MBE Advance Access published April 6, 2005

Expanding Southwest Pacific mitochondrial haplogroups P and Q

Jonathan Friedlaender[1*], Theodore Schurr[2¶], Fred Gentz[1§], George Koki[3], Françoise Friedlaender*, Gisele Horvat, Paul Babb[2¶]Sal Cerchio[4§], Frederika Kaestle[5§], Moses Schanfield[6], Ranjan Deka[7], Ric Yanagihara[8], and D. Andrew Merriwether[4§]

[1]Anthropology Department, Temple University, Philadelphia, PA

[2]Anthropology Department, University of Pennsylvania, Philadelphia, PA

[3]Papua New Guinea Institute for Medical Research, Goroka, PNG

[4]Anthropology Department, University of Binghamton, Binghamton, NY

[5]Anthropology Department, University of Indiana, Bloomington, IN

[6]Forensics Department, George Washington University, Washington, DC

[7]Department of Environmental Health, University of Cincinnati, Cincinnati, OH

[8]Department of Pediatrics, John Burns School of Medicine, University of Hawaii, Honolulu, HI

Key Words: mtDNA Melanesia Pacific phylogeography

Running Title: Melanesian mtDNA haplogroups P and Q

Corresponding Author:

Jonathan S. Friedlaender,

Department of Anthropology, Temple University,

1115 West Berks Mall, Philadelphia, PA 19106.

Telephone: (215) 351-4567; Fax: (215) 204-4501

Email: jfriedla@temple.edu

* Anthropology Department, Temple University, Philadelphia, PA

¶ Anthropology Department, University of Pennsylvania, Philadelphia, PA

§ Anthropology Department, University of Michigan, Ann Arbor, MI

Abstract

Modern humans have occupied New Guinea and the nearby Bismarck and Solomon archipelagos of Island Melanesia for at least 40,000 years. Previous mitochondrial DNA (mtDNA) studies indicated that two common lineages in this region, haplogroups P and Q, were particularly diverse, with the coalescence for P considered significantly older than that for Q.

In this study, we expand the definition of haplogroup Q so that it includes three major branches, each separated by multiple mutational distinctions (Q1, equivalent to the earlier definition of Q, plus Q2 and Q3). We report three whole mtDNA genomes that establish Q2 as a major Q branch. In addition, we describe 314 control region sequences that belong to the expanded haplogroups P and Q from our Southwest Pacific collection. The coalescence dates for the largest P and Q branches (P1 and Q1) are similar to each other (~ 50,000 years old), and considerably older than prior estimates. Newly identified Q2, which was found in Island Melanesian samples just to the east, is somewhat younger, by more than 10,000 years. Our coalescence estimates should be more reliable than prior ones because they were based on significantly larger samples, as well as complete mtDNA coding region sequencing. Our estimates are roughly in accord with current suggested dates for the first settlement of New Guinea/Sahul.

The phylogeography of P and Q indicates almost total (female) isolation of ancient New Guinea/Island Melanesia from Australia that may have existed from the time of first settlement. While Q subsequently diversified extensively in New Guinea/Island Melanesia, it has not been found in Australia. The only shared mtDNA haplogroup between Australia and New Guinea identified to date remains one minor branch of P.

Introduction

This paper on variation in two major mtDNA haplogroups that are restricted to Southwest Pacific populations is pertinent to three questions on the prehistory of Sahul (the ancient continent that encompassed Australia and New Guinea) and Island Melanesia. First, are there any clear signatures in these data reflecting differences in population dynamics across the region (i.e., in ancient population expansions or dispersals)? Second, how limited has contact been between populations in New Guinea/Island Melanesia and Australia during the Holocene and Upper Pleistocene? Finally, are there any indications that the initial settlers of northern and southern Sahul were part of a single population, or instead were composed of multiple founding groups?

The current dates from excavated habitation sites could be interpreted to suggest the Australian section of Sahul was settled first. While there have been claims for settlement dates >100,000 BP, the earliest generally accepted dates from Australia fall into the 50,000 to 60,000 year range and are based on thermoluminescence dating (O'Connor and Chappell 2003). However, some maintain that the dates before 42-45,000 BP are not well supported there (O'Connell and Allen 2004). The proposed entry point for any southern voyaging was via the Nusa Tenggara island chain through Timor, as first suggested by Birdsell (1967). The earliest accepted dates for the occupation of New Guinea/Island Melanesia, where no deposits amenable to thermoluminscence dating have been identified, are at the limits of radiocarbon reliability ~ 40,000 BP for both the Huon Peninsula in eastern New Guinea (Groube 1986) and New Ireland (Leavesley et al. 2002), with dates from the Moluccas being no earlier (Bellwood et al. 1998). This temporal framework suggests that a second contemporaneous settlement route originating in

the Moluccas and progressing eastward along the ancient New Guinea north coast into the Bismarck Archipelago is very likely (Allen 2003; Yen 1995).

Subsequent to initial settlement, populations appear to have followed different resource exploitation strategies in the two regions (more terrestrial in Australia vs. mixed maritime and terrestrial in Melanesia - Leavesley et al. 2002) and to have experienced different expansion rates (more numerous sites in Pleistocene Australia - Chappell 2000). By 20,000 years ago, a number of novel cultural developments are detectable, particularly in the Island Melanesian record (Allen 2003), and this pattern of episodic change intensifies in the Holocene (Green 2003). There are competing colonization scenarios for Pleistocene Australia (Bowdler 1993; Flannery 1994; Horton 1999; O'Connell and Allen 1998), with some indication of more fluid population movements in the interior (O'Connor and Chappell 2003).

Relevant to more recent times, historical linguists have not found any connections between Aboriginal Australian and New Guinea/Island Melanesian languages, implying little to no significant contact occurred during the Holocene. In New Guinea/Island Melanesia, one language family (a branch of Austronesian) was introduced from the west only ~3,500 years ago (Kirch 1997; Pawley 2003; Spriggs 1997). The hundreds of other heterogeneous languages spoken there are referred to as non-Austronesian or Papuan, and fall into more than 20 different families which are not demonstrably related (Dunn, Terrill and Reesink 2002; Ross 2001, in press). The largest of these is the Trans New Guinea Phylum, which has been reconstructed for all the languages of the central cordillera of New Guinea, or approximately 70% of the languages of the interior. It is thought to be at least 8 - 12,000 years old (Pawley in press). The diverse languages in north New Guinea and the interiors of Island Melanesia suggest the presence of even older linguistic strata (Pawley in press). The Aboriginal Australian languages are less

diverse. The one major family, Pama-Nyungan, that may be derived from a single proto-language spoken ~10,000 years ago (Bowern and Koch 2004), includes all the Australian languages outside the northwest, and one speculation associated it with the spread of the Small-Tool Tradition (Evans and Jones 1997). In sum, the linguistic and archaeological data suggest little to no contact between the two regions subsequent to their initial settlement, with considerably more inland mobility in Australia than within New Guinea.

Past genetic surveys, particularly those involving the HLA and α-globin loci, indicated simple isolation by distance relationships across the region, with some weak associations reported between Australian Aborigines and highland New Guinea populations (Gao and Serjeantson 1991; Kirk 1989; Roberts-Thomson et al. 1996; Tsintsof et al. 1990). Pertinent mtDNA data have been equivocal in this regard (Forster et al. 2001; Huoponen et al. 2001; Ingman and Gyllensten 2003; Ingman et al. 2000; Redd and Stoneking 1999; Redd et al. 1995; Stoneking et al. 1990). Haplogroups P and Q (as defined by Forster et al. 2001) were especially common in New Guinea/Island Melanesia, and other variants that belong to P were also found in Australia (B^, H^, and I^ of Huoponen et al. 2001; 1b, 1c, and 1d of Ingman and Gyllensten 2003). Previous estimates of their coalescence times and ages of expansion have not been consistent, although haplogroup P appeared to be considerably older than Q as then defined.

Here we report complementary analyses of these two mtDNA haplogroups from our large Southwest Pacific sample set, which clarify their definitions, distributions and ancient expansions. The first is a phylogenetic study of the two haplogroups based on whole mtDNA genome sequences (from published data plus three of our own) that clarifies and modifies published reports on P and expands the definition of Q. The second covers the phylogeography of these P and Q branches (from the total of 886 matrilineages identified in our Southwest

Pacific samples) on the basis of mtDNA hypervariable segment 1 and 2 sequences (HVS1 and HVS2, from np 16000 to 408). The results of these analyses show that haplogroup Q in its expanded definition is complex, with at least two major star-like branches and occurs primarily in populations in New Guinea/Island Melanesia. It has not been found in Australia. P1 and Q1, which have their greatest diversities in New Guinea, have approximately the same coalescence times, while that of the Island Melanesian branch Q2 is smaller (younger).

Materials and Methods

Sample Collection

Our core data set derived from more than 1,500 blood samples collected over the last six years from the Bismarck Archipelago (New Britain, New Ireland, New Hanover/Lavongai and Mussau) and Bougainville Island, Papua New Guinea. Information collected on survey subjects included language, residence, and genealogy. The core sample was supplemented with plasmas from older collections belonging to J. S. Friedlaender, L. Weitkamp, B. S. Blumberg, M. Schanfield, R. Deka, B. Zemel, and A. Damon, and with urine samples obtained from R. Yanagihara and V. Nerukar. Details on sample provenience and collection source are given in the Supplementary Materials, table 1. As detailed there, identification of residence and familial relationships were more problematic for individuals in some of these sets, as they came from a variety of locations in New Guinea, Island Melanesia, Polynesia, and Micronesia. The core samples were collected, and all samples were analyzed, with the approval of the appropriate Human Subjects Ethical Committees of Papua New Guinea, the University of Michigan, and Temple University.

Sample Extraction, Amplification and Analysis

DNA was extracted from 100 to 200 microliters of plasma, red cells, buffy coat or urine (depending on the source of the sample) by using either the quanidine-silica based IsoQuick extraction kit (Orca Scientific) or the column-based Qiagen extraction kit (Qiagen, Inc).

The analysis of these mtDNAs occurred in three steps. The first was the sequencing of the HVS1 and HVS2 from the 886 different matrilineages identified from survey records. The HVS1/2 sequences that could not be definitely assigned to known clades were then screened for the two mutations defining macrohaplogroup M (DdeI 10394, AluI 10397), and then, depending on the presence or absence of these sites, were surveyed for additional RFLPs identified in Malaysian Aborigines, Orang Asli and Malays (Ballinger et al. 1992), PNG populations (Stoneking et al. 1990) and Aboriginal Australians (Cann et al. 1987). Finally, sequencing of the coding region was performed on representative samples from each of the major branches identified by PCR-RFLP analysis using the PCR conditions and primer sets of Rieder et al. (1998).

For control region sequencing, DNA was PCR amplified following standard protocols, using primers spanning nucleotide pairs (nps) 15938 to 00429 and Platinum Taq Polymerase (Invitrogen). Successful amplification was verified by electrophoresis on a 1% ethidium bromide stained agarose gels. Samples were prepared for sequencing by an ExoI digest followed by filtration through a Millipore 96-well filter plate (Millipore, Inc) to remove single stranded DNA and unincorporated nucleotides, and then sequenced using BigDye Terminator 3.0 Ready-reaction kits (ABI). The sequences were then purified with Centri-Sep columns (Princeton Separations), and sequence data collected on an ABI 377XL automated sequencer. Sequence alignment on the coding region was accomplished with Sequencher: Forensic Version (GeneCodes Inc), and, for the control region, Sequence Navigator.

Median joining networks were generated from the control region sequences (manually converted to binary format) in Network 4.0 (Bandelt et al. 1995). The measures of haplotype diversities of the star phylogeny networks, their standard errors and associated expansion age estimates were calculated as specified in Saillard et al. (2000), and as modified in the Appendix to cover HVS1 plus HVS2 (nps16090 -16365 and 035 - 315) .

Results

As defined by Forster et al. (2001), haplogroups P and Q form deep branches within the two major mtDNA clades that exist outside Africa, with P being part of the R division of the N clade, and Q being part of the M clade. The accepted schematic relationships of P and Q to the other mtDNA branches of the African root, L3, are shown in figure 1. Earlier studies indicated these two haplogroups constituted approximately 90% of the observed mtDNA haplotypes in New Guinea (Forster et al. 2001; Tommaseo-Ponzetta et al. 2002).

Haplogroup Q

Figure 2 shows the phylogeny of Q based on whole mtDNA genome sequences. This analysis enlarged the prior definition of Q (now equivalent to Q1) by adding a newly identified branch (Q2), and confirming a third branch (Q3) that had been previously reported in this region with different nomenclatures (Ingman et al. 2003; Redd and Stoneking 1999; Tomasseo-Ponzetta et al. 2001). As shown here, all Q branches shared distinctive transitions at the base, including nps 4117, 5843, 8790, 12940, 13500, 16129, and 16241. Each branch also had its own long set of defining mutations. Interestingly, eight of the ten mutations that define Q1 were found in the control region.

Figure 3 shows the median joining network derived from the haplogroup Q control region sequences. In this network, Q1, Q2, and Q3 all had long branch lengths, and both Q1 and Q2 had

clear “star-like” expansion signatures. The internal diversity of Q1 was greater than Q2. Q3 was represented by only nine total samples (five of our own), and therefore remained poorly delineated. Because of the clear separation of the Q branches, the Q mismatch distribution was bimodal, with a large mean (Supplementary Materials, figure 1).

Haplogroup P

The phylogenetic tree for haplogroup P in figure 4 incorporated published whole mtDNA genome sequences (Ingman and Gyllensten 2003; Ingman et al. 2000; Kivisild et al. 2002). The specific internal branching order of P was ambiguous for two reasons. First, there were many apparent parallel mutations in the separate long branches, probably because of their great age and the loss of intermediate states. Second, the major branches of P were only linked together by a single coding region mutation at np 15607. This is not a common mutation, but it has occurred independently in two West Eurasian haplogroups (J and T).

P3 was the only haplogroup that contained both New Guinea and Australian branches (as reported with different methods and terminology in Ingman and Gyllensten 2003 and Huoponen et al. 2001). Ingman and Gyllensten (2003) proposed a second New Guinea-Australian connection (their branch 1d), which corresponds to our P2 plus P5 branches. However, this tie is not convincing since it is based on the single frequently occurring shared transition at np 1438 (which has occurred independently within haplogroups L1a, L1b, D, H, I and R).

The median joining network for P based on control region sequences (in figure 5) was not as clearly resolved as that for Q, as it contained a number of reticulations. Still, the P1 branch had a clear star-like quality suggesting its great age, with long branches and an under populated central node. The pairwise mismatch distribution for P was smooth and unimodal, with a mean of 7.076 (see also Supplementary Materials, figure 1).

Coalescence Times for P and Q

Table 1 presents the coalescence estimates and associated ages for the "star-like" branches of P and Q. Comparable results from Forster et al. (2001) are paired with them. For expansion age calculations, we had to estimate an average substitution rate per nucleotide per year across both HVS1 and HVS2, which was one per 24,880 years (see the Appendix). The calculated expansion ages for Q1 and P1 from our dataset were large and essentially the same, ~50,000 years. As shown, Forster et al. (2001) reported much smaller coalescence values for their Q1 and P1 equivalents. However, their calculations were based on a set of RFLPs from a very small sample, as opposed to our much larger sample of control region sequences (>4 times greater for P1 and >5 times greater for Q1). Our data also indicated that Q2 expanded ~36,000 years ago, more than 10,000 years later than Q1 and P1.

Geographic Distributions of P and Q

The distributions of P and Q in New Guinea/Island Melanesia are given in table 2 and figure 6 (the much younger haplotype B data are also included but not discussed). The distributions of P and Q do not completely overlap. Q is the more common of the two, and, within New Guinea, the only place that Q has not been found is the South Coast. Haplogroup P is infrequent in the western half of New Guinea (Tommaseo-Ponzetta et al. 2002), and very unevenly distributed through the eastern half of the island. East of New Guinea, P is rare in Island Melanesia except in certain islands of the Louisiade Archipelago off the Papuan Tip.

The Island Melanesian distribution of these two haplogroups is particularly heterogeneous (see table 3 and figure 7). Q2 is found primarily in certain Papuan-speaking groups of New Britain, which indicates its origin lies in this part of Island Melanesia. In New Ireland, P and Q are both very rare, with Q2 present in low frequency in the Madak, who used to

speak a Papuan language. Q1 occurs frequently in most, but not all Papuan-speaking areas reported in table 3.

Further details of the P and Q haplotype subdivisions on the total sample set are given in the Supplementary Materials, tables 1-3. These show that specific haplotypes, especially within Q branches, often have extremely restricted distributions, and sometimes are found only in particular island sections. P haplotypes occur even more sporadically in sections of Island Melanesia than New Guinea, with one subbranch (P1e) being specific to New Britain, while one specific P2 haplotype was found in Fiji, New Caledonia and also Misima Island - all Austronesian-speaking groups. The structured branching implies a relatively undisturbed accumulation of mutational diversity in time and space, with the deepest branches generally being more widely dispersed geographically than the shallower, more recent ones.

Both P and Q were carried further to the southeast (see table 3 and figure 6), to Santa Cruz (Friedlaender et al. 2002), Vanuatu (Cox 2003; Hagelberg et al. 1999; Sykes et al. 1995), and to Fiji (present study), while only Q has been found in Polynesia, in frequencies less than 0.05 (Sykes et al. 1995). Also, both haplogroups have been found in Micronesia to the north in low frequencies. Out of 584 samples tested, ten Q1s and 47 P1s were reported there by Lum and Cann (2000).

To the west of New Guinea, both P and Q are rare. Out of more than 1100 mtDNAs reported from Island Southeast Asia, only seven Qs and three Ps have been identified, all but one of these occurring in central and eastern Indonesia (Ballinger et al. 1992; Cox 2003; Redd and Stoneking 1999; Tajima et al. 2004). We suggest these rare occurrences are most likely the result of a few women moving to those regions from New Guinea, rather than this region of Island Southeast Asia being the ancient homeland of the haplogroups P and Q.

Discussion

These results illustrate a number of problems concerning coalescent statistics and associated age estimates, but also their utility. One issue is that mutation rates vary for different mtDNA regions and nucleotide sites and also among clades (Forster et al. 2002; Macaulay et al. 1999; Parsons et al. 1997). In our study, the proportion of mutations that occurred in the control region was particularly large in the base stem of Q1, compared to other parts of the P and Q trees (compare figures 2 and 4). If, instead, this distortion had happened within Q1, control region sequence and RFLP data would yield divergent coalescence estimates for Q1. Fortunately for our comparisons, this disproportion is not apparent in the expansions within Q1, Q2, and P1 (figures 2 and 4). A second more obvious issue is that reliable coalescence estimates depend on adequate sample sizes as well as clear expansion signatures. The discrepancies between our results and those of Forster et al. (2001) for P1 and Q1 in table 1 suggest that, especially when analyzing more expanded, older networks, very large samples are required.

When considered together, haplogroups P and Q and their coalescence estimates do inform questions on ancient population dynamics, settlement and isolation in this region. The star-like networks of P1 and Q1 independently suggest the same ancient population expansion in New Guinea subsequent to first settlement ~50,000 years ago, followed by the expansion of Q2 in adjacent Island Melanesia somewhat more than 10,000 years later. In addition, the extremely localized distributions of specific haplotypes within the branches of Q and P are consistent with very restricted female movement within the region over the following millennia. The absence of Q in Australia, plus the very separate branch distributions of P in Australia and in New Guinea, indicate an almost complete (female) isolation between the two regions. The single shared haplogroup (P3) only occurs in New Guinea in a restricted southwestern region. Also the New

Guinea branch of P3 is distinctive, suggesting its Australian connection is very old. The first female settlers of Sahul might have effectively been members of the same population, possibly even entering at one place – this cannot be ascertained - but, if that was the case, then they split into two groups shortly afterwards, and remained effectively isolated thereafter.

Specific links of haplogroups P and Q to other branches within macrohaplogroups M and N remain unresolved. The best candidates for close Q relationships are some other Island Melanesian M haplogroups designated "other" in the last two tables and figures (Friedlaender et al. 2002; Friedlaender et al. in press), as well as the very limited M haplotypes reported in Aboriginal Australian populations (Redd and Stoneking 1999; Huoponen et al. 2001; Holst Van Pellekaan et al. 1998; Ingman and Gyllensten 2003). No convincing ties to other particular branches of M or N outside the Southwest Pacific have yet been presented, which means that the ancient Eurasian origins of these peoples remain an open question.

GenBank Accessions. Accession numbers AY956412, AY956413, and AY956414 have been assigned to the whole mtDNA genome Q2 sequences (see figure 2).

Acknowledgements

The National Science Foundation, the Wenner-Gren Foundation for Anthropological Research, and The National Geographic Exploration Fund, and internal funds from Temple University, The University of Michigan, and The University of Pennsylvania, supported this research. We especially thank Heather Norton, Dan Hrdy, Charles Mgone, Babette Zemel, Baruch Blumberg, Andrew Pawley, Glenn Summerhayes, and the people of Bougainville, New Ireland, Lavongai, and New Britain who participated in the study and our collaborators at the Papua New Guinea Institute for Medical Research. We also thank two anonymous reviewers for their constructive suggestions.

Appendix

An average substitution rate per np per year for HVS1 (more accurately nps 16090-16365) has been estimated as one per np per 20,180 years (Forster et al. 1996), but we extrapolated this substitution rate to include HVS2 (information available on nps 35-315) by comparing known variability in HVS1 to HVS1 plus HVS2. First, we estimated a weighted substitution rate per np across both HVS1 and HVS2 from data in table 2 of Forster et al. (2002) on Eurasian mtDNA trees, giving less or no weight to the most mutable nps. Five hypermutable sites were excluded (nps16189, 16362, 146, 152, and 195) and the 21 sites with three to six independent substitutions were assigned half weights (six substitutions - at nps16129 and 16311; five – nps16093, 16304, 150, and 200; four – nps16126, 16274, 16278, and 228; three – nps 93, 151, 189, 194, 199, 204, 16179, 16234, 16256, 16266, and 16357). The resulting weighted substitution rates per np were calculated as follows: HVS 1 = 0.40; HVS 2 = 0.26; and HVS 1 and 2 combined = 0.353. The ratio of the HVS 1 rate to that for the entire control region is then 1.2329, or ~20% faster. Therefore, as the estimated average substitution rate for HVS 1 (nps16090 -16365) is one per 20,180 years, the average substitution rate for HVS1 plus HVS2 is 1 per 24,880 years.

Literature Cited

Allen, J. 2003. Discovering the Pleistocene in Island Melanesia. Pp. 33-42 *in* C. Sand, ed. Pacific Archaeology: Assessments and Anniversary of the First Lapita Excavation (July 1952) Koné, Nouméa, 2002. Le Cahiers de l'Archéologie en Nouvelle-Calédonie, Nouméa, New Caledonia.

Ballinger, S. W., T. G. Schurr, A. Torroni, Y. Y. Gan, J. A. Hodge et al. 1992. Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient Mongoloid migrations. Genetics **130:** 139-152.

Bandelt, H. J., P. Forster, B. C. Sykes and M. B. Richards. 1995. Mitochondrial portraits of human populations using median networks. Genetics **141:** 743-753.

Bellwood, P., G. Nitihaminoto, G. Irwin, A. Gunadi, Waluyo et al. 1998. 35,000 years of prehistory in the northern Moluccas. Pp. 233-275 *in* G.-D. Bartstra, ed. Bird's Head Approaches. Iryan Jaya Studies - A Programme for Interdisciplinary Research. Balkema, Rotterdam.

Birdsell, J. B. 1967. Preliminary data on the trihybrid origin of the Australian Aborigines. Arch. and Phys. Anth. in Oceania **2:** 100-155.

Bowdler, S. 1993. Sunda and Sahul: a 30KYR BP culture area? Pp. 60-70 *in* M. A. Smith, Spriggs, M., and Fankhauser B., ed. Sahul in Review: Pleistocene Archaeology in Australia, New Guinea and Island Melanesia. Department of Prehistory, Research School of Pacific Studies, Australian National University, Canberra.

Bowern, C., and H. Koch (Editors), 2004 Australian languages: classification and the comparative method. John Benjamins, Amsterdam.

Chappell, J. 2000. Pleistocene seed beds of western Pacific maritime cultures and the importance of chronology. Pp. 77-98 *in* S. O'Connor and P. Veth, eds., East of Wallace's Line: Studies of Past and Present Maritime Cultures of the Indo-Pacific region. Balkema, Rotterdam.

Cox, M. P. 2003. Genetic Patterning at Austronesian Contact Zones, pp. 270. University of Otago, Dunedin, New Zealand (doctoral dissertation).

Dunn, M., A. Terrill and G. Reesink. 2002. The East Papuan languages: A preliminary typological appraisal. Ocean. Ling. **41:** 28-62.

Evans, N., and R. Jones 1997. The cradle of the Pama-Nyungans: linguistic and archaeological speculations`. Pp. 376-384 *in* P. McConvell and N. Evans, eds., Archaeology and Linguistics: Global Perspectives on Ancient Australia. Oxford University Press, Melbourne.

Flannery, T. F. 1994. The Future Eaters: an Ecological History of the Australian Lands and People. Reed Books, Chatswood, Australia.

Forster, L., P. Forster, S. Lutz-Bonengel, H. Willkomm and B. Brinkmann. 2002. Natural radioactivity and human mitochondrial DNA mutations. Proc. Natl. Acad. Sci. USA **99:** 13950-13954.

Forster, P., R. Harding, A. Torroni and H.-J. Bandelt. 1996. Origin and evolution of Native American mtDNA variation: a reappraisal. Am. J. Hum. Genet. **59:** 935-945.

Forster, P., A. Torroni, C. Renfrew and A. Rohl. 2001. Phylogenetic star contraction applied to Asian and Papuan mtDNA evolution. Mol. Biol. Evol. **18:** 1864-1881.

Friedlaender, J. S., F. Gentz, K. Green and D. A. Merriwether. 2002. A cautionary tale on ancient migration detection: mitochondrial DNA variation in Santa Cruz Islands, Solomon Islands. Hum. Biol. **74:** 453-471.

Friedlaender, J. S., F. Gentz, F. Thompson, F. Kaestle, T. G. Schurr et al. in press. Mitochondrial genetic diversity and its determinants in Island Melanesia *in* A. Pawley, R. Attenborough, J. Golson and R. Hyde, eds., Papuan Pasts: Studies in the cultural, linguistic and biological history of the Papuan speaking peoples. Pacific Linguistics, Canberra.

Gao, X., and S. W. Serjeantson. 1991. Diversity in HLA-DR4-related DR,DQ haplotypes in Australia, Oceania, and China. Hum. Immunol. **32:** 269-276.

Green, R. C. 2003. The Lapita horizon and traditions - Signature for one set of Oceanic migrations. Pp. 95-120 *in* C. Sand, ed. Pacific Archaeology: Assessments and Anniversary of the First Lapita Excavation (July 1952) Koné, Nouméa, 2002. Le Cahiers de l'Archéologie en Nouvelle-Calédonie, Nouméa, New Caledonia.

Groube, L. e. a. 1986. A 40,000 year old occupation site at Huon Peninsula, Papua New Guinea. Nature. Nature **324:** 453-455.

Hagelberg, E., M. Kayser, M. Nagy, L. Roewer, H. Zimbahl et al. 1999. Molecular genetic evidence for the human settlement of the Pacific: analysis of mitochondrial DNA, Y chromosome and HLA markers. Phil.Trans. Royal Soc. London (B) **354:** 141-152.

Horton, D. R. 1999. The Pure State of Nature: Sacred Cows, Destructive Myths and the Environment. Allen and Unwin, Sydney.

Huoponen, K., T. G. Schurr, Y. Chen and D. C. Wallace. 2001. Mitochondrial DNA variation in an aboriginal Australian population: evidence for genetic isolation and regional differentiation. Hum. Immunol. **62:** 954-969.

Ingman, M., and U. Gyllensten. 2003. Mitochondrial genome variation and evolutionary history of Australian and New Guinean aborigines. Genome Research **13:** 1600-1606.

Ingman, M., H. Kaessmann, S. Paabo and U. Gyllensten. 2000. Mitochondrial genome variation and the origin of modern humans. Nature **408:** 708-713.

Kirch, P. V. 1997. The Lapita People: Ancestors of the Oceanic world. Blackwell Scientific, Cambridge MA.

Kirk, R. L. 1989. Population genetic studies in the Pacific: red cell antigen, serum protein, and enzyme systems. Pp. 60-119 *in* A. V. S. Hill and S. W. Serjeantson, eds., Colonization of the Pacific: A Genetic Trail. Clarendon Press, Oxford.

Kivisild, T., H. V. Tolk, J. Parik, Y. Wang, S. S. Papiha et al. 2002. The emerging limbs and twigs of the East Asian mtDNA tree. Mol Biol Evol **19:** 1737-1751.

Leavesley, M. G., M. I. Bird, L. K. Fifield, P. A. Hausladen, G. M. Santos et al. 2002. Buang Merabak: Early evidence for human occupation in the Bismarck Archipelago, Papua New Guinea. Aust. Arch. **54:** 55-57.

Lum, J. K., and R. L. Cann. 2000. mtDNA lineage analyses: origins and migrations of Micronesians and Polynesians. Am. J. Phys. Anthropol. **113:** 151-168.

Macaulay, V., M. Richards, E. Hickey, E. Vega, F. Cruciani et al. 1999. The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. Am. J. Hum. Genet. **64:** 232-249.

O'Connell, J. F., and J. Allen. 1998. When did human first arrive in greater Australia and why is it important to know? Evolutionary Anthropology **6:** 132-146.

O'Connell, J. F., and J. Allen. 2004. Dating the colonization of Sahul (Pleistocene Australia-New Guinea): a review of recent research. Journal of Archaeological Science **31:** 835-853.

O'Connor, S., and J. Chappell 2003. Colonisation and coastal subsistence in Australia and Papua New Guinea: different timing, different modes? Pp. 17-32 *in* C. Sand, ed. Pacific Archaeology: Assessments and Anniversary of the First Lapita Excavation (July 1952) Koné, Nouméa, 2002. Le Cahiers de l'Archéologie en Nouvelle-Calédonie, Nouméa, New Caledonia.

Parsons, T. J., D. S. Muniec, K. Sullivan, N. Woodyatt, R. Alliston-Greiner et al. 1997. A high observed substitution rate in the human mitochondrial DNA control region. Nat. Genet. **15:** 363-368.

Pawley, A. 2003. Locating Proto-Oceanic. Pp. 17-34 *in* A. P. Malcolm Ross, and Meredith Osmond, ed. The Lexicon of Proto Oceanic: The Physical Environment. Research School of Pacific and Asian Studies, The Australian National University, Canberra.

Pawley, A. in press. The chequered career of the Trans New Guinea Phylum: Recent historical research and its implications *in* A. Pawley, R. Attenborough, J. Golson and R. Hide, eds., Papuan Pasts: Investigations into the cultural, linguistic and biological history of the Papuan speaking peoples. Pacific Linguistics, Canberra.

Redd, A. J., and M. Stoneking. 1999. Peopling of Sahul: mtDNA variation in Aboriginal Australian and Papua New Guinean populations. Am. J. Hum. Genet. **65:** 808-828.

Redd, A. J., N. Takezaki, S. T. Sherry, S. T. McGarvey, A. S. M. Sofro et al. 1995. Evolutionary History of the COII/tRNALys Intergenic 9 Base Pair Deletion in Human Mitochondrial DNAs from the Pacific. Mol. Biol. Evol. **12:** 604-615.

Rieder, M. J., S. L. Taylor, V. O. Tobe and D. A. Nickerson. 1998. Automating the identification of DNA variations using quality-based fluorescence re-sequencing: analysis of the human mitochondrial genome. Nucleic Acids Res. **26:** 967-973.

Roberts-Thomson, J., J. Martinson, J. Norwich, R. Harding, J. Clegg et al. 1996. An ancient common origin of Aboriginal Australians and New Guinean Highlanders is supported by alpha-globin haplotype analysis. Am. J. Hum. Genet. **58:** 1017-1024.

Ross, M. 2001. Is there an East Papuan Phylum? Evidence from Pronouns. Pp. 301-321 *in* A. Pawley, Ross, Malcolm, and Tryon, Darryl, ed. The Boy from Bundaberg: Studies in Melanesian Linguistics in Honour of Tom Dutton. Pacific Linguistics, Canberra.

Ross, M. in press. Pronouns as a preliminary diagnostic for grouping Papuan languages *in* A. Pawley, R. Attenborough, J. Golson and R. Hide, eds., Papuan Pasts: Investigations into the cultural, linguistic and biological history of the Papuan speaking peoples. Pacific Linguistics, Canberra.

Saillard, J., P. Forster, N. Lynnerup, H. J. Bandelt and S. Norby. 2000. mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. Am. J. Hum. Genet. **67:** 718-726.

Spriggs, M. J. T. 1997. The Island Melanesians. Blackwell Scientific, Cambridge MA.

Stoneking, M., L. B. Jorde, K. Bhatia and A. C. Wilson. 1990. Geographic variation in human mitochondrial DNA from Papua New Guinea. Genetics **124:** 717-733.

Sykes, B., A. Lieboff, J. Low-Beeer, S. Tetzner and M. Richards. 1995. The origins of the Polynesians: An interpretation from mitochondrial lineage analysis. Am. J. Hum. Genet. **57:** 1463-1475.

Tajima, A., M. Hayami, K. Tokunaga, T. Juji, M. Matsuo et al. 2004. Genetic origins of the Ainu inferred from combined DNA analyses of maternal and paternal lineages. J. Hum. Genet. **49:** 187-193.

Tommaseo-Ponzetta, M., M. Attimonelli, M. De Robertis, F. Tanzariello and C. Saccone. 2002. Mitochondrial DNA variability of West New Guinea populations. Am. J. Phys. Anthropol. **117:** 49-67.

Tsintsof, A. S., M. S. Hertzberg, J. F. Prior, K. N. Mickleson and R. J. Trent. 1990. Alpha-globin gene markers identify genetic differences between Australian aborigines and Melanesians. Am. J. Hum. Genet. **46:** 138-143.

Weitkamp, L., Shreffler, DC, and Saave, JJ. 1969. Serum albumin variants in New Guinea indigenes. . Vox Sang **17:** 237-240.

Yen, D. 1995. The development of Sahul agriculture with Australia as a bystander. Pp. 831-847 *in* J. Allen and J. F. O'Connell, eds., Transitions: Pleistocene to Holocene in Australia and New Guinea. Antiquity 69, Special Number 265.

Figure Legends

Figure 1. Schematic mtDNA phylogeny for Eurasia. The African L3 root splits into the Eurasian N and Asian M branches. Haplogroups P and Q are ancient subdivisions of each major branch.

Figure 2. Haplogroup Q branching from the African L3 root, based on whole mtDNA sequencing. Note the deep stem from M, the long primary branch lengths, and the short terminal branches. No Qs have been found in Australia. Q as defined by Forster et al. (2001) is boxed as the "Original definition." Provenience abbreviations at the top are as follows: New Guniea (NG); New Britain (NB); New Ireland (NI); and Nasioi (of Bougainville). GenBank accession numbers are listed with the AY and AF prefixes. Other designations are from Redd and Stoneking (1999), Stoneking et al. (1990), Ingman and Gyllensten (2003) or the current series (UV). PNG C89 was previously identified as a Q by Kivisild et al. (2002). Sub-branches of Q refer to those shown in table 2. Control region mutations are in bold, transversions have a base suffix, recurrent mutations in the same tree are underlined, back mutations are italicized, and 16519 is ignored.

Figure 3. A haplogroup Q median network phylogeny, based on control region sequences of the 210 Q samples from the current study. The major Q1 and Q2 branches are star-like in their configurations, with the majority of samples outside the centers, suggesting ancient population expansion(s). Node sizes are proportional to the number of haplotypes contained within each. Nucleotide pair (np) numbers are transitions from the CRS, with bases noted for transversions. The star indicates the branching from ancestral M.

Figure 4. Haplogroup P branching from the African root L3, based on whole genome sequencing. In contrast to Q, the different branches of P share only one defining mutation

(15607), the terminal branches are generally very long, and only P1 from Melanesia has extensive internal branching. Abbreviations follow figure 2, with the addition of Australian Aboriginals (AUS). H35, H6, and C80 were previously identified as Ps by Kivisild et al. (2002). Half the sequences are Australian, and only P3 occurs there and New Guinea.

Figure 5. Haplogroup P median network phylogeny, based on control region sequences on the 104 P samples from the current study. Only P1 (from New Guinea) has a star-like array, although its core is almost empty, consistent with its great age. Other branches are indeterminate in part because relatively few control region mutations distinguish them. The star indicates the origin from R.

Figure 6. Population locations and frequencies of P and Q haplotypes in the Southwest Pacific. Numbers are given in table 2.

Figure 7. Population locations and frequencies of P and Q haplotypes in the Bismarck Archipelago and Bougainville Island. Numbers are given in table 3. Underlined names are the Papuan-speaking locations.

Table 1. **Coalescence Estimates for the main P and Q branches**

Haplogroup	ρ[a]	ρ (Years)	σ	σ (years)	*n*	Source
Q1	2.0214	49,856[b]	0.4353	10,832	192	This study
"Q"[c]	0.7027	15,300[d]	0.1622	3,500	37	Forster et al. 2001
Q2	1.4839	36,599	0.4977	11,617	87	This study
P1	2.0806	51,316	0.3861	9,827	90	This study
"P/p94"	1.5263	33,300	0.3722	8,100	19	Forster et al. 2001

Total P and Q Mean Pairwise Differences[e]

Q	7.936
P	7.076

[a] As defined in Saillard et al. (2000), ρ is the mean mutational distance from the haplogroup node of interest (i.e. its center), and σ is its standard error. The value of σ does not include uncertainty concerning mutation rates.
[b] Based on one mutation per 24,880 years across HVS1 and 2 (see Appendix for the rate estimation).
[c] Quotation marks denote equivalent haplogroup and estimates from high resolution RFLP analysis (Forster et al. 2001).
[d] Based on one mutation per 21,800 for coding region RFLP data from (Forster et al. 2001, Table 2).
[e] From mismatch distributions presented in Supplemental Materials Figure 1, after the method of Rogers and Harpending (1992).

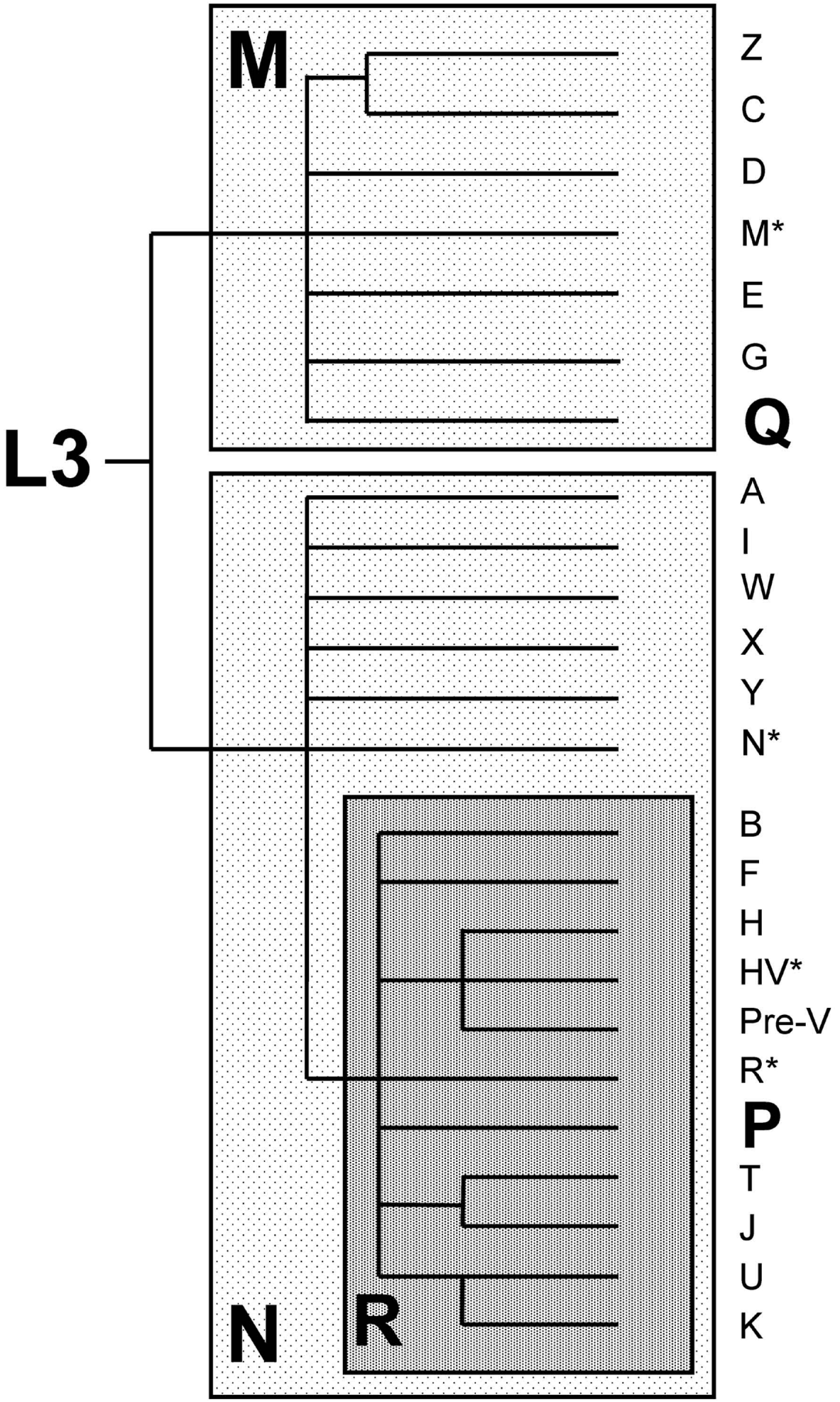
Table 2. Distribution of Control Region Samples from Haplogroups P and Q in the Southwest Pacific

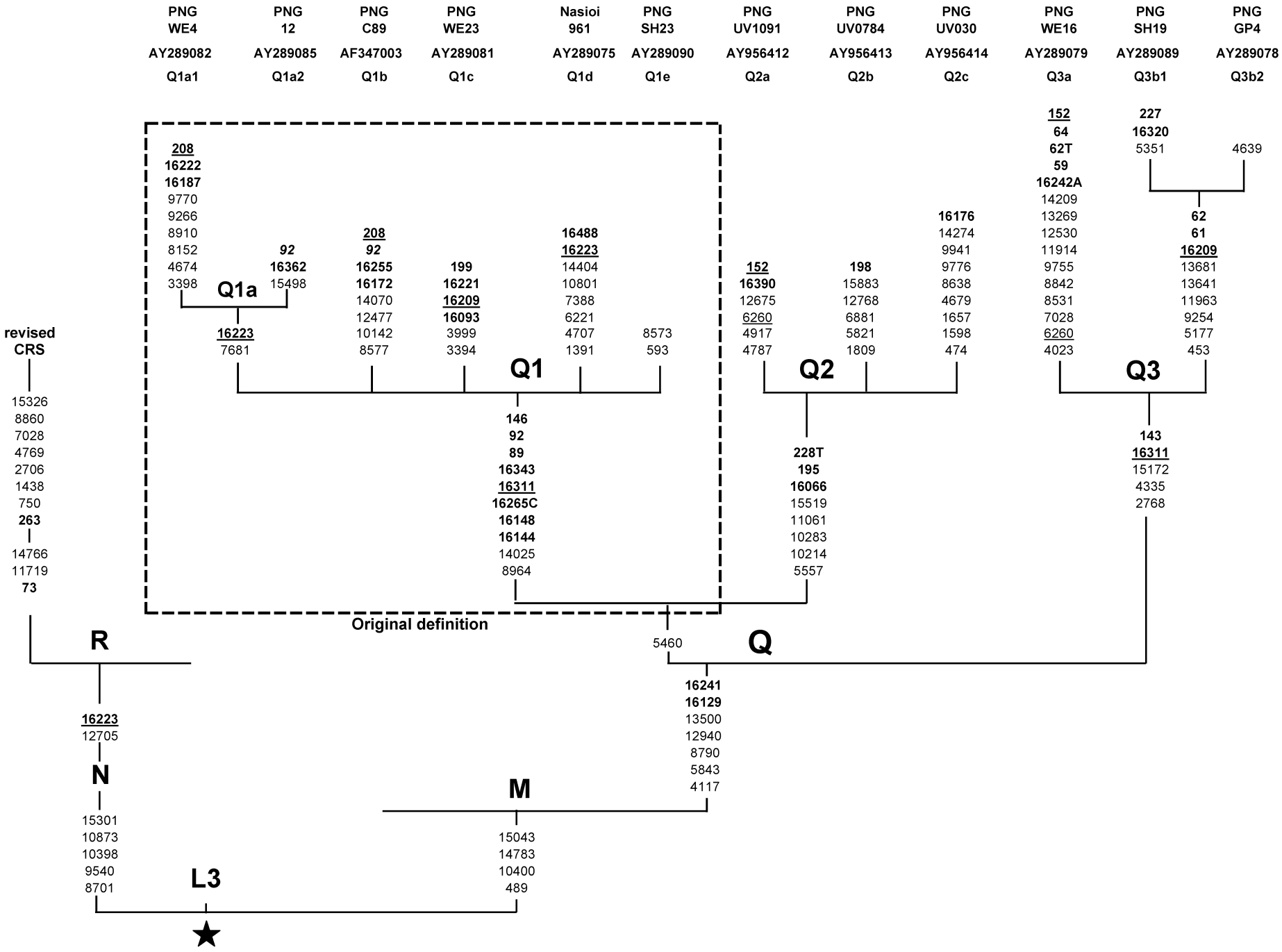
Major Region	Region	Population	Haplogroups							
			P		Q			B	Other	Total
			P1	P2	Q1	Q2	Q3			
New Guinea	West Papua	Southwest Riverine	7	6	19			1		33
		Lowland Riverine	3		22				1	26
	PNG	PNG	2		4		1	7	5	19
	PNG Highlands	West Highlands	10	1	6				1	18
		Fringe Highlands	3		2		2		8	15
		East Highlands	2		1				1	4
		Morobe Highlnds	5	1	1			7	6	20
	PNG Coastal	Sepik	5		4		2	4		15
		Markham	9	5	12	10		22	9	67
		South Coast	1					10	2	13
	PNG islands	Misima	4	2					1	7
		Rossel	3						2	5
Island Melanesia	New Britain	West New Britain			11	14		17	60	102
		East New Britain	3		9	21		12	87	132
	Mussau	Mussau				1		1		2
	New Ireland	New Ireland	1			5		72	13	91
	Bougainville	North Bougainville	4		38			17	16	75
		South Bougainville						24	7	31
	Solomon Is.	Solomon Islands	1					18	7	26
		Santa Cruz	5		12	7		23	22	69
	Vanuatu	Vanuatu	3			2		8	9	22
	New Caledonia	New Caledonia	2	8		1		6	14	31
	Fiji	Fiji		1	2	1		9	3	16
Micronesia		Kiribati						4	2	6
		Ontong Java						16	8	24
Polynesia		Samoa						4		4
		Tonga						3		3
		Wallis						2		2
Totals			74	30	143	62	5	286	290	878

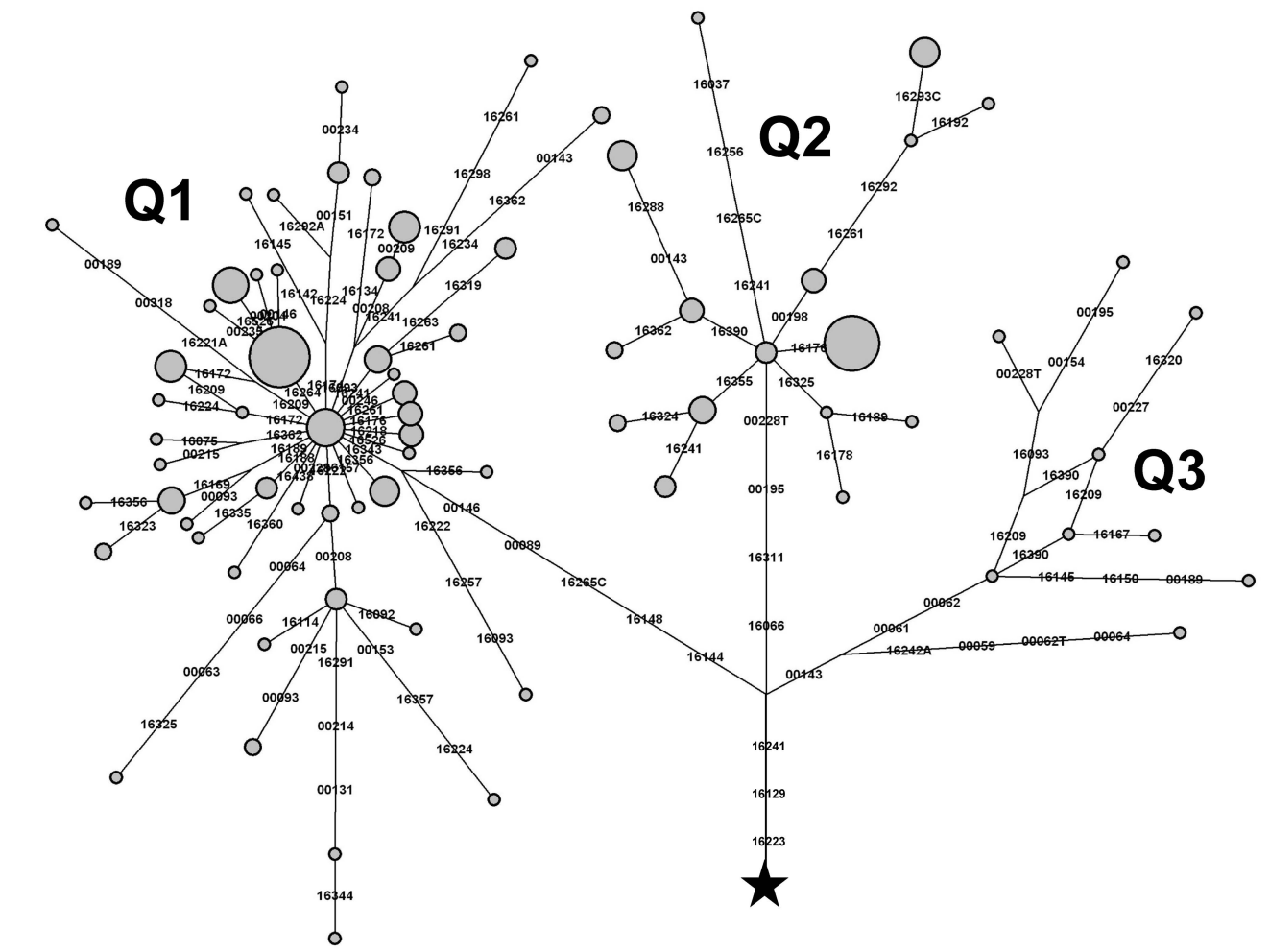
Table 3. Distribution of haplogroups P and Q in populations from the Bismarck Archipelago and Bougainville Island.

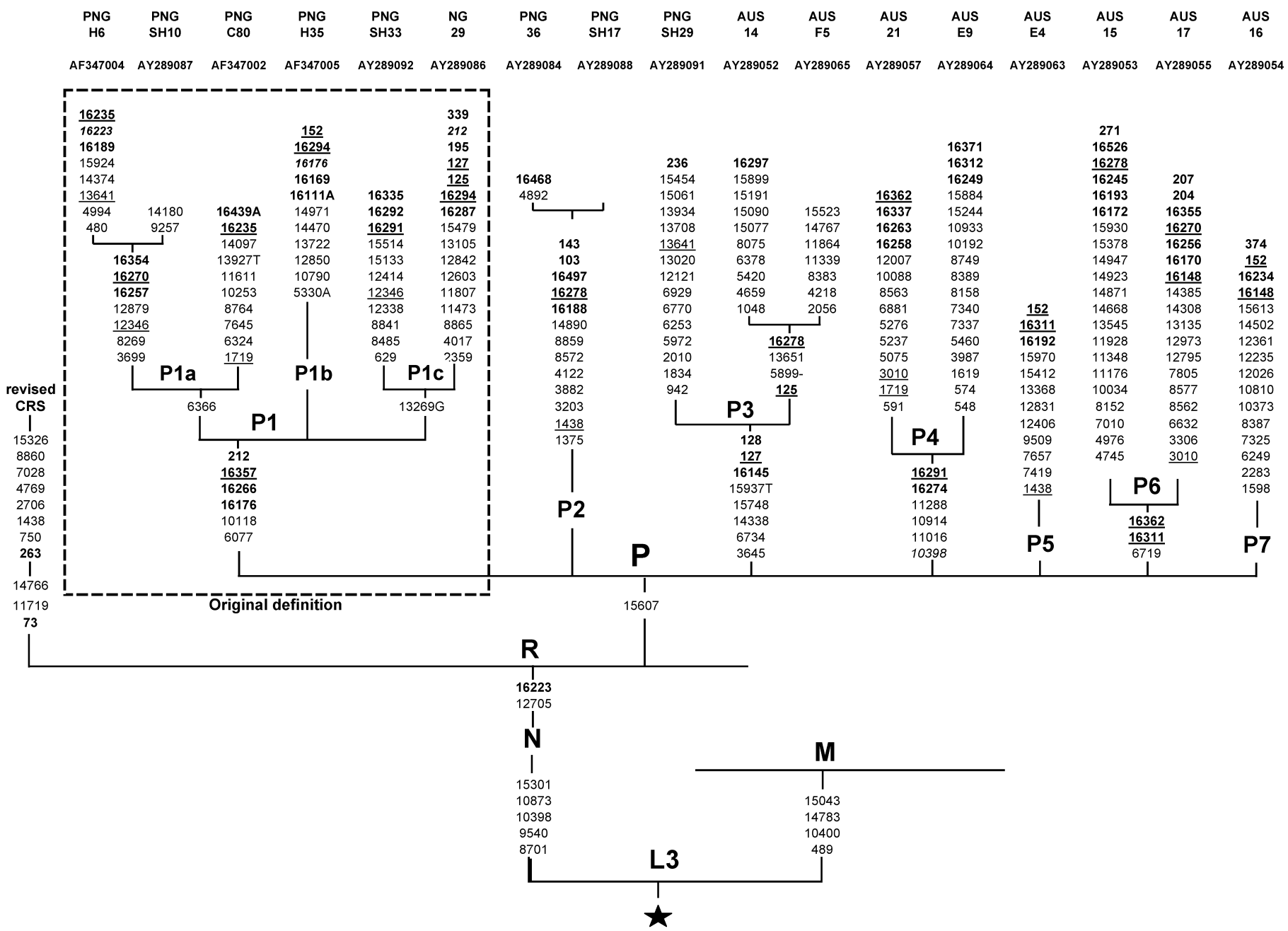
Island	Island section	Population	Haplogroups					
			P1	Q1	Q2	B	Other	Total
New Britain	West New Britain	Ata (Uasilau)[f]			5		12	17
		Ata (Lugei)		1	3		12	16
		Kol		7	6	14	35	62
		Mengen (Pomio)		3		3	1	7
	East New Britain	Baining (Marabu)	2	1	21		34	58
		Baining (Malasait)					19	19
		Sulka (Ganai)		3		2	5	10
		Sulka (Watwat)		2		2	6	10
		Tolai (Kabakada)		3		1	8	12
		Tolai (Vunairoto)				4	3	7
		Tolai (Matupit)	1			1	3	5
		Tolai				2	9	11
Mussau		Kapugu			1	1		2
New Ireland		Kavieng	1			1		2
		Kuot (Lamalaua)				9		9
		Kuot (Kabil)				38	3	41
		Kuot				9		9
		Madak			5	15	10	30
Bougainville	North Bougainville	Aita	4	31		1	2	38
		Rotokas		5			6	11
		Eivo		2		13	4	19
		Simeku				3	4	7
	South Bougainville	Nasioi				1	6	7
		Torau				4	1	5
		Nagovisi				15		15
		Siwai				4		4
Totals			8	58	41	143	170	433

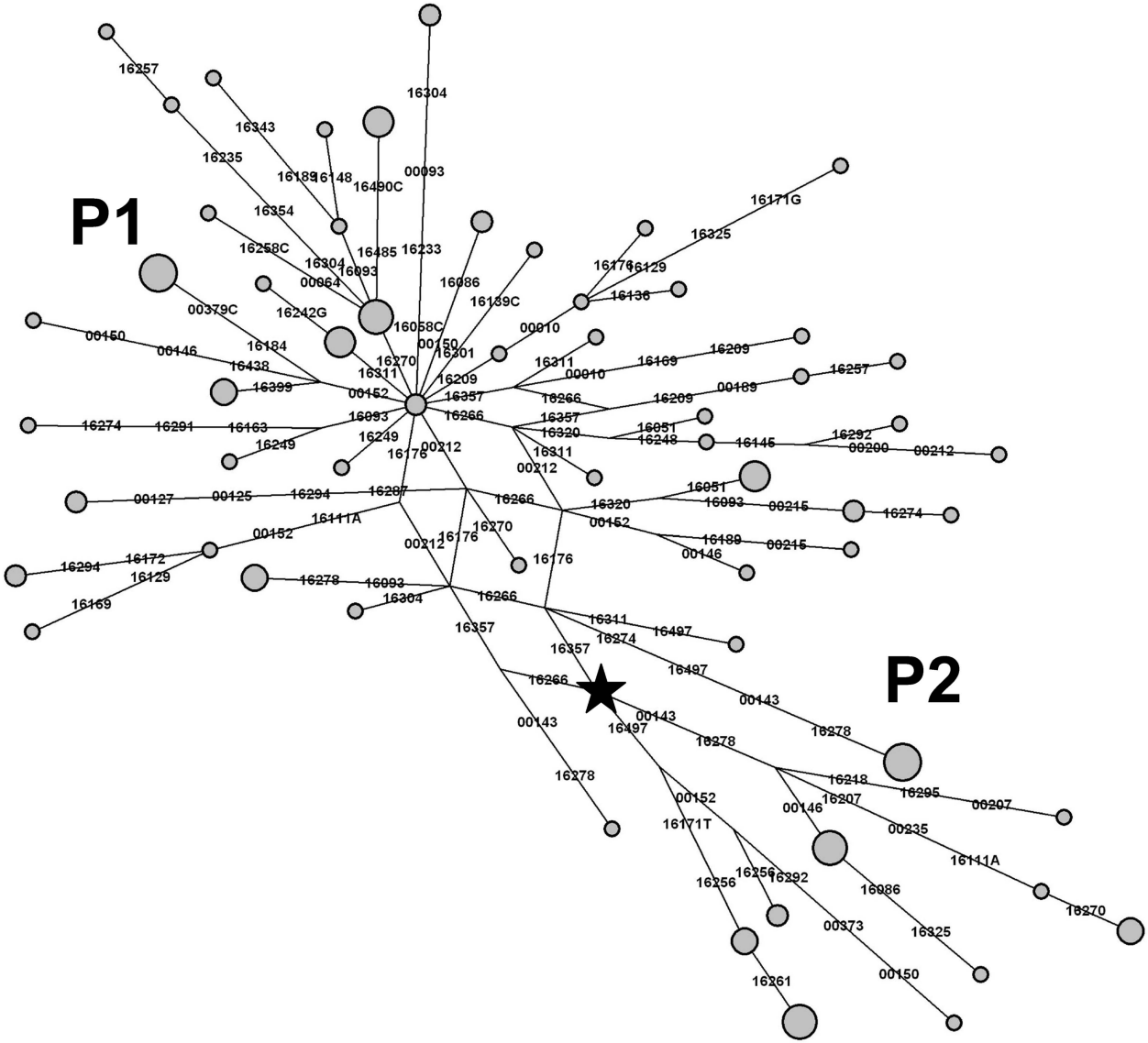
[f] Papuan speaking groups are underlined

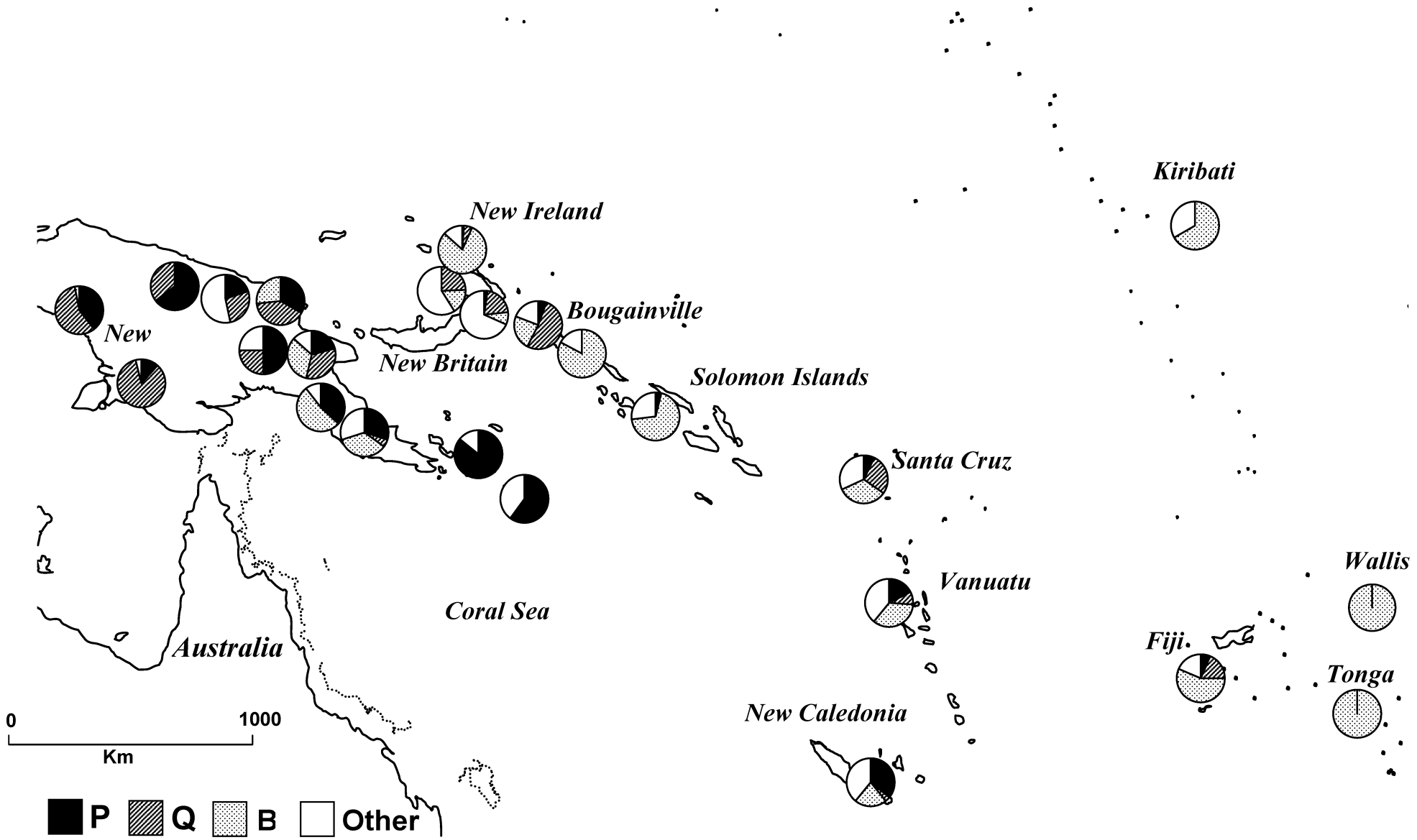

M
L3
N
R
Z
C
D
M*
E
G
Q
A
I
W
X
Y
N*
B
F
H
HV*
Pre-V
R*
P
T
J
U
K

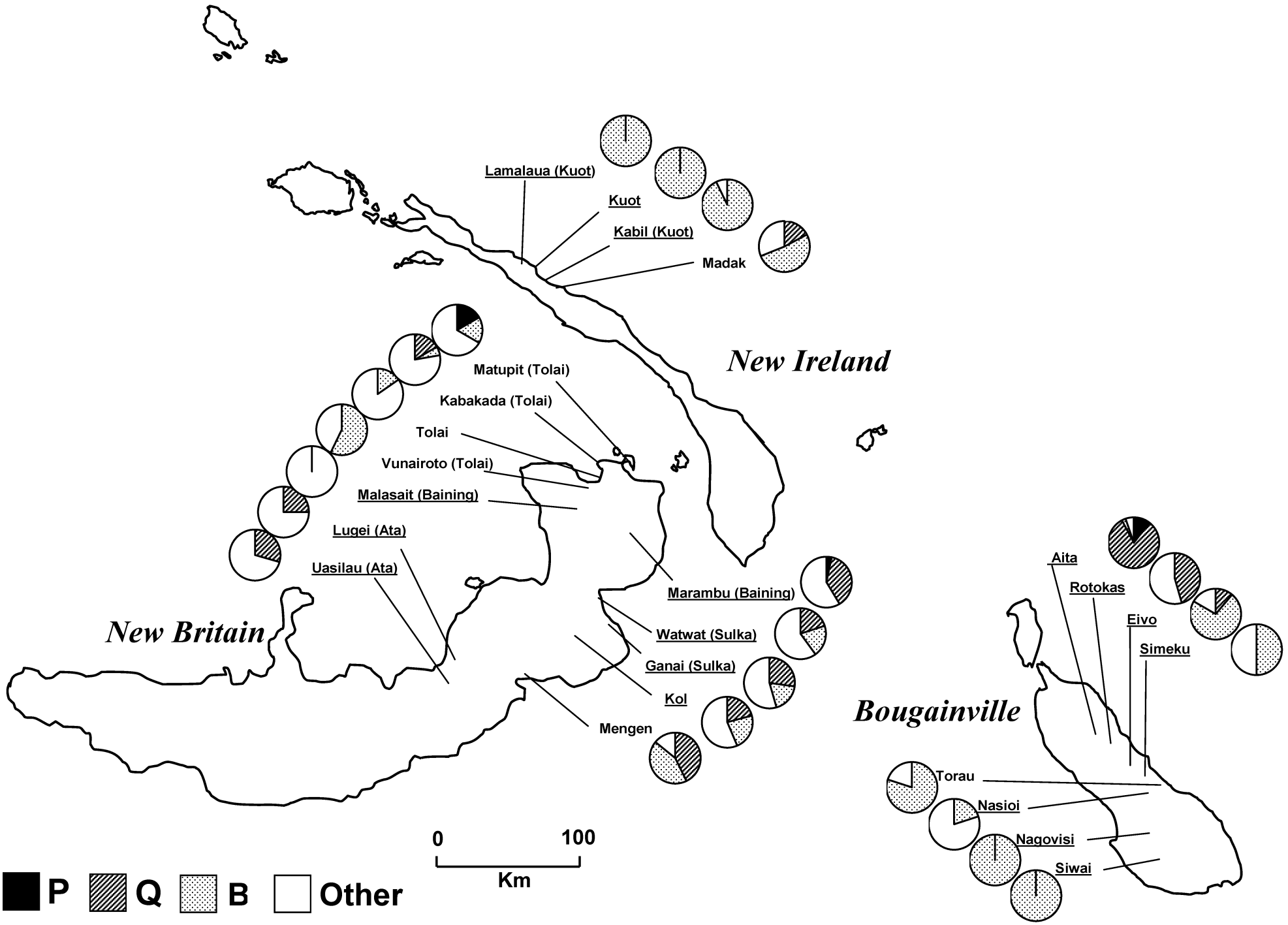

PNG WE4 AY289082 Q1a1
PNG 12 AY289085 Q1a2
PNG C89 AF347003 Q1b
PNG WE23 AY289081 Q1c
Nasioi 961 AY289075 Q1d
PNG SH23 AY289090 Q1e
PNG UV1091 AY956412 Q2a
PNG UV0784 AY956413 Q2b
PNG UV030 AY956414 Q2c
PNG WE16 AY289079 Q3a
PNG SH19 AY289089 Q3b1
PNG GP4 AY289078 Q3b2
revised CRS
15326 8860 7028 4769 2706 1438 750 263 14766 11719 73
R
16223 12705
N
15301 10873 10398 9540 8701
L3
★
208 16222 16187 9770 9266 8910 8152 4674 3398
92 16362 15498
Q1a
16223 7681
208 92 16255 16172 14070 12477 10142 8577
199 16221 16209 16093 3999 3394
16488 16223 14404 10801 7388 6221 4707 1391
8573 593
Q1
146 92 89 16343 16311 16265C 16148 16144 14025 8964
Original definition
152 16390 12675 6260 4917 4787
198 15883 12768 6881 5821 1809
16176 14274 9941 9776 8638 4679 1657 1598 474
Q2
228T 195 16066 15519 11061 10283 10214 5557
5460
Q
16241 16129 13500 12940 8790 5843 4117
M
15043 14783 10400 489
152 64 62T 59 16242A 14209 13269 12530 11914 9755 8842 8531 7028 6260 4023
227 16320 5351
4639
62 61 16209 13681 13641 11963 9254 5177 453
Q3
143 16311 15172 4335 2768


Q1
Q2
Q3
16223
16129
16241
16311
16066
16144
16148
16265C
00089
00146
16222
16356
16257
16093
16357
16224
00131
16344
00214
16225
00063
00066
16114
16092
00064
00208
16360
16335
00093
16323
16165
16356
00215
16075
16189
16362
16472
16209
16224
16221A
00318
00189
16145
16292A
00151
00234
16261
16298
00143
16362
16291
16234
16319
16263
16241
00208
16134
16224
16142
00209
16172
16390
16355
16325
16324
16241
00228T
16189
16178
16176
00198
16288
16256
16037
16265C
16292
16261
16293C
16192
00143
00195
16320
00227
00154
00228T
16093
16390
16209
16167
16209
16390
16145
16150
00189
00062
00061
16242A
00059
00062T
00064
00143


PNG H6
PNG SH10
PNG C80
PNG H35
PNG SH33
NG 29
PNG 36
PNG SH17
PNG SH29
AUS 14
AUS F5
AUS 21
AUS E9
AUS E4
AUS 15
AUS 17
AUS 16
AF347004
AY289087
AF347002
AF347005
AY289092
AY289086
AY289084
AY289088
AY289091
AY289052
AY289065
AY289057
AY289064
AY289063
AY289053
AY289055
AY289054
16235
16223
16189
15924
14374
13641
4994
480
14180
9257
16354
16270
16257
12879
12346
8269
3699
16439A
16235
14097
13927T
11611
10253
8764
7645
6324
1719
152
16294
16176
16169
16111A
14971
14470
13722
12850
10790
5330A
16335
16292
16291
15514
15133
12414
12346
12338
8841
8485
629
339
212
195
127
125
16294
16287
15479
13105
12842
12603
11807
11473
8865
4017
2359
P1a
P1b
P1c
P1
6366
13269G
212
16357
16266
16176
10118
6077
16468
4892
143
103
16497
16278
16188
14890
8859
8572
4122
3882
3203
1438
1375
236
15454
15061
13934
13708
13641
13020
12121
6929
6770
6253
5972
2010
1834
942
16297
15899
15191
15090
15077
8075
6378
5420
4659
1048
16278
13651
5899-
125
15523
14767
11864
11339
8383
4218
2056
16362
16337
16263
16258
12007
10088
8563
6881
5276
5237
5075
3010
1719
591
16371
16312
16249
15884
15244
10933
10192
8749
8389
8158
7340
7337
5460
3987
1619
574
548
152
16311
16192
15970
15412
13368
12831
12406
9509
7657
7419
1438
271
16526
16278
16245
16193
16172
15930
15378
14947
14923
14871
14668
13545
11928
11348
11176
10034
8152
7010
4976
4745
207
204
16355
16270
16256
16170
16148
14385
14308
13135
12973
12795
7805
8577
8562
6632
3306
3010
374
152
16234
16148
15613
14502
12361
12235
12026
10810
10373
8387
7325
6249
2283
1598
P2
P3
P4
P5
P6
P7
P
15607
16145
15937T
15748
14338
6734
3645
16291
16274
11288
10914
11016
10398
16362
16311
6719
revised CRS
15326
8860
7028
4769
2706
1438
750
263
14766
11719
73
Original definition
R
16223
12705
N
15301
10873
10398
9540
8701
M
15043
14783
10400
489
L3


P1
P2
16257
16304
16343
16235
16189
16148
16490C
00093
16354
16258C
16485
16233
16304
16093
00064
16086
16189C
16242G
00379C
16058C
00150
00146
16184
16438
16399
16270
16301
16311
16209
00152
16357
16266
16093
16274
16291
16163
16249
16176
00212
16311
16320
16248
16145
16292
00200
00212
00127
00125
16294
16287
16111A
16270
16172
16129
16169
16278
16304
16357
16274
16497
16143
16278
16218
16295
00207
16207
00235
16171T
16256
16292
16986
16270
00373
16325
16261
00150
16171G
16325
16176
16129
16136
00010
16169
16209
16257
00189
16051
00215
16189
00146


Kiribati
New Ireland
Bougainville
New
New Britain
Solomon Islands
Santa Cruz
Wallis
Vanuatu
Coral Sea
Fiji.
Tonga
Australia
New Caledonia
0
1000
Km
P
Q
B
Other


Lamalaua (Kuot)
Kuot
Kabil (Kuot)
Madak
New Ireland
Matupit (Tolai)
Kabakada (Tolai)
Tolai
Vunairoto (Tolai)
Malasait (Baining)
Lugei (Ata)
Uasilau (Ata)
Marambu (Baining)
New Britain
Watwat (Sulka)
Ganai (Sulka)
Kol
Mengen
Aita
Rotokas
Eivo
Simeku
Bougainville
Torau
Nasioi
Nagovisi
Siwai
0
100
Km
P
Q
B
Other