*Proc. Natl. Acad. Sci. USA*
Vol. 95, pp. 9047–9052, July 1998
Anthropology, Evolution

# Testing migration patterns and estimating founding population size in Polynesia by using human mtDNA sequences

**(hypervariable region I/human evolution/New Zealand Maori/sensitivity analysis)**

ROSALIND P. MURRAY-MCINTOSH*†, BRIAN J. SCRIMSHAW‡, PETER J. HATFIELD§, AND DAVID PENNY*

*Institute for Molecular BioSciences, Massey University, P.O. Box 11222, Palmerston North, New Zealand; ‡Institute of Environmental Science and Research, Wellington, New Zealand; and §Renal Unit, Wellington Hospital, Wellington, New Zealand

*Communicated by Patrick V. Kirch, University of California at Berkeley, Stanford, CA, March 30, 1998 (received for review July 10, 1997)*

**ABSTRACT** **The hypervariable 1 region of human mtDNA shows markedly reduced variability in Polynesians, and this variability decreases from western to eastern Polynesia. Fifty-four sequences from New Zealand Maori show that the mitochondrial variability with just four haplotypes is the lowest of any sizeable human group studied and that the frequency of haplotypes is markedly skewed. The Maori sequences, combined with 268 published sequences from the Pacific, are consistent with a series of founder effects from small populations settling new island groups. The distributions of haplotypes were used to estimate the number of females in founding population of New Zealand Maori. The three-step simulation used a randomly selected founding population from eastern Polynesia, an expansionary phase in New Zealand, and finally the random selection of 54 haplotypes. The results are consistent with a founding population that includes ≈70 women (between 50 and 100), and sensitivity analysis shows that this conclusion is robust to small changes in haplotype frequencies. This size is too large for models postulating a very small founding population of "castaways," but it is consistent with a general understanding of Maori oral history as well as the results of recent canoe voyages recreating early trans-oceanic voyages.**

The hypervariable region 1 of human mtDNA now is used extensively to study the origins and migration of modern humans, *Homo sapiens sapiens* (see, for example, refs. 1–3). Relatively consistent patterns of human origin and dispersal are emerging, but it is desirable to investigate in more detail the processes that have led to the current global distribution.

Polynesia is the most suitable place to study human migration into previously uninhabited regions. The earliest migrations are recent (3,200–800 BP) (4, 5), and therefore evidence of migration patterns and their consequences are easier to find. For example, most of the earliest archaeological sites were not submerged by sea level rises at the end of the last ice age and common patterns of extinctions can be found (6). Because the arrival of Polynesians is so recent, it is relatively easy to sequence DNA from bones of animals that have become extinct since human impact (7, 8), and DNA can be sequenced from old human bones in the area, such as on Easter Island (9). Polynesians are also well known from both anthropological and archaeological viewpoints (see, for example, refs. 5 and 10–12) as well as genetically (see refs. 13 and 14). In addition, the migrations are retained in the strong oral histories of Polynesian people, and it has been possible to replicate, based on traditional knowledge and skills, long canoe voyages between major island groups (15, 16).

Polynesians are speakers of a subgroup of Austronesian languages, and their origins are suggested to extend back into mainland Asia (17, 18) with some Melanesian genetic admixture. Pre-Polynesians are thought to have occupied the eastern islands of South-East Asia at ≈2,000 BC, their Lapita culture with its characteristic pottery expanding rapidly through Melanesia and out from the Solomon Islands into the western islands of Polynesia (such as Tonga and Samoa) by ≈1,200–1,000 BC (19, 20). Expansion to the eastern islands of the Pacific occurred largely between AD 1 and 1,000 (20, 21). Eastern Polynesia includes Tahiti (Society Islands), Easter Island, Hawaii, Marquesas, Northern and Southern Cooks, the Australs, Pitcairn, and New Zealand. A colonizer model of rapid exploration and settlement of the uninhabited eastern zone has been proposed based on searching by sailing upwind with a relatively safe downwind return (19).

Polynesian populations are relatively homogeneous phenotypically and genetically. Because they moved into unoccupied areas of the Pacific, for some 3,200 years they were less affected by admixture with other populations. Strong effects from the small size of founder populations and genetic drift are expected, as well as environmental and cultural selection (22). Population bottlenecks have been inferred from minisatellite data (23). Within eastern Polynesia, the last major settlement was New Zealand (Aotearoa) with archaeological sites dating back only to 800 BP (24) although there may be, as judged by release of the Pacific food-rat (*Rattus exulans*), earlier human contact by 1,800 BP (25).

Hawaii, Easter Island, and Aotearoa (New Zealand) are at the end of chains of migrations. For the settling of Aotearoa, there are two major theories that represent ends of a continuum. The most widely accepted theory is a planned settlement with multiple voyages over time bringing people, animals, plants, and cultural artefacts. This fits with our understanding of Maori oral traditions (elaborated by 19th century ethnologists) describing epic voyages requiring great navigational skills. Long sea voyages of Polynesians to New Zealand are supported by recent canoe voyages using only traditional navigational knowledge; these voyages have supported the oral tradition on such matters as seasonal timing, the setting of courses, and pohutukawa trees (*Metrosideros*) in bloom at arrival (16, 26). Traditional stories of return voyages from New Zealand to eastern Polynesia are supported by the discovery of New Zealand obsidian in the Kermadec Islands (27).

Within this framework, McGlone *et al.* (28) propose that up to 500 settlers arriving over several generations around 800 BP would have provided both the necessary numbers to occupy the early coastal sites found throughout the country and the necessary skills, traditions, and knowledge for successful colonization. This number provides the basis for an initial rapid population growth expected in an environment previously uninhabited and therefore extremely rich in seals, large birds (including flightless moa), fish, and shellfish. Population from ≈600 BP is postulated to have grown more slowly to an estimate of 115,000 at 300 years ago (12).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.



Abbreviation: BP, before present.

†To whom reprint requests should be addressed. e-mail: rosmm@compuserve.com.

At the other end of the spectrum are theories suggesting that islands were settled accidentally by random voyages made by one canoe with perhaps 10–20 people (or at most a few canoes) drifting before the wind or current–or by people forced to leave their home island, blindly searching for a new island on which to settle (29). Other authors (see, for example, ref. 30) also restrict the number of original canoes by suggesting that the oral tradition of founding canoes really includes later subsidiary voyages within New Zealand. Such theories are parsimonious with respect to the number of long ocean voyages (see discussion in ref. 16) and have had some acceptance as minimalist theories by which to interpret cultural development in isolation, without multiple interchange of people, ideas, language, artefacts, animals, and plants.

It is now possible to test these models by using mtDNA. The expectation in the settlement of Pacific Islands is a series of founder effects, modified by immigration and emigration over several generations. Under this model, Maori and Hawaii are expected to be a subset of the genetic variability in central eastern Polynesia, which in turn is expected to be a subset from western Polynesia, which itself is a subset of Melanesia. A series of founder effects is expected to lead to low diversity (31).

We report 31 mtDNA sequences in addition to the 23 already reported from New Zealand, and, after combining them with a further 268 sequences from other parts of the Pacific (2–4), we show that the model of settlement of New Zealand developed from archaeology, anthropology, and a general understanding of oral tradition is strongly supported. Given 10–20 people per canoe (32, 33), the estimate of the number of founding women settlers in New Zealand as 70 (50–100) contradicts colonization by a single canoe or indeed by any model with a very small number of settlers.

## MATERIALS AND METHODS

**DNA Extraction.** Samples of single plucked head hairs or venous blood were provided by volunteers in a study approved by the Ethics Committee of the Wellington Area Health Board. Volunteers identified themselves as Maori and recorded anonymously their *iwi* (tribal) affiliations and place of birth for themselves, their mothers, and in most cases maternal grandmothers. Hair shafts were extracted by a standard proteinase K digestion and organic extraction followed by concentration by spin-filtration (34). Extract (1–5 $\mu$l) was used in each amplification reaction. For blood samples (frozen), 200 $\mu$l of each was mixed with 800 $\mu$l of 170 mM $NH_4Cl$ and centrifuged. The pellet was washed four times in 10 mM NaCl/10 mM EDTA (pH 8.0), resuspended in 500 $\mu$l of 50 mM NaOH, heated to 95°C for 15 min, and neutralized with 1 M Tris (pH 7.5). Five microliters of this extract was used in each amplification reaction.

**DNA Amplification and Sequencing.** A region of the human mtDNA control region ≈440 bp in length was amplified by PCR by using primers H16401 (5′-TGATTTCACGGAGGATG-GTG-3′) and L15997 (5′-biotinCACCATTAGCACCCA-AAGCT-3′). Final concentration of reaction components was 50 mM KCl, 10 mM Tris·HCl (pH 9.0), 0.1% Triton X-100, 2.5 mM $MgCl_2$, 0.2 $\mu$M each primer, 0.2 mM dNTPs (Boehringer Mannheim), and 2.5 units of *Taq* DNA polymerase (Promega)/100-$\mu$l reaction. Cycling parameters were 30 cycles of 94°C, 55°C, and 72°C, each step taking 60 s. Control samples without added DNA were processed as a check for contamination. The number of copies of the 9-bp repeat in the COII/$tRNA^{Lys}$ intergenic region was determined by using the specific primers and conditions described in ref. 35. The length of the PCR product was compared with standards having either 111 or 120 bp after electrophoresis on a 5% agarose gel followed by ethidium bromide staining.

Single-stranded DNA sequencing templates were prepared from double-stranded PCR amplification products by binding streptavidin-conjugated beads (Dynal, Oslo, Norway) to biotinylated (L15997) DNA strands and removing nonbiotinylated strands by alkali denaturation, according to manufacturer's instructions. Sequencing was carried out by dideoxynucleotide chain termination technology (36) by using either the Sequenase version 2.0 kit (United States Biochemicals), the Amplicycle sequencing kit (Perkin–Elmer), or an automated sequencing unit (Applied Biosystems model 373) of the Centre for Gene Technology, University of Auckland. Because length heteroplasmy in a G-C rich section in the control region of most Polynesian mtDNA generates blurred sequence readout subsequently (37), only the "L" strand of each sample was sequenced. Changes in sequence at bases numbered lower than ≈16,189 therefore are not reported. Each sample was repeated two or more times until there were no ambiguities in base assignment.

A three-step simulation based on the observed frequencies of haplotypes (Table 1) in eastern Polynesia (2, 3) and in New Zealand (ref. 3; this study) was used to estimate the numbers of females founding the Maori population in New Zealand. The first step selected randomly, with replacement, a maternal founding population of from 4 to 250 people from the frequencies of the 11 haplotypes observed in eastern Polynesia (Table 1, column 2[EP], 108 sequences, frequencies 69, 16, 12, 3, 2, 1, 1, 1, 1, 1, 1). The second step allowed this maternal founding population to expand randomly and exponentially over 30 generations to 50,000 [corresponding to a population of ≈100,000 (32, 38)] and thus to give the haplotypes expected to be present in New Zealand Maori at the time of European settlement. (Mutation was not included within New Zealand because all of the haplotypes observed also occur in other parts of Polynesia.) The third step of the simulation randomly selected 54 haplotypes from this expanded population to compare with the observed frequency of haplotypes in Table 1, column 1[Ma], frequencies 47,5,1,1. [Our sample has 54 Maori haplotypes, 31 from this study and 23 from Sykes *et al.* (3).]

The number and frequency of haplotypes was recorded at all three stages, the founding population, the estimated present population, and the final sample of 54 individuals. The simulations were repeated 20,000 times for each of 46 founding population sizes ranging from 4 to 250. A population size of four is the minimum for four haplotypes, and 250 is the number of females in the largest population size suggested (28).

## RESULTS

Sequences from 31 unrelated New Zealand Maori were identical to Polynesian sequences already known (2, 3) and had the characteristic one copy of the 9-bp repeat in the COII/$tRNA^{Lys}$ intergenic region. We found sequence 1 (Table 1) 27 times, sequence 11 once, and sequence 15 three times. Combined with the 23 samples from Maori described in ref. 3, there is a total of 54 samples, but these still contain only four distinct sequences (haplotypes) between bases 16,189 to 16,388 (40). On the combined data set, one haplotype (sequence 1) occurs in 47 of the 54 people and sequence 15 five times. The frequencies of all haplotypes reported in different parts of Polynesia, together with some Melanesian samples, are shown in Table 1. In some cases, because not all publications report the same length of sequence, two or more haplotypes are identical for the region 16,189 to 16,388. Sequence 1 was found in 87% of the Maori (Ma) samples, and it occurs in 64% of the samples from eastern Polynesia (EP), in 56% of the western Polynesian (WP) samples, and in 23% of Melanesians (MN) sampled. The apparent heterozygosities [$h = (1 - \Sigma x_i^2)$ (41)] for Maori, other eastern Polynesians, and western Polynesians are 0.233, 0.564, and 0.659, respectively, over this region of the mitochondrial genome.

Given the generally low diversity in Polynesians (2, 3), a low number of variants is anticipated in Maori, although the similarity was greater than expected. The lack of variability does not appear to be a sampling artefact. For the present sample, the mother's and maternal grandmother's *iwi* (tribal affiliation) and place of birth were recorded and showed a wide distribution from throughout New Zealand (Table 2) and from all but two of the major tribal areas (38, p. 52). Of these two areas, at least one would be well represented in the sample of ref. 3, which was

Table 1. Variable sites in hypervariable region I (16,189–16,365) of combined Polynesian mtDNA data sets

| Seq. no. | Ma | EP | WP | MN | Anderson no. 16,1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3<br>8 1 1 1 1 2 4 4 6 6 7 9 9 9 9 0 0 1 1 4 4 5 5 5 6 6 6<br>9 3 4 5 7 3 1 7 1 5 4 1 3 4 8 2 4 1 2 2 3 4 5 7 1 2 5<br>TGCATCAACAGCACTATTATACCTGTC | L* | S† |
|---|---|---|---|---|---|---|---|
| Group I‡ | | | | | | | |
| 1. | 47 | 69 | 41 | 20 | C...C..GT.. | 2 | 1 |
| 2. | 0 | 0 | 1 | 0 | C...C..GT.....A.. | – | – |
| 3. | 0 | 0 | 0 | 1 | C...C..GT.....C..... | – | – |
| 4. | 0 | 0 | 0 | 1 | C..GC..GT.. | – | – |
| 5. | 0 | 0 | 0 | 1 | C.T.C..GT.. | – | – |
| 6. | 0 | 0 | 0 | 1 | C...C..GT.....G..... | – | – |
| 7. | 0 | 0 | 1 | 0 | C...C..GT.....A..... | – | 14 |
| 8. | 0 | 0 | 1 | 0 | C...C..GT.C.. | – | 15 |
| 9. | 0 | 0 | 1 | 0 | C...C..GT.....C..... | – | 16 |
| 10. | 0 | 0 | 0 | 1 | C...C..GT...C..... | – | 17 |
| 11. | 1 | 1 | 0 | 0 | C...C..G.. | – | 19 |
| 12. | 0 | 1 | 2 | 0 | C...C..GT..T..... | 1 | 13 |
| 13. | 0 | 1 | 0 | 0 | C.....GT.. | 4 | – |
| 14. | 0 | 0 | 2 | 0 | C...C..GT.....C. | – | 18 |
| 15. | 5 | 12 | 10 | 9 | C...C...T.. | 5 | 6 |
| 16. | 0 | 0 | 0 | 1 | C...C...T.....G..... | – | – |
| 17. | 0 | 0 | 2 | 0 | C...C...T.....C..... | – | – |
| 18. | 0 | 0 | 0 | 1 | C...CT..T.. | – | – |
| 19. | 0 | 1 | 0 | 0 | C...C...T.....G..... | – | 10 |
| 20. | 0 | 0 | 2 | 0 | CA..C...T.. | 7 | – |
| 21. | 0 | 0 | 1 | 0 | ....C...T.....C. | 8 | – |
| 22. | 0 | 0 | 1 | 0 | C...C..... | 11 | 4 |
| Group II | | | | | | | |
| 23. | 0 | 1 | 0 | 0 | .....TG..CA.....C..... | 39 | – |
| 24. | 0 | 0 | 1 | 0 | .....TG..C..G....C..... | 40 | – |
| 25. | 1 | 16 | 1 | 0 | .....TG..C..G....C..G..... | – | 20 |
| 26. | 0 | 0 | 2 | 4 | .....TG..C.....C..G..... | | 21&24 |
| 27. | 0 | 3 | 0 | 0 | .....TG..CA.....C..G..... | – | 22 |
| 28. | 0 | 0 | 2 | 0 | .....TG..C.....C..... | – | 23 |
| 29. | 0 | 0 | 0 | 1 | .....T...C.....C..G..... | – | 25 |
| 30. | 0 | 0 | 0 | 2 | .....T.....C.....C. | – | 27 |
| 31. | 0 | 0 | 0 | 8 | .....T.....C. | | 28&31 |
| 32 | 0 | 0 | 0 | 1 | .....T.....T..C. | – | 29 |
| 33. | 0 | 0 | 0 | 1 | .....T.....T..CT | – | 30 |
| 34. | 0 | 0 | 0 | 2 | .....C.....C..G..... | – | 26 |
| 35. | 0 | 0 | 0 | 2 | .....T...C.....C..C....C. | 41 | – |
| Group III | | | | | | | |
| 36. | 0 | 0 | 1 | 0 | .....T...C.....C. | 20 | – |
| 37. | 0 | 2 | 0 | 0 | .....C..... | – | 34 |
| 38. | 0 | 1 | 0 | 0 | C.....C..... | – | 35 |
| 39. | 0 | 0 | 1 | 0 | .....CC..... | – | 36 |
| Totals (including sequences 40–63 in MN from ref. 5) | | | | | | | |
| | **54** | **108** | **73** | **87** | | | |

Ma, Maori; EP, eastern Polynesia; WP, western Polynesia; MN, Melanesia.
L*, sequence no. in ref. 2.
S†, sequence no. in ref. 3. Sequence no. equivalents in ref. 4 are: 1 = 11,4,6,8,10,12,14; 2 = 2; 3 = 9; 4 = 11; 5 = 3; 6 = 14; 15 = 21,22; 16 = 17; 17 = 18; 18 = 25.
‡Groups as defined in ref. 2.

collected in Auckland but without recorded *iwi* affiliation. In addition, the two groups of samples were collected in centers 600 km apart and both have a similar preponderance of the same sequence. Thus, we consider the sample to be representative.

Table 1 includes the overall frequency of haplotypes from two data sets for New Zealand Maori (54 sequences), two data sets for other eastern Polynesians (108 sequences), and three data sets for western Polynesia (73 sequences) and Melanesia (57 sequences of 87 total are included in Table 1; others are not directly relevant to this study).

A standard method for illustrating the level of diversity in a population is the pairwise distance distribution (or mismatch distribution) (42). Fig. 1 shows the results for Maori, other eastern Polynesians, western Polynesians, and Melanesians. New Zealand Maori sequences are the most similar to each other, then eastern Polynesian. Melanesian sequences show the highest diversity and are comparable with those from other regions of the world (see below). With one exception, Maori sequences are identical or have just 1 bp different. A single sequence (from ref. 3) lacks the 9-bp deletion and gives a small peak at 10 differences in this region of the D-loop. This bimodal distribution is more marked in the eastern Polynesian data and more so again in western Polynesian data, again resulting from the differences between sequences with and without the 9-bp deletion. The Melanesian sequences are more diverse again and have considerable diversity within groups lacking the 9-bp deletion.

The bimodal distribution could be interpreted as a "ragged" distribution (43) and could be used to favor a stable population size over a long time period. However, in the present work, given that humans have not been in Polynesia for a long time period and given the evidence for a small founding population relative to the present population size, an alternative hypothesis is more reasonable. This would be a fusion of two groups of people, probably in Melanesia (20) between the Austronesian-speaking ancestors of Polynesians (with a 9-bp deletion) and an earlier Melanesian population (without this feature).

Comparison of mitochondrial diversity with other parts of the world is shown in Fig. 2 for five groups: Maori, Polynesian, Asian, African, and the Turkana of East Africa. All major groups have higher diversity in sequences than Polynesians in general and Maori in particular. Overall, the results are consistent with the findings that mitochondrial diversity in humans is highest in East Africa and decreases steadily away from that center. The Turkana of East Africa have the highest sequence diversity of any group yet studied (44) yet have a similar population size to the Maori of New Zealand. Population size alone cannot be an explanation for the level of diversity of sequences of either group. The origin of a people and their migration history must be considered.

Simulation estimates of the numbers of women who originally arrived in New Zealand (the founding population) are shown in Fig. 3. The size of this founding population was varied from 4 (the minimum to give four haplotypes) to 250. For each population size, the number of times the final sample of 54 people would have 1 to 11 different haplotypes is shown. With small founding population sizes, only one or two distinct sequences are expected in the final sample. At higher founding population sizes, five, six, or more different sequences are expected, but even with a founding population of 250, only 5 or 6 of the 11 haplotypes in eastern Polynesia are expected. For a founding population size from ≈40 to 90, it is most likely that four distinct haplotypes will be found, and this is the "best estimate" when considering just the number, not the frequency, of haplotypes. This result has a wide confidence interval, so a more strict criterion then was used.

A tighter lower bound on the numbers is given in Fig. 4 where the frequency of finding four haplotypes with the very unequal distribution of frequencies is shown. This more strict criterion includes simulations where the product of the frequencies of the four haplotypes is ≤235 ($47 \times 5 \times 1 \times 1$). A distribution of 46,6,1,1 also was allowed. The results (Fig. 4) give a slightly larger estimate of the founding population with the most likely number of women being 70. By using a one-tailed test, we found that the 95% confidence interval is 50 or more women and the 99% confidence interval is 30 or more (Table 3). The additional factor in this calculation (compared with Fig. 3) is that, with larger founding populations, the distributions of the 54 sampled haplotypes are more unequal than for small founding populations (data not shown). Thus, the two factors leading to the result in Fig. 4 are the frequency of finding four haplotypes in the final sample (Fig. 3) and an increased chance of an unequal distribution of haplotypes when the founding population is larger.

A more relaxed criterion also was simulated that allowed any sample with four haplotypes in which one of the rare haplotypes occurs twice and the other just once. These results are similar to Fig. 4 (data not shown) and support a similar size for the number of females in the founding population. It is apparent that the highly unequal distribution of haplotype frequencies is an unusual feature of the data. The main feature of the simulation results is that they are inconsistent with a very small population from just

Table 2. Distribution of samples by major tribal areas

| Major tribal area | Population in 1801*, % | Wellington sample, % | Likely presence in Auckland sample | No. with Seq. 11 | No. with Seq. 15 |
|---|---|---|---|---|---|
| Arawa | 7 | 7 | + | | |
| Manawatu | 10 | 17 | | | 1‡ |
| Mataatua | 14 | 10 | + | | |
| Tai Tokerau | 18 | 13 | + | | |
| Tainui | 17 | 0 | ++ | | |
| Tairawhiti† | 8 | 17 | | | |
| Takitimu† | 3 | 27 | | | 2 |
| Taranaki | 10 | 7 | | | |
| Wanganui | 12 | 0 | | | |
| Te wai pounamu | – | 3 | | 1 | |

*Population estimates from ref. 38, p. 52.
†By 1840, Tairawhiti and Takitimu combined contained 33% of the Maori population.
‡Incomplete data on *iwi* affiliation.

one or a small number of canoes but support most strongly a founding female population size of 50–100 and possibly more.

It is important to have an estimate of the sensitivity of the model to variations in the parameters; those in which small changes in their value lead to significant differences in output require close attention (39). For the sensitivity analysis, simulations were re-run: omitting one sequence from each of the four haplotypes observed in Maori; allowing one additional sequence for each haplotype; and adding a fifth haplotype. Each case was simulated 5,000 times, the most likely number of women was determined for the strict criterion (that includes frequencies of haplotypes), and the 1% and 5% cutoffs were estimated (Table 3). As expected, the most sensitive parameter is the number of rare haplotypes (see rows 4 and 8 in Table 3), but even if the Maori sample, just by chance, had not included one rare haplotype, the results still reject a single, or small number of, canoes. Similarly, if a fifth allele had been observed in Maori, it would increase the expected founding population size to over 100. This point is interesting because, during sequencing, one sample indicated an additional allele, but there was insufficient DNA to verify all positions in the sequence, so it was omitted from the analysis. Again, as expected, changes to the Maori sample (with four haplotypes) had more effect than gaining or losing a haplotype from eastern Polynesia (with 11 haplotypes), as shown by the last two rows in Table 3. Thus, the sensitivity analysis has been especially important in showing that our conclusions about the size of the maternal founding population are robust with respect to the frequencies of haplotypes in the samples.

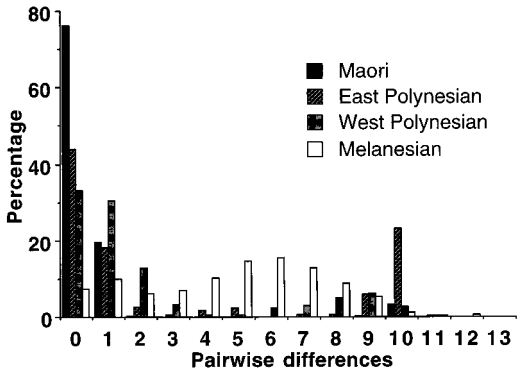


FIG. 1. Distribution of distances between pairs of individuals within Polynesian populations and within Melanesia. From left to right for each set, New Zealand Maori, other eastern Polynesians, western Polynesians, and Melanesians. There is decreasing mitochondrial diversity going from Melanesia to New Zealand Maori and a bimodal distribution.

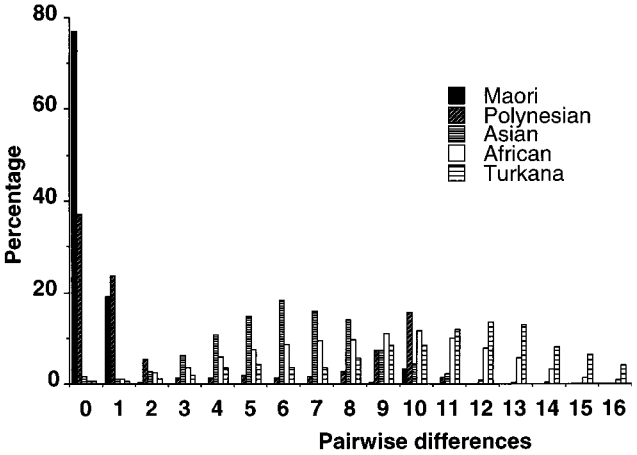


FIG. 2. Distribution of distances within: New Zealand Maori; all other Polynesians; Asians; Africans; and the Turkana of East Africa. (A few (≈6%) of the values of differences between Turkana sequences are not shown, lying between 17 and 20 bases.) Maori and Turkana were similar medium-sized populations but have, respectively, the lowest and highest mitochondrial diversity of populations studied at present.

## DISCUSSION

As expected, only a subset of eastern (and western) Polynesian genetic diversity is represented in New Zealand Maori. Although a minimum of four founding female settlers would be needed for four haplotypes, it is most unlikely that all four haplotypes would survive during the phase of population expansion with sufficient frequency to occur in the final sample of 54. Many more settlers appear to have been necessary to explain the observed skewed distribution of the four haplotypes. By comparing the genetic diversity in eastern Polynesian, and in Maori, we estimate that the female founding population of New Zealand included ≈50–100 women. Because of the skewed distribution in Fig. 4 the upper limit (100) is less clear, but given the difficulty in sailing canoes to New Zealand, the value is reasonable [however, Anderson (27) has briefly considered a large founding population]. Our estimates are quite inconsistent with models (29, 30) that assume that only one canoe, or a very small number of canoes, arrived. In contrast, it is in very good agreement with a common understanding of the Maori oral tradition of 8–10 canoes with a small number of people (10–20 people, thus probably 5–10 females) per canoe (see refs. 32 and 33). The conclusion is consistent with more recent work that has tested the possibility of deliberate exploration and migration (16, 19).

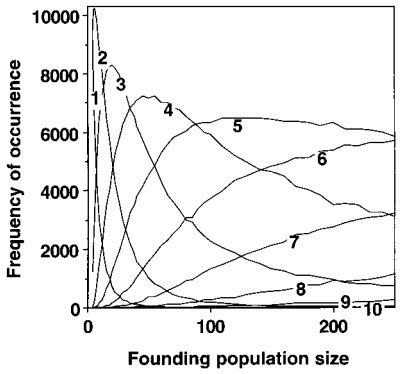


FIG. 3. Effect of the founding maternal population size on the expected number of haplotypes for the New Zealand Maori sample. The size of the founding population from eastern Polynesian is indicated on the *x* axis; the *y* axis shows the frequency of 20,000 simulations that the indicated number of alleles (1–10) are present in a sample of 54 sequences after the founding population has expanded to 50,000 females is on the *y* axis. The estimated founding population size on this criterion is 40–90 females to give four haplotypes as observed.

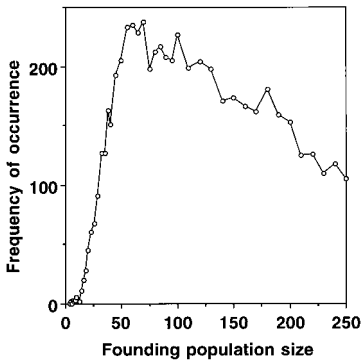


FIG. 4. The frequency, relative to founding maternal population size, of final samples with a highly unequal distribution of four haplotypes (two haplotypes occur only once in the 54 sequences). Larger founding population sizes are more likely to lead to an unequal distribution of haplotypes, and the estimated founding population size is now ≈70 (50–100) females. The 95% confidence interval on a one-tailed test excludes a founding population size of less than ≈50 (see Table 3).

Our estimates of the founding female population assume that the samples have no hidden biases favoring any particular haplotype. As discussed earlier (45), the mitochondrial genome is expected to be at least as well dispersed as nuclear alleles in that, in nearly all earlier societies, women transferred much more frequently between groups and settlements than men (46). In addition, adoption of children has been reported to be widespread throughout Polynesia (10, 11), and this also would be expected to reduce any local concentrations in the distribution of particular sequences. It is likely that several women in a canoe would be related maternally, but this too would make it less likely that a small number of founders would give the observed distribution, including the skewed distribution. The presence of older women in canoes also would tend to increase the size of our estimate of founding population size. Overall, our estimate of 50–100 women seems conservative in several respects.

Could there have been earlier settlement by a small population that grew more slowly? A founding population of 50 has been suggested to increase to ≈400 after 200–400 years, allowing for a longer prehistory than proposed (32). There is no good evidence yet for permanent earlier settlement (see the Discussion in ref. 33), and given an abundant food supply for the first settlers and no new indigenous diseases, an initial rapid population expansion may be expected until resources became short (28, 47). The recent discovery of bones of kiore (the Pacific food-rat) dating from 1,800 BP (25) indicates that there may have been early Polynesian contact 1,000 years before permanent human habitations are represented in the archaeological record (24).

Table 3. Sensitivity analysis of the model

| | Haplotypes* | Optimum† | 99%‡ | 95%‡ |
|---|---|---|---|---|
| 1 | 47, 5, 1, 1 | 72 | 29 | 49 |
| **2** | **46, 5, 1, 1** | **80** | **28** | **50** |
| **3** | **47, 4, 1, 1** | **77** | **26** | **46** |
| **4** | **47, 5, 1** | **45** | **14** | **28** |
| 5 | 48, 5, 1, 1 | 86 | 28 | 48 |
| 6 | 47, 6, 1, 1 | 81 | 26 | 45 |
| 7 | 47, 5, 2, 1 | 78 | 28 | 50 |
| 8 | 47, 5, 1, 1, 1 | 119 | 45 | 78 |
| **9** | **47, 5, 1, 1, (40 gen.)** | **114** | **35** | **60** |
| 10 | 69,16,12, 3, 2,1 × 5(EP) | 110 | 37 | 62 |
| 11 | 69,16,12, 3, 2,1 × 7(EP) | 75 | 35 | 55 |

*Numbers in the first row are the population sizes with the observed frequency of Maori haplotypes. The next three rows omit one sequence of each haplotype; then four rows add one additional sequence. The ninth row allows the founding population to expand more slowly over 40 generations, rather than 30. Rows 10 and 11 omit, or add, one unique sequence from the eastern Polynesian sample and are comparable to rows 4 and 8. Reducing diversity in eastern Polynesia has less effect on population sizes than increasing the observed Maori diversity.

†The most likely number of women in the founding population for the simulation.

‡Founding populations of this number of women, or fewer, can be excluded at the 99 and 95% confidence intervals under the model tested. Simulations were carried out 5,000 times except for the first row with 20,000 runs.

However, with the present data, an early settlement combined with a slower population increase makes it even less likely that there was a small founding population—rare alleles are more likely to be lost during a slower increase in population than during a faster increase. This point is shown in row 9 of Table 3, where the founding population was allowed to expand over 40, rather than 30, generations; a greater number of women would be required to give the observed Maori diversity. So, if there were earlier settlement, it would require a founding population of over 100 women, but this would be more likely to leave an archaeological record. Thus, we consider a small, early population to be incompatible with the data.

Several explanations could be proposed for the very unequal frequencies of the four haplotypes (namely 47, 5, 1, 1). These include one or more factors involving chance, continuing migration, and genetic and/or cultural selection as follows:

1. The group of women setting sail for Aotearoa may, by chance, have had a different frequency of haplotypes from the population in the islands of origin.
2. Diversity of the mtDNA studied in Maori may reflect truly the diversity found in central eastern Polynesia at the time of migration. However, subsequent migrations from western to eastern Polynesia may have produced the higher diversity currently shown, for example, in the Cook Islands.
3. Canoes with females with the rare haplotypes may have arrived in New Zealand later, and consequently they expanded less than haplotypes arriving earlier. Maori oral tradition (quoted in ref. 48) records that some canoes arrived after earlier settlement. For example "Wairake had sailed with her father. . . and she and her descendants had intermarried with 'the early tribes of Tuhoeland' to establish a new tribe. . . . " Our simulations did not include the effect of differing times of arrival.
4. Lineages with other mtDNA sequences found in eastern Polynesia may not have survived either the journey to New Zealand or conditions after reaching it. Selection favoring the common haplotype may have occurred. There have been suggestions that certain genetic variants may be advantageous in surviving long canoe voyages, a version of the "thrifty gene" hypothesis (49, 50). In addition, mitochondrial variants have been linked to some forms of diabetes (reviewed in ref. 51), which might play such a role.
5. High ranking females may have carried the common haplotype in the original population, and there was differential survival of offspring favoring their children. This would give a cultural linkage to survival, rather than a genetic advantage.

Similar explanations could be invoked for the increase in frequency of 9-bp deletions across Polynesia; from Table 1, Maori mtDNA has 98% sequences with the deletion, eastern Polynesia 80%, western Polynesia 89%, and the Melanesian sample 41%.

The extent of genetic variability and admixing of populations is also of interest in relation to exploring genetic contributions and linkages to nonsusceptibility and susceptibility to specific diseases. Maori have a lower incidence than non-Maori for some diseases, for example skin cancer (52), phenylketonuria (53), and cancer of the large bowel. Examples of higher incidence include some infections, inflammatory and immune diseases, cancers, and metabolic diseases (for example, pneumonia, hepatitis, tuberculosis, ear diseases, nephritis, asthma, rheumatic and hypertensive heart disease, cancers of the stomach, liver, lung and cervix, and diabetes, gout, and obesity) (52). In general, it is

thought that the higher incidences are caused by socioeconomic factors and that there is negligible contribution from genetic susceptibility (52). However, few genetic investigations have been carried out with Maori samples. These include studies related to the immune system [T cell receptor polymorphism (54) and HLA polymorphism (see, for example, refs. 21, 55, and 56)], globin gene markers (57), hemophilia (58), minisatellites (23), and variable numbers of tandem repeats (59). In general, these studies show genes with restricted diversity and types, combinations, and/or frequencies different from Caucasians. There is as yet little specific information on whether genetic features may influence susceptibilities to major diseases.

As a general comment, we note that genetic data, including interpretation and analysis of DNA sequences, are just one aspect of the multifaceted and profound issues of relationship, kinship, and differences between human groups, their cultures, and perceptions of self-identity. Our data and the interpretation thereof do not include specifics of traditional knowledge and genealogies and therefore offer neither affirmation nor denial. Experts in historical and cultural matters can use, re-interpret, reject, or add to our analysis according to their different perspectives. In particular, we welcome additional information refining our understanding.

Although this work is based on just one section of Polynesia, this region is perhaps the only place at present where hypotheses about the processes of migration of early humans into previously uninhabited regions can be tested quantitatively. For example, models for the founding populations of Australia/New Guinea (60) and the Americas (see, for example, ref. 61) also assume small founding populations and bottlenecks, but these models are not yet quantitative. Polynesia is an excellent region to test models of earlier human expansions.

We thank Professor Paul Harvey for computer facilities for analyses, Z. Velickovic for some 9-bp analyses and for collecting some samples, Elizabeth Watson, John Holloway, and Gina Lento, who assisted with methodology, Elizabeth Watson for supplying comparative data, Professor Mason Durie for initial consultation and interest and for commenting on the final manuscript, and particularly the enthusiastic volunteers who gave samples and expressed strong interest in this study. We acknowledge salary for R.P.M.-M. from the Health Research Council of New Zealand.

1. Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K. & Wilson, A. C. (1991) *Science* **253,** 1503–1507.
2. Lum, J. K., Rickards, O., Ching, C. & Cann, R. L. (1994) *Hum. Biol.* **66,** 567–590.
3. Sykes, B., Leiboff, A., Low-Beer, J., Tetzner, S. & Richards, M. (1995) *Am. J. Hum. Gen.* **57,** 1463–1475.
4. Redd, A. J., Takezaki, N., Sherry, S. T., Mcgarvey, S. T., Sofro, A. S. M. & Stoneking, M. (1995) *Mol. Biol. Evol.* **12,** 604–615.
5. Bellwood, P. S. (1987) *The Polynesians: Prehistory of an Island People* (Thames and Hudson, London), 2nd Ed.
6. Steadman, D. W. (1995) *Science* **267,** 1123–1131.
7. Cooper, A., Mourer-Chauvire, C., Chambers, G. K., von Haeseler, A., Wilson, A. C. & Pääbo, S. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 8741–8744.
8. Cooper, A., Rhymer, J., James, H. F., Olson, S. L., McIntosh, C. E., Sorenson, M. D. & Fleischer, R. C. (1996) *Nature (London)* **381,** 484.
9. Hagelberg, E., Quevedo, S., Turbon, D. & Clegg, J. B. (1994) *Nature (London)* **369,** 25–26.
10. Sahlins, M. D. (1958) *Social Stratification in Polynesia* (Univ. of Washington Press, Seattle).
11. Goldman, I. (1970) *Ancient Polynesian Society* (Univ. of Chicago Press, Chicago).
12. Kirch, P. V. (1984) *The Evolution of the Polynesian Chiefdoms* (Cambridge Univ. Press, Cambridge, U.K.).
13. Hill, A. V. S. & Serjeantson, S. W., eds. (1989) *The Colonization of the Pacific: A Genetic Trail* (Clarendon, Oxford).
14. Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. (1994) *The History and Geography of Human Genes* (Princeton Univ. Press, Princeton).
15. Finney, B. R. (1977) *Science* **196,** 1277–1285.
16. Finney, B. (1994) in *The Origins of the First New Zealanders*, ed., Sutton, D. G. (Auckland Univ. Press, Auckland), pp. 52–76.
17. Serjeantson, S. W. & Hill, A. V. S. (1989) in *The Colonization of the Pacific: A Genetic Trail*, eds. Hill, A. V. S. & Serjeantson, S. W. (Clarendon, Oxford), pp. 286–294.
18. Melton, T., Peterson, R., Redd, A. J., Saha, N., Sofro, A. S. M., Martinson, J. & Stoneking, M. (1996) *Am. J. Hum. Genet.* **57,** 403–414.
19. Irwin, G. J. (1989) *J. Polyn. Soc.* **98,** 167–206.
20. Kirch, P. V. (1997) *The Lapita Peoples* (Blackwell, Oxford).
21. Walter, R. (1994) in *The Origins of the First New Zealanders,* ed. Sutton, D. G. (Auckland Univ. Press, Auckland), pp. 182–203.
22. Serjeantson, S. W. (1989) in *The Colonization of the Pacific: A Genetic Trail*, eds. Hill, A. V. S. & Serjeantson, S. W. (Clarendon, Oxford), pp. 120–173.
23. Flint, J., Boyce, A. J., Martinson, J. J. & Clegg, J. B. (1989) *Genetics* **83,** 257–263.
24. McFadgen, B. G., Knox, F. B. & Cole, T. R. L. (1994) *Radiocarbon* **36,** 221–236.
25. Holdaway, R. N. (1996) *Nature (London)* **384,** 225–226.
26. Buck, P. H. (Te Rangi Hiroa) (1926) *J. Polyn. Soc.* **35,** 181–203.
27. Anderson A. J. & McFadgen, B. (1990) *Archeol. Oceania* **25,** 37–42.
28. McGlone, M. S., Anderson, A. J. & Holdaway, R. N. (1994) in *The Origins of the First New Zealanders*, ed. Sutton, D. G. (Auckland Univ. Press, Auckland), pp. 136–163.
29. Sharp, A. (1957) *Ancient Voyagers in Polynesia* (Penguin, Harmondsworth, U.K.).
30. Orbell, M. R. (1985) *Hawaiki: A new Approach to Maori Tradition* (Univ. of Canterbury Press, Christchurch, U.K.).
31. Maruyama, T. & Fuerst, P. A. (1985) *Genetics* **111,** 691–703.
32. Brewis, A. A., Molloy, M. A. & Sutton, D. G. (1990) *Am J. Phys. Anthrol.* **81,** 343–356.
33. Sutton, D. G. (1994) in *The Origins of the First New Zealanders,* ed. Sutton, D. G. (Auckland Univ. Press, Auckland), pp. 243–258.
34. Higuchi, R., Von Beroldingen, C. H., Sensabaugh, G. F. & Erlich, H. A. (1988) *Nature (London)* **332,** 543–546.
35. Wrischnik, L. A., Higuchi, R. G., Stoneking, M., Erlich, H. A., Arnheim, N. & Wilson, A. C. (1987) *Nucleic Acid Res.* **15,** 529–524.
36. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
37. Bendall, K. E. & Sykes, B. C. (1995) *Am. J. Hum. Genet.* **57,** 248–256.
38. Pool, I. (1991) *Te Iwi Maori: A New Zealand Population, Past, Present and Projected* (Auckland Univ. Press, Auckland).
39. McIntosh, J. E. A. & Murray-McIntosh, R. P. (1980) *Mathematical Modelling and Computers in Endocrinology* (Springer-Verlag, Berlin), pp. 23 and 224.
40. Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., *et al.* (1981) *Nature (London)* **290,** 457–465.
41. Nei, M. (1987) *Molecular Evolutionary Genetics* (Columbia Univ. Press, New York), p. 177.
42. Rogers, A. R. & Harpending, H. (1992) *Mol. Biol. Evol.* **9,** 552–569.
43. Harpending, H. C., Sherry, S. T., Rogers, A. R. & Stoneking, M. (1993) *Curr. Anthrol.* **34,** 483–496.
44. Watson, E. E., Foster, P., Richards, M. & Bamdelt, H.-J. (1997) *Am. J. Hum. Gen.* **61**, 694–704.
45. Penny, D., Steel, M. A., Waddell, P. J. & Hendy, M. D. (1995) *Mol. Biol. Evol.* **12,** 863–882.
46. Murdock, G. P. (1960) *Social Structure* (Macmillan, New York), 2nd ed.
47. Flannery, T. F. (1995) *The Future Eaters: An Ecological History of the Australasian Lands and People* (G. Braziller, New York).
48. Sissons, J. (1991) *Te Waimana: The Spring of Mana* (Univ. of Otago Press, Dunedin, New Zealand), p. 7.
49. Neel, J. V. (1962) *Am. J. Hum. Genet.* **14,** 353–362.
50. Dowse, G. & Zimmet, P. (1993) *Br. Med J.* **306,** 532–533.
51. Gerbitz, K. D., Gempel, K. & Brdiczka, D. (1996) *Diabetes* **45,** 113–126.
52. Pomare, E., Keefe-Ormsby, V., Ormsby, C., Pearce, N., Reid, P., Robson, B. & Watene-Haydon, N. (1995) *Hauora: Maori Standards of Health III* (Maori Health Research Centre, Wellington, New Zealand).
53. Hertzberg, M., Jahromi, K., Ferguson, V., Dahl, H. H. M., Mercer, J., Mickleson, K. N. P. & Trent, R. J. (1989) *Am. J. Hum. Genet.* **44,** 382–387.
54. Abbott, W. G. H., Tan, P. L. J., Skinner, M. A., Marbrook, J., Peake, J. S., Woodfield, D. G. & Geursen, A. (1993) *Immun. Cell Biol.* **71,** 543–548.
55. Gao, X., Zimmett, P. & Serjeantson, S. W. (1992) *Hum. Immunol.* **34,** 153–161.
56. Henry, S., Perry, H., Roberts, M. & Woodfield, G. (1994) *N. Z. Med. J.* **107,** 119–121.
57. Hertzberg, M. S., Mickleson, K. N. P. & Trent, R. J. (1992) *Hum. Hered.* **42,** 157–161.
58. Van-de-Water, N. S., Ridgway, D. & Ockelford, P. A. (1991) *J. Med. Genet.* **28,** 171–176.
59. Clark, A. G., Hamilton, J. F. & Chambers, G. K. (1995) *Genetica* **96,** 37–49.
60. Roberts-Thomson, J. M., Martinson, J. J., Norwich, J. T., Harding, R. M., Clegg, J. B. & Boettcher, B. (1996) *Am. J. Hum. Genet.* **58,** 1017–1024.
61. Merriwether, D. A., Hall, W. W., Vahlne, A. & Ferrell, R. E. (1996) *Am. J. Hum. Genet.* **59,** 204–212.