum Genet. 19:224–230.

Donohue M, Denham T. 2010. Farming and language in island southeast Asia. Curr Anthropol. 51:223–256.

Dunn M, Reesink G, Terrill A. 2002. The East Papuan languages: a preliminary typological appraisal. Ocean Linguist. 41:28–62.

Dunn M, Terrill A, Reesink G, Foley RA, Levinson SC. 2005. Structural phylogenetics and the reconstruction of ancient language history. Science 309:2072–2075.

Dureau C. 2000. Skulls, mana, and causality. J Polyn Soc. 109:71–97.

Elbert S. 1962. Phonemic expansion in Rennellese. J Polyn Soc. 71:25–31.

Elbert S, Monberg T. 1965. From the two canoes. Oral traditions of Rennell and Bellona Islands. Honululu (HI): University of Hawaii Press.

Excoffier L, Laval G, Schneider S. 2005. Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evol Bioinform Online. 1:47–50.

Fenner JN. 2005. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. Am J Phys Anthropol. 128:415–423.

Friedlaender JS, Friedlaender FR, Hodgson JA, Stoltz M, Koki G, Horvat G, Zhadanov S, Schurr TG, Merriwether DA. 2007. Melanesian mtDNA complexity. PLoS One. 2:e248.

Friedlaender JS, Friedlaender FR, Reed FA, et al. (12 co-authors). 2008. The genetic structure of Pacific Islanders. PLoS Genet. 4:e19.

Friedlaender JS, Gentz F, Green K, Merriwether DA. 2002. A cautionary tale on ancient migration detection: mitochondrial DNA variation in Santa Cruz Islands, Solomon Islands. Hum Biol. 74:453–471.

Goedbloed M, Vermeulen M, Fang RN, et al. (17 co-authors). 2009. Comprehensive mutation analysis of 17 Y-chromosomal short tandem repeat polymorphisms included in the AmpFlSTR® Yfiler® PCR amplification kit. Int J Legal Med. 123:471–482.

Gray R, Drummond A, Greenhill J. 2009. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. Science 323:479–483.

Green R. 1995. Linguistic, biological and cultural origins of the initial inhabitants of Remote Oceania. N Z J Archaeol. 17:5–27.

Green R. 2000. Lapita and the cultural model for intrusion, integration and innovation. In: Anderson A, Murray T, editors. Australian archaeologist: collected papers in honour of Jim Allen. Canberra (Australia): Coombs Academic Publishing, Australian National University. p. 372–392.

Green RC. 1991. A reappraisal of the dating for some Lapita sites in the ReeF/Santa Cruz Group of the southeast Solomons. J Polyn Soc. 100:197–207.

Groube L, Chappell J, Muke J, Price D. 1986. A 40,000 year-old human occupation site at Huan Peninsula, Papua New Guinea. Nature 324:453–455.

Hage P, Marck J. 2003. Matrilineality and the Melanesian origin of Polynesian Y chromosomes. Curr Anthropol. 44:121–127.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser. 41:95–98.

Hill C, Soares P, Mormina M, et al. (11 co-authors). 2007. A mitochondrial stratigraphy for island southeast Asia. Am J Hum Genet. 80:29–43.

Hunley K, Dunn M, Lindström E, Reesink G, Terrill A, Healy ME, Koki G, Friedlaender FR, Friedlaender JS. 2008. Genetic and linguistic coevolution in northern island Melanesia. PLoS Genet. 4:e1000239.

Hurles ME, Nicholson J, Bosch E, Renfrew C, Sykes BC, Jobling MA. 2002. Y chromosomal evidence for the origins of Oceanic-speaking peoples. Genetics 160:289–303.

Karafet TM, Hallmark B, Cox MP, Sudoyo H, Downey S, Lansing JS, Hammer MF. 2010. Major east–west division underlies Y chromosome stratification across Indonesia. Mol Biol Evol. 27:1833–1844.

Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF. 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. Genome Res. 18:830–838.

Katoh K, Kuma K-i, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 33:511–518.

Kayser M. 2010. The human genetic history of Oceania: Near and Remote views of dispersal. Curr Biol. 20:R194–R201.

Kayser M, Brauer S, Cordaux R, et al. (15 co-authors). 2006. Melanesian and Asian origins of Polynesians: mtDNA and Y chromosome gradients across the Pacific. Mol Biol Evol. 23:2234–2244.

Kayser M, Brauer S, Weiss G, Underhill PA, Roewe L, Schiefenhövel W, Stoneking M. 2000. Melanesian origin of Polynesian Y chromosomes. Curr Biol. 10:1237–1246.

Kayser M, Choi Y, van Oven M, Mona S, Brauer S, Trent RJ, Suarkia D, Schiefenhövel W, Stoneking M. 2008. The impact of the Austronesian expansion: evidence from mtDNA and Y chromosome diversity in the Admiralty Islands of Melanesia. Mol Biol Evol. 25:1362–1374.

Kayser M, Lao O, Saar K, Brauer S, Wang X, Nurnberg P, Trent R, Stoneking M. 2008. Genome-wide analysis indicates more Asian

than Melanesian ancestry of Polynesians. *Am J Hum Genet.* 82:194–198.

Kirch P. 1997. The Lapita peoples: ancestors of the Oceanic World. Oxford: Blackwell.

Kirch PV. 1984. The evolution of Polynesian Chiefdoms. Cambridge: Cambridge University Press.

Kirch PV. 2010. Peopling of the Pacific: a holistic anthropological perspective. *Annu Rev Anthropol.* 39:131–148.

Kivisild T, Tolk H-V, Parik J, Wang Y, Papiha S, Bandelt H-J, Villems R. 2002. The emerging limbs and twigs of the east Asian mtDNA tree. *Mol Biol Evol.* 19:1737–1751 (erratum 1720:1162).

Melton T, Peterson R, Redd AJ, Saha N, Sofro ASM, Martinson J, Stoneking M. 1995. Polynesian genetic affinities with Southeast Asian populations as identified by mtDNA analysis. *Am J Hum Genet.* 57:403–414.

Melton T, Stoneking M. 1996. Extent of heterogeneity in mitochondrial DNA of ethnic Asian populations. *J Forensic Sci.* 41:591–602.

Miller S, Dykes D, Polesky H. 1988. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* 16:1215.

Mona S, Grunz KE, Brauer S, et al. (11 co-authors). 2009. Genetic admixture history of eastern Indonesia as revealed by Y-chromosome and mitochondrial DNA analysis. *Mol Biol Evol.* 26:1865–1877.

Næss Å. 2006. Bound nominal elements in Äiwoo (Reefs): a reappraisal of the "Multiple Noun Class Systems". *Ocean Linguist.* 45:269–296.

Næss Å, Boerger BH. 2008. Reefs-Santa Cruz as Oceanic: evidence from the verb complex. *Ocean Linguist.* 47:185–212.

Nasidze I, Quinque D, Rahmani M, Alemohamad SA, Asadova P, Zhukova O, Stoneking M. 2009. mtDNA and Y-chromosome variation in the Talysh of Iran and Azerbaijan. *Am J Phys Anthropol.* 138:82–89.

Polzina T, Daneshmand SV. 2003. On Steiner trees and minimum spanning trees in hypergraphs. *Oper Res Lett.* 31:12–20.

Poulsen J. 1972. Outliers archaeology: Bellona. *Archaeol Phys Anthropol Oceania.* 7:184–205.

Pugach I, Matveyev R, Wollstein A, Kayser M, Stoneking M. 2011. Dating the age of admixture via wavelet transform analysis of genome-wide data. *Genome Biol.* 12:R19.

Redd A, Takezaki N, Sherry S, McGarvey S, Sofro A, Stoneking M. 1995. Evolutionary history of the COII/tRNALys intergenic 9 base pair deletion in human mitochondrial. *Mol Biol Evol.* 12:604–615.

Ricaut F-X, Thomas T, Mormina M, Cox MP, Bellatti M, Foley RA, Mirazon-Lahr M. 2010. Ancient Solomon Islands mtDNA: assessing holocene settlement and the impact of European contact. *J Archaeol Sci.* 37:1161–1170.

Rivers WHR. 1914. The history of Melanesia. Cambridge: Cambridge University Press.

Ross M, Næss Å. 2007. An Oceanic origin for Äiwoo, the language of the Reef Islands? *Ocean Linguist.* 46:456–498.

Ross MD. 1989. Early Oceanic linguistic prehistory: a reassessment. *J Pac Hist.* 24:135–149.

Scheinfeldt L, Friedlaender F, Friedlaender J, Latham K, Koki G, Karafet T, Hammer M, Lorenz J. 2006. Unexpected NRY chromosome variation in northern island Melanesia. *Mol Biol Evol.* 23:1628–1641.

Seielstad MT, Minch E, Cavalli-Sforza LL. 1998. Genetic evidence for a higher female migration rate in humans. *Nat Genet.* 20:278–280.

Sengupta S, Zhivotovsky LA, King R, et al. (15 co-authors). 2006. Polarity and temporality of high-resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of central Asian pastoralists. *Am J Hum Genet.* 78:202–221.

Sheppard PJ, Walter R. 2006. A revised model of Solomon Islands culture history. *J Polyn Soc.* 115:47–76.

Shi H, Dong Y-l, Wen B, Xiao C-J, Underhill PA, Shen P-d, Chakraborty R, Jin L, Su B. 2005. Y-chromosome evidence of southern origin of the east Asian specific haplogroup O3-M122. *Am J Hum Genet.* 77:408–419.

Smith N. 1995. An annotated list of creoles, pidgins, and mixed languages. In: Arends J, Muysken P, Smith N, editors. Pidgins and Creoles: an introduction. Philadelphia (PA): John Benjamins Publishing Company. p. 331–374.

Soares P, Rito T, Trejaut J, et al. (16 co-authors). 2011. Ancient voyaging and Polynesian origins. *Am J Hum Genet.* 88:239–247.

Specht J. 2005. Revisiting the Bismarcks: some alternative views. In: Pawley A, Attenborough R, Golson J, Hide R, editors. Papuan pasts: cultural, linguistic and biological histories of Papuan-speaking peoples. Canberra (Australia): Pacific Linguistics. p. 235–288.

Specht J. 2009. The aceramic to ceramic boundary in the Bismarck archipelago. In: Sheppard P, Thomas T, Summerhayes G, editors. Lapita: ancestors and descendants, New Zealand Archaeological Association Monograph 28. Auckland (New Zealand): Publishing Press Ltd. p. 11–34.

Spriggs M. 1995. The Lapita culture and Austronesian prehistory in Oceania. In: Bellwood P, Fox J, Tryon D, editors. The Austronesians: historical and comparative perspectives. Canberra (Australia): Research School of Pacific Studies, Australian National University. p. 119–142.

Spriggs M. 1997. The Island Melanesians. Oxford: Blackwell.

Spriggs M. 2003. Chronology of the neolithic transition in southeast Asia and the western Pacific: a view from 2003. *Rev Archaeol.* 24:57–80.

StatSoft, Inc. 2007. STATISTICA (data analysis software system). Tulsa (OK): StatSoft Inc.

Su B, Jin L, Underhill P, et al. (11 co-authors). 2000. Polynesian origins: insights from the Y chromosome. *Proc Natl Acad Sci U S A.* 97:8225–8228.

Su B, Xiao J, Underhill P, et al. (21 co-authors). 1999. Y-chromosome evidence for a northward migration of modern humans into eastern Asia during the last ice age. *Am J Hum Genet.* 65:1718–1724.

Summerhayes G. 2001. Defining the chronology of Lapita in the Bismarck Archipelago. In: Clark G, Anderson A, Vunidilo T, editors. The archaeology of Lapita dispersal in Oceania. Terra Australis 17. Canberra (Australia): Pandanus p. 25–38.

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

Sykes B, Leiboff A, Low-Beer J, Tetzner S, Richards M. 1995. The origins of the Polynesians: an interpretation from mitochondrial lineage analysis. *Am J Hum Genet.* 57:1463–1475.

Terrell J. 2009. Return to the entangled bank: deciphering the Lapita cultural series. In: Sheppard P, Thomas T, Summerhayes G, editors. Lapita: ancestors and descendants, New Zealand Archaeological Association Monograph 28. Auckland (New Zealand): Publishing Press Ltd. p. 255–269.

Thomas T. 2009. Communities of practice in the archaeological record of New Georgia, Rendova and Tetepare. In: Sheppard P, Thomas T, Summerhayes G, editors. Lapita: ancestors and descendants, New Zealand Archaeological Association Monograph 28. Auckland (New Zealand): Publishing Press Ltd. p. 119–145.

Torrence R, Neall V, Doelman T, et al. (12 co-authors). 2004. Pleistocene colonisation of the Bismarck archipelago: new evidence from west New Britain. *Archaeol Oceania.* 39:101–130.

Trejaut JA, Kivisild T, Loo JH, Lee CL, He CL, Hsu CJ, Li ZY, Lin M. 2005. Traces of archaic mitochondrial lineages persist in Austronesian-speaking Formosan populations. *PLoS Biol.* 3:e247.

Tryon D, Hackman B. 1983. Solomon Islands languages: an internal classification. Canberra (Australia): Australian National University.

van Oven M, Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat.* 30:E386–E394.

Walter R, Sheppard P. 2009. A review of Solomon Island archaeology. In: Sheppard P, Thomas T, Summerhayes G, editors. Lapita: ancestors and descendants, New Zealand Archaeological Association Monograph 28. Auckland (New Zealand): Publishing Press Ltd. p. 35–72.

Wickler S, Spriggs M. 1988. Pleistocene human occupation of the Solomon Islands, Melanesia. *Antiquity* 62:703–706.

Wilson IJ, Weale ME, Balding DJ. 2003. Inferences from DNA data: population histories, evolutionary processes and forensic match probabilities. *J R Stat Soc Ser A.* 166:155–201.

Wollstein A, Lao O, Becker C, Brauer S, Trent RJ, Nürnberg P, Stoneking M, Kayser M. 2010. Demographic history of Oceania inferred from genome-wide data. *Curr Biol.* 20:1983–1992.

Wurm S. 1978. Reefs-Santa Cruz: Austronesian, but . . . !. In: Carrington L, Wurm SA, editors. Second International Conference on Austronesian Linguistics: Proceedings; 1978 Jan 4–12. Canberra (Australia): Pacific Linguistics. p. 969–1010.

Xue Y, Zerjal T, Bao W, et al. (12 co-authors). 2006. Male demography in east Asia: a north-south contrast in human population expansion times. *Genetics* 172:2431–2439.