

Reconstructing the origin of the Lapita Cultural Complex: mtDNA analyses of East Sepik Province, PNG

Miguel G. Vilar · Akira Kaneko ·
Francis W. Hombhanje · Takahiro Tsukahara ·
Ilomo Hwaihwanje · J. Koji Lum

Received: 3 December 2007 / Accepted: 29 April 2008 / Published online: 23 May 2008
© The Japan Society of Human Genetics and Springer 2008

Abstract The colonization of Oceania occurred in two waves. By 32,000 BP, humans had reached New Guinea and settled all intervisible islands east to the Solomon Islands. Around 3,500 BP, a distinct intrusive group from Southeast Asia reached coastal New Guinea, integrated their components with indigenous resources, and gave rise to the Lapita Cultural Complex. Within 2,500 years, Lapita and its descendant cultures colonized the Pacific. To uncover the origin of the Lapita Cultural Complex, we

analyzed the hypervariable region I of the mitochondrial deoxyribonucleic acid (mtDNA) in 219 individuals from eight East Sepik Province villages: two villages in each of four environmental zones. Same-zone villages spoke different languages: one Austronesian and three Papuan (Arapesh, Abelam, and Boiken). Our analysis examined whether language or geography better predicted gene flow. In general, language better predicted genetic affinities. Boiken villages across all four zones showed no significant genetic difference (F_{ST} P value > 0.05). In contrast, the Austronesian village was significantly different to most other villages ($P < 0.05$). Only the mountains and coast showed zonal gene flow ($P > 0.05$). We interpret the data to reflect limited gene flow inland by Austronesians overshadowed by a regional displacement by inland Boiken speakers migrating seaward. These results are consistent with oral histories and ethnographic accounts.

GenBank accession numbers: EU312981–EU313199.

M. G. Vilar · J. K. Lum
Laboratory of Evolutionary Anthropology and Health,
Binghamton University, Binghamton, NY, USA

M. G. Vilar (✉) · J. K. Lum
Department of Anthropology, Binghamton University,
Binghamton, NY 13902-6000, USA
e-mail: mvilar1@binghamton.edu

A. Kaneko
Malaria Research Unit, Unit for Infectious Diseases,
Department of Medicine, Karolinska University Hospital,
Stockholm, Sweden

A. Kaneko · T. Tsukahara
Department of International Affairs and Tropical Medicine,
Tokyo's Women's Medical University, Tokyo, Japan

F. W. Hombhanje
Faculty of Health Sciences, Divine Word University,
Madang, Papua New Guinea

I. Hwaihwanje
Modilon Teaching Hospital, Madang, Papua New Guinea

J. K. Lum
Department of Biological Sciences, Binghamton University,
Binghamton, NY, USA

Keywords Austronesian · Papuan · Near Oceania ·
Papua New Guinea · Boiken · mtDNA · Gene flow ·
Migration · Origin · Haplogroup

Introduction

The birth of the Lapita Cultural Complex in the south-western Pacific approximately 3,500 years before present (ybp) enabled the discovery and/or settlement of every island of the world's largest ocean by 1,000 ybp. The Lapita Cultural Complex evolved from the interaction of two distinct cultures, both capable of horticulture and advanced interisland transport. How these two cultures collided, integrated local and intrusive materials and technology, and generated the innovative Lapita Cultural Complex (Green 1991a) along the north coast and offshore islands of

New Guinea has yet to be examined at the local level of intervillage genetic exchange.

Human settlement of New Guinea began in the late Pleistocene. The earliest archeological sites confirm human presence on the island by 40,000 ybp (Groube et al. 1986). Settlers first reached the Bismarck Archipelago between 35–39,000 ybp (Allen et al. 1988; Leavesely and Chappell 2004) and the northern Solomon Islands by 32,000 ybp (Wickler and Spriggs 1988). By 20,000 ybp, humans were transporting resources, including obsidian and tree kangaroos, among New Guinea, the Bismarck Archipelago, and the northern Solomon Islands (Summerhayes and Allen 1993). Furthermore, by 20,000 ybp, the island of Manus was settled (Ambrose 2002) in the first blind ocean crossing in human prehistory, and by 9,000 ybp, horticulture was developed independently in the Highlands of New Guinea (Allen 1970).

The Pleistocene-settled area extending from New Guinea to the Solomon Islands, known collectively as Near Oceania (Green 1991b), is coincident with the region's current distribution of Papuan languages, a noncohesive group of 23 language families (Ross 2005) consisting of some 950 mutually unintelligible languages (Gordon 2005).

By 3,500 ybp, a second wave of migrants arrived in Near Oceania from Island Southeast Asia (ISEA) (Diamond 1988; Kirch and Hunt 1988). With the arrival of this intrusive group came the introduction of new domestic animals from Southeast Asia, including the dog, pig, chicken, and rat (Matisoo-Smith and Robins 2004; Lum et al. 2006; Matisoo-Smith 2007; Storey et al. 2007). The intrusive Holocene settlers incorporated horticultural (e.g., tree and root crops) and technological achievements of the original Pleistocene settlers, giving rise to the innovative Lapita Cultural Complex that would eventually settle the rest of the Pacific. Evidence of this innovation is seen in the first appearance of Lapita pottery, a red slipped, lime-filled ceramic thought to be derived from similar forms found in ISEA (Kirch 2000), but stylistically unique to the southwestern Pacific. The earliest Lapita pottery was found in the Bismarck Archipelago (Kirch and Hunt 1988; Summerhayes 2001), but within a few generations, nearly identical pottery was distributed as far east as the Tongan and Samoan archipelagoes (Hendrick 1971; Green 1976; Davidson et al. 1990; Sand 1997). From the Bismarck Archipelago to western Polynesia, the distribution of Lapita pottery coincides with the current distribution of Austronesian languages, a cohesive family of languages with origins in ISEA, but spoken today as far west as Madagascar and as far east as Rapa Nui (Easter Island) (Kirch and Hunt 1988; Bellwood 1989; Spriggs 1989). The Austronesian language family includes more than 1,200 unambiguously related languages (Gordon 2005) thought

to have been dispersed with oceanic voyages within the past 5,000 years (Pawley and Green 1973; Bellwood 1991).

Since the mid-Holocene arrival of the presumed Austronesian speakers and the birth and spread of the Lapita Cultural Complex, more recent localized migrations have taken place. Boiken speakers from the Sepik plain are thought to have expanded northward, displacing some previous occupants of the Prince Alexander Mountains and the coastal regions in today's East Sepik Province, Papua New Guinea (PNG) (Roscoe 1989). Evidence supporting this expansion also comes from linguistic and archeological data (Swadling and Hide 2005).

Previous genetic studies of mitochondrial deoxyribonucleic acid (DNA) (mtDNA) from Pacific Island populations show that three major haplogroups (B, P, and Q; formerly I, IV, II, respectively; Lum et al. 1994; Lum and Cann 2000) characterize the majority of people from Melanesia (PNG, the Solomon Islands, Vanuatu, New Caledonia, and Fiji). Two of these haplogroups (Q and P) decrease in frequency as one moves east from New Guinea to Polynesia, whereas the third (B) increases along the same trajectory. Haplogroups P and Q account for approximately 98% of lineages in the Highlands of PNG (Stoneking et al. 1990), decreasing in frequency as one moves east through Melanesia, and ultimately accounting for approximately 3.5% (0–13%) of Polynesians (Lum et al. 1994; Redd et al. 1995; Sykes et al. 1995; Lum and Cann 1998, 2000). Haplogroups P and Q are likewise very rare (< 3%) in ISEA (Hill et al. 2007).

Haplogroup B frequencies range from 2% to 45% throughout ISEA (Sykes et al. 1995; Hill et al. 2007), accounting for 34% of Aboriginal Taiwanese (Sykes et al. 1995; Trejault et al. 2005). Throughout coastal and island Melanesia, haplogroup B frequencies range from 10% to 25% (Stoneking et al. 1990; Lum and Cann 1998; Friedlaender et al. 2007) yet are absent in the New Guinea Highlands (Stoneking et al. 1990; Lum and Cann 1998; Forster et al. 2001). In Fiji, the easternmost archipelago of Melanesia, haplogroup B accounts for 64% of lineages (Lum and Cann 1998). Throughout Melanesia, haplogroup B lineages are found predominantly in coastal regions and offshore islands associated with prehistoric Lapita settlements and the current distribution of Austronesian languages (Merriwether et al. 1999). Haplogroup B approaches fixation throughout Polynesia, as well as on the central-eastern Micronesian islands derived from the expansion of the Lapita Cultural Complex (Lum et al. 1994; Sykes et al. 1995; Lum and Cann 1998, 2000), with variable frequencies observed in western Micronesia (Lum and Cann 1998, 2000; Lynch et al. 2008).

The formation of the Lapita Cultural Complex and the consequent settlement of the Pacific are the result of the

innovations born from the cultural and biological integration of the Pleistocene settlers and the Holocene intruders (Green 1991a). Previous studies compared genetic data from Austronesian and Papuan-speaking populations dispersed across the Pacific to construct scenarios of prehistoric cultural interactions. Here we examined Austronesian- and Papuan-speaking populations separated by realistic prehistoric interaction distances along a transect from off-shore islands to the Sepik Plain. Our study addresses the following questions, previously unanswerable through larger scale studies: (1) Do genetic barriers remain between neighboring Austronesian and Papuan speakers after 3,500 years of potential gene flow? (2) Do linguistic similarities predict genetic affinities? (3) Does geographic proximity predict genetic affinity? (4) To what extent do current genetic patterns reflect the recent migration of Boiken speakers from the East Sepik Plain, and did this migration obscure previous population movements?

Materials and methods

To address these questions, we analyzed the mtDNA sequences from 219 volunteers from eight villages in the Wewak and Yangoru districts of East Sepik Province on the north coast of PNG (Fig. 1a). Two villages were selected from each of four ecological zones (off-shore islands, coast, Prince Alexander Range, and Sepik Plain) along a 69-km transect from 3.15°S to 4°S, with intervillage distances averaging 21 km. Four indigenous languages from three linguistic families, one Austronesian and two Papuan (Torricelli and Sepik-Ramu, Fig. 1b), are spoken in our study area. Two languages are spoken in only one village of our study area: Kairiru is the Austronesian (A) language of St. Martin, and Arapesh is the Torricelli (T) language of Dagua. Abelam is the Sepik-Ramu (S-R) language spoken in Wingei and Witupe. Boiken, also an S-R language, is spoken in four villages, one village from each ecological zone: Walis, Boiken, Warabung, and Kিনিambu (Fig. 1).

Approval for this study was obtained from the National Department of Health Medical Research Advisory Committee of PNG, the Tokyo Women's Medical University Ethical Committee, and the Binghamton University Human Subjects Research Review Committee. In addition, the research participants gave consent prior to their inclusion in this study.

Laboratory methods

DNA was extracted from 219 blood samples using chaotropes guanidium thiocyanate with size-fractionated silica

particles using an abbreviated protocol from Boom et al. (1990) (Tsukahara et al. 2006). A 405 base pair (bp) segment of the mtDNA hypervariable region 1 (HVS1) was amplified by polymerase chain reaction (PCR) using primers L15996 and H16401 as described previously (Vigilant et al. 1989; Lum and Cann 2000). The PCR product was purified using a Manu03050 Filter Plate (Millipore Corporation, Billerica, MA, USA) and sequenced in both directions with the Big Dye Terminator Kit version 3.1 in an Applied Biosystems 3730xl 96-capillary DNA analyzer (Applied Biosystems, Foster City, CA, USA).

Sequence analysis

In total, 333 bp of unambiguous mtDNA HVS1 sequence was determined from each participant and used for phylogenetic analyses. The 219 sequences (Genbank accession numbers EU312981-EU313199) were aligned using ClustalX (Thompson et al. 1997). A median-joining network was constructed from the sequences using Network 4.500 (www.fluxus-engineering.com), and times of coalescence were estimated using a mutation rate of one substitution every 20,180 years (Richards et al. 1998).

The estimated time of coalescence derived from the median-joining network of all individuals was 123,700 years, a time depth approximately three times greater than the initial peopling of New Guinea (Groube et al. 1986; Hudjashov et al. 2007) and approximately twice the estimated age of the dispersal out of Africa by anatomically modern humans (Mellars 2006). The times of coalescence for both haplogroups P and Q roughly matched the archeological dates of the first arrival of humans to New Guinea, whereas the estimated age for haplogroup B was consistent with the more recent arrival of the presumed Austronesian speakers and the consequent expansion of the descendant Lapita Cultural Complex. Furthermore, the events we are interested in occurred within the past 3,500 years, less than one fifth the average time inferred for single nucleotide polymorphisms (SNP) to evolve in the HVS1 region of mtDNA (Richards et al. 1998). Given the recent timeframe of the formation of the Lapita Cultural Complex and the Boiken expansion, we treated distinct mtDNA HVS1 lineages as alleles and estimated F_{ST} genetic distances among villages from frequencies of shared alleles (Lum and Cann 2000).

Statistical analysis

The 219 sequences were first sorted into haplogroups B, P, and Q by suites of shared mtDNA mutations, and the frequencies of these haplogroups were determined (Table 1). Intervillage F_{ST} genetic distances were estimated from

Fig. 1 **a** Map of Papua New Guinea and the East Sepik Province. The eight villages in our study are located two each in four ecological zones: islands (*circles*), coast (*squares*), mountains (*triangles*), and plains (*diamonds*). **b** The topology of the four languages in our study area: Kairiru (*white*), Arapesh (*striped*), Abelam (*gray*), and Boiken (*black*)

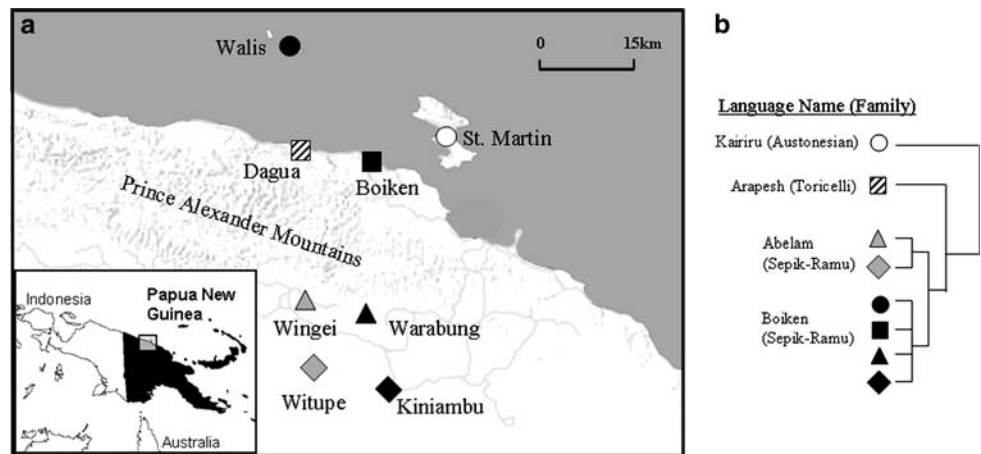


Table 1 Geographic location, language spoken, number of participants, and haplogroup frequencies of the eight studied villages

Ecological zone	Village name	Language	Language family	Number	Haplogroup			
					B	P	Q	Other
Island	Walis	Boiken	Sepik-Ramu	30	0.07	0.23	0.67	0.03
	St. Martin	Kairiru	Austronesian	31	0.33	0.13	0.43	0.10
Coast	Dagua	Arapesh	Toricelli	22	0.18	0.09	0.59	0.14
	Boiken	Boiken	Sepik-Ramu	22	0.09	0.14	0.64	0.14
Mountain	Wingei	Abelam	Sepik-Ramu	27		0.30	0.70	
	Warabung	Boiken	Sepik-Ramu	26		0.23	0.73	0.04
Plain	Witupe	Abelam	Sepik-Ramu	29		0.17	0.65	0.17
	Kiniambu	Boiken	Sepik-Ramu	32	0.12	0.25	0.63	
	Total			219	0.11	0.20	0.63	0.07

frequencies of shared unique lineages using Arlequin 2.0 (Schneider et al. 2000) and were used to create a neighbor-joining tree (Fig. 3a) (Felsenstein 2005). Individuals were permuted randomly between populations and F_{ST} values were recalculated for each of these 10,000 simulated random migrations (Schneider et al. 2000). The extent of gene flow was inferred from the proportion of permutations of individuals between populations, resulting in a higher F_{ST} value than observed in the original data set. Seven out of 28 pairs of populations (25%) were not statistically different at the 0.1 level and were thus categorized as having high gene flow ($0.44 > P > 0.10$) (Lum et al. 2002). Lines representing high intervillage gene flow ($P > 0.10$) were drawn onto a map of the East Sepik (Fig. 3b). In contrast, populations were considered to be significantly distinct and genetically isolated when $F_{ST} P$ values were < 0.05 . The pairwise F_{ST} matrix is available from the authors upon request.

The correlation between haplogroup B frequency and south latitude was evaluated using Spearman’s rank order correlation (Fig. 4). Lastly, the 219 sequences were compared to other previously published sequences in GenBank

to provide a regional context to our local interpretations. Lineages were considered ancestral if they were found at high frequencies throughout the region and derived if found in low frequencies in restricted geographic locations.

Results

Haplogroups and lineages

Of the 219 participants, 93% ($n = 203$) were grouped into one of the three haplogroups mentioned earlier: B (11%, $n = 23$), P (20%, $n = 43$), and Q (63%, $n = 137$). Haplogroup frequencies varied substantially across the eight villages (Table 1). Other haplogroups commonly found in ISEA were also present but in low numbers (e.g., E: $n = 2$, M: $n = 1$) (Trejault et al. 2005; Meriwether et al. 2005; Hill et al. 2007).

The 219 sequences yielded 91 unique lineages (Table 2). The median-joining network diagram (Fig. 2) of the 91 lineages was characterized by three nonoverlapping monophyletic clusters corresponding to the three main

Table 2 List of 91 unique lineages and their distribution by village

Lineage	Hap	HVS1 mutations (+16000)	Shared node	Wa	SM	Da	Bo	Wn	Wg	Wt	Ki	Total
1	B	189, 217, 247, 261	A	1	9	3	2				3	18
2		189, 217, 261	A	1	2						1	4
3		189, 217, 234, 247, 261				1						1
4	P	176, 224, 266, 293	A	5		1					1	7
5		176, 223, 224, 266, 293									4	4
6		176, 224, 266, 293, 348	D			1	1					2
7		176, 266, 293		1								1
8		176, 266									1	1
9		176, 266, 270		1							1	3
10		176, 264, 266, 270						2	1	2		5
11		176, 222, 264, 266, 270								1		1
12		172, 176, 266, 270									1	1
13		93, 176, 266, 270						1				1
14		266, 270			1				2			3
15		93, 176, 214, 266			2							2
16		93, 214, 249, 266						1				1
17		93, 176, 249, 266						2				2
18		26, 93, 176, 249, 266						2				2
19		176, 249, 266					1		2	2		5
20		176, 249, 264, 266			1							1
21		209, 266					1					1
22	Q	129, 144, 148, 223, 241, 265t, 311, 343	A	5			3	3	1		5	17
23		129, 144, 148, 192, 241, 265t, 311, 343	A	2		1		2	3	3		11
24		129, 148, 189, 192, 223, 241, 265t, 274, 311, 343	D	1		1						2
25		129, 144, 148, 189, 192, 223, 241, 265t, 311, 343							2			2
26		129, 144, 148, 192, 241, 265t, 286, 311, 343							1			1
27		94, 129, 144, 148, 223, 241, 265t, 286, 311, 343		1								1
28		26, 129, 144, 148, 192, 223, 241, 265t, 311, 343						1				1
29		26, 129, 144, 148, 223, 241, 265t, 311, 343						1				1
30		129, 144, 148, 241, 265t, 311, 343	A			1		1	2		1	5
31		129, 144, 148, 222, 241, 265t, 311, 343	A	1	1		1	2		3		8
32		129, 144, 148, 192, 222, 241, 265t, 343					1					1
33		129, 148, 222, 241, 265t, 311, 343						1				1
34		222, 241, 265t, 311, 343								1		1
35		129, 144, 148, 209, 222, 241, 265t, 311, 343							1			1
36		111, 129, 144, 148, 209, 222, 241, 265t, 311, 343			2							2
37		129, 144, 148, 153, 209, 222, 241, 265t, 311, 343									1	1
38		129, 144, 148, 222, 241, 265t, 292, 311, 343				1						1
39		93, 129, 144, 148, 222, 241, 265t, 311, 343									1	1
40		93, 129, 144, 148, 241, 265t, 311, 343	A		1	2	1	1		1	3	9
41		93, 129, 144, 148, 223, 241, 265t, 311, 343									2	2
42		93, 129, 144, 148, 223, 241, 265t, 271, 311, 343		1								1
43		80, 93, 129, 144, 148, 241, 265t, 311, 343								1		1
44		129, 144, 148, 223, 241, 243, 265t, 311, 343			1		2					3
45		129, 144, 148, 223, 243, 265t, 311, 343			2							2
46		129, 144, 148, 223, 243, 265t, 290, 311, 343		3								3

Table 2 continued

Lineage	Hap	HVS1 mutations (+16000)	Shared node	Wa	SM	Da	Bo	Wn	Wg	Wt	Ki	Total
47		129, 144, 148, 241, 265t, 311, 332t, 343						1				1
48		129, 144, 148, 216, 223, 241, 265t, 311, 343		1								1
49		129, 144, 148, 218, 241, 265t, 311, 343							1			1
50		129, 144, 148, 223, 241, 265t, 291, 311, 343			3							3
51		129, 144, 148, 189, 223, 241, 265t, 311, 343			1						3	4
52		129, 144, 148, 183, 189, 223, 241, 265t, 311, 343	D		1	1						2
53		129, 144, 148, 189, 223, 241, 265t, 278t, 311, 343		1								1
54		129, 144, 148, 189, 223, 241, 265t, 274, 311, 343								1		1
55		113t, 129, 144, 148, 189, 223, 241, 265t, 311, 343									1	1
56		36, 129, 144, 148, 189, 223, 241, 265t, 311, 343									1	1
57		129, 144, 148, 189, 241, 311, 343									1	1
58		97t, 129, 144, 148, 164t, 241, 265t, 311, 343						1				1
59		129, 144, 148, 164t, 223, 241, 265t, 311, 343				1						1
60		114, 129, 144, 148, 223, 241, 265t, 311					1					1
61		129, 144, 148, 209, 221t, 223, 241, 265t, 311, 343	D			1			1			2
62		129, 144, 148, 223, 241, 261, 265t, 304, 311, 343				1						1
63		129, 144, 148, 223, 241, 265t, 304, 311, 343						1	1			2
64		129, 148, 223, 241, 265t, 311, 343								1		1
65		129, 144, 148, 173, 223, 241, 265t, 294, 311, 343	D						1	1		2
66		129, 144, 148, 193, 223, 241, 265t, 311, 343					3					3
67		129, 144, 148, 223, 241, 263, 265t, 311, 330t, 343						1				1
68		41, 129, 144, 148, 159, 223, 234, 241, 265t, 311, 343						1				1
69		129, 144, 148, 223, 261, 265t, 311, 343	D			2		1		3		6
70		129, 144, 148, 223, 241, 261, 265t, 311, 343			1							1
71		85, 95, 129, 144, 148, 223, 231, 249, 265t, 293, 311, 343							1			1
72		85, 129, 144, 148, 223, 231, 249, 265t, 293, 311, 343							1			1
73		85, 129, 223, 265t, 293, 311, 343				1						1
74		129, 223, 241, 311			2				1	1		4
75		129, 223, 241, 311, 320	D	4			2					6
76		129, 223, 241, 294, 311,								1		1
77		129, 223, 241, 266, 311							1			1
78		176, 249, 265t, 266, 311, 343								1		1
79		129, 176, 223, 241, 266, 311								1		1
80		129, 144, 148, 249, 266									1	1
81	Other	51, 86, 129, 148, 223	D					1	1			2
82		51, 223		1								1
83		126, 129, 192, 223, 304				1						1
84		184, 223, 274, 278							1			1
85		69, 113t, 126, 163					1					1
86		69, 126, 163					1					1
87		278, 293	D		1		1					2
88		278								1		1
89		261, 278, 291								3		3
90		239, 319				2						2
91		319, 342								1		1
		Total		30	31	22	22	27	26	29	32	219

Wa Walis, SM St. Martin, Da Dagua, Bo Boiken, Wn Wingei, Wg Warabung, Wt Witupe, Ki Kiniambu; Hap haplogroup

The inferred ancestral (A) or derived (D) status of corresponding shared lineages is also noted

Fig. 2 Median-joining network of the 91 lineages: haplogroup B ($n = 3$), haplogroup P ($n = 18$), haplogroup Q ($n = 59$), and Others ($n = 11$). Lineages found in more than one village (shared) have assigned numbers corresponding to Table 2

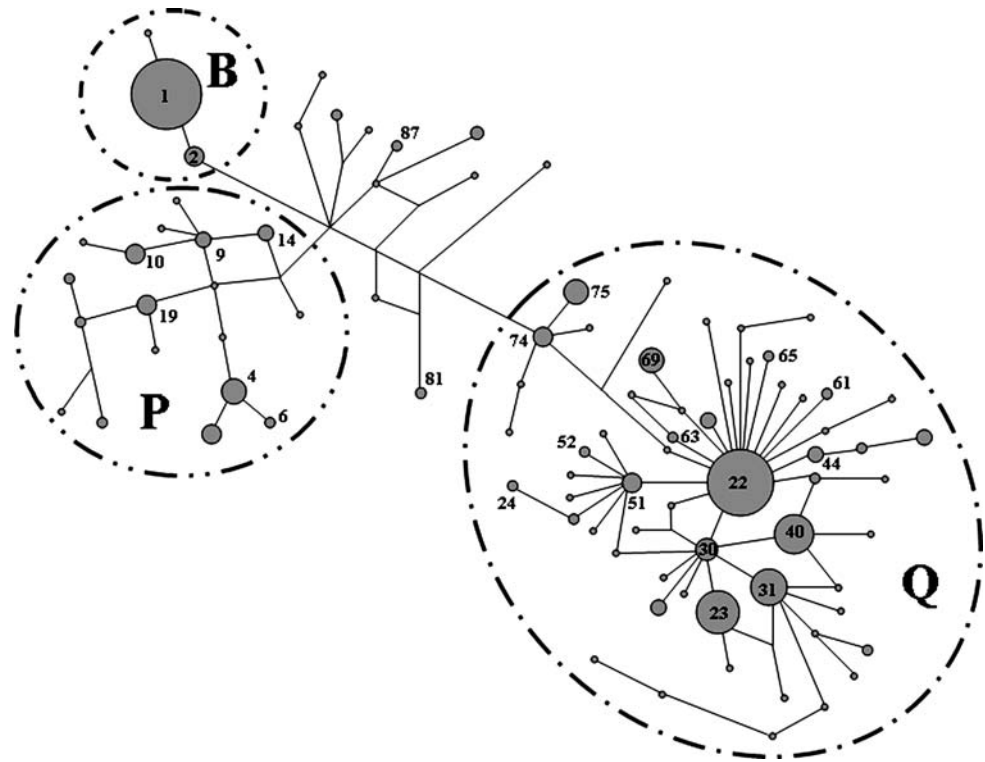
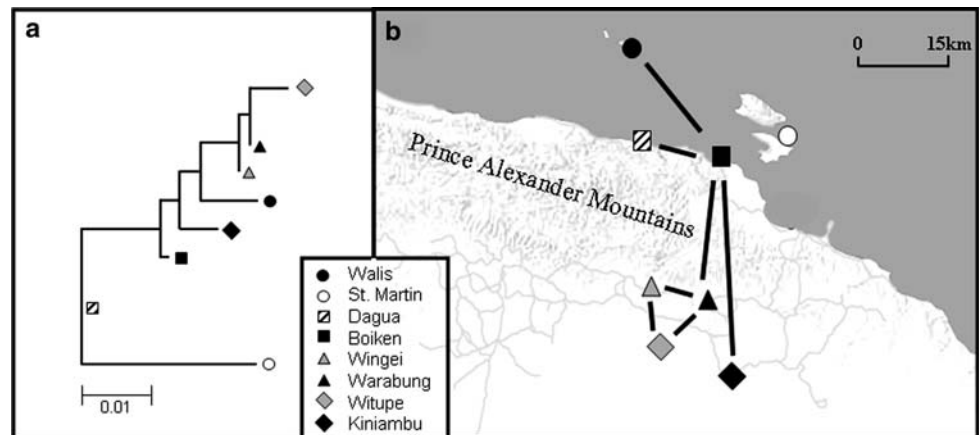


Fig. 3 a Neighbor-joining tree based on F_{ST} genetic distances. The scale represents the proportion of difference among populations. **b** Bold lines correspond to high intervillage gene flow inferred from pairwise F_{ST} genetic distances determined to be highly nonsignificant ($P > 0.10$)



haplogroups (B, P, and Q). Haplogroups B and Q were characterized by at least one high-frequency node ($n > 7$). These larger nodes are ancestral and were found in either three or all four ecological zones (Table 2). Haplogroup Q also had multiple terminal tips of derived lineages branching from several large serially ancestral nodes. The smaller tips were derived relatively recently from the larger ancestral nodes, and thus expected to be geographically restricted. This pattern of large ancestral nodes with multiple radiating tips is consistent with the constant growth of an expanding population. In contrast, haplogroup P formed a lattice of dispersed nodes of relatively similar frequencies. This suggests that haplogroup P had a restricted or

reduced population size over time. The times of coalescence for each individual monophyletic cluster were $42,237 \pm 13,818$ years for P, $37,052 \pm 10,022$ years for Q, and $4,386 \pm 3,617$ years for B. Of the 91 unique lineages, 25 were shared between villages. All villages shared at least one lineage with all other villages, and some village pairs shared more than one. Nine of the 25 shared lineages were inferred to be derived branch tips from the pattern observed in the network diagram (Fig. 2) of small localized tips radiating from larger, more ubiquitous, nodal lineages one or two polymorphisms away. Additionally, eight of the nine lineages did not match any other sequences previously submitted to GenBank, whereas the large ancestral nodes

did match sequences previously found outside our study area. These patterns suggest that the network tips were localized and unique and therefore derived and indicative of recent gene flow when shared among populations.

Inter-village F_{ST} values

The village of Boiken (S-R) and the three other Boiken-speaking villages were linked by high gene flow ($P > 0.10$). Likewise, the two Abelam-speaking villages (Wingei and Witupe) were not significantly distinct ($P > 0.05$) from each other or from Warabung (S-R), the neighboring ridge village (Fig. 3b), suggesting high gene flow among villages where the same and related Sepik-Ramu languages are spoken (Fig. 3a). St. Martin (A) was significantly different ($P < 0.05$) from all other villages except the coastal village of Dagua (T).

F_{ST} genetic distances between neighboring villages on the coast were not significantly distinct ($P > 0.05$). In contrast, the two Sepik Plain villages only 13 km apart, Witupe (S-R) and Kiniambu (S-R), were significantly distinct from each other ($P < 0.05$).

Haplogroup B and distance inland

A significant inverse correlation was observed between haplogroup B frequencies and south latitude ($r = -0.7$, $P < 0.05$) (Fig. 4), reflecting the higher frequency of this haplogroup in the coastal and offshore island villages and its absence in three of the four inland villages.

Discussion

In this study, we addressed several questions pertinent to the formation of the Lapita Cultural Complex, the settlement of the Pacific, and regional migrations within northern New Guinea: (1) Do genetic patterns today still

reflect the integration of the Pleistocene and the Holocene populations that gave rise to the Lapita Cultural Complex? (2) Do linguistic similarities predict genetic affinities? (3) Does geographic proximity predict genetic affinity? (4) And to what extent do genetic patterns reflect regional migratory events, such as the seaward expansion of the Boiken speakers of the Sepik Plain?

Gene flow and language

Overall, language was a good predictor of genetic affinities (Figs. 1b, 3b). F_{ST} genetic distances between the two Abelam-speaking villages and between the village of Boiken (S-R) and the three other Boiken-speaking villages spread across a 69-km north–south transect were not significantly distinct ($P > 0.05$), indicating substantial gene flow within both language groups. In contrast, St. Martin (A) was significantly different ($P < 0.05$) to all Sepik-Ramu villages, suggestive of genetic isolation. These patterns of significant gene flow among the Sepik-Ramu villages and limited gene flow between the Austronesian-speaking village and each Sepik-Ramu village suggest that linguistic similarities have encouraged gene flow, whereas linguistic barriers have restricted gene flow among these populations. Thus, linguistic similarities predicted genetic affinities among these closely spaced villages in addition to populations dispersed at greater distances across the Pacific (Lum and Cann 1998; Lum et al. 1998; Merriwether et al. 1999).

Gene flow and geography

We observed substantial gene flow between neighboring villages in two of our four ecological zones. F_{ST} genetic distances between coastal villages of Dagua (T) and Boiken (S-R) and mountain villages of Wingei (S-R) and Warabung (S-R) were not significantly distinct ($P > 0.10$), suggesting substantial gene flow between populations speaking different languages. The two mountain villages shared six lineages (Table 2), two of which were derived and inferred to result from recent gene flow. In contrast, the F_{ST} genetic distances between St. Martin (A) and Walis (S-R), the two island villages, as well as between Kiniambu (S-R) and Witupe (S-R), the plains villages, were significantly distinct ($P < 0.01$). These last two villages only shared one high frequency, presumably ancestral Q lineage (Table 2). This restricted gene flow between the two villages of the Sepik Plain is consistent with oral traditions that describe generational animosity between the people of Kiniambu (S-R) and Witupe (S-R) (Kazim 1983). This animosity likely reflected resource competition in one of the island’s most densely populated areas and may have fomented the expansion north and seaward by the Sepik Plain’s Boiken speakers. Within our study area, substantial

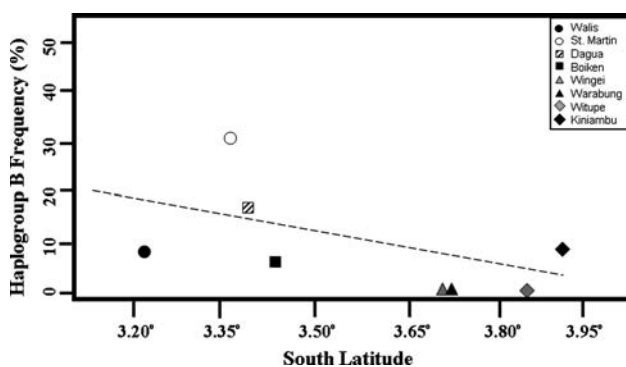


Fig. 4 Significant inverse correlation ($r = -0.7$, $P < 0.05$) between south latitude (distance inland) and haplogroup B frequencies

gene flow was observed in half of the ecological zones; thus, genetic affinity was random with respect to geographic proximity contra Terrell et al. (1997, 2001).

Gene flow and the Boiken expansion

The Boiken-speaking villages of Walis (S-R) and Kiniambu (S-R), the two most distant sites of our study (69 km apart), shared five lineages (Table 2) and had the most similar haplogroup frequencies of any two villages (Table 1). These villages were also not significantly distinct ($P > 0.05$) from the coastal village of Boiken (S-R) as evaluated by the F_{ST} genetic distances. These data suggest a genetic continuity among the Boiken-speaking villages, even across great distances and ecological barriers (mountains and ocean). This pattern is consistent with archeological and linguistic data describing an expansion of the Boiken speakers (Roscoe 1989) from the inland plain out to the offshore islands. It is also consistent with an oral account of the settlement of the island of Walis. The story tells of an intrusive movement from the East Sepik Plain [specifically the Kiniambu (S-R) area] across the Prince Alexander Mountains to the coasts and islands, displacing and replacing people along the way (Anurim and Kabaru 1980).

Haplogroup B and south latitude

The significant inverse correlation between haplogroup B frequencies and south latitude (distance inland) ($r = -0.7$, $P < 0.05$) is consistent with a coastal dispersal of haplogroup B followed by limited dispersal inland. This pattern parallels the restricted distribution of Austronesian languages on offshore islands and along the north coast of New Guinea. Tsukahara et al. (2006) examined two loci from a subset of the same populations and observed a similar pattern. Similar to haplogroup B, the Southeast Asian Ovalocytosis 27 bp deletion is thought to have been introduced with the mid-Holocene arrival of Austronesian speakers from ISEA (Tsukahara et al. 2006). In both studies, alleles and lineages thought to have been introduced by the mid-Holocene settlers remained predominantly limited to the coast and islands, only occurring inland in Boiken-speaking Kiniambu (S-R). This exception may be the result of gene flow from the offshore islands and coast to the Sepik Plain during the Boiken expansion and of limited gene flow within the Sepik Plain.

Data in local and regional context

A GenBank blast search of our 91 unique lineages identified eight exact matches with previously published sequences. Six haplogroup P and Q lineages had been

previously identified in high numbers in other Melanesian populations (Stoneking et al. 1990; Redd and Stoneking 1999; Tomasseo-Ponzetta et al. 2002; Ingman and Gyllenstein 2003; Friedlaender et al. 2005). Four of those six P and Q lineages had also been found in small numbers outside Melanesia, specifically in eastern Indonesia (Hill et al. 1997), Polynesia (Lum et al. 1994; Sykes et al. 1995), and Micronesia (Lum et al. 2000). Those Q lineages found outside Melanesia were consistently the larger, presumably ancestral, nodes ($n > 7$) at the center of the network clusters (Fig. 2). In contrast, only one of the nine shared tips inferred to be recently derived matched a previously published GenBank sequence, a single individual from the Bismarck Archipelago of PNG (Friedlaender et al. 2007).

Two haplogroup B lineages (lineages 1 and 2) were also found in high frequency throughout the Pacific. Lineage 1, referred to as the Polynesian Motif (Redd et al. 1995) and the Austronesian Motif (Lum and Cann 2000), is the most common Micronesian and Polynesian lineage, whereas lineage 2 has also been found across Asia and the Americas. This pattern of dispersed central/ancestral lineages and localized derived tip lineages (private polymorphisms) is consistent with an early expansion across the Pacific followed by more restricted movement and the accumulation of locally specific variants.

Conclusion

Analyses of mtDNA sequences from the East Sepik province of PNG allowed us to reconstruct patterns of gene flow associated with several events taking place over the past 3,500 years. We inferred from our coalescent time estimates that haplogroups P and Q arrived in New Guinea in the Pleistocene, potentially at different times and experiencing distinct demographic histories. In the mid-Holocene, we see the intrusion of haplogroup B from ISEA. This haplogroup, like its associated Austronesian languages, was limited to the coast and islands where it did not replace, but instead integrated with, the gene pool of the Pleistocene settlers. The resulting combination of haplogroups (B, P, and Q) were then dispersed with the Lapita Cultural Complex, resulting in high frequencies of B lineages with traces of Q and P lineages throughout Polynesia and Micronesia (Sykes et al. 1995; Lum et al. 1998, 2000). In the past 3,000 years, the Boiken expansion (Roscoe 1989) created a zone of genetic and linguistic homogeneity from the Sepik Plain to the offshore islands. Localized gene flow, inferred from both shared derived/tip lineages and nonsignificant F_{ST} genetic distances ($P > 0.05$), followed the expansion within two of the four ecological zones (coast and mountain). This

localized gene flow was absent in the Sepik plain, as inferred from within zone F_{ST} genetic distances ($P < 0.01$) and from the retention in Kiniambu (S-R) of Boiken-shared lineages from all three haplogroups. Thus, genetic patterns associated with linguistic differences have been maintained for more than 3,000 years allowing us to reconstruct the patterns of biological interactions that resulted in the Lapita Cultural Complex and the settlement of the Pacific.

The patterns observed and the conclusions inferred from this study are based on maternally inherited mtDNA. If greater male gene flow is associated with warfare, differential migration, patterns of land tenure, and regional conquest, as suggested by oral histories of the settlement of Walis (Anurim and Kabaru 1980), we would expect to see even stronger genetic affinities among the Boiken speakers based on paternally inherited Y-chromosome diversity. To explore these possibilities, we have begun analysis of Y-chromosome SNP variations from these and five additional East Sepik Province villages. The forthcoming analyses will help complete the picture of regional gene flow between the two parental populations that birthed the Lapita Cultural Complex.

Acknowledgments We thank the field workers from the eight villages in the East Sepik Province, as well as the health center and aid posts workers and the staff at the Wewak General Hospital. Funds for this study came from the Japan International Cooperation Agency (JICA) under the partnership programme.

References

- Allen J (1970) Prehistoric agricultural systems in the Wahgi Valley: A further note. *Mankind* 7:177–183
- Allen J, Gosden C, Jones R, White JP (1988) Pleistocene dates for the human occupation of New Ireland, northern Melanesia. *Nature* 331:707–709
- Ambrose W (2002) From very old to new; obsidian artifacts in the Admiralty Islands. In: Kaufmann C, Kocher Schmid C, Ohnemus S (eds) *Admiralty Islands: Art from the South Seas*. Museum Rietberg, Zurich, pp 67–72
- Anurim B, Kabaru C (1980) How people came to Tarawai and Walis Islands. *Oral Hist* 8:81–83
- Bellwood P (1989) The colonization of the Pacific: some current hypotheses. In: Hill AV, Serjeantson SW (eds) *The colonization of the Pacific: a Genetic Trail*. Oxford Scientific Publications, Clarendon Press, Oxford, UK, pp 1–59
- Bellwood P (1991) The Austronesian dispersal and the origins of languages. *Sci Am* 265:88–93
- Boom R, Sol CJ, Salimans MM, Jansen CL, Wertheim-van Dillen PM, van der Noordaa J (1990) Rapid and simple method for purification of nucleic acids. *J Clin Microbiol* 28:495–503
- Davidson J, Hinds E, Holdaway S, Leach F (1990) The Lapita site of Natunuku, Fiji. *N Z J Archaeol* 12:121–155
- Diamond J (1988) Express train to Polynesia. *Nature* 336:307–308
- Felsenstein J (2005) PHYLIP (Phylogeny Inference Package) v. 3.6. Department of Genome Sciences, University of Washington, Seattle
- Forster P, Torroni A, Renfrew C, Rohl A (2001) Phylogenetic star contractions applied to Asian and Papuan mtDNA evolution. *Mol Biol Evol* 18:1864–1881
- Friedlaender J, Schurr T, Gentz F, Koki G, Friedlaender F, Horvat G, Babb P, Cerchio S, Kaestle F, Schanfield M, Dekka R, Yanagihara R, Merriwether DA (2005) Expanding Southwest Pacific mitochondrial haplogroups P and Q. *Mol Biol Evol* 22(6):1506–1517
- Friedlaender J, Friedlaender FR, Hodgson J, Stoltz M, Koki G, Horvat G, Zhadanov S, Schurr TG, Merriwether DA (2007) Melanesian mtDNA complexity. *PLOS* 2, e248:1–13
- Gordon RG (ed) (2005) *Ethnologue: languages of the World*, fifteenth edition online. SIL International, Dallas, TX
- Green RC (1976) Lapita sites in the Santa Cruz group. In: Green RC, Creswell MM, Allen J (eds) *Southeast Solomon Islands cultural history: a preliminary survey*. Royal Society of New Zealand, Wellington, NZ, pp 245–265
- Green RC (1991a) The Lapita Cultural Complex: current evidence and proposed models. In: Bellwood P (ed) *Indo-Pacific Prehistory 1990*, vol.2. Bulletin of the Indo-Pacific Prehistory Association. Australian National University, Canberra, 11:295–305
- Green RC (1991b) Near and remote Oceania- disestablishing “Melanesia” in culture history. In: Pawley A (ed) *Man and a Half: Essays on Pacific Anthropology and Ethnobiology in honour of Ralph Bulmer*. The Polynesian Society, Auckland, NZ, pp 491–502
- 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
- Hedrick J (1971) Lapita style pottery from Malo Island. *J Polyn Soc* 80(1):5–19
- Hill C, Soares P, Mormina M, Macaulay V, Clarke D, Blumbach PB, Vizuette-Forster M, Forster P, Bullbeck D, Oppenheimer S, Richards M (2007) A mitochondrial stratigraphy for island Southeast Asia. *Am J Hum Genet* 80:29–43
- Hudjashov G, Kivisild T, Underhill PA, Endicott P, Sanchez JJ, Lin AA, Shen P, Oefner P, Renfrew C, Villems R, Forster P (2007) Revealing the prehistoric settlement of Australia by Y chromosome and mtDNA analysis. *Proc Natl Acad Sci* 104:8726–8730
- Ingman M, Gyllenstein U (2003) Mitochondrial genome variation and evolutionary history of Australian and New Guinean aborigines. *Genome Res* 13:1600–1606
- Kazim B (1983) In: Slone TH (ed) *One thousand and one Papua New Guinean nights*. Masalai Press, PNG, pp 194–195
- Kirch PV (2000) On the road of the winds: an archaeological history of the Pacific islands before European contact. University of California Press, Berkeley, CA
- Kirch PV, Hunt TL (1988) Radiocarbon dates from the Mussau Islands and the Lapita colonization of the Southwest Pacific. *R.A.C.A.A.* 30:161–169
- Leavesley MG, Chapell J (2004) Buang Merabak: additional early radiocarbon evidence of the colonization of the Bismarck Archipelago, Papua New Guinea. *Antiquity Project Gallery*, <http://antiquity.acuk/ProjGall/leavesley/indexhtml> 78
- Lum JK, Cann RL (1998) MtDNA and language support a common origin of Micronesians and Polynesians in island Southeast Asia. *Am J Phys Anthropol* 105:109–119
- Lum JK, Cann RL (2000) mtDNA lineage analyses: origins and migrations of Micronesians and Polynesians. *Am J Phys Anthropol* 113:151–168
- Lum JK, Rickards O, Ching C, Cann RL (1994) Polynesian mitochondrial DNAs reveal three deep maternal lineage clusters. *Hum Biol* 66:567–590
- Lum JK, Cann RL, Martinson JJ, Jorde LB (1998) Mitochondrial and nuclear genetic relationships among Pacific Island and Asian Populations. *Am J Hum Genet* 63:613–624

- Lum JK, Jorde LB, Schiefenhover W (2002) Affinities among Melanesians, Micronesians, and Polynesians: a neutral, biparental genetic perspective. *Hum Biol* 74:413–430
- Lum JK, McIntyre JK, Greger DL, Huffman KW, Vilar MG (2006) Recent Southeast Asian domestication and Lapita dispersal of sacred male pseudohermaphroditic “tuskers” and hairless pigs of Vanuatu. *Proc Natl Acad Sci* 103:17190–17195
- Lynch D, Wanglund C, Spathis R, Chan CW, Reiff DM, Lum JK, Garruto RM (2008) The contribution of mitochondrial dysfunction to a gene-environment model of Guamanian ALS and PD. *Mitochondrion* 8(2):109–116
- Matisoo-Smith E (2007) Animal translocations, genetic variation and the human settlement of the Pacific. In: Friedlaender J (ed) *Genes, language, and culture history in the Southwest Pacific*. Oxford University Press, Oxford, UK, pp 157–170
- Matisoo-Smith E, Robins JH (2004) Origins and dispersals of Pacific peoples: evidence from mtDNA phylogenies of the Pacific rat. *Proc Natl Acad Sci* 101:9167–9172
- Mellars P (2006) Why did modern human populations disperse from Africa ca. 60,000 years ago? A new model. *Proc Natl Acad Sci* 103:9381–9386
- Merriwether DA, Friedlaender JS, Mediavilla J, Mgone C, Gentz F, Ferrell RE (1999) Mitochondrial DNA variation is an indicator of Austronesian influence in island Melanesia. *Am J Phys Anthropol* 110:243–270
- Merriwether DA, Hodgson JA, Friedlaender FR, Allaby R, Cerchio S, Koki G, Friedlaender JS (2005) Ancient mitochondrial M haplogroups identified in the Southwest Pacific. *Proc Natl Acad Sci* 102:13034–13039
- Pawley A, Green RC (1973) Dating the dispersal of the Oceanic languages. *Ocean Linguist* 12:1–67
- Redd AJ, Stoneking M (1999) People of Sahul: mtDNA variation in Aboriginal Australian and Papua New Guinean populations. *Am J Hum Genet* 65:808–828
- Redd AJ, Takezaki N, Sherry ST, McGarvey ST, Sofro ASM, Stoneking M (1995) Evolutionary history of the COII/tRNA^{Lys} intergenic 9 base pair deletion in human mitochondrial DNAs from the Pacific. *Mol Biol Evol* 12:604–615
- Richards MB, Macaulay VA, Bandelt H-J, Sykes BC (1998) Phylogeography of mitochondrial DNA in western Europe. *Ann Hum Genet* 62:241–260
- Roscoe BR (1989) The flight from the fen: the prehistoric migrations of the Boiken of the East Sepik Province, Papua New Guinea. *Oceania* 60:139–154
- Ross M (2005) Pronouns as a preliminary diagnostic for grouping Papuan languages. In: Pawley A, Attenborough R, Hide R, Golson J (eds) *Papuan Pasts*. Pacific Linguistics, Australian National University, Canberra, pp 15–66
- Sand C (1997) The chronology of Lapita ware in New Caledonia. *Antiq* 71:539–547
- Schneider S, Roessli D, Excoffier L (2000) Arlequin: A software for population genetics data analysis. Ver 2.000. Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva, Switzerland
- Spriggs M (1989) The dating of the island Southeast Asia Neolithic: an attempt at chronometric hygiene and linguistic correlation. *Antiq* 63:587–613
- Stoneking M, Jorde LB, Bhatia K, Wilson AC (1990) Geographic variation in human mitochondrial DNA from Papua New Guinea. *Genetics* 124:717–733
- Storey AA, Ramírez JM, Quiroz D, Burley DV, Addison DJ, Walter R, Anderson AJ, Hunt TL, Athens JS, Huynen L, Matisoo-Smith EA (2007) Radiocarbon and DNA evidence for a pre-Columbian introduction of Polynesian chickens to Chile. *Proc Natl Acad Sci* 104:10335–10339
- Summerhayes GR (2001) Defining the chronology of Lapita in the Bismarck Archipelago. In: Clark GR, Anderson AJ, Vunidilo T (eds) *The chronology of Lapita dispersal in Oceania*. Terra Australis 17, Pandanus Books, Canberra, pp 25–38
- Summerhayes GR, Allen J (1993) The transport of Mopir obsidian to late Pleistocene New Ireland. *Archaeol Ocean* 28:144–148
- Swadling P, Hide R (2005) Changing landscape and social interaction: looking at agricultural history from a Sepik-Ramu perspective. In: Pawley A, Attenborough R, Hide R, Golson J (eds) *Papuan Pasts*, Pacific Linguistics, Australian National University, Canberra, pp 289–328
- 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 JE, Hunt TL, Gosden C (1997) The dimensions of social life in the Pacific. *Curr Anthropol* 38:155–195
- Terrell JE, Kelly KM, Rainbird P (2001) Foregone conclusions? In search of “Papuans” and “Austronesians.” *Curr Anthropol* 42:97–124
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876–4882
- Tommaseo-Ponzetta M, Attimonelli M, De Robertis M, Tanzariello F, Saccone C (2002) Mitochondrial DNA variability of West New Guinea populations. *Am J Phys Anthropol* 117:49–67
- Trejault 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(8):e247:1–11
- Tsukahara T, Hombhanje FW, Lum JK, Hwaihwanje I, Masta A, Kaneko A, Kobayakawa T (2006) Austronesian origin of the 27 base pair deletion of the erythrocyte band 3 gene in Papua New Guinea inferred from mtDNA analysis. *J Hum Genet* 51:244–248
- Vigilant L, Pennington R, Harpending H, Kocher TD, Wilson AC (1989) Mitochondrial DNA sequences in single hairs from a southern African population. *Proc Natl Acad Sci* 86:9350–9354
- Wickler S, Spriggs M (1988) Pleistocene human occupation of the Solomon Islands, Melanesia. *Antiq* 62:703–706