A Polynesian Motif on the Y Chromosome: Population Structure in Remote Oceania

MURRAY P. COX,1,2 ALAN J. REDD,3 TATIANA M. KARAFET,1 CHRISTINE A. PONDER,4 J. STEPHEN LANSING,2,5 HERAWATI SUDOYO,5 AND MICHAEL F. HAMMER1

Abstract  The Polynesian motif, a mitochondrial DNA marker of ancestral Polynesian communities, has filled a critical role in reconstructions of remote Oceanic history. Although the motif provides an effective narrative for Polynesian females, no equivalent male history is available from paternal lineages. Here, we describe a Y-chromosome binary polymorphism with absolute Polynesian affinity. We illustrate its unique spatial and temporal connections to early Polynesian communities, and through an analysis of associated short tandem repeat variation, we describe the first clear genealogic structure within Polynesia. Unlike the eastern and western regions advocated by archeology, we identify a tripartite structure comprising interaction spheres in the west (Tonga and Samoa), center (Tahiti), and east (Rapanui/Easter Island). Such patterning, a product of early regional contact and subsequent isolation, signals the conflicting roles of mobility and seclusion in Polynesian prehistory.

A mitochondrial DNA (mtDNA) lineage found at high frequencies in Polynesians but uncommon elsewhere was identified early in the 1980s (Cann and Wilson 1983; Cann et al. 1987; Hertzberg et al. 1989). Initially characterized by a 9-bp deletion in hypervariable region V (Wrischnik et al. 1987) and later defined more explicitly with three hypervariable region I polymorphisms (T16217C, A16247G, and C16261T) (Hagelberg and Clegg 1993; Lum et al. 1994), this lineage was quickly dubbed the Polynesian motif (Melton et al. 1995; Redd et al. 1995). Its high frequency in Polynesia, where it is fixed in many populations, coupled with its low frequency elsewhere gives the lineage unique status as a migrational signal of ancestral Polynesians.

The Polynesian motif (or rather, the suite of closely related lineages termed B4a) has in turn become a genetic proxy for the extraordinary cultural processes

1Arizona Research Laboratories, University of Arizona, Tucson, AZ 85721.
2Santa Fe Institute, Santa Fe, NM 87501.
3Department of Anthropology, University of Kansas, Lawrence, KS 66045.
4Department of Genetics and Development, Columbia University, New York, NY 10032.
5Department of Anthropology, University of Arizona, Tucson, AZ 85721.
6Eijkman Institute for Molecular Biology, Jakarta 10430, Indonesia.

Copyright © 2007 Wayne State University Press, Detroit, Michigan 48201-1369

KEY WORDS: OCEANIA, INDONESIA, MELANESIA, MICRONESIA, POLYNESIA, POLYNESIAN PHYLOGEOGRAPHY, MALE HISTORY, POLYNESIAN MOTIF, Y CHROMOSOME, SHORT TANDEM REPEATS, SINGLE NUCLEOTIDE POLYMORPHISMS.
underlying the recent and rapid colonization of the Pacific region. Although Australia, New Guinea, and adjacent landmasses were settled 60,000–30,000 years before present (b.p.) (O’Connell and Allen 2004), the small and widely scattered islands of Polynesia were breached only in the late Holocene, ca. 2,900 years b.p. (Burley and Dickinson 2001). Ancestral Polynesians eventually landed in New Zealand; the last island they would settle, around 800 years b.p. (Higham et al. 1999). The B4a assemblage in Polynesia, variously interpreted as Taiwanese (Cox 2005; Trejaut et al. 2005) or East Indonesian in origin (Richards et al. 1998), therefore marks one of the world’s most impressive settlement ventures: a quarter of the globe’s surface in just over 2,000 years.

Despite considerable research on paternally inherited Y-chromosome diversity in Oceania (Capelli et al. 2001; Cox and Lahr 2006; Hurles et al. 2002; Karafet et al. 2005; Kayser et al. 2000, 2001, 2003, 2006; Lucotte 2003; Redd et al. 2002; Scheinfeldt et al. 2006; Underhill et al. 2001; Wilder and Hammer 2007), no comparable narrative is available for Polynesian males. Here, we identify a novel Y-chromosome marker with absolute Polynesian affinity, a paternal counterpoint to mitochondrial evidence, that bears on early Polynesian colonization and subsequent internal developments.

Materials and Methods

Y-chromosome markers were screened in 1,295 men from four broad Indo-Pacific regions: Indonesia (Mentawai, Flores, Sumba, Timor, Sulawesi, Maluku Islands), Melanesia (Highland and Coastal New Guinea, New Britain, Bougainville, Solomon Islands, Vanuatu), Micronesia, and Polynesia (Tonga, Samoa, Tahiti, Rarotonga/Easter Island). Samples from Melanesia, Micronesia, Polynesia, Timor, and the Maluku Islands have been described in previous studies (Cox and Lahr 2006; Hammer et al. 2001; Karafet et al. 2005; Redd et al. 2002). Indonesian samples from Mentawai, Sulawesi, Flores, and Sumba were collected by two of us (Herawati Sudoyo and J. Stephen Lansing). All samples were collected from volunteer donors with written informed consent and appropriate permits from the Indonesian government. The University of Arizona Human Subjects Committee approved sampling protocols. In all cases, care was taken to exclude individuals related within the last three generations, as determined by detailed personal genealogies.

Three Y-chromosome single nucleotide polymorphisms (SNPs)—M38, M208, and P33—on the C-RPS4Y background (Cox 2006b) were typed in hierarchical fashion (Hammer et al. 2001). M38 is a T → G transition in a single-copy Y-chromosome region at position 20,201,546 (University of California, Santa Cruz, Genome Browser, March 2006 assembly; Hinrichs et al. 2006). M208 is a C → T transition in a single-copy Y-chromosome region at position 14,085,597. P33 is a T → C transition at one of three paralogous regions on the Y chromosome. Appearing together with the ancestral state when derived, P33 can be traced to Y-chromosome position 23,397,374 or 25,031,122 or 25,750,056. All polymorphism
Polynesian Motif on the Y Chromosome  / 527

states were initially determined by allele-specific PCR; M208 and P33 polymorphism states were later confirmed by direct DNA sequencing.

C-M38* (xM208, P33) and C-M208* (xP33) individuals were screened with a suite of 12 Y-chromosome short tandem repeats (STRs): DYS385a, DYS385b, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS426, DYS438, and DYS457. C-P33 individuals were screened with 11 additional Y-chromosome STRs: DYS442, DYS446, DYS447, DYS453, DYS454, DYS455, DYS456, DYS458, DYS570, DYS576, and DYS607. The entire set of 23 Y-chromosome STRs was used for analyses of P33-derived individuals.

The link between Polynesian ancestry and lineage frequencies was determined in a Bayesian framework using MCMC calculations (Bandelt et al. 1999). Only nearest C-M38* (xM208, P33) relatives are illustrated for clarity.

TMRCA (time to the most recent common ancestor) dates were inferred using Y-chromosome STR data under a geometric-geometric mutation model (Watkins 2007). Analysis was conditioned on three parameters: $\alpha (= 0.1)$, which determines the geometric distribution of Y-chromosome STR step changes; $\beta (= 1.05)$, which sets the likelihood of mutation to a geometric dependence on microsatellite length; and $k (= 13)$, which represents the mean Y-chromosome STR repeat length in the data set. Confidence intervals (95%) were applied as $1/n^{1/2}$, where $n$ is the number of Y-chromosome STR loci. Unscaled coalescent time units were converted to chronological dates using a mean male intergeneration interval of 31 years (Fenner 2005) and a mean Y-chromosome STR mutation rate (averaged across multiple studies) of $2.1 \times 10^{-8}$ per generation (Gusmão et al. 2005).

Results and Discussion

Previous interpretations of Polynesian Y-chromosome diversity have depended on Y-chromosome markers that were not ascertained directly in Polynesian populations. By seeking region-specific variants directly in a subsample of Pacific individuals (Hammer et al. 2001), we identified a novel T $\rightarrow$ C transition on the nonrecombining portion of the Y-chromosome (NRY). This paralogous sequence variant (Sanchez et al. 2004), labeled P33 (Y Chromosome Consortium 2002), occurs within a duplicated region, one of the Y chromosome's many segmental duplications (Skaletsky et al. 2003). The polymorphism is found only on derived C-RPS4Y backgrounds, and its exact location on the tree of Y-chromosome variation was determined by typing additional NRY markers. Although previously ambiguous (Jobling and Tyler-Smith 2003), the correct mutational topology of the P33 lineage (incorporating two additional polymorphisms, M38 and M208) has now been determined (Figure 1).

M38, M208, and P33 were screened hierarchically in 1,295 men from 17
Indo-Pacific populations (Table 1). These lineages show a high degree of geographic specificity (see Figure 1). C-M38* (xM208, P33) is a widely dispersed Oceanic variant, but it occurs most frequently in eastern parts of the Indonesian archipelago. Conversely, C-M208* (xP33) is rare. We identified five men with this lineage (coastal New Guinea, Vanuatu, Samoa, and Tahiti), which occurs no farther west than two highland populations in central West New Guinea (Kayser et al. 2003). Finally, the derived C-P33 lineage was detected in 40 men with Samoan, Tongan, Tahitian, or Rapanui ancestry. We infer that P33 is restricted to descendants of ancestral Polynesian communities because the marker is absent in an additional data set of 16,163 men worldwide (unpublished data).

To emphasize the dichotomy in P33 frequencies between Polynesia and the wider Indo-Pacific region, we compared published mtDNA B4a frequencies from the same populations (Cox 2003, 2005; Hagelberg et al. 1994, 1999; Lum et al. 1998; Lum and Cann 2000; Melton et al. 1995; Redd et al. 1995; Scheinfeldt et al. 2006; Sykes et al. 1995). B4a lineages (which include the Polynesian motif sequence and allied haplotypes) are often considered markers of Austronesian-speaking, particularly Polynesian, populations (Merriwether et al. 1999). P33 might be expected to follow a similar distributional pattern. Indeed B4a and P33 frequencies are correlated (Pearson’s two-sided $r^2 = 0.79; t_{15} = 5.0$, $p = 1.6 \times 10^{-4}$) but are not overly strong predictors of each other (64% explanatory power). Although P33 is restricted to the Polynesian triangle (bounded by New Zealand, Hawaii, and Rapanui), B4a has a broader Indo-Pacific distribution.
Table 1. Geographic Distribution of the C-M38* (xM208, P33), C-M208* (xP33), and C-P33 Lineages

<table>
<thead>
<tr>
<th>Region</th>
<th>Sampling Location</th>
<th>N</th>
<th>C-M38* (xM208, P33)</th>
<th>C-M208* (xP33)</th>
<th>C-P33</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polynesia</td>
<td>Rapanui (Easter Island)</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>Tahiti</td>
<td>25</td>
<td>0</td>
<td>1</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Samoa</td>
<td>18</td>
<td>0</td>
<td>1</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Tonga</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Subtotal</td>
<td>64</td>
<td>0</td>
<td>2</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>Melanesia</td>
<td>Vanuatu (Maewo)</td>
<td>44</td>
<td>9</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Solomon Islands</td>
<td>32</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Bougainville</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>New Britain</td>
<td>50</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Coastal New Guinea</td>
<td>21</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Highland New Guinea</td>
<td>51</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Subtotal</td>
<td>208</td>
<td>11</td>
<td>3</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Micronesia</td>
<td>Maluku Islands</td>
<td>16</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sulawesi</td>
<td>36</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Timor</td>
<td>48</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Sumba</td>
<td>12</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Flores</td>
<td>352</td>
<td>202</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Mentawai</td>
<td>485</td>
<td>81</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Subtotal</td>
<td>1,007</td>
<td>304</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1,295</td>
<td>316</td>
<td>5</td>
<td>40</td>
<td></td>
</tr>
</tbody>
</table>

Frequencies of C-M38* (xM208, P33), C-M208* (xP33), and C-P33 carriers among 1,295 individuals sampled from 17 locations across the Indo-Pacific region. Lineages consistent with European, Japanese, or South American ancestry were considered evidence of recent admixture and were excluded from the counts (Burles et al. 2003).

B4a frequencies clearly differ between Polynesian and non-Polynesian populations \( (F_{1.15} = 14.3, p = 1.8 \times 10^{-3}, \eta^2 = 0.49) \), but the P33 mutation is far more closely associated with Polynesia \( (F_{1.15} = 54.6, p = 1.5 \times 10^{-3}, \eta^2 = 0.85) \). Not all Polynesian men carry the P33 marker, but Polynesian ancestry is strongly indicated in those who do.

Given its Polynesian specificity, the P33 mutation probably arose just before, or during, the region’s initial settlement. Archeological remains of a founding Polynesian settlement at Nukulaeke, Tonga, were radiocarbon-dated to about 2,900 years B.P. cal. (Burley and Dickinson 2001). Although the Tongan and Samoan archipelagos of Western Polynesia were colonized rapidly, there was a hiatus before settlement farther eastward (Kirch 2000). The age of the P33 mutation should match this timeframe within the current limits of genetic dating. We inferred the TMRCA of extant P33 lineages using variation in 23 Y-chromosome STRs under a coalescent-based geometric-geometric mutation model (Watkins 2007). The P33 transition likely arose within a broad temporal window from 7,500 to 1,500 years.
B.P. (95% confidence interval; mean 4,500 years B.P.). Although encompassing a large chronological spectrum, this interval is consistent with radiocarbon estimates for the earliest Polynesian settlements.

The spatial and temporal distributions of ancestral haplogroups also enlighten the early history of ancestral Polynesians. A relative paucity of C-M208* (xP33) suggests that the marker may have developed not long before P33. Expanding Austronesian-speaking populations probably assimilated M208 individuals (Green 1991, 2003) as they moved eastward along New Guinea’s northern coast about 3,500 years B.P. (Spriggs 2003). Descendants of these admixed communities later swept Melanesian markers, such as the lineages considered here, to high frequency in the remote islands of Polynesia (Cox 2006a; Cox and Lahr 2006). Consequently, M208 must have arisen before (Kayser et al. 2003) or during this population expansion. Using variation in 12 Y-chromosome STRs in M208-derived individuals, we infer an origin for the M208 mutation between 12,300 and 3,900 years B.P. (95% confidence interval; mean 8,100 years B.P.), consistent with earlier Bayesian estimates (95% confidence interval 19,700 to 2,800 years B.P.; mean 6,900 years B.P.; Kayser et al. 2003).

The distribution of C-M38* (xM208, P33) is centered on eastern Indonesia and Melanesia (Underhill et al. 2001) and has been equated with the C-RPS4Y/DYS390.3del variant in earlier studies (Kayser et al. 2003). Bayesian dates for this mutation yielded upper confidence limits with considerable antiquity: 30,300 to 4,500 years B.P. (95% confidence interval; mean 10,600 years B.P.) (Kayser et al. 2003). Despite a reduced chronological range, we also infer a late Pleistocene ancestor for M38-derived lineages, between 37,200 and 28,800 years B.P. (95% confidence interval; mean 33,000 years B.P.). Coupled with this marker’s spatial distribution, there seems little doubt that the ultimate paternal ancestors of Polynesian P33 carriers once lived in Melanesia or its immediate environs (Kayser et al. 2003).

We traced individual histories of C-P33 lineages using Y-chromosome STR variation to tease apart recent Polynesian history. Clear spatial structuring is observable in the network (Figure 2). From a broad geographic perspective, the most closely related C-M38* (xM208, P33) individuals were found on the islands of Flores and Sulawesi in eastern Indonesia. The next most closely related men, C-M208* (xP33) carriers, inhabit coastal Papua New Guinea. The most basal P33 lineage was found in Samoa, but it may represent a generic Western Polynesian founding lineage given the limited number of Tongans in our sample (n = 12). For similar reasons, related P33 lineages may some day be identified in Fiji (cf. Figure 2; Kayser et al. 2006), a likely staging ground for the settlement of Polynesia. Regardless, this overall pattern fits a movement of ancestral Polynesians trending generally eastward from (or through) Island Southeast Asia, along coastal New Guinea, and out into the wider Pacific (Hurles et al. 2002; Kayser et al. 2003). We predict that similar evolutionary patterns may be observed in other Y-chromosome lineages carried by Polynesians, such as O-M122.

Although Pacific haplogroup distributions are increasingly well mapped, no
clear signals of population structure have been identified within Polynesia itself. Two reasons seem immediately apparent. First, population structure might not be detectable in Polynesia if the settlement period was too short relative to genetic mutation rates (even for fast-mutating Y-chromosome STRs). Second, Polynesia may not show population structure if frequent interisland mobility effectively created a single, panmictic community. More realistically, some structuring should be expected within Polynesia, if only because most island groups were geographically isolated during later Polynesian prehistory (Rolett 2002; Weisler 2004).

In practice, we find all these processes written in the genetic record. The Y-chromosome STR network depicts clear substructuring within Polynesia but not the simple starlike pattern (Slatkin and Hudson 1991) so characteristic of rapid mitochondrial expansions in Oceania (Cox 2005). Instead, lineages tend to cluster into western (Tongan and Samoan), central (Tahitian), and eastern (Rapanui) groups. Western Polynesian men dominate the network’s center, consistent with their island homes being stepping-stones to Central and Eastern Polynesia (Hurles et al. 2002). More surprisingly, the central and eastern clusters are not nested but
instead derive independently from Western Polynesian precursors. Most likely, we are observing a two-pronged settlement of remote Polynesia: one group of people moving southward through the southern Cook Islands toward French Polynesia (represented here by Tahiti) and another group moving northward through the northern Cook Islands toward the Marquesas archipelago, ultimately settling Rapanui to the southeast. Marquesan and Rapanui populations have been linked archeologically (Finney 2001; but see Bahn and Flenn 2003); and Tahiti and Rapanui may have possessed fewer direct connections despite their relative geographic propinquity. Interestingly, Rapanui P33 carriers are still represented by at least two disparate Y-chromosome STR lineages (gray in Figure 2), including a Tahitian connection, despite extreme population collapses following European contact (González-Martín et al. 2006).

Alternatively, colonization of the greater Pacific may have entailed a single, concerted settlement process followed by increasing isolation during recent Polynesian prehistory and a resultant clustering of regional Y-chromosome STR lineages. Both models may be partly correct. The central and eastern Y-chromosome STR clusters derive independently from unrelated Samoan haplotypes (Figure 2), thus suggesting at least two independent excursions from Western Polynesia. However, the central and eastern Y-chromosome STR clusters also consist primarily of closely allied lineages, often differing by a single mutational change. The restricted spatial distribution of such closely related lineages favors the development of these haplotypes during a period of strong regional isolation. Consequently, we appear to be observing a complex interplay between two conflicting forces, mobility and seclusion, throughout the course of Polynesian prehistory.

Here, we have identified a paternal equivalent to the mitochondrial Polynesian motif, with stronger Polynesian affinity and clearer regional structuring. For the first time, genetic data reflect both early migrational processes and, later, an increasingly isolated Pacific. Although archeology traditionally divides Polynesia into western and eastern regions, we recognize a tripartite division, indicating substantial regional isolation following an earlier dispersal phase. Nevertheless, we show that Polynesian mtDNA and Y-chromosome patterns are related, and similar demographic forces helped shape them.

Acknowledgments We would like to thank the Pacific peoples and communities who participated in this study. We also thank J. Watkins and B. Hallmark (University of Arizona) for assistance with genetic dating and J. Schoenfelder (University of California, Los Angeles) for Geographic Information System (GIS) support. This study was partly funded by the National Science Foundation through grant BCS-0432262 ("Austronesian societies: Reading social structure from the genome").

Received 29 December 2006; revision received 4 June 2007.
Polynesian Motif on the Y Chromosome / 533

Literature Cited


Polynesian Motif on the Y Chromosome / 535


