

ORIGINAL ARTICLE

Origins of the Moken Sea Gypsies inferred from mitochondrial hypervariable region and whole genome sequences

Kelsey Needham Dancause^{1,2}, Chim W Chan^{1,2}, Narumon Hinshiranan Arunotai³ and J Koji Lum^{1,2,4}

The origins of the Moken ‘Sea Gypsies,’ a group of traditionally boat-dwelling nomadic foragers, remain speculative despite previous examinations from linguistic, sociocultural and genetic perspectives. We explored Moken origin(s) and affinities by comparing whole mitochondrial genome and hypervariable segment I sequences from 12 Moken individuals, sampled from four islands of the Mergui Archipelago, to other mainland Asian, Island Southeast Asian (ISEA) and Oceanic populations. These analyses revealed a major (11/12) and a minor (1/12) haplotype in the population, indicating low mitochondrial diversity likely resulting from historically low population sizes, isolation and consequent genetic drift. Phylogenetic analyses revealed close relationships between the major lineage (MKN1) and ISEA, mainland Asian and aboriginal Malay populations, and of the minor lineage (MKN2) to populations from ISEA. MKN1 belongs to a recently defined subclade of the ancient yet localized M21 haplogroup. MKN2 is not closely related to any previously sampled lineages, but has been tentatively assigned to the basal M46 haplogroup that possibly originated among the original inhabitants of ISEA. Our analyses suggest that MKN1 originated within coastal mainland SEA and dispersed into ISEA and rapidly into the Mergui Archipelago within the past few thousand years as a result of climate change induced population pressure.

Journal of Human Genetics (2009) 54, 86–93; doi:10.1038/jhg.2008.12; published online 16 January 2009

Keywords: anthropology; mitochondrial DNA; Moken; phylogeography; Southeast Asia

INTRODUCTION

The sea nomads of Southeast Asia (SEA) are divided linguistically, culturally and geographically into three groups: the Moken and the related Moken, the Orang Suku Laut (literally, the ‘Tribe of Sea People’) and the Bajau Laut.¹ The Moken, also known as the sea gypsies of the Andaman Sea, now number from 2000 to 3000 individuals.² They inhabit the Mergui Archipelago off the coast of Myanmar and Thailand (Figure 1), where until recently they spent most of the year on their boats (kabang) and subsisted through maritime foraging. During the wet monsoon season (mid-May to October), they built temporary shelters on the coast, and relied on food gathering in the forests and trading for subsistence.^{3–7} Ethnographic records of the Moken are relatively abundant due to extensive surveys during the British colonial period of the nineteenth and twentieth centuries, but their origins remain mostly speculative.

The Moken language belongs to the Malayo-Polynesian branch of the Austronesian language family.^{8,9} Although anatomically modern humans first reached and settled island Southeast Asia (ISEA) during the Pleistocene, between 45 000 and 50 000 years ago, the ancestry of most modern populations in ISEA is linguistically linked to the mid-Holocene range expansion of the Austronesian language family. Today,

the Austronesian languages are spoken throughout ISEA and the Pacific: west to east, from Madagascar to Rapa-nui (Easter Island) and, north to south, from Taiwan to New Zealand.¹⁰ According to archaeo-linguistic reconstructions, pre-Austronesian languages likely originated in the southern Chinese mainland 5000–6000 years ago among agriculturalists who migrated to Taiwan and there fully developed the Austronesian languages.¹¹ This argument is supported by the diversity of Austronesian languages within Taiwan, where 9 of the 10 branches of the language family are spoken. Aided by rice cultivation and skilled maritime navigation,¹² Austronesian speakers are suggested to have spread rapidly from Taiwan to the Philippines, Indonesia, parts of Melanesia and the rest of the Pacific beginning 4000 years ago, biologically assimilating to various degrees with long resident indigenous populations along the way.¹³

Recent studies of mitochondrial DNA (mtDNA) variation have greatly expanded upon this two-tiered model of ISEA colonization by emphasizing the movement of indigenous populations in Sundaland, the continent exposed by low sea levels during the Pleistocene that encompassed the Malay Peninsula, Sumatra, Java and Borneo. These migrations were likely prompted by rising sea levels that drowned up to half of the region’s land area between 15 000 and 7000 years

¹Department of Anthropology, Binghamton University, Binghamton, NY, USA; ²Laboratory of Evolutionary Anthropology and Health, Binghamton University, Binghamton, NY, USA; ³Social Research Institute, Chulalongkorn University, Bangkok, Thailand and ⁴Department of Biological Sciences, Binghamton University, Binghamton, NY, USA
Correspondence: Dr JK Lum, Department of Anthropology, Binghamton University, PO Box 6000, Binghamton, NY 13902-6000, USA.
E-mail: klum@binghamton.edu

Received 4 December 2008; accepted 5 December 2008; published online 16 January 2009

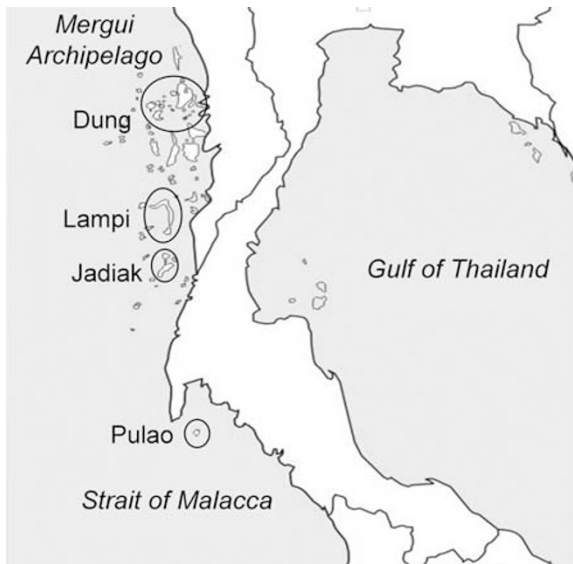


Figure 1 Map showing the location of the Mergui Archipelago within Southeast Asia and sampling locations.

ago.^{14–16} The high level of mtDNA diversity in this region today is characterized by a large number of indigenous clades, of which up to 20% were dated to the initial Pleistocene colonization about 50 000 years ago, and another 20% to the mid-Holocene Out-of-Taiwan expansion supported by linguistic and archeological evidence. The remaining majority of mtDNA lineages were dated between 5000 and 15 000 years ago and associated with the expansion and movement of indigenous populations within ISEA and from mainland SEA into ISEA.¹⁵ This movement coincides with the period of postglacial rising sea levels that flooded Sundaland, prompting migration to higher altitudes or the adoption of stilt or boat housing and sea-based subsistence strategies for those who remained at lower altitudes. If the homeland of the early Austronesian speakers lies not in Taiwan but in SEA or ISEA, with Taiwan representing an early offshoot of Austronesian voyagers and the linguistic diversity there reflecting not language origins but isolation, rising sea levels also provide an explanation for the dispersal of the Austronesian languages.^{15–18}

Moken ethnohistory asserts their ancestral ties to mainland SEA. According to these accounts, Moken ancestors lived in settlements of the Myanmar–Malaya mainland and practiced agriculture, but were driven to the coast by the Burmese to the north and the Malays to the south, and subsequently settled in the Mergui Archipelago.³ Continuous raidings by pirates further forced the Moken to adopt a sea-based lifestyle to avoid capture.⁵ Alternatively, Ivanoff⁶ has suggested that the Moken originated in China 4000 years ago and eventually split off from other migrating groups as late as the early seventeenth century. In contrast, subsistence pattern similarities among the Moken and other sea nomads of the region led Sopher⁴ to propose an ancient common origin of these groups in ISEA, which implies a northward maritime migration of the Moken to their present location in the Mergui Archipelago.⁵ Initial studies of Moken underwater visual acuity suggested long-term adaptations consistent with an extensive history of maritime subsistence,¹⁹ but more recent work indicates that these skills reflect short-term acclimation²⁰ and so could also coincide with a recent adoption of a sea-based lifestyle. Thus, the origins of the Moken remain unresolved.

To investigate the maternal origin of the Thai Moken, we analyzed mtDNA variation from both the hypervariable segment I (HVSI) and

whole genome sequences. Available mitochondrial sequences of neighboring mainland SEA and ISEA populations from recent studies were included for comparative purposes.^{14–16,21–31}

MATERIALS AND METHODS

Subjects

Twelve Moken adults contacted in early 1994 on the Surin Islands were included in this study. These individuals were born on four different islands of the Mergui Archipelago: Dung ($n=6$), Lampi ($n=1$), Jadiak ($n=4$) and Pulao ($n=1$) (Figure 1). At the time of collection, there were 201 Moken individuals from 45 households on the Surin Islands, coming from various upper islands in the Mergui Archipelago.⁷ The sample included two paternally related individuals, but participants are believed to be maternally unrelated based on self-reported pedigrees. Plucked hair samples were collected with informed consent for DNA extraction. This study was approved by Binghamton University's Human Subjects Research Review Committee.

DNA extraction

DNA was extracted from plucked hair using an abbreviated silica protocol³² previously described by Lum *et al.*³³ Briefly, hair roots were washed with ethanol, rinsed with distilled water and incubated in 40 μ l of silica and 900 μ l of L6 buffer (4.5 M guanidinium thiocyanate, 0.1 M Tris–Cl (pH 6.4), 20 mM EDTA and 1.1% Triton X-100) at room temperature for 10 min. After centrifugation, the pellet was washed twice with 900 μ l L2 buffer (5 M guanidinium thiocyanate, 0.1 M Tris–Cl (pH 6.4)), twice with 500 μ l 70% ethanol and once with 500 μ l acetone. After drying, DNA was eluted from silica in 1 ml Tris–EDTA buffer (10 mM Tris–Cl (pH 8.0) and 1 mM EDTA).

Mitochondrial HVSI sequencing and phylogenetic analysis

The mitochondrial HVSI was amplified using primers L15996 and H16401.³⁴ PCR products were purified using the Millipore Manu03050 Filter Plate (Millipore, Billerica, MA, USA), and were sequenced in both directions with the BigDye Terminator Kit v3.1 on an ABI 3730xl DNA Analyzer (Applied Biosystems, Foster City, CA, USA). Sample sequences were compared and aligned to the revised Cambridge Reference Sequence³⁵ using SeqScape v2.1.1 (Applied Biosystems).

Unique Moken HVSI lineages ($n=2$) were identified based on 355 base pairs (bp) of sequences (nucleotide positions (np) 16024–16378). Unbiased heterozygosity (h) was calculated using the equation $h=n(1-\sum x_i^2)/(n-1)$, where x_i is the frequency of the i th haplotype and n is the number of individuals in the sample.³⁶ Haplogroup designation for both Moken lineages was not unambiguously resolved based on HVSI polymorphisms alone, thus whole mitochondrial genome sequencing (see below) was carried out. For each of the Moken mtDNA HVSI haplotypes, a haplogroup-specific median-joining network diagram³⁷ was constructed using Network 4.2.0.1 (<http://fluxus-engineering.com>) to illustrate the relationship of the Moken to their neighbors^{14,15,22,24,27} (Figures 2 and 3).

Whole mitochondrial genome sequencing and phylogenetic analysis

On the basis of the assumption that all individuals sharing identical HVSI haplotypes were likely to have inherited identical polymorphisms in the rest of the mitochondrial genome, whole mitochondrial genome sequencing was performed on two Moken samples bearing the two distinct HVSI haplotypes observed (see above). Twenty-four overlapping fragments of approximately 800 bp were amplified using primers and conditions described by Rieder *et al.*³⁸ PCR products were purified and sequenced, and polymorphisms were identified as described above for HVSI. These sequence data appear in GenBank under accession numbers FJ442938–FJ442939.

Sequences for all whole mitochondrial genome sequences from Asia, Melanesia, Micronesia, Polynesia and South Asia of the Human Mitochondrial Genome Database³⁹ and other regional sequences deposited in GenBank were obtained for initial analyses. Sequences were aligned using the MAVID/AMAP multiple alignment server.⁴⁰ Phylogenetic analysis of the whole genome data was carried out using PHYLIP 3.6.⁴¹ Pairwise genetic distance between the two

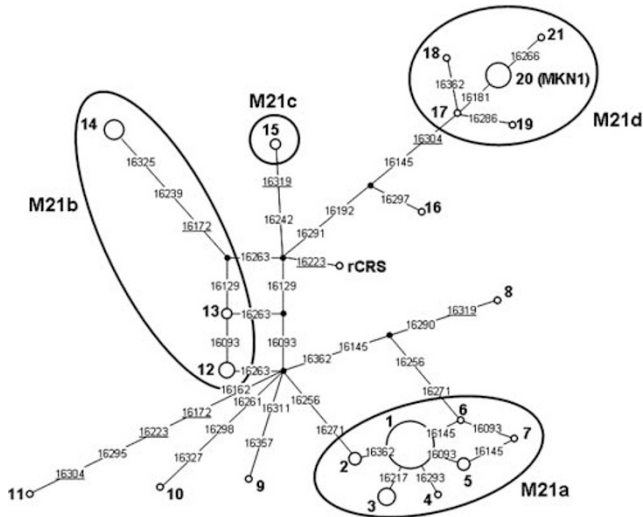


Figure 2 A median-joining³⁷ network diagram consisting of the major Moken hypervariable segment I (HVSI) (nucleotide positions 16024–16378) lineage (MKN1) and other haplogroup M21 sequences found in Southeast Asia (SEA).^{14,15,24,27}

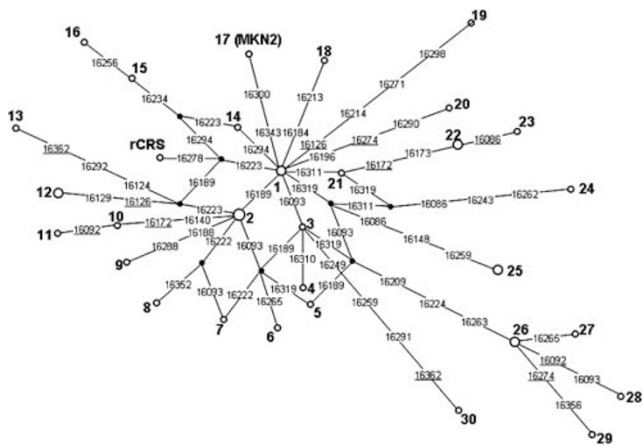


Figure 3 A median-joining³⁷ network diagram consisting of the minor Moken hypervariable segment I (HVSI) (nucleotide positions 16024–16378) lineage (MKN2) and other SEA M46 sequences found in Southeast Asia (SEA).¹⁵

Moken mtDNA genomes and others from the region^{14,16,21,23,25–29,31,35,42} were estimated using the Kimura two-parameter model with a transition to transversion ratio of 2:1 of the program DNADIST. The 50 most closely related sequences were then retained for further analyses. Bootstrap consensus phylogenies (1000 replicates) were generated for both whole mitochondrial genome sequences and for the coding region only using the SEQBOOT program. Consensus trees were then generated with the CONSENSE program using the extended Majority Rule method. The two consensus trees showed similar topologies, so only the tree generated from the coding region is shown in Figure 4. Although the structure of all nodes is shown, only bootstrap values greater than 50% are presented.

RESULTS

Despite including Moken individuals born on four islands spanning the range of the Mergui Archipelago, only two unique mtDNA HVSI lineages were observed (Table 1), resulting in low mtDNA haplotype diversity ($h=0.167$). The major lineage, MKN1 was shared by 11 of 12

Moken individuals surveyed and was found on all four islands sampled. The minor lineage, MKN2, was identified in only one individual, who was born in Dung. According to the nomenclature of Hill *et al.*,¹⁵ MKN1 belongs to mitochondrial haplogroup M21d, and MKN2 was tentatively assigned to M46.

The relationship between MKN1 and other haplogroup M21 sequences found in SEA^{14,15,22,24,27} based on HVSI is shown in Figure 2 and Table 2. MKN1 was shared by one individual from Bali, and clustered with another from Bali (ISEA) and sequences from mainland SEA (Zhuang from Guangxi, Dai from Yunnan, northern Thai and Phuthai from northeastern Thailand).

The relationship between MKN2 and other M46 sequences, based on HVSI, is shown in Figure 3 and Table 3. Although MKN2 was not found in other ISEA populations, it was most closely related to the presumed ancestral lineage of sequences found in Borneo and Sumatra (Sundaland). In contrast to the relatively wide geographic range of haplogroup M21, M46 has been found almost exclusively in ISEA.¹⁵

The Moken whole mitochondrial genome sequences revealed one previously unreported control region polymorphism and five previously unreported coding region polymorphisms; all of the latter are synonymous changes (Table 1). Phylogenetic analysis of whole mitochondrial genome sequences is shown in Figure 4. Unambiguous resolution of the affinities of MKN2 awaits more extensive sampling of regional mitochondrial genomes. There is support, however, for the clustering of MKN1 with a Semelai aboriginal Malay sequence, a reflection of the relationships between subclades of the M21 haplogroup.

DISCUSSION

We identified two mitochondrial HVSI haplotypes among the Thai Moken. The low haplotype diversity ($h=0.167$) is consistent with previous analyses that describe relatively low diversities of both mtDNA sequences (0.009 ± 0.007) and nuclear short-tandem repeats (0.650 ± 0.030),³³ moderate inferred gene flow with populations from the Philippines and low inferred gene flow with the Marianas and Borneo.⁴³ This limited diversity is likely the consequence of genetic drift associated with historically small population size and exacerbated by the recent split with the Moklen. The Moklen have recently taken up permanent residence along the west coast of Thailand and diversified their resource acquisition to include coastal fisheries, pararubber tree plantations and daily wage labor. Despite contrasting subsistence patterns between the Moken and Moklen, the fission is considered recent because of mutually intelligible languages.⁵ The major legends and folklore of the Moken and Moklen, although reflecting important differences (for example, the Moken consider ‘Sipian and Gaman’ to be the legend portraying their origin, whereas the Moklen consider ‘Grandfather Sampan’ to be the legend of their origin), share similarities that further point to a recent split of the two groups. Ferrari *et al.*⁴⁴ speculated that some Moken were captured as slaves to work on the mainland, and have since gradually absorbed local cultural traits and consequently distinguished themselves as Moklen.

Trade relations between the Moken and many other populations are documented and sex-biased marriages primarily between male mainland Asian traders and Moken women are suggested to have introduced characteristic glucose-6-phosphate dehydrogenase (G6PD) alleles into the population. The three variants observed link the Moken to Chinese and Burmese populations, and to other mainland Asian groups through the Thai.⁴⁵ The relatively large G6PD diversity compared with mtDNA variation likely reflects both malaria-selective pressure and low female gene flow into the Moken. Despite these documented trade networks, Sather¹² suggests that compared with

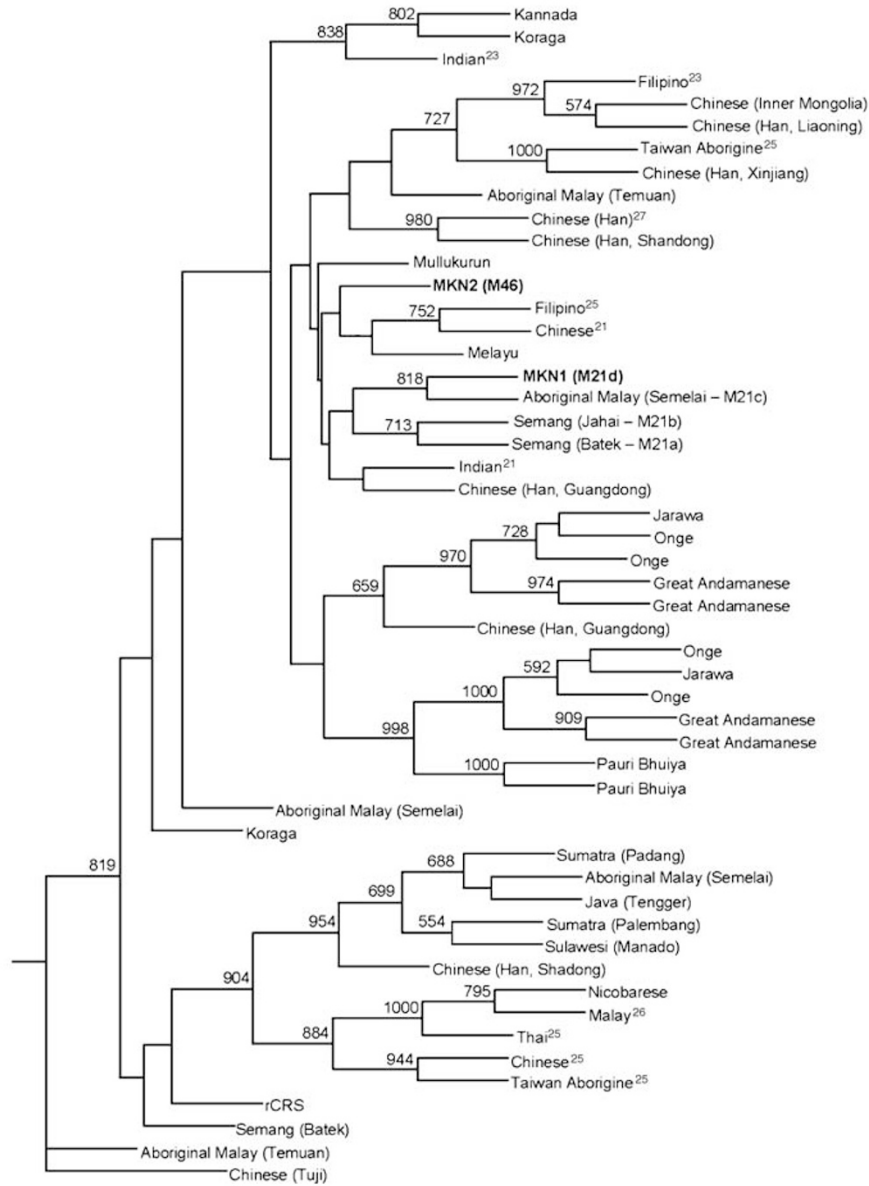


Figure 4 Consensus phylogram illustrating relationships of Moken to Asian and Island Southeast Asian (ISEA) populations, generated from 1000 replicate bootstrap phylogenies of mitochondrial genome-coding region sequences. Original references are indicated for sequences lacking precise geographical identifiers: Ingman and Gyllensten²⁵; Ingman *et al.*²¹; Maca-Meyer *et al.*²³; Macaulay *et al.*²⁷; Mishmar *et al.*²⁶

Table 1 Polymorphisms identified in two Moken mitochondrial genomes

Lineage	Control region polymorphisms (16024-576)	Coding region polymorphisms (577-16023); previously unreported polymorphisms in bold	Haplogroup
MKN 1	16145 16181 16192 16223 16291 16304 16519 73 210 263 489	709 1438 1598 1763 2080 2706 3663 3819 3915 4769 4985 5108 6231 7028 7861 8509 8701 8860 9540 10398 10400 10873 10876 11482 11719 12705 14783 15043 15236 15301 15326	M21d
MKN 2	16223 16278 16300 16343T 16362 16519 73 146 152 263 489	1393 1438 2706 3588 4394 4769 6032 6253 7028 7828 8701 8860 8886 9115 9540 10130 10398 10400 10873 11008 11151 11335 11719 12705 13434 14783 15043 15301 15326	M46

Nucleotide position and variants are identified with respect to the revised Cambridge reference sequence (rCRS).³⁵ Transitions are denoted by the nucleotide position only, whereas the lone transversion is denoted by the suffix after the nucleotide position.

Table 2 HVSI (nucleotide positions 16024–16378) sequence polymorphisms of the major Moken lineage (MKN1) and other SEA haplogroup M21 lineages^{14,15,22,24,27}

Lineage	HVSI polymorphisms (+16 000)	Mainland Southeast Asia							Island Southeast Asia					
		Moken (12)	Semang (112)	Senoï (52)	Aboriginal Malay (96)	Melayu Malay (109)	Dai (38)	Zhuang (83)	Tu (35)	Phuthai (25)	Thai (32)	Borneo (157)	Sumatra (180)	Bali (82)
1	093-129-223-256-271-362		39	1	1									
2	093-129-223-256-271				2						1			
3	093-129-217-223-256-271-362		6											
4	093-129-223-256-271-293-362		1											
5	129-223-256-271-362		3											
6	093-129-145-223-256-271-362			1										
7	129-145-223-256-271-362			1										
8	093-129-145-223-290-319-362					1								
9	093-129-223-311-357						1							
10	093-129-223-261-298-327							1						
11	093-129-162-172-295-304						1							
12	093-129-223-263		3	1										
13	129-223-263										2			
14	172-223-239-263-325				6							1		1
15	223-242-319				2									
16	192-223-291-297						1							
17	145-192-223-291-304						1							
18	145-192-223-291-304-362								1					
19	145-192-223-286-291-304									1				
20	145-181-192-223-291-304	11											1	
21	145-181-192-223-266-291-304												1	

Abbreviations: HVSI, hypervariable segment I; SEA, Southeast Asia.

^aOne Melayu Malay individual out of 109 sampled (0.9%) has been categorized as M21d.¹⁵ The sequence for this individual is unavailable and so it is not included in the network diagram. Total size of each sample from which the lineages were drawn is indicated in parentheses.

other ISEA sea nomad populations, the Moken were largely self-sufficient in sea foraging and were thus less influenced by external markets of agriculturalists than other sea-dwelling populations. Frequent attacks by pirates and slave raiders have conditioned the Moken to avoid contact with unfamiliar outsiders as part of the survival strategy. In more recent times, the Moken pattern of semi-nomadic subsistence has become the focal point for discrimination and marginalization by the sedentary Thai majority.⁵ All of these factors would promote isolation consistent with the paucity of maternal genetic diversity we observe among the Moken.

The major Moken lineage was assigned to haplogroup M21d. M21 is an ancient (about 57 000 years old) yet localized haplogroup that appears only distantly related to other M types found outside SEA/ISEA.^{14,27} Hill *et al.*¹⁵ suggest it might be traced back to the first anatomically modern humans who settled the region at least 45 000 years ago, with the derived subclades representing deep Upper Pleistocene ancestry. M21a is most common among the Semang, and M21b is found in both the Semang and Senoi. M21c has been identified in only two Semelai individuals.¹⁴ It is unclear whether haplogroup M21d is indigenous to ISEA. Our coding region phylogram reflects the relationships between these M21 lineages. MKN1 clusters with one of the M21c Semelai sequences, and the branch clustering the Semelai and MKN1 sequences with Semang Batek (M21a) and Jahai (M21b) sequences, although supported by a relatively low bootstrap value, defines the M21 haplogroup from others.

Because of these limitations in available whole mitochondrial sequence data, analyses based on more extensively sampled HVSI sequences are more useful in revealing affinities within the M21d

haplogroup. HVSI analyses have identified M21d in several individuals from ISEA and South China and at high frequencies among the Burmese Moken. Hill *et al.*¹⁵ found frequencies of 0.9% among Melayu Malays (1 out of 109 samples), 2.4% in Bali (2 out of 82 samples) and 0.2% across the ISEA samples. One of these Balinese sequences was identical to our MKN1 HVSI sequence; the other differed at one nucleotide position. The MKN1 and Balinese sequences cluster with mainland SEA Dai and Thai HVSI sequences in our median-joining network diagram analyses.

Two different patterns of movement could explain these relationships. If the M21d haplogroup is indigenous to ISEA, it is possible that the pressure of rising postglacial sea levels prompted some populations to disperse from coastal areas and adopt a sea-based subsistence strategy, whereas genetically related groups chose instead to move to higher altitudes. The presence of the M21d haplogroup among the Moken then reflects dispersal from the coast, although its presence in Bali is a reflection of indigenous carriers who remained. Limited dispersal with early voyaging groups or genetic admixture through later trade relations is implied by the presence of the lineage among several individuals in ISEA.

The presence of the M21d lineage in South China, however, is not well explained by this pattern of movement. A more parsimonious explanation is that MKN1 and the related Balinese sequences share common ancestry in a coastal mainland SEA population. Both MKN1 and the related Balinese sequences cluster with the ethnic Dai of Yunnan province in the south. The Dai were originally distributed across southeastern China⁴⁶ and trace their ancestry to the ancient Pai-Yuei tribe, which also contributed to the ancestral gene pool of the Thai.²⁴ Genetic phylogenies constructed from microsatellite data

Table 3 HVSI (nucleotide positions 16024–16378) sequence polymorphisms of the minor Moken lineage (MKN2) and other SEA haplogroup M46 lineages¹⁵

Lineage	HVSI polymorphisms (+16 000)	Moken (12)	Borneo (157)	Sumatra (180)	Java (46)	Sulawesi (237)	Lombok (44)	Ambon (43)	Alor (45)	Bali (82)
1	223-278		1	1						
2	189-223-278			1		2	1			
3	093-223-278		1							
4	093-223-278-310		1							
5	093-189-223-278-319			1						
6	093-189-223-265-278							1		
7	093-189-222-223-278						1			
8	189-222-223-278-352						1			
9	188-189-223-278-288								1	
10	140-172-189-223-278			1						
11	092-140-172-189-223-278						1			
12	126-129-189-278		3							
13	124-189-278-292-362									1
14	223-278-294									1
15	234-278-294					1				
16	234-256-278-294									1
17	223-278-300-343T	1								
18	184A-213-223-278								1	
19	126-214A-223-271-278-298			1						
20	196-223-274-278-290					1				
21	223-278-311		1							
22	172-173-223-278-311		2			1				
23	086-172-173-223-278-311			1						
24	086-223-243-262-278-311-319									1
25	086-148-223-259-278-319		1		2					
26	092-209-223-224-263-278-319			2						
27	093-209-223-224-263-265-278-319					1				
28	092-209-223-224-263-278-319			1						
29	093-209-223-224-263-274-278-319-356			1						
30	093-223-249-259-278-291-362					1				

Abbreviations: HVSI, hypervariable segment I; SEA, Southeast Asia.
Total size of each sample from which the lineages were drawn is indicated in parentheses.

clustered the Dai and neighboring southern Chinese groups with three Taiwanese Aborigine groups, reflecting a possible origin of early Austronesian dispersal.⁴⁷ The Dai, Thai, Balinese and Moken might thus ultimately share ancestry among proto-Austronesian populations of coastal east Asia.

The ancestral population of the Dai and Thai then dispersed further inland, whereas the Moken and Balinese ancestral population likely dispersed coastally into ISEA. As temperatures rose in the late Pleistocene and early Holocene, the pressure of rising sea levels and of other expanding ethnic groups forced rapid movement of this population in two directions, toward both the Mergui and Balinese coasts. Genetic drift associated with low population size and isolation has pushed the M21d lineage to high frequencies among the Moken; at the same time, the complex history of population movement and genetic admixture documented in Bali⁴⁸ have contributed to its relatively low frequency there. The absence of the M21d lineage among other populations in ISEA reflects the rapidity of dispersal, extreme genetic drift and a lack of adequate sampling given the high levels of diversity within the region.

Linguistic patterns suggest contact between the Moken and populations indigenous to the Malay Peninsula consistent with early trade relations. Although the Moken language is classified as Malayo-Polynesian within the Austronesian language family, it exhibits influ-

ences from Austroasiatic Mon-Khmer languages and limited, but striking, similarities to the Aslian branch of these languages.^{8,49} The Austroasiatic languages originated in South China around 7000 years ago and today are spoken across SEA and parts of South Asia.¹¹ The Semang, Senoi and Semelai Orang Asli speak Aslian languages (other Aboriginal Malays speak Malay dialects, which are in the Austronesian language family).¹⁴ The linguistic influences observed suggest contact between proto-Moken–Moklen speakers and early Mon-Khmer speakers of the Malay Peninsula, likely prompted by trade,⁴⁹ which could coincide with dispersal from SEA toward ISEA and into the Mergui Archipelago at the end of the sea level rises.

The minor Moken lineage was tentatively assigned to the newly defined basal haplogroup M46, which is found almost exclusively in ISEA. M46 has an estimated age of $62\,700 \pm 12\,400$ years and is postulated to be indigenous to ISEA.¹⁵ The affinities of our MKN2 lineage could not be ascertained unambiguously from our coding region phylogram, which again reflects limitations based on the number of whole mitochondrial genome sequences available, as well as the basal nature and relative rarity of the haplogroup. In multiple analyses, the MKN2 sequence clustered with low bootstrap support to others falling into various basal M haplogroups, including sequences from Chinese, Filipino, Melayu and South Indian (Koraga and Kannada) populations. HVSI analyses indicate that the lineage is

most closely related to the presumed ancestral sequence of those currently found in Borneo and Sumatra. Its introduction to the Moken could have occurred early with the dispersal of Sundaland populations carrying the lineage at the time of sea level rises, or more recently through genetic admixture fostered by trade relations.

The sharing of Moken lineages with individuals from Bali and the clustering of these sequences with ethnic Dai and Thai populations are most parsimoniously accommodated by a recent coastal mainland SEA origin, dispersal into ISEA and rapid movement into the Mergui Archipelago with pressure from expanding ethnic groups and rising sea levels. Small-scale maternal gene flow from ISEA is suggested by the presence of the M46 lineage MKN2, thought to be derived from the first anatomically modern human settlers of the region. Low population sizes, as well as isolation driven by a history of slavery and marginalization, have contributed to considerable genetic drift, which is evident in the low mtDNA diversity among the individuals sampled. Within the past few decades, the Moken have experienced population declines and increasingly intense pressure to settle into permanent villages, which will further influence the population's genetic structure. Studies of Moken subsistence, language, traditions—and origins—are of even greater interest as the Moken struggle to maintain their lifestyle and identity in the face of rapid cultural change.

- 1 Chou, C. Research trends on Southeast Asian Sea nomads. *Kyoto Rev Southeast Asia* (2006).
- 2 Ivanoff, J. *The Moken Boat* (White Lotus Press, Bangkok, 1999).
- 3 White, W. G. *The Sea Gypsies of Malaya* (JB Lippincott Co., Philadelphia, 1922).
- 4 Sopher, D. E. *The Sea Nomads: a Study Based on the Literature of the Maritime Boat People of Southeast Asia* (Authority: Lim Bian Han Government Printer, Singapore, 1965).
- 5 Hinshiranan, N. *The Analysis of Moken Opportunistic Foragers' Intragroup and Intergroup Relations*. Dissertation, University of Hawai'i, Manoa, HI (1996).
- 6 Ivanoff, J. Sea Gypsies of Myanmar. *Natl Geographic* April 36–55 (2005).
- 7 UNESCO. Bridging the gap between the rights and needs of indigenous communities and the management of protected areas: case studies from Thailand. UNESCO Bangkok (2007) http://www2.unesco.org/elib/publications/116/indigenous_communities.pdf.
- 8 Larish, M. D. *The Position of Moken and Moklen within the Austronesian Language Family*. Dissertation, Department of Linguistics, University of Hawai'i, Manoa, HI (1999).
- 9 Gordon, R. G. Jr (ed.) *Ethnologue: Languages of the World* 5th edn (SIL International, Dallas, TX, 2005). Online version: <http://www.ethnologue.com/>.
- 10 Tyrone, D. Proto-Austronesian and the major Austronesian subgroups. in *The Austronesians: Historical and Comparative Perspectives* (eds Bellwood, P., Fox, J. J. & Tryon, D.) 19–41 (Research School of Pacific and Asian Studies Publication, Department of Anthropology, Australian National University, Canberra, 1995).
- 11 Bellwood, P. Austronesian prehistory in Southeast Asia: homeland, expansion and transformation. in *The Austronesians: Historical and Comparative Perspectives* (eds Bellwood, P., Fox, J.J. & Tryon, D.) 103–118 (Research School of Pacific and Asian Studies Publication, Department of Anthropology, Australian National University, Canberra, 1995).
- 12 Sather, C. A. Sea nomads and rainforest hunter-gatherers: foraging adaptations in the Indo-Malaysian Archipelago. in *The Austronesians: Historical and Comparative Perspectives* (eds Bellwood, P., Fox, J.J. & Tryon, D.) 229–268 (Research School of Pacific and Asian Studies Publication, Department of Anthropology, Australian National University, Canberra, 1995).
- 13 Diamond, J. M. Express train to Polynesia. *Nature* **336**, 307–308, doi:10.1038/336307a0 (1988).
- 14 Hill, C., Soares, P., Mormina, M., Macaulay, V., Meehan, W., Blackburn, J. *et al*. Phylogeography and ethnogenesis of aboriginal Southeast Asians. *Mol. Biol. Evol.* **23**, 2480–2491, doi:10.1093/molbev/msl124 (2006).
- 15 Hill, C., Soares, P., Mormina, M., Macaulay, V., Clarke, D., Blumbach, P. B. *et al*. A mitochondrial stratigraphy for Island Southeast Asia. *Am. J. Hum. Genet.* **80**, 29–43, doi:10.1086/510412 (2007).
- 16 Soares, P., Trejaut, J. A., Loo, J. H., Hill, C., Mormina, M., Lee, C. L. *et al*. Climate change and postglacial human dispersals in Southeast Asia. *Mol. Biol. Evol.* **25**, 1209–1218, doi:10.1093/molbev/msn068 (2008).
- 17 Oppenheimer, S. *Eden in the East: the Drowned Continent of Southeast Asia* (Weidenfeld & Nicolson, London, 1998).
- 18 Lum, J. K. The colonization of remote Oceania and the drowning of Sundaland. in *Genetic, Linguistic and Archaeological Perspectives on Human Diversity in Southeast Asia* (eds Seielstad, M., Jin, L. & Xiao, C.) 147–170 (World Scientific Publishing Co., Singapore, 2001).
- 19 Gislén, A., Dacke, M., Kröger, R. H., Abrahamsson, M., Nilsson, D. E. & Warrant, E. J. Superior underwater vision in a human population of sea gypsies. *Curr. Biol.* **13**, 833–836 (2003).
- 20 Gislén, A., Warrant, E. J., Dacke, M. & Kröger, R. H. Visual training improves underwater vision in children. *Vision Res.* **46**, 3443–3450 (2006).
- 21 Ingman, M., Kaessmann, H., Paabo, S. & Gyllensten, U. Mitochondrial genome variation and the origin of modern humans. *Nature* **408**, 708–713, doi:10.1038/35047064 (2000).
- 22 Fuchareon, G., Fuchareon, S. & Horai, S. Mitochondrial DNA polymorphisms in Thailand. *J. Hum. Genet.* **46**, 115–125, doi:10.1007/s100380170098 (2001).
- 23 Maca-Meyer, N., Gonzalez, A. M., Larruga, J. M., Flores, C. & Cabrera, V. M. Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet.* **2**, 13, doi:10.1186/1471-2156-2-13 (2001).
- 24 Yao, Y. G., Nie, L., Harpending, H., Fu, Y. X., Yuan, Z. G. & Zhang, Y. P. Genetic relationship of Chinese ethnic populations revealed by mtDNA sequence diversity. *Am. J. Phys. Anthropol.* **118**, 63–76 (2002).
- 25 Ingman, M. & Gyllensten, U. Mitochondrial genome variation and evolutionary history of Australian and New Guinean aborigines. *Genome Res.* **13**, 1600–1606, doi:10.1101/gr.686603 (2003).
- 26 Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A. G., Hosseini, S. *et al*. Natural selection shaped regional mtDNA variation in humans. *Proc. Natl Acad. Sci. USA* **100**, 171–176, doi:10.1073/pnas.0136972100 (2003).
- 27 Macaulay, V., Hill, C., Achilli, A., Rengo, C., Clarke, D., Meehan, W. *et al*. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* **308**, 1034–1036, doi:10.1126/science.1109792 (2005).
- 28 Rao, V. R. Structuring the reconstructed origin of Andaman Islanders: Jarawa. GenBank Direct Submission (16 July 2005). Culture, Anthropological Survey of India, Government of India, 27, Jawaharlal Nehru Road, Kolkata, West Bengal 700 016, India (2005).
- 29 Thangaraj, K., Chaubey, G., Kivisild, T., Reddy, A. G., Singh, V. K., Rasalkar, A. A. *et al*. Reconstructing the origin of Andaman Islanders. *Science* **308**, 996, doi:10.1126/science.1109987 (2005).
- 30 Trejaut, J. A., Kivisild, T., Loo, J. H., Lee, C. L., He, C. L., Hsu, C. J. *et al*. Traces of archaic mitochondrial lineages persist in Austronesian-speaking Formosan populations. *PLoS Biol.* **3**, e247, doi:10.1371/journal.pbio.0030247 (2005).
- 31 Rao, V. R. Deconstructing the genetic pre-history of the Andaman Islands: inter and intra regional mtDNA perspectives from South and South-East Asia. GenBank Direct Submission (28 September 2006). Culture, Anthropological Survey of India, 27, Jawaharlal Nehru Road, Kolkata, West Bengal 700 016, India (2006).
- 32 Boom, R., Sol, C. J. A., Salimans, M. M. M., Jansen, C. L., Wertheim-van Dillen, P. M. E. & van der Noordaa, J. Rapid and simple method for purification of nucleic acids. *J. Clin. Microbiol.* **28**, 495–503 (1990).
- 33 Lum, J. K., Cann, R. L., Martinson, J. J. & Jorde, L. B. Mitochondrial and nuclear genetic relationships among Pacific Island and Asian populations. *Am. J. Hum. Genet.* **63**, 613–624 (1998).
- 34 Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K. & Wilson, A. C. African populations and the evolution of human mitochondrial DNA. *Science* **253**, 1503–1507 (1991).
- 35 Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowlers, R. N., Turnbull, D. M. & Howell, N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **23**, 147, doi:10.1038/13779 (1999).
- 36 Nei, M. *Molecular Evolutionary Genetics* (Columbia University Press, New York, 1987).
- 37 Bandelt, H. J., Forster, P. & Rohl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**, 37–48 (1999).
- 38 Rieder, M. J., Taylor, S. L., Tobe, V. O. & Nickerson, D. A. Automating the identification of DNA variations using quality-based fluorescence re-sequencing: analysis of the human mitochondrial genome. *Nucleic Acids Res.* **26**, 967–973 (1998).
- 39 Ingman, M. & Gyllensten, U. mtDB: Human Mitochondrial Genome Database, a resource for population genetics and medical sciences. *Nucleic Acids Res.* **34**, D749–D751, doi:10.1093/nar/gkj010 (2006).
- 40 Bray, N. & Pachter, L. MAVID: constrained ancestral alignment of multiple sequences. *Genome Res.* **14**, 693–699, doi:10.1101/gr.1960404 (2004).
- 41 Felsenstein, J. *PHYLIP (Phylogeny Inference Package)* version 3.6. Distributed by the author (Department of Genome Sciences, University of Washington, Seattle, 2005).
- 42 Kong, Q. P., Yao, Y. G., Sun, C., Bandelt, H. J., Zhu, C. L. & Zhang, Y. P. Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences. *Am. J. Hum. Genet.* **73**, 671–676 (2004).
- 43 Lum, J. K., Jorde, L. B. & Schiefelshovel, W. Affinities among Melanesians, Micronesians, and Polynesians: a neutral, biparental genetic perspective. *Hum. Biol.* **74**, 413–430 (2002).
- 44 Ferrari, O., Utupuy, K., Hinshiranan, N. & Ivanoff, J. *Turbulence on Ko Phra Thong* (Coll Kélos Anthropologie Maritime, Paris, 2006).
- 45 Ninokata, A., Kimura, R., Samakkarn, U., Settheetham-Ishida, W. & Ishida, T. Coexistence of five G6PD variants indicates ethnic complexity of Phuket islanders, Southern Thailand. *J. Hum. Genet.* **51**, 424–428, doi:10.1007/s10038-006-0380-y (2006).
- 46 Ding, Y. C., Wooding, S., Harpending, H. C., Chi, H. C., Li, H. P., Fu, Y. X. *et al*. Population structure and history in East Asia. *Proc. Natl Acad. Sci. USA* **97**, 14003–14006, doi:10.1073/pnas.240441297 (2000).
- 47 Chu, J. Y., Huang, W., Kuang, S. Q., Wang, J. M., Xu, J. J., Chu, Z. T. *et al*. Genetic relationships of populations in China. *Proc. Natl Acad. Sci. USA* **95**, 11763–11768 (1998).

- 48 Karafet, T. M., Lansing, J. S., Redd, A. J., Reznikova, S., Watkins, J. C., Surata, S. P. K. *et al*. Balinese Y-chromosome perspective on the peopling of Indonesia: genetic contributions from pre-Neolithic hunter-gatherers, Austronesian farmers, and Indian traders. *Hum. Biol.* **77**, 93–114 (2005).
- 49 Larish, M. D. *Possible ProtoAsian Archaic Residue and the Statigraphy of Diffusional Cumulation in AustroAsian Languages*. Paper presented at Tenth International Conference on Austronesian Linguistics, 17–20 January 2006, Puerto Princesa City, Palawan, Philippines (2006); <http://www.sil.org/asia/philippines/ical/papers.html>.
- 50 Bulbeck, D. Indigenous traditions and exogenous influences in the early history of Peninsular Malaysia. in *Southeast Asia: from Prehistory to History* (eds Glover, I. & Bellwood, P.) 314–336 (Routledge Curzon, New York, 2004).
- 51 Horridge, A. The Austronesian conquest of the sea—upwind. in *The Austronesians: Historical and Comparative Perspectives* (eds Bellwood, P., Fox, J. J. & Tryon, D.) 143–160 (Research School of Pacific and Asian Studies Publication, Department of Anthropology, Australian National University, Canberra, 1995).