BMC Biology

RESEARCH

Open Access

Maternal genetic origin of Chao Lay coastal maritime populations from Thailand

Wibhu Kutanan[1,2*], Wipada Woravatin[3], Metawee Srikummool[4], Chatmongkon Suwannapoom[5], Alexander Hübner[6], Jatupol Kampuansai[7], Chawalit Khaokiew[8], Helmut Schaschl[9], Sanni Översti[10,11], Duy Duc La[12], Leonardo Arias[6,13] and Mark Stoneking[6,14]

Abstract

Background The Chao Lay, also known as sea nomads, include the Austronesian-speaking Moken, Moklen, and Urak Lawoi, who traditionally inhabit the coastal regions and islands of the Andaman Sea in southern Thailand. Their maritime lifestyle has attracted significant interest in their genetic origins and relationships with other sea nomad groups in Island Southeast Asia (ISEA); however, comprehensive genetic data on these communities remain scarce. Here, we generated complete mitochondrial genome sequences from Moken and Moklen groups, along with the Tai-Kadai-speaking southern Thai population and additional Austroasiatic-speaking Maniq samples (hunter-gatherer) from southern Thailand.

Results Our findings indicate that the Chao Lay display lower genetic diversity compared to the majority of southern Thai populations. Furthermore, the results suggest the absence of recent maternal expansions among the Chao Lay. Notably, haplogroups D4e1a, E1a1a1a, M21b2, M46a, M50a1, and M71c are predominant among the Chao Lay, underscoring their genetic distinctiveness. Bayesian coalescent age estimates of clades characteristic to Chao Lay for these haplogroups point to the time associated with the Austronesian expansion period.

Conclusions The Chao Lay populations were closer to each other than to other groups and exhibited more genetic connections to Mainland Southeast Asian (MSEA) populations than ISEA populations. However, we do not exclude potential origins of the Chao Lay in ISEA or Taiwan, as it is possible that ancestral Chao Lay males incorporated MSEA females into their communities upon arriving in Thailand. Further studies on genome-wide and Y chromosome data would provide more insights into their genetic history.

Keywords Chao Lay, Southern Thailand, Mitochondrial genome, Moken, Moklen, Urak Lawoi

Background

Southern Thailand is a part of the Malay Peninsula and is bordered by Central Thailand to the north, the Andaman Sea to the west, the Gulf of Thailand to the east, and western Malaysia to the south (Fig. 1). Among the ~9.16 million people living in southern Thailand [1], two minority groups stand out: (1) the Austroasiatic (AA)-speaking Maniq, one of the indigenous people of Southeast Asia (SEA) who share physical characteristics with other such nomadic hunter-gatherer groups, e.g. the Andamanese and the Malaysian Semang (e.g. Jehai, Batek, Kintaq and Mendriq) [2, 3]; and (2) the "Chao Lay," also referred to as sea nomads, which include the Moken, Moklen, and Urak Lawoi, who traditionally inhabit the west coast of southern Thailand. The Moken continue to live as sea-faring nomads, preserving much of their traditional culture. In contrast, the Moklen and Urak Lawoi have settled in seaside villages and have largely assimilated, losing much of their unique cultural identity [4,

*Correspondence:
Wibhu Kutanan
wibhuk@nu.ac.th
Full list of author information is available at the end of the article

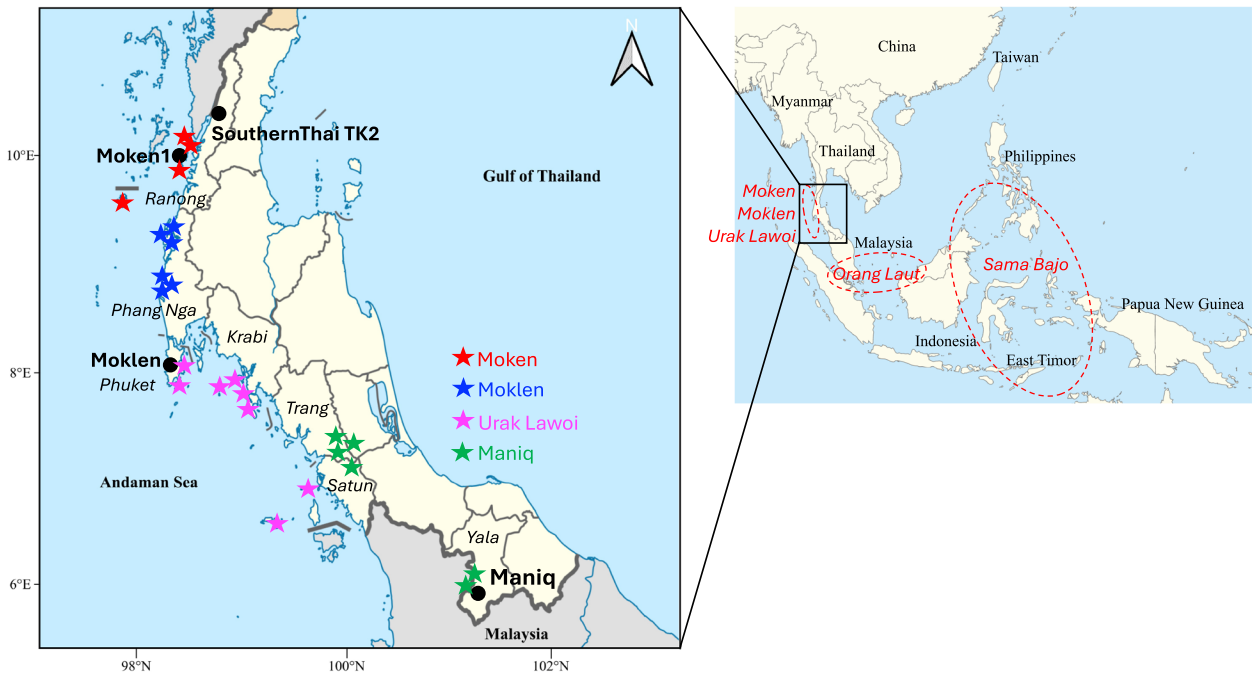

Fig. 1 Geographical locations of the newly studied populations, namely Moken1, Moklen, SouthernThai TK2 and Maniq, highlighting the approximate distribution of Chao Lay and Maniq groups, indicated by asterisks for those in Southern Thailand. The dashed circles represent the broader presence of sea nomads across Southeast Asia

5]. There are ~41 Chao Lay communities, comprising approximately 2,758 households and a total population of around 12,000, including 2,100 for the Moken, 3,700 for the Moklen, and 6,200 for the Urak Lawoi [5]. Their geographic distribution varies: the Moken are found further to the north, near southern Myanmar; the Urak Lawoi (meaning Orang Laut) are located further to the south near the Malaysian border; and the Moklen are in between (Fig. 1) [6].

Sea nomads are known throughout SEA, and based on ethnolinguistic and geographic criteria, they have been categorized into three groups: the Moken and the related Moklen; the Orang Laut; and the Bajo [7]. The Moken and the related Moklen live in the Mergui Archipelago in the Andaman Sea off the coast of southern Myanmar and southern Thailand; the Orang Laut groups live in the Riau Archipelago at the south of Singapore and Sumatra; and the Bajo groups live scattered in the large area between northeast Borneo, the Sulu Archipelago, and northwest Papua (Fig. 1). Historical evidence indicates that sea nomads were a part of a trade network in the maritime SEA (a maritime Silk Road) linking the South China Sea with India, documented since around 400 BCE [8, 9]. The Orang Laut maintained a political and economic relationship with a Malay trading polity during the emergence of the Srivijaya kingdom, a Malay Buddhist maritime empire based in Sumatra, around the seventh century CE [10–12]. At the height of its power, Srivijaya expanded its influence across the region, with its territory encompassing areas such as Cambodia, southern Thailand, the Malay Peninsula, Sumatra, and parts of Java. The kingdom effectively controlled much of SEA's maritime trade, leveraging the expertise of the sea nomads. In exchange, the Orang Laut gained access to vital technologies, including iron tools and advanced boat-building methods, which enabled them to broaden the scope and variety of their trading activities. Their deep social and economic integration with Malay trading polities is evident in both historical texts and contemporary ethnographic studies [10].

Although the Chao Lay groups share an Austronesian (AN) language and cultural background, several differences exist among them. Linguistic studies have shown that the Moken and Moklen languages are more closely related to each other, whereas the Urak Lawoi show more difference, with their language categorized within the Malayo-Chamic branch of AN languages and closely related to Malay [1, 13]. Furthermore, while the Moken and Moklen rely on seagoing vessels for their livelihoods, their activities are confined to limited areas, primarily involving the exchange of maritime goods for essential items. In contrast, the Urak Lawoi engage in more extensive trading activities [14]. The sea-oriented lifestyle of SEA sea nomads, including fishing and facilitating

inter-regional trade, has drawn significant interest in studying their genetic background [15, 16] and genetic adaptations [17]. In Thailand, previous research on mitochondrial DNA (mtDNA) from the hypervariable region 1 (HVR1) of the Moken population (12 samples), originally from southern Myanmar but now residing in Thailand, revealed low genetic diversity [18]. There were only two haplotypes identified from those 12 Moken samples, corresponding to the M21 and M46 haplogroups [18]. Further analysis of the HVR1 sequences of Moken from Chang Island, Ranong Province, revealed four mtDNA haplogroups: M21b2, D4e1a, M46a, and F1a1c1, and highlighted genetic relationships between Moken groups in the Mergui Archipelago, while Y-chromosomal short tandem repeats (STRs) demonstrated a distinctive genetic structure of the Y-chromosomal lineages [19]. Autosomal STR data further revealed unique genetic patterns among the Urak Lawoi of Lipe Island, Satun Province, and evidence of mixed ancestries in the Moklen population from Phuket Province [20]. However, genetic studies on sea nomad populations in Thailand have largely been limited to mtDNA HVR1 sequences and Y-chromosomal and autosomal STR analyses.

In this study, we present newly generated complete mtDNA sequences from Moken and Moklen groups, alongside Tai-Kadai (TK)-speaking southern Thai populations and additional AA-speaking Maniq samples from southern Thailand. Published mtDNA data from other Moken, Urak Lawoi, and TK- and AN-speaking populations in southern Thailand were incorporated to create a comprehensive southern Thai mtDNA dataset. Additionally, we included mtDNA sequences and haplogroup frequencies from populations across South Asia (SA), Mainland Southeast Asia (MSEA), Island Southeast Asia (ISEA), East Asia (EA), Oceania, and Australia. Our findings support a MSEA origin for Chao Lay, with evidence of some maternal genetic interactions with ISEA populations dating back to around the Austronesian expansion period.

Results

Genetic diversity within Southern Thai populations

We generated 98 complete mtDNA sequences belonging to 4 populations: 37 Moken1, 27 Moklen, 5 Maniq and 29 SouthernThai TK2 with mean coverages ranging from 11 ×to 7844 ×(Additional File 2: Table S1. Table S1–[Sequencing coverage and mtDNA haplogroup of each individual]). We combined the newly generated data with published data from other southern Thai groups: Moken2, Urak Lawoi, Maniq, SouthernThai TK1 and SouthernThai AN for a total 205 sequences from 8 populations (Additional File 2: Table S2. Table S2–[General information of studied populations and genetic diversity of mtDNA results]). There are 111 haplotypes and 68 haplogroups, of which 28 haplogroups were not observed before in Southern Thailand (Additional File 2: Table S3. Table S3–[Haplogroup frequency in eight populations from Southern Thailand]). Genetic diversity values were lowest in the Moken1 (haplotype diversity (h) = 0.72 ± 0.06; haplogroup diversity = 0.65 ± 0.07; mean number of pairwise difference (MPD) = 25.36 ± 11.39) (Additional File 2: Table S2). High genetic diversity was observed in both SouthernThai TK populations (h = 1.00 ± 0.01 and 1.00 ± 0.01; haplogroup diversity = 1.00 ± 0.01 and 0.99 ± 0.01; MPD = 37.00 ± 16.70 and 37.60 ± 16.84 in SouthernThai TK1 and SouthernThai TK2, respectively) (Additional File 2: Table S2). Tajima's D value, which is indicative of changes in population size [21], was not significantly different from zero in all Chao Lay and Maniq groups (Additional File 2: Table S2), suggesting no recent maternal expansion or contraction. In contrast, all southern Thai groups showed significant negative Tajima's D values (Additional File 2: Table S2), suggesting recent maternal expansions in these groups. In general, there are differences in genetic diversity values between the minority groups (Chao Lay and Maniq) and majority TK- and AN- Southern Thai groups: the former show lower diversity and no signal of population expansion.

Haplogroups characteristic of Chao Lay and Maniq and divergence estimates

Within the southern Thai dataset, ten of the 68 haplogroups occur in at least five individuals and together account for 60.98% of the 205 sequences; these are D4e1a, E1a1a1a, F1a1a1, F1a1c1, M17a, M21a, M21b2, M46a, M50a1 and M71c (Additional File 2: Table S3). These common haplogroups are mostly prevalent in Chao Lay groups (D4e1a, F1a1c1, M21b2 and M46a in Moken1, 97.30%; F1a1a1, F1a1c1, M21b2, and M50a1 in Moken2, 84.00%; E1a1a1a, F1a1a1, M21b2 and M50a1 in Moklen, 92.59%; E1a1a1a, F1a1a1, M46a, M50a1 and M71c in Urak Lawoi, 76%) and the Maniq (M17a and M21a, 75.00%). These very distinct haplogroup distributions further emphasize the genetic distinctiveness of the Chao Lay and Maniq from other southern Thai populations.

Among the five newly obtained Maniq sequences, four belong to the M21a haplogroup and one to the B4 g haplogroup. M21a is prevalent among the indigenous hunter-gatherers and Proto-Malay from the Malay Peninsula (37.5% in Jehai, 43.2% in Kensui and 50% in Temuan). A network of M21a sequences showed closer genetic relatedness of Maniq to the populations from the Malay Peninsula (Jehai, Temuan and SouthernThai AN) than Mon, Bamar and Karen, reflecting interactions within the Peninsula, whereas the network of B4 g indicated that the Maniq sequence was diverged from TK- and

Hmong-Mien-speaking groups from MSEA, supporting genetic influences from other EA populations into the Maniq (Additional File 1: Fig. S1. Fig. S1–[Networks of haplogroup M21a, M17a, B4 g and R21 that were found in the Maniq]).

For the haplogroups that are prevalent in Chao Lay, there are star-like structures in the networks for haplogroups E1a1a, F1a1a and B5a1 d, suggesting population expansions [22] (Fig. 2). The Chao Lay sequences of haplogroups E1a1a were ancestral to a central haplotype of AN groups from Taiwan and the Philippines. However, when the ancient mtDNA from Liangdao Man, which is dated to 8,060–8,320 Cal BP and is ancestral to haplogroup E1, was included [23], the ancient mtDNA was ancestral to the central haplotype and the Chao Lay haplotypes were diverged from the central nodes, with no shared haplotypes between Chao Lay and other AN-speaking groups (Additional File 1: Fig. S2. Fig. S2–[Network of haplogroups E1a1a1, M21b C7a, F1a1c, and R6 that were found in the Chao Lay. For haplogroups E1a1a1 and M21b, ancient mtDNA sequences were included (Liangdao Man from Taiwan for haplogroup E1 and Gua Cha Cave from Malaysia for M21b)]). In haplogroup B5a1 d, Urak Lawoi haplotypes were diverged from the central nodes while the Moken haplotype was related to sequences from TK-speaking southern Thais. There are two groups of Chao Lay sequences in haplogroup F1a1a (Fig. 2); one contains Moken, Moklen and Urak Lawoi sequences and is distinct from other haplotypes, while another one branches from the central node and is most closely related to a Cambodian sequence. D4e1 is abundant in Moken1 and the network of D4e1 sequences (Fig. 2) shows that Chao Lay haplotypes are more related to haplotypes of Mon, Central Thai and Indian groups than to other MSEA groups.

M21b is prevalent in both Moken populations and Moklen, and the M21b network shows genetic relatedness between haplotypes of Chao Lay and central Thai and Bamar (Fig. 3). The ancient Hoabinhian mtDNA from Gua Cha Cave, located on the Malay Peninsula and dated to 3,872 ± 33 BP [24], does not show any close genetic connection to the Chao Lay sequences (Additional File 1: Fig. S2). The network of M50 sequences, which was found in all Chao Lay groups and at high frequency in Moklen (Additional File 2: Table S3) indicates close genetic relationships between Chao Lay sequences and TK-speaking central Thai and southern Thai sequences (Fig. 3). In addition, the network of M71 sequences reveals closely related sequences between Chao Lay and central Thai, while haplotypes of Moken and Urak Lawoi belonging to haplogroup M46, also found in Bamar and Cambodia, are closely related to Bamar haplotypes (Fig. 3). In conclusion, the Chao Lay from southern Thailand overall show closer genetic relationships to MSEA than ISEA populations.

To estimate the coalescent ages, we combined more sequences from the larger dataset to obtain more reliable ages in each haplogroup and their sublineages. In general, the coalescent ages of basal M lineages range from ~15.19 (M21a) to ~49.83 (M71) kya (Table 1 and Additional File 1: Fig. S3–S4. Fig. S3–[The MCC trees of haplogroups M21a, M17a and B4 g that were found in the Maniq]. Fig. S4–[The MCC trees of haplogroups F1a1a1, D4e11, C7a, E1a1a1, B5a1 d, F1a1c, R6, M50, M71, M21b and M46 that were found in the Chao Lay]). The ages of haplogroups frequently observed in the Chao Lay are ~22.03 kya for M21b2, ~23.04 kya for M50a1, ~13.98 kya for M46a, ~4.88 kya for M71c, ~29.65 kya for D4e1a and ~13.96 kya for F1a1c. Other non-basal M haplogroups with expansion signals (Fig. 2) were dated during the beginning of Holocene: ~13.65 kya, for F1a1a1 and ~14.17 kya for E1a1a1 (Table 1). As demonstrated in Additional File 1: Figs. S3–S4, for certain haplogroups we observed clades primarily composed of samples from Chao Lay. Interestingly, the coalescent ages for these"Chao Lay"clades, are in the range from ~1.43 (M71c) to ~5.23 (F1a1c1) kya (Table 1 and Additional File 1: Figs. S3–S4), which is concurrent with the arrival of Austronesians in the area [25].

Past effective population size

To assess the changes in past effective population size (N_e), we constructed Bayesian skyline plots (BSPs) for specific populations and haplogroups of primary interest. For the populations from Southern Thailand, there were two different trends (Fig. 4). The N_e of major Southern Thai groups (SouthernThai TK1, SouthernThai TK2, and SouthernThai AN) remained relatively stable over an extended period, whereas the other groups (Moken1, Moken2, Moklen, Urak Lawoi, and Maniq) exhibited signs of population decline around ~3–2 kya. While this distinction may provide valuable insights into the differing demographic histories of these populations, it is important to note that a similar pattern of decline in more recent times does not necessarily indicate an actual population decline. Instead, it may be an artifact of the methodology used [26]. Additionally, BSP reconstructions for several populations (i.e. Moken2, Urak Lawoi, SouthernThai TK1, SouthernThai TK2, and SouthernThai AN) suggested population size increases approximately 40–50 kya. However, this should be interpreted with caution, as the histories of these individual populations do not extend this far back. Based on archaeological evidence and historical records, these individual populations likely emerged much later, possibly during or after the AN or TK expansion ~3–2 kya [24, 25, 27]. Thus,

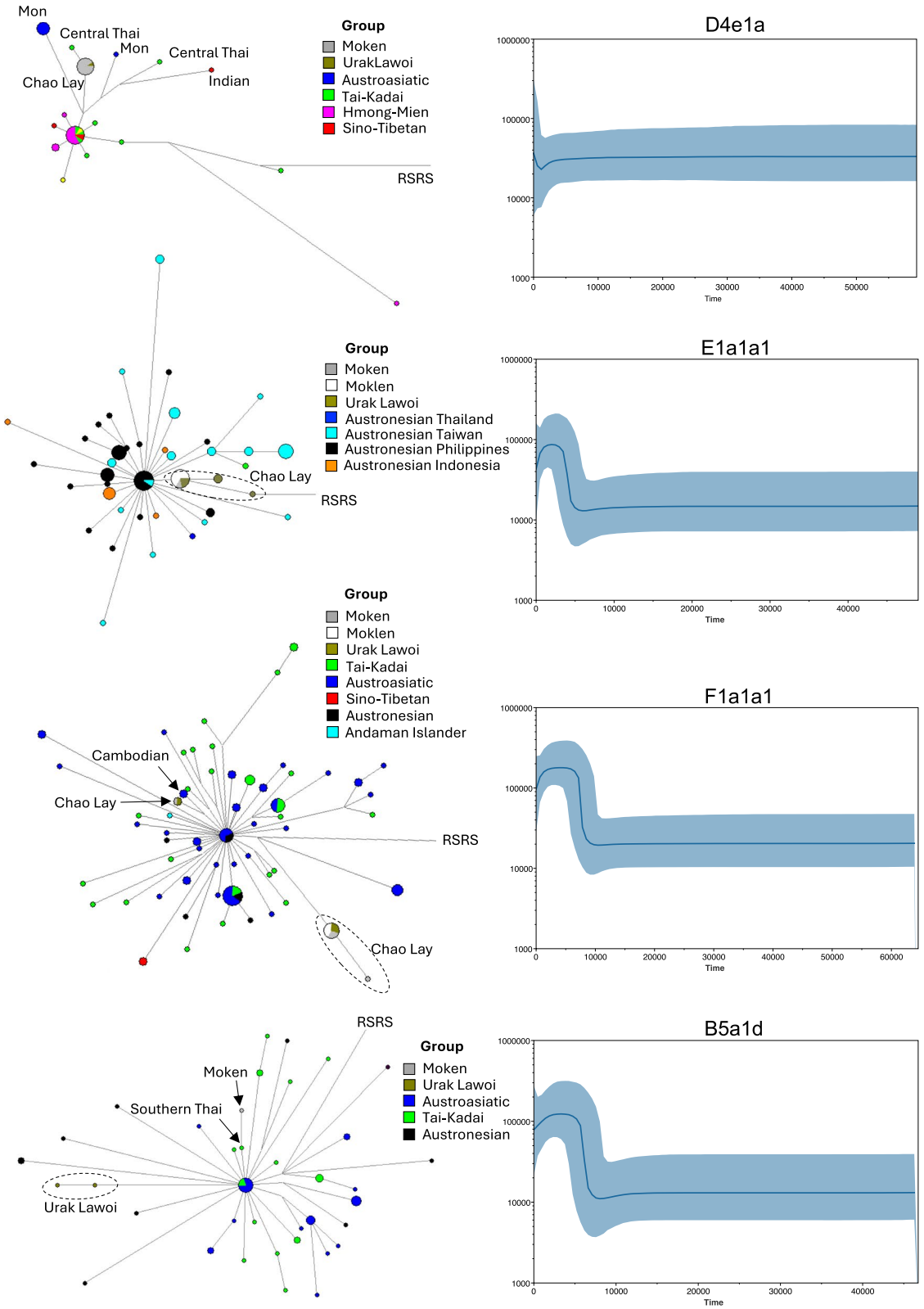

Fig. 2 Networks and BSPs of each major non-basal M haplogroup found in Chao Lay, namely D4e1a, E1a1a1, F1a1a1 and B5a1 d. For the BSPs, dark blue lines are the estimated effective population size on a logarithmic scale (*y* axis) through time from the present in years (*x* axis). The 95% highest posterior density limits are indicated by the blue shaded area

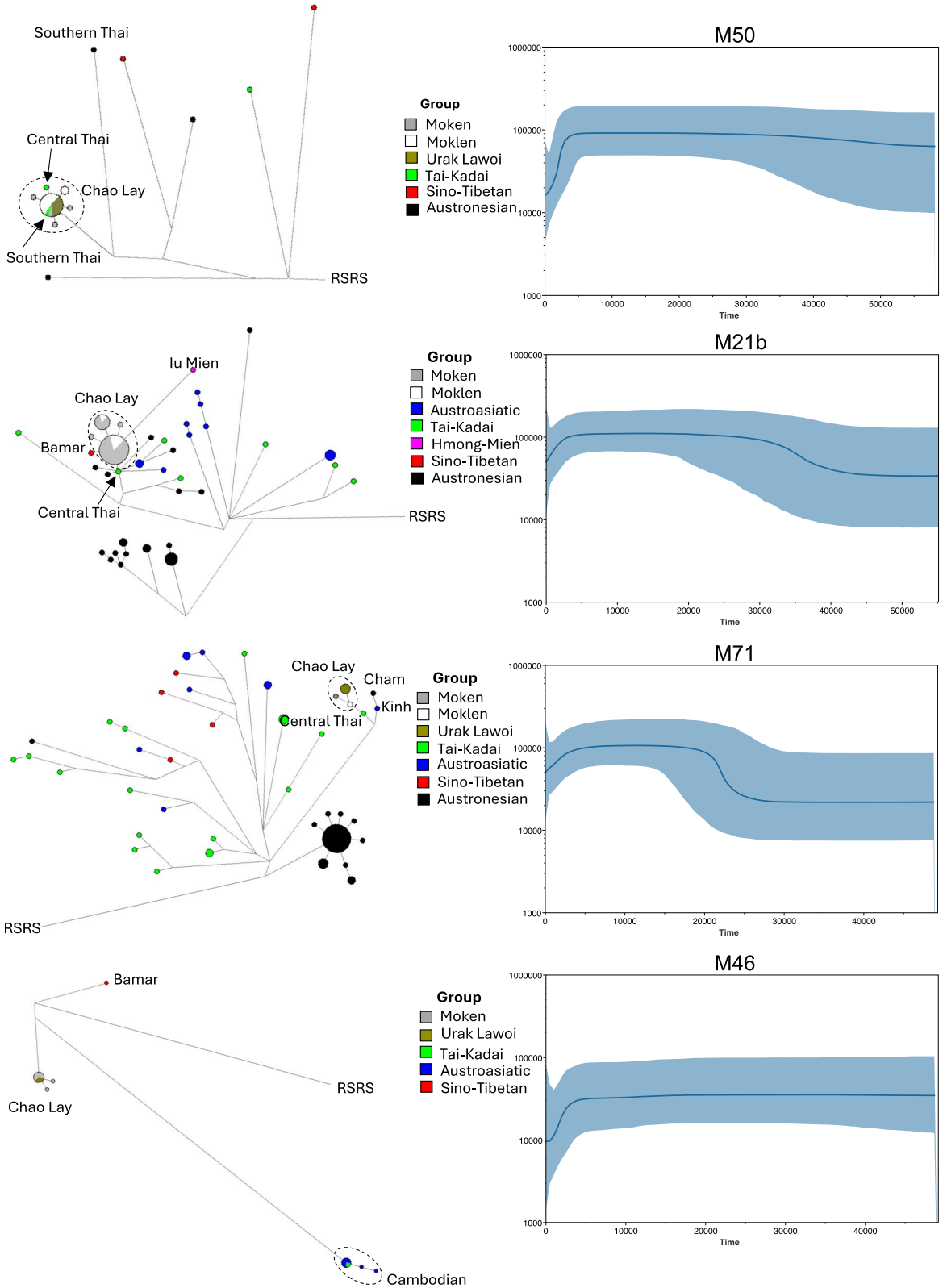

Fig. 3 Networks and BSPs of each major basal M haplogroups that found in Chao Lay, namely M50, M21b, M71 and M46. For the BSPs, dark blue lines are the estimated effective population size on a logarithmic scale (*y* axis) through time from the present in years (*x* axis). The 95% highest posterior density limits are indicated by blue shaded area

Table 1 Coalescent ages based on Bayesian estimation with 95% highest posterior density (HPD) interval. All estimates are given in 'years before present'

Haplogroup	Sample size (All clades/ Clades characteristic to the Chao Lay)	All clades			Clades characteristic to the Chao Lay		
		Age	Lower 95% HPD	Upper 95% HPD	Age	Lower 95% HPD	Upper 95% HPD
F1a1c	12/0	13645	7545	20318	-	-	-
F1a1c1	0/9	-	-	-	5226	1604	9099
F1a1a1	102/8	13957	9216	18929	2715	523	5902
E1a1a1	84/0	14172	8526	20474	-	-	-
E1a1a1a	36/11	6672	3881	7750	2755	895	4683
E1a1a1b	12/0	3828	1537	5928	-	-	-
D4e1a	43/12	29646	19252	41261	2448	631	4869
C7a	25/0	11594	6951	16619	-	-	-
B5a1 d	60/0	11511	7481	16090	-	-	-
B4 g	27/0	21789	14495	28917	-	-	-
B4 g1	15/0	16539	10816	22751	-	-	-
B4 g1a	9/0	10821	6478	15489	-	-	-
B4 g2	12/0	12689	7547	18049	-	-	-
R6	9/0	41237	31877	51955	-	-	-
M71	75/0	30692	23558	38244	-	-	-
M71a	13/0	23732	18183	29965	-	-	-
M71b	3/0	5399	1389	10117	-	-	-
M71c	8/5	4881	1729	8809	1427	280	2867
M71 + 151 (AN clade)	36/0	4901	1963	8182	-	-	-
M71 + 151 (non-AN clade)	15/0	22913	17357	28588	-	-	-
M50	29/0	49834	38558	61185	-	-	-
M50a	26/0	34433	25116	44984	-	-	-
M50a1	24/23	23038	14500	31694	3566	1117	6682
M50a2	2/0	22481	13766	31526	-	-	-
M46	16/0	42603	29026	57221	-	-	-
M46a	9/8	13985	6610	21699	2387	397	4938
M21b	76/0	45486	35790	56709	-	-	-
M21b2	49/34	22025	15243	29574	4671	1947	8072
M21a	32/0	15195	7744	22814	-	-	-
M17	17/0	43919	30813	57148	-	-	-
M17a	16/0	22859	16407	29589	-	-	-

the observed population size increases likely reflect the demographic history of a hypothetical ancestral population from which these groups may have descended, rather than the histories of the populations themselves.

We also plotted the BSPs of published mtDNA genome sequences from AN-speaking populations from MSEA to compare with the AN-speaking populations from Thailand (Additional File 1: Fig. S5–[The BSPs of AN-speaking populations in MSEA: Cham from Cambodia (Cham C) and Vietnam (ChamV1 and ChamV2), Churu, Rhade1, Rhade2, Jarai1, Jarai2 and Raglay from Vietnam and Bidayuh, Selatar and Temuan from Malaysia]). Nearly all populations – ChamC, ChamV2, Churu, Raglay, Rhade1, Rhade2, Jarai1, Jarai2, Bidayuh, Seletar, and Temuan – exhibited signals of population decline around ~5–2 kya, mirroring the pattern observed for the minor groups from Southern Thailand (Moken1, Moken2, Moklen, Urak Lawoi, and Maniq)). In contrast, ChamV1 displayed a relatively stable population size, resembling the trends seen in the major Southern Thai groups.

The BSPs for the dominant haplogroups found among the Chao Lay and Maniq exhibited patterns similar to those observed for the individual populations (Figs. 2 and 3). While some subhaplogroups of M (i.e. M21b, M71, and M17) showed indications of an increase in N_e in the distant past, the 95% highest posterior density intervals overlapped significantly, making this increase uncertain (Fig. 3). In contrast, the BSPs for haplogroups B5a1 d and

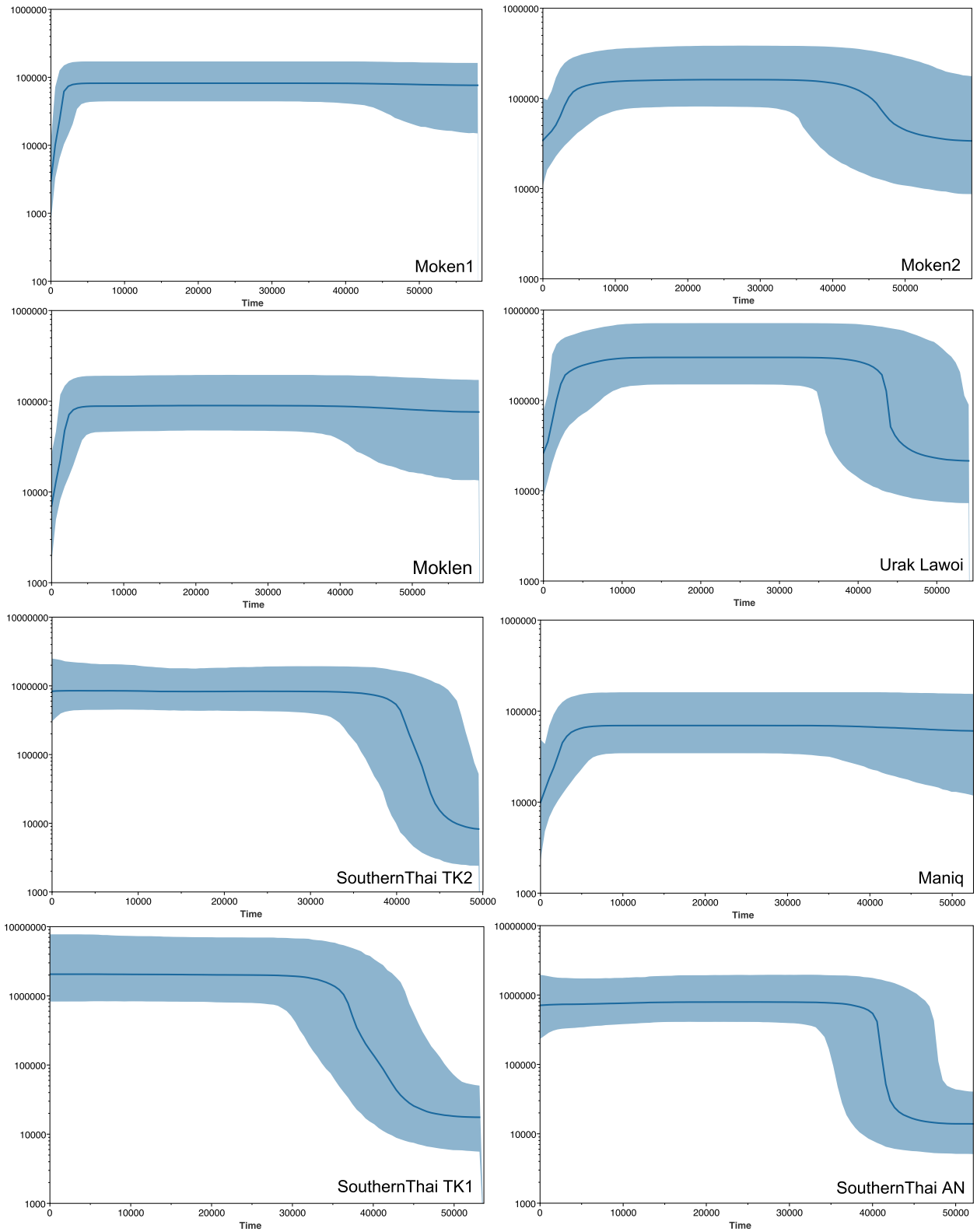

Fig. 4 The BSPs of 8 ethnicities from Southern Thailand: Moken1, Moken2, Moklen, Urak Lawoi, Maniq, SouthernThai TK2, SouthernThai TK1, SouthernThai AN. Dark blue lines are the estimated effective population size on a logarithmic scale (*y* axis) through time from the present in years (*x* axis). The 95% highest posterior density limits are indicated by the blue shaded area

F1a1a suggest a more pronounced rise in effective population size around ~10–5 kya, while haplogroup E1a1a1 shows a similar trend around ~5–2.5 kya (Fig. 2). These observed increases coincide temporally with the beginning of agriculture in EA [28] and the subsequent agricultural expansion in SEA [29], respectively.

Genetic diversity and genetic relationships among populations

When combined with the dataset of all Asian populations, there are a total of 3,775 sequences from 86 populations. In total there are 2,568 haplotypes and 662 haplogroups (Additional File 2: Table S4. Table S4–[Haplogroup frequency for all haplogroups of entire dataset of 86 populations]). The analysis of molecular variance (AMOVA) results indicate that the variation among populations accounts for 10.94% of the total genetic variance (Table 2). When studied populations were grouped based on geography, the SA group shows greater among-population variation (15.46%, $P < 0.01$) than the EA group (8.75%, $P < 0.01$), indicating that the latter is more homogeneous than the former. Within the EA group, the AN group shows the greatest genetic heterogeneity among populations (13.59%, $P < 0.01$), followed by the ST (7.53%, $P < 0.01$) and AA groups (5.30%, $P < 0.01$); the TK group shows the lowest among-population variation (1.01%, $P < 0.01$). The AN groups from ISEA (12.28%, $P < 0.01$) and MSEA (12.24%, $P < 0.01$) show slightly lower among-population variation than the total AN group, while the Chao Lay group has slightly lower variation (11.76%, $P < 0.01$) than the average of AN from MSEA. The highest level of genetic heterogeneity is observed among the indigenous hunter-gatherer groups of Southeast Asia, which have an among-population variation of 23.56% ($P < 0.01$). This could reflect significant genetic drift among these smaller, more isolated populations, or perhaps the grouping of populations as indigenous hunter-gatherers, based as it is on lifestyle characteristics, does not reflect shared genetic connections or ancestry among these groups.

A direct comparison of Chao Lay vs. other groups showed no significant differences between Chao Lay vs. a group of populations originating from Myanmar (Bamar and Karen). The genetic difference between Chao Lay

Table 2 AMOVA results

Group	No. of groups	No. of populations	Percent variation		
			Within populations	Between populations within groups	Among groups
Total	1	86	89.06	10.94**	
East Asia (EA)	1	70	91.25	8.75**	
South Asia (SA)	1	16	84.54	15.46**	
Total Austronesian-speaking populations from East Asia (AN-EA)	1	41	86.41	13.59**	
Austronesian-speaking populations from Island Southeast Asia (AN-ISEA)	1	25	87.72	12.28**	
Austronesian-speaking populations from Mainland Southeast Asia (AN-MSEA)	1	16	87.76	12.24**	
Austroasiatic-speaking populations from East Asia (AA-EA)	1	9	94.7	5.30**	
Sino-Tibetan-speaking populations from East Asia (ST-EA)	1	5	92.47	7.53**	
Tai-Kadai-speaking populations from East Asia (TK-EA)	1	11	98.99	1.01**	
Chao Lay	1	4	88.24	11.76**	
Indigenous hunter-gatherers	1	5	76.44	23.56**	
Chao Lay vs. AA-EA	2	13	87.13**	5.64**	7.23**
Chao Lay vs. SA (without Andaman Islander)	2	19	78.8**	12.9**	8.30**
Chao Lay vs. AN-ISEA	2	29	79.19**	11.05**	9.76**
Chao Lay vs. AN-MSEA	2	16	84.28**	9.25**	6.48**
Chao Lay vs. Indigenous hunter-gatherers	2	9	75.49**	16.89**	7.63*
Chao Lay vs. ST-EA	2	9	85.68**	7.91**	6.42**
Chao Lay vs. TK-EA	2	15	88.72**	1.93**	9.35**
Chao Lay vs. Southern Thai (TK + AN)	2	7	87.25**	7.2**	5.5*
Chao Lay vs. ST-speaking Bamar and Karen	2	6	84.01**	12.31**	3.68

* indicates $P < 0.05$

** indicate $P < 0.01$

and southern Thai groups (5.5%, $0.01 < P < 0.05$) is lower than that observed with other groups (Table 2). This suggests a closer genetic relationship between Chao Lay and populations from southern Thailand and Myanmar.

To further investigate genetic relatedness among populations, we identified share haplotypes within and between populations. Shared haplotypes within populations indicate a smaller population size and increased relatedness among individuals, whereas shared haplotypes between populations suggest recent shared ancestry or contact. There were shared haplotypes within all Chao Lay populations (particularly high in Moken1), Maniq and SouthernThai AN (Fig. 5), but there were no shared haplotypes within either of the TK-speaking Southern Thai groups. The Moklen shared haplotypes with SouthernThai TK2 and Moken2 with Urak Lawoi,

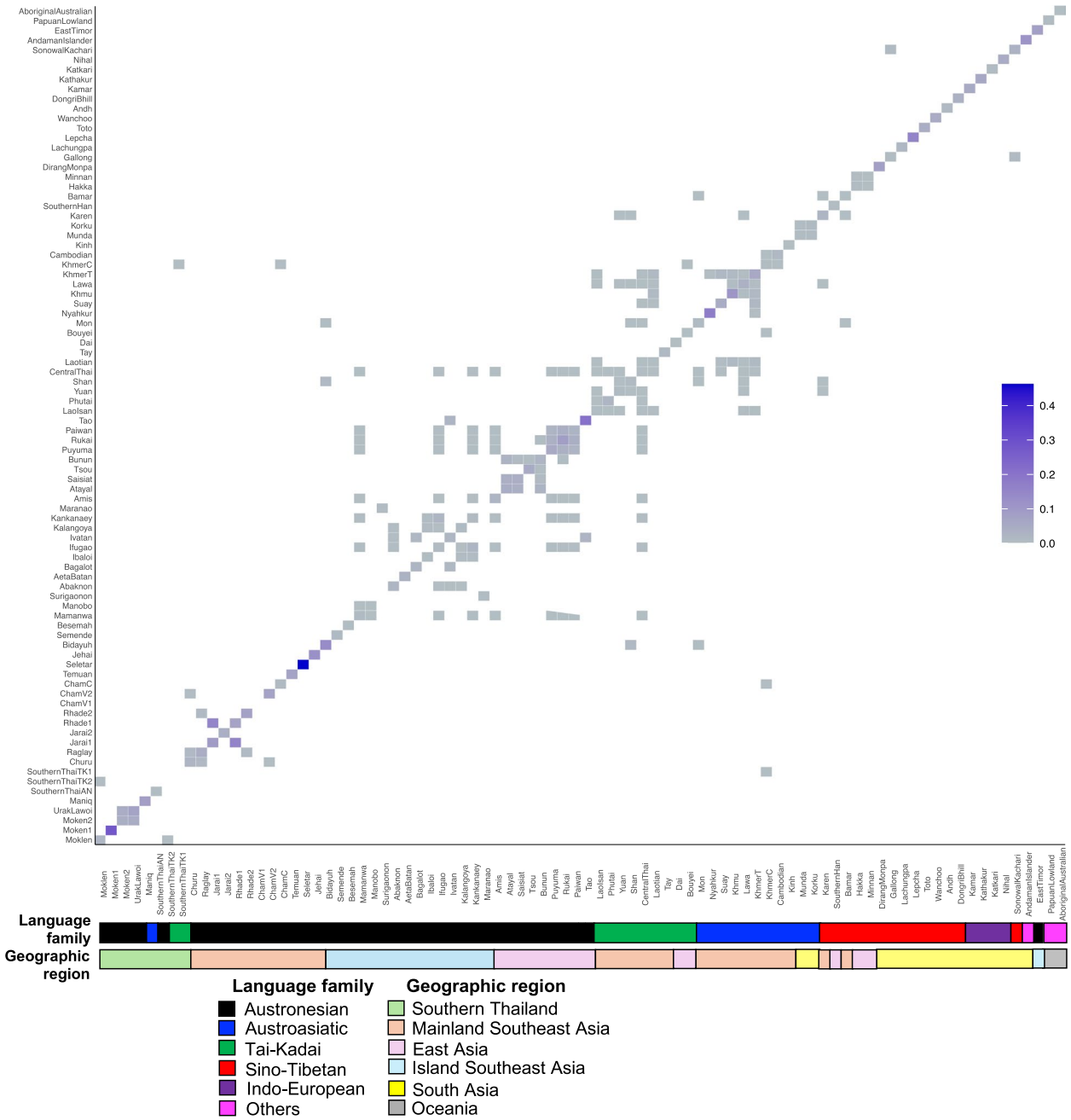

Fig. 5 Frequency of shared haplotypes within and between the entire set of 86 populations. White squares indicate no sharing, while the color of other squares indicates the percentage of shared haplotypes according to the scale

while SouthernThai TK1 shared haplotypes with Khmer from Cambodia, suggesting some recent shared ancestry or contact. The Moken1, Maniq and SouthernThai AN did not share any haplotypes with any other populations. In general, haplotypes were frequently shared among Vietnamese AN-speaking populations, among ISEA AN-speaking populations, and among MSEA TK- and AA- speaking populations. These patterns reflect genetic differences between ISEA and MSEA populations.

Genetic distances (Φ_{st} values) indicate non-significant genetic differences between Moklen and Urak Lawoi and between the hunter-gatherer groups – the Maniq from Thailand and the Jehai from Malaysia (Fig. 6). The Chao Lay populations showed closer genetic relationships to each other than with other populations, although the Moken1 are somewhat more distantly related to the other Chao Lay groups. Overall, the MSEA AN-speaking populations showed closer genetic relatedness to AA- and TK-speaking groups in MSEA than to AN-speaking populations from ISEA, while the AN groups from Taiwan were distantly related to AN groups from the Philippines and Indonesia.

An MDS analysis, based on the Φ_{st} values, was used to further visualize the population relationships. In dimension 1 of the MDS plot (Fig. 7 and Additional File 1: Fig. S6–[A heat plot of the normalized MDS multidimensional scaling values of the entire set of 86 populations based on Φ_{st} values]), SA populations were clustered on the left side and AN-speaking populations from Taiwan were on the right, with all Chao Lay, some AN-speaking from MSEA, Bamar, Maniq, Andaman Islander and Cambodian shifting to the SA side, suggesting some Indian relatedness of these populations. In dimension 2, both Moken populations (MK1 and MK2) and Moklen were along the margin of the plot, reflecting their genetic distinction. Dimension 3 shows the genetic separation of ST-speaking populations from SA to other SA groups; Moken 1 (MK1) shows genetic distinction but is closer to SA ST-speaking populations (Fig. 7 and Additional File 1: Fig. S6). Overall, the Chao Lay exhibit more genetic affinity to MSEA populations than to ISEA populations.

To broaden the analysis of genetic relatedness between the Chao Lay of Thailand and other populations, haplogroup distributions from 86 populations were consolidated into 45 major haplogroups, each consisting of more than 20 sequences (Additional File 2: Table S5. Table S5–[Haplogroup frequency for major haplogroups of entire dataset of 89 populations]). Additionally, we incorporated haplogroup frequency data from the Bajo, an Indonesian sea nomad group residing in Kendari Province, southeastern Sulawesi, as well as from the Ma'anyan and Lebbo' groups of Borneo, Indonesia [15]. These datasets (Additional File 2: Table S5) were subsequently analyzed using correspondence analysis (CA) and multidimensional scaling (MDS) based on haplogroup frequency. In dimension 1 and 2 of the CA plot (Additional File 1: Fig. S7. Fig. S7–[The CA plots of dimension 1 vs. dimension 2 (A) and dimension 3 vs. dimension 4 (B) based on major haplogroup frequency among the set of 89 populations]), the East Timor, Papuan lowland, Aboriginal Australian and Aeta Bataan from Philippines show genetic distinctiveness with high frequencies of haplogroups P, Q, S and M42. Given the strong effect of these four outliers on the analysis, we removed the populations from Oceania and Australia and repeated the analysis. The results reveal the distinctiveness of SA populations in the first dimension and genetic separation between AN-speaking groups from ISEA and Taiwan and MSEA in the second dimension (Fig. 8). The third dimension distinguishes SA ST-speaking groups whereas further dimensions distinguish all Chao Lay, based on haplogroups M21b and M50, and Maniq, Jehai (hunter-gatherer) and Temuan (Proto-Malay) based on haplogroup M21a (Fig. 8). In dimension 1 of the MDS plots (Additional File 1: Fig. S8 Fig. S8–[The MDS plot based on haplogroup frequencies for the entire set of 89 populations for dimension 1 vs. dimension 2 (A) and dimension 1 vs. dimension 3 (B)]), non-ST-speaking SA populations and the Bamar group appeared on the right, while SEA, EA, and Oceanian populations clustered on the left. Meanwhile, SA ST-speaking populations and some MSEA populations—such as the Moken1 (MK1), Mon, Cambodian groups (KHC and AAC), Suay, Semande, Ragley, Churu, Jarai (Jr2), Cham (Chv1), and both TK-speaking Southern Thais (BST1 and BST2)—were distributed in between, suggesting a degree of SA-related ancestry in these MSEA populations. In dimension 2, groups like East Timor, Papuan Lowland, Aboriginal Australians, Aeta Bataan, Indian Toto, and Seletar were positioned along the plot's margins, reflecting their genetic distinctiveness. Dimension 3 highlighted the genetic separation of the Moken (MK1 and MK2), Moklen, and Indian Toto. Overall, the findings suggest that the Thai sea nomads are more closely related to MSEA populations, whereas the Indonesian sea nomads show closer genetic ties to ISEA populations.

Discussion

The maritime lifestyle including traditional fishing and long-distance trade networks of sea nomads has attracted interest about their origins, genetic structure, adaptations, and relationships with other populations in SEA. However, there is a paucity of genetic data on the sea nomads from Thailand (Chao Lay), and the few previous genetic studies have largely been limited to mtDNA HVR1 sequences and Y-chromosomal and autosomal STR analyses [18–20]. Here, we sequenced and

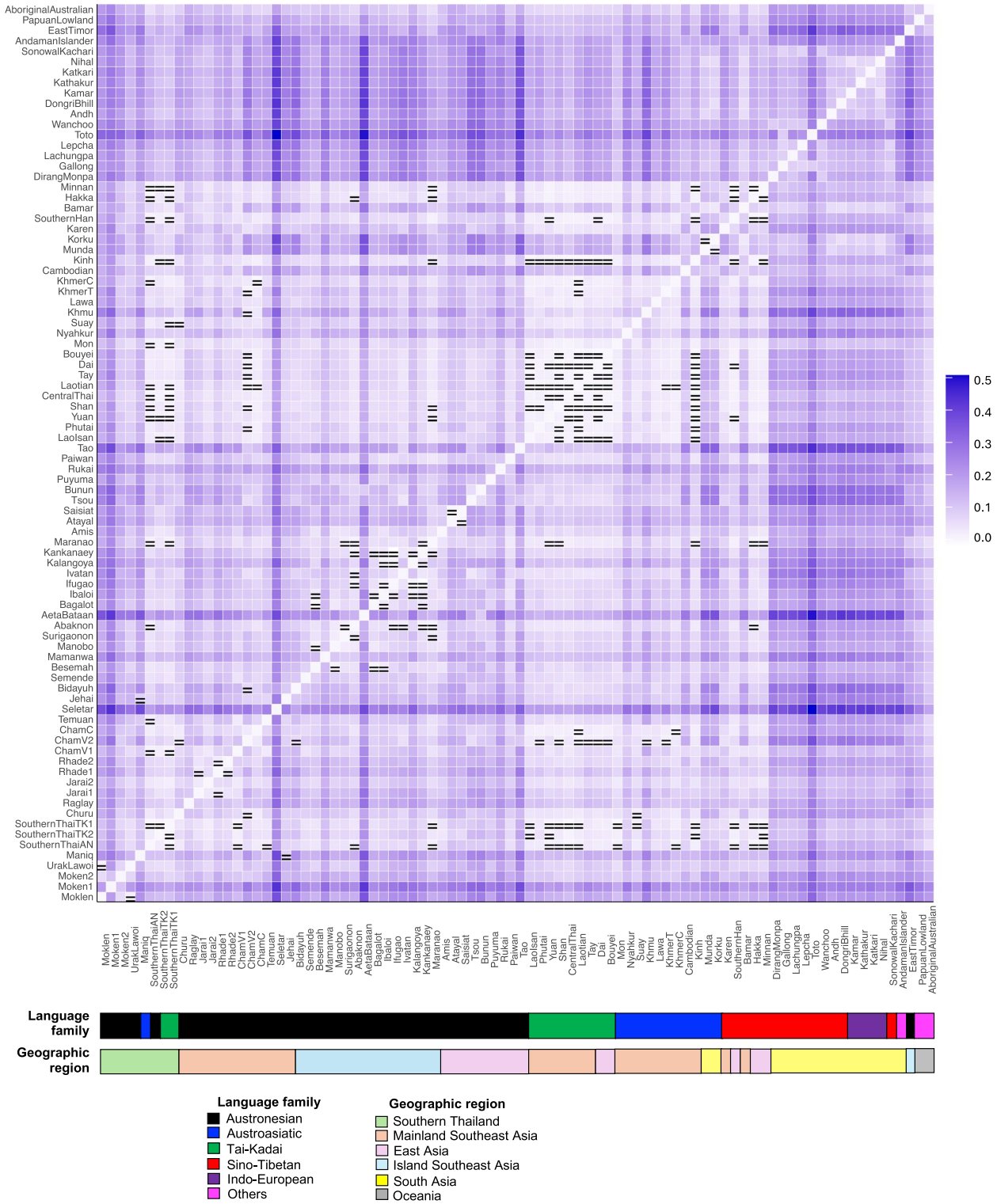

Fig. 6 Heat plot of Φ_{st} values based on mtDNA haplotypes between the entire set of 86 populations. The "=" symbol indicates Φ_{st} values that are not significantly different from zero ($P > 0.05$)

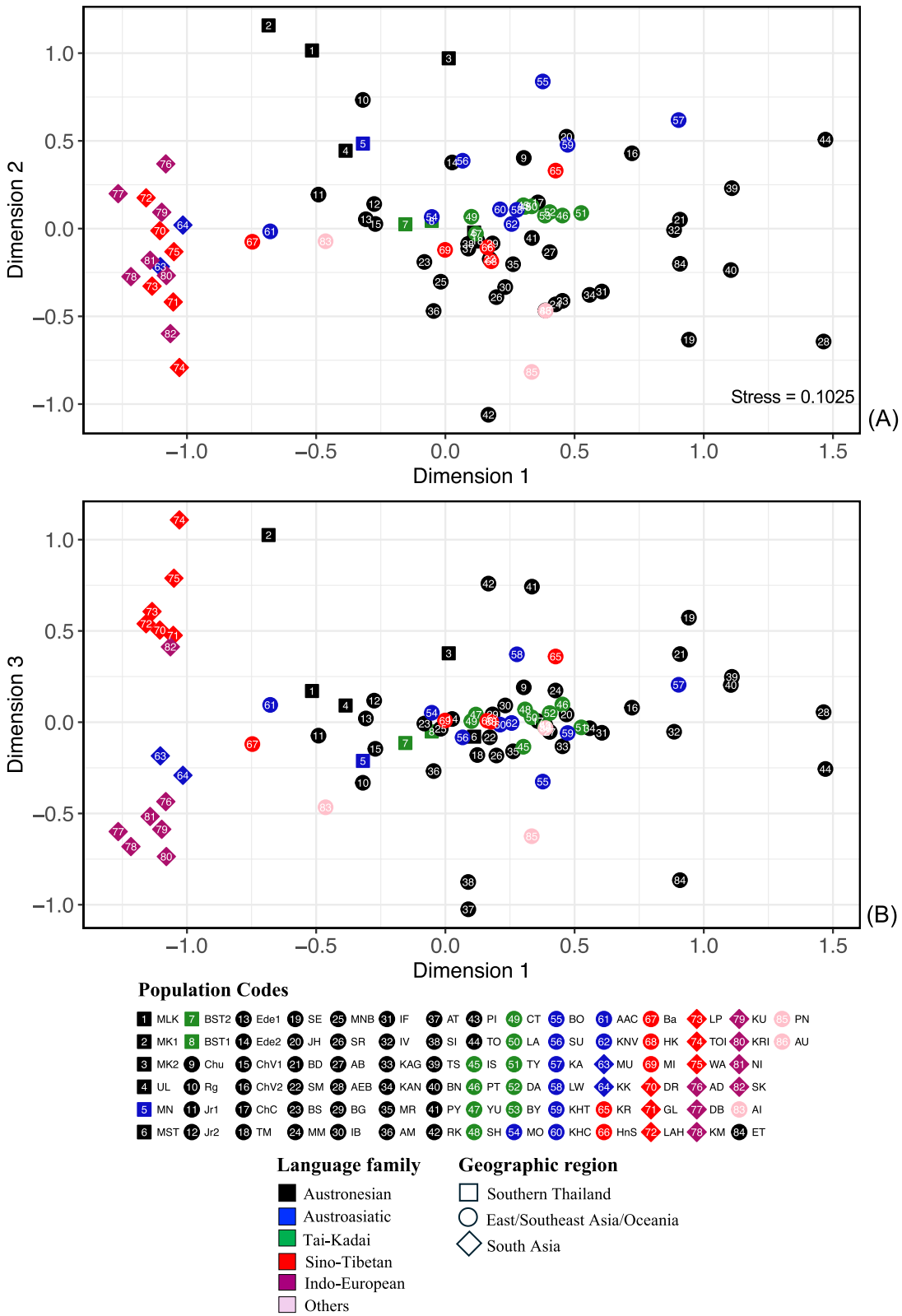

Fig. 7 The three-dimensional MDS plot of the entire set of 86 populations (population codes are given in Additional File 2: Table S1) for dimension 1 vs. dimension 2 (**A**) and dimension 1 vs. dimension 3 (**B**). The stress value is 0.1025

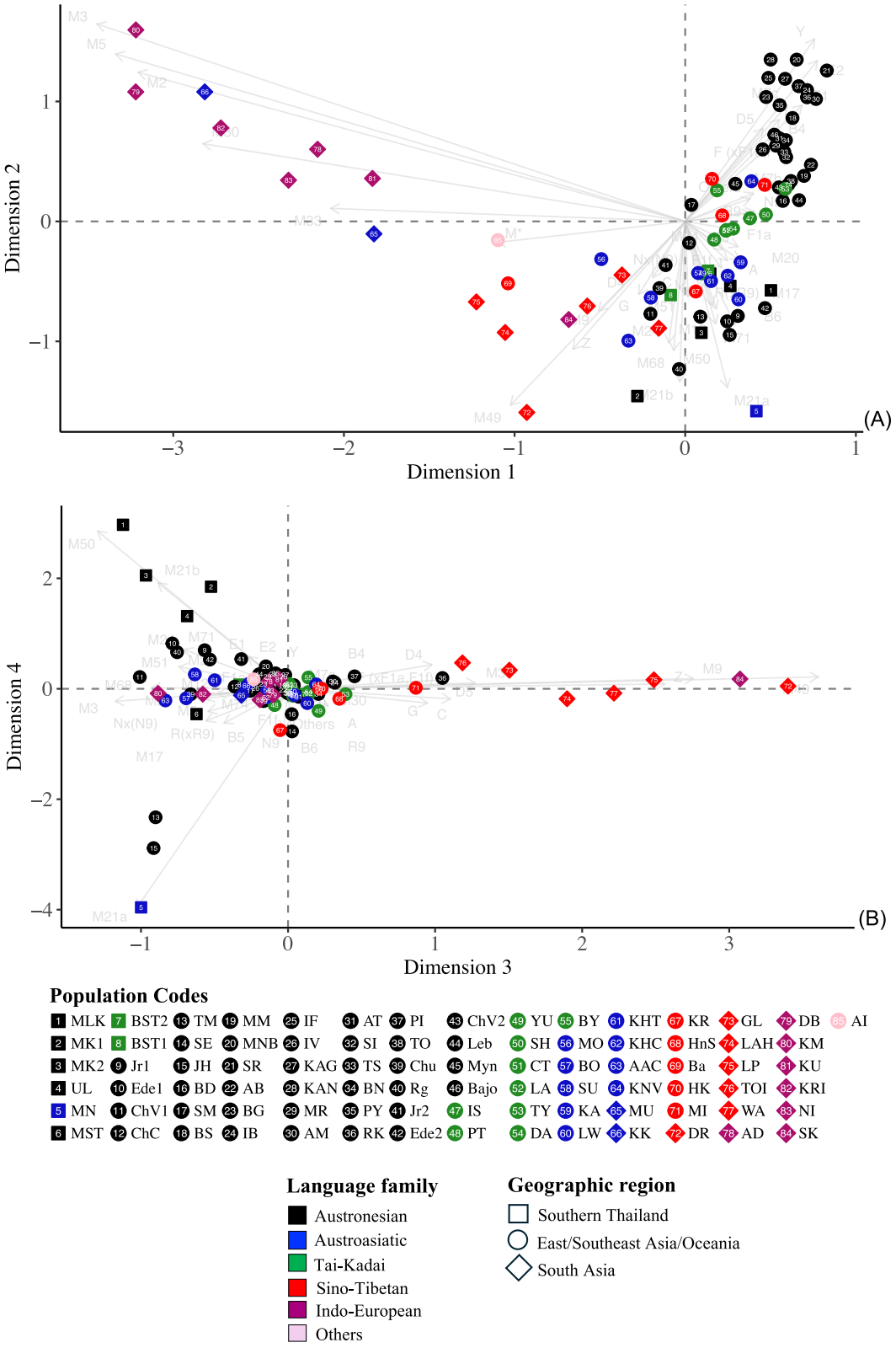

Fig. 8 The CA plot of dimension 1 vs. dimension 2 (**A**) and dimension 3 vs. dimension 4 (**B**) based on major haplogroup frequencies among the set of 85 populations after removal of four outliers (AU, PN, ET and AEB). Population abbreviations are provided in Additional File 2: Table S2

investigated the complete mitochondrial genomes of AN-speaking sea nomads from southern Thailand to explore their origins, genetic heritage, and relationships with other groups in MSEA and ISEA (including other sea nomad groups in ISEA).

The genetic diversities within each Chao Lay group, especially the Moken1 from Chang Island, Ranong Province, were lower than for other populations (Additional File 2: Table S2). The genetic differentiation of Moken1 is driven by a relatively high frequency of haplogroup D4e1a (24.32%). Haplogroup D4e1 and sublineages are found in MSEA and China (Additional File 2: Table S4); the network and MCC tree of D4e1 indicates close relationships among Chao Lay and central Thai haplotypes with a coalescent age ~2.45 kya (Fig. 2 and Additional File 1: Fig. S4). M21b2 is another haplogroup with a high frequency in Moken1 (54.05%) and accounts for the differentiation of Moken1 in the CA plot (Fig. 8); this haplogroup was also found in lower frequency in Moken2 (32.00%) and Moklen (18.52%) (Additional File 2: Table S4). M21b is at higher frequency in AN-speaking groups from MSEA than in other MSEA groups (Additional File 2: Table S5). The subclade of M21b characteristic to Chao Lay, is dated to ~3.57 kya (Table 1 and Additional File 1: Fig. S4). The absence of haplogroups M50a, M71c and F1a1a1 in Moken1, which were found in other Chao Lay populations, also differentiates Moken1. The reduced diversity (Additional File 2: Table S2), high amount of shared haplotypes within the population (Fig. 5) and large genetic distance from the other groups (Figs. 6, 7 and 8) may reflect the small population size, with only ~200 Mokens living on this island. Interestingly, the smaller census sizes of the Chao Lay (~ 2,100 for the Moken, ~3,700 for the Moklen, and ~6,200 for the Urak Lawoi) [5] and the Maniq (~ 250 individuals) compared to the major Southern Thai groups (~ 4.5 and ~3.6 million) [1] correspond to their lower estimated effective population sizes (N_e). Moreover, past N_e trends indicate that only the Chao Lay and Maniq experienced a decline around ~3–2 kya, whereas the three major Southern Thai groups did not show such a decline (Fig. 4).

Although the Φ_{st} value between both Moken populations is significantly greater than 0 (Fig. 6) the genetic distance values are smaller between Moken1 and Moken2 (Φ_{st}= 0.1052) than between either Moken group and other Asian populations (Additional File 2: Table S6. Table S6–[The values of Φ_{st} among entire dataset of 86 populations]), suggesting some genetic relatedness within the Moken ethnic group. Both Moken populations are significantly different from Moklen and Urak Lawoi, which are not significantly different from each other (Fig. 6). Haplogroup E1a1a1a likely drives the differentiation of Moken from Moklen and Urak Lawoi, as only one individual of Moken2 was assigned to E1a1a1a while this haplogroup was found at frequencies of 22.22% in Moklen and 20.00% in Urak Lawoi (Additional File 2: Table S3). Haplogroup E1a1a1 is one of the AN-specific haplogroups that is abundant in Taiwan and ISEA [23, 30–32]. The network of haplogroup E1a1a1 sequences showed a star-like structure (Fig. 2 and Additional File 1: Fig. S2) and the BSP plot of E1a1a1 showed an increase in effective population size during ~5–2.5 kya (Fig. 2), suggesting a lineage expansion probably associated with the spread of AN languages. The MCC tree showed the bifurcation of the Chao Lay clade of E1a1a ~2.76 kya (Additional File 1: Fig. S4), which is in the same range with a previous study ~2.35 kya [30]. However, we do not observe other AN-specific haplogroups within the sea nomads, e.g. Y2, B4a1a and M7c3c [30, 31].

The present results reveal that Moklen and Urak Lawoi are more closely related to each other than to the Moken, which is not consistent with their linguistic affiliation. The Moklen and Moken languages are related to one another and ultimately to the Malayic group of AN languages, either as a member of it or as a parallel branch of western Malayo-Polynesian [33, 34]. Both have been heavily influenced by AA languages (and more recently, Thai) [33]. Moken/Moklen has many changes in common with the Chamic languages of Vietnam, and it has been suggested that these two languages split from Chamic around 2,000 years ago [33]. By contrast, the Urak Lawoi language belongs to the dialect cluster of Orang Laut languages. Linguistic data indicates affiliation of Urak Lawoi with the Malays and the language shows little influence from AA languages [35, 36]. However, the differentiation of Moken from Moklen and Urak Lawoi is in agreement with their lifestyle. The Moken still retain a more traditional lifestyle relating to the sea than do the Moklen and Urak Lawoi, who are permanently settled along coastal areas or inland and also work in coastal fisheries, gardens, and other employment. Moreover, they have adopted other cultures and often refer to themselves as Thai Mai (meaning"New Thai") [37]. Many of the Moklen refer to themselves as “Chao Bok” (coastal people) since they are no longer sea nomads and their traditional culture is disappearing [38, 39]. One distinguishing characteristic of the Moken compared to other Chao Lay groups is their boat type. Regarding the boats used for their livelihoods, the Moklen and Urak Lawoi use dredge boats with plank gunwales, while the Moken use the *Kabang*, a boat made from wintergreen wood. The *Kabang* features a roof, as it is designed for traveling further out to sea than the boats used by the other two sub-groups [5].

The Chao Lay practice monogamy, and after marriage, the husband typically moves into the wife's house for a period before the couple eventually establishes a separate residence, though this practice is not strictly followed. In the past, intermarriage within Chao Lay sub-groups (e.g., Moken marrying Moklen) was occasionally observed. However, intermarriage between Chao Lay and non-Chao Lay groups has become more common in recent years [40]. The sharing of haplogroup E1a1a1 indicates some contact between Moklen and Urak Lawoi with ISEA groups. In addition, both Moklen and Urak Lawoi shared haplotypes with SouthernThai TK2 and Moken2, respectively (Fig. 5), reflecting recent contacts within the region.

Nevertheless, although there are some differences among the Chao Lay from Thailand, they still tend to cluster genetically (Figs. 7 and 8). Moreover, the Chao Lay show a different maternal genetic profile from the Indonesian Bajo (Additional File 1: Figs. S7-S8). In the Bajo, major haplogroups (55.55%) were reported to be related to the AN expansion, e.g. B4a1a1, M7c1a4, M7b1a1i and E1a, followed by haplogroups that are prevalent in MSEA, i.e. B4a, M73, R22 (37.03%); haplogroup Q1 was also found in the Kendari Bajo (0.07%), indicating a link with New Guinea or eastern Indonesia [15]. In contrast, the predominant haplogroups in the Chao Lay (84.21%) were D4e1a, F1a1a1, F1a1c1, M21b2, M46a, M50a1 and M71c, which are prevalent in MSEA populations; the Chao Lay also have a lower frequency of haplogroup E1a1a1a (10.52%) which links them with AN in Taiwan and ISEA. The haplogroup profile overall thus indicates an MSEA origin of Chao Lay from Thailand and a different maternal history from the Indonesian sea nomads.

Previously, three hypotheses for the origin of the Moken were proposed [18]: 1) from the Myanmar–Malaya mainland where their ancestors practiced agriculture but later adopted a sea-based lifestyle and finally settled in the Mergui Archipelago [41, 42]; 2) from China, migrating southward during the Neolithic Period and later separating from other migrating groups as late as the early seventeenth century [43]; and 3) from ISEA with other sea nomads groups and later northward migration to the Mergui Archipelago [42, 44]. Since the Moken and Moklen are ethnolinguistically related, these three hypotheses could also hold for Moklen. In addition, the Urak Lawoi probably originate from the Malay stock but with a separate history for some hundreds of years [45]. Overall, our results indicate all Chao Lay populations were closer to each other than to other groups and exhibited more genetic connections to MSEA populations than ISEA populations (Figs. 6, 7 and 8). Both sea nomads and the most closely-related MSEA populations additionally share SA ancestry; the overall closer genetic relationship between Chao Lay and populations from Myanmar and southern Thailand suggests a Myanmar–Malaya mainland maternal origin (Table 2).

Although the main focus of our study is on the Chao Lay groups, we obtained interesting results for the other two groups. The four additional Maniq sequences are assigned to two lineages: a basal M21a lineage which is dated to ~ 15.19 kya (Table 1); and one B4g sequence that is sporadically distributed in SEA populations and was not found in our previous study [46]. This B4g sequence suggests some admixture of Maniq with other East Asian groups, consistent with a previous genome-wide study that indicate ~35% East Asian-related ancestry in the Maniq [3]. The Maniq overall showed the closest genetic relatedness with the Jehai, another indigenous hunter-gatherer group in the Malay Peninsula, and greater differences with other such groups from the Philippines, Andaman Islands, Papua New Guinea, and Australia (Figs. 6 and 7 and Additional File 1: Figs. S6-S7). The SouthernThai TK2 and both southern Thai groups (SouthernThai TK1 and SouthernThai AN) from our previous study [47] showed similar genetic backgrounds, e.g. small genetic distance to each other, signals of population expansion, high genetic diversity and South Asian influence (Figs. 4, 6 and Additional File 2: Table S2,S6). Their close genetic relatedness is in agreement with previous studies of autosomal STR variation [20, 48].

Conclusions

Our findings suggest that the Chao Lay likely originated from MSEA and have a different maternal history from Indonesian sea nomads, who show closer ties to ISEA. Alternatively, they may have roots in Taiwan or ISEA but incorporated MSEA females into their communities upon their arrival in Thailand. This hypothesis could be further investigated through genome-wide and Y chromosome studies. In addition, we caution that our conclusions are based on a relatively limited sampling of the extensive Chao Lay communities along the coastal area of the Andaman Sea and are restricted to the mtDNA data. Further studies of Chao Lay from different communities, and of additional genetic markers, would provide more insights into the genetic origin of the Thai sea nomads and the impact of the Austronesian expansion in MSEA.

Methods

Samples

We generated complete mtDNA sequences for 27 Moklen samples and 37 Moken (Moken1) samples from previous studies [19, 20], and an additional 34 newly-collected samples of Maniq (5 samples) and SouthernThai TK2 (29 samples), with genomic DNA isolated from buccal swabs using the Gentra Puregene Buccal Cell Kit

(Qiagen, Germany). Sample donors were interviewed to screen volunteers who were healthy and unrelated for at least two generations and then written informed consent was obtained. The protocol on human subjects was approved by the Khon Kaen University Ethic Committee (Protocol No. HE622223) and Naresuan University (COA No.250/2023).

MtDNA sequencing and data processing

The sequencing of saliva sample extracts and subsequent data processing were performed as described in previous study [49]. In brief, we prepared genomic libraries with double indices and enriched for full mtDNA genomes using a hybridization-capture method described previously [50, 51]. The enriched libraries were sequenced on the Illumina HiSeq 2500 platform, generating 2× 76 bp paired-end reads. After standard Illumina base calling, adapters were trimmed, and completely overlapping paired sequences were merged using leeHOM [52]. The sequencing data were de-multiplexed using deML [53],and the sequences were aligned to the human reference genome 19 using BWA's *aln* algorithm [54]. We extracted all sequencing reads from each library that aligned to the mitochondrial genome or to a published list of nuclear copies of mtDNA (NUMTs) [55] and re-aligned them against the revised Cambridge Reference Sequence (rCRS) [56] using BWA's *mem* algorithm [57] with the non-default parameters"-B 6 -O 3 -E 2."After left-aligning indels with GATK IndelRealigner [58], we performed variant calling using BCFtools *call* [59] with an assumed ploidy of 1 and inferred the consensus sequence from the resulting variant file. The consensus sequences were based on a mean sequence coverage ranging from 11-fold to 7,844-fold across the mitochondrial genome. We then manually checked and manipulated sequences with Bioedit (www.mbio.ncsu.edu/BioEdit/bioedit.html).

Statistical analysis

The newly-generated 98 mtDNA sequences from Southern Thailand (Moken1, Moklen, Maniq and SouthernThai TK2) were combined with two datasets. The first is the previously published southern Thai sequences: Moken, Urak Lawoi, Maniq, SouthernThai TK1 and SouthernThai AN [46, 47, 60] for a total of 205 sequences from 8 southern Thai populations (Additional File 2: Table S2). To obtain a broader picture of genetic structure and population relationships, the southern Thai dataset was combined with 3,570 sequences from 78 populations from MSEA, ISEA, EA, SA, Oceania and Australia. with the number of samples per population varying from 16 to 248 [23, 30, 61–82] for a total of 3,775 sequences belonging to 86 populations (Additional File 2: Table S2). All sequences were aligned to the Reconstructed Sapiens Reference Sequence (RSRS) [83] using the Multiple Alignment with Fast Fourier Transform (MAFFT) online service with default options [84]. Haplogroup assignment was performed by Haplogrep 3 [85] with PhyloTree mtDNA tree Build 17 (http://www.phylotree.org) [86].

The genetic diversity within populations (summary statistics), i.e. number of haplotypes, haplotype diversity (h) [87], haplogroup diversity, nucleotide diversity (π) and mean number of pairwise differences (MPD), as well as Tajima's D value [21], were calculated using Arlequin 3.5.2.2 for Windows [88]. PGDSpider [89] was used to convert the fasta file into an Arlequin project file (*.arp) with manually modifications according to the Arlequin user manual (https://cmpg.unibe.ch/software/arlequin35/man/Arlequin35.pdf). In the graphical interface of Arlequin during execution, calculation settings were chosen as following: Haplotype inference (estimate haplotype frequencies by counting and search for shared haplotypes); Molecular diversity indices (both standard diversity indices and molecular diversity indices with pairwise differences as a molecular distance); and Neutrality tests (Tajima's D with the infinite site model). The equations for all statistics can be found in the Arlequin user manual.

To compare the genetic variation among populations, Arlequin 3.5.2.2 [88] was also used to measure the genetic distance (Φ_{st}) between pairs of populations and generate the Φ_{st} distance matrix. The calculation settings were: "Population comparisons" with Compute pairwise difference (pi); No. of permutations: 1,000; and Significance level: 0.05. To visualize population relatedness, an R package [90] was used to carry out nonparametric MDS analysis (based on the Φ_{st} distance matrix and using R function: isoMDS package: MASS, and to construct heat plots of the Φ_{st} distance matrix (R function: ape, pegas, adegenet and ggplot2 packages). The matrix of shared haplotypes was also constructed using the R package (R function: ape, pegas, adegenet and ggplot2 packages). STATISTICA 13.0 (StatSoft, Inc., USA) was used to carry out a correspondence analysis (CA) based on haplogroup frequencies.

Genetic variance at three hierarchical subdivisions (within individuals of a population, among populations within a group, and among groups of populations) was assessed using the analysis of molecular variance (AMOVA) procedure [91], as implemented in Arlequin 3.5.2.2 [88]. Their indices (Φ statistics) can be considered the molecular equivalent of Wright's F statistics, as they summarize genetic diversity at various levels of hierarchical subdivisions [92]. In Arlequin execution, AMOVA was run under standard haplotypic conditions, with the distance matrix of pairwise difference computed from the data. The statistical significance of each variance

component and Φ value was evaluated using 1,000 permutations [91]. In this analysis, the studied populations were grouped based on both geographic and linguistic categories (Table 2).

Median-joining networks [93] of haplogroups D4e1a, E1a1a1, F1a1a1, B5a1 d, M50, M21b, M71, M46, C7a, F1a1c, R6, M21a, M17a, B4 g, and R21 were performed by Network (www.fluxus-engineering.com) and visualized in Network publisher 2.1.2.5. For haplogroup E1a1a1, we included ancient mtDNA from Liangdao Man in Taiwan, dated to 8,060–8,320 cal BP (GenBank accession number KF540505) [94], with an mtDNA coverage of 245 ×. This haplogroup was reported as ancestral to E1 [23]. For haplogroup M21b, we included ancient mtDNA from a Hoabinhian sample (Ma911) from Gua Cha Cave, Malaysia, dated to 3,872 ± 33 BP [24]. We processed the sequences as previously described [95], and used an in-house pipeline (https://github.com/alexhbnr/mitoBench-ancientMT) to reconstruct the mtDNA sequence. This yielded 18,451 mtDNA reads and allowed to reconstruct the mtDNA sequence with a coverage of 44.6 ×.

Through Bayesian phylogenetic analysis, we estimated haplogroup divergence times and past population sizes for the haplogroups observed in sea nomad populations, i.e. F1a1a1, D4e1a, C7a, E1a1a1, B5a1 d, F1a1c, R6, M71, M50, M21b and M46, and those found in the Maniq (B4 g, M21a and M17). For each haplogroup, estimates were determined based on all available sequences belonging to the haplogroup, including those from other populations. The number of samples per analysis is provided in Table 1. We also determined effective population size (N_e) estimates for the populations of interest (Additional File 2: Table S7. Table S7–[Parameters for BEAST runs]). All Bayesian phylogenetic analysis were performed with BEAST v2.7.6 software package [96], with RSRS [83] as an outgroup. Data were partitioned into coding (nucleotides 577–16,023) and noncoding regions (nucleotides 16,024–576), with the best-fit model of sequence evolution for both partitions determined using bModelTest [97]. Haplogroup divergence and N_e estimates were inferred by utilising the Bayesian skyline plot (BSP) model [98] and we assumed stepwise-constant changes for the N_e estimates. A strict molecular clock was assumed across branches within each tree. Given that the dataset consisted solely of contemporary mitochondrial genomes, we fixed the evolutionary rates to 1.708×10^{-8} substitutions/site/year and 9.883×10^{-8} substitutions/site/year for coding and noncoding regions, respectively [99] to allow for absolute timescale inference. The Markov Chain Monte Carlo (MCMC) chain length was adjusted based on sample size (10,000,000 to 100,000,000 generations) (Additional File 2: Table S7), and 10,000 trees were generated per run. Convergence of each MCMC chain was confirmed using Tracer v1.7.2 [100], ensuring effective sample size (ESS) values over 200 for each parameter. Tracer v1.7.2 was furthermore used to visualize the BSP. The posterior distributions of inferred trees were summarized into maximum clade credibility (MCC) trees utilising TreeAnnotator v2.7.6 with a 20% burn-in. MCC trees were visualized with FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/, last visited 29 October 2024).

Abbreviations

MSEA	Mainland Southeast Asia
ISEA	Island Southeast Asia
SEA	Southeast Asia
SA	South Asia
EA	East Asia
AN	Austronesian
AA	Austroasiatic
TK	Tai-Kadai
ST	Sino-Tibetan
BSP	Bayesian Skyline Plot
mtDNA	Mitochondrial Deoxyribonucleic Acid
MCC	Maximum Clade Credibility
RSRS	Reconstructed Sapiens Reference Sequence
HVR1	Hypervariable Region 1
STRs	Short Tandem Repeats
AMOVA	Analysis of Molecular Variance

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12915-025-02252-5.

Additional File 1.

Additional File 2.

Acknowledgements

We would like to thank coordinators who assisted with sample recruitment, and we thank all participants who donated their biological samples. We greatly appreciate Roland Schröder and Apisit Cheukaw for technical assistance. We thank the Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany, for hosting laboratory facilities.

Authors' contributions

W.K. conceived the study. W.K. M.S1., J.K., C.S. collected samples. W.K., W.W., L.A., A.H. and M.S2. were involved with generating data. W.K., W.W., S.O. and D.L. were involved with data analyses. W.K. and M.S2. interpreted the analyses and wrote the manuscript with input from A.H., C.K., H.S. and S.O. All authors read and approved the final manuscript.

Funding

This study was supported by Naresuan University (NU) and National Science, Research and Innovation Fund (NSRF) (Grant No: R2568B017), the Max Planck Society and the Institute of Suvanabhumi Studies under TASSHA (Thailand Academy of Social Sciences, Humanities and Arts), Ministry of Higher Education, Science, Research and Innovation. C.S. was supported by The Unit of Excellence 2025 on Biodiversity and Natural Resources Management, University of Phayao. W.K. and M.Sr. were also partially supported by The Global and Frontier Research University Fund, Naresuan University (Grant number: R2566 C051).

Data availability

Sequence data that support the findings of this study have been deposited in GenBank (Accession numbers PV067297-PV067394).

Declarations

Ethics approval and consent to participate
Written informed consent was obtained from sample donors. The protocol on human subjects was approved by the Khon Kaen University Ethic Committee (Protocol No. HE622223) and Naresuan University (COA No.250/2023).

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Biology, Faculty of Science, Naresuan University, Phitsanulok, Thailand. [2]Center of Excellence for Innovation and Technology for Detection and Advanced Materials (ITDAM), Naresuan University, Phitsanulok, Thailand. [3]Department of Biology, Faculty of Science, Khon Kaen University, Khon Kaen, Thailand. [4]Department of Biochemistry, Faculty of Medical Science, Naresuan University, Phitsanulok, Thailand. [5]School of Agriculture and Natural Resources, University of Phayao, Muang Phayao, Thailand. [6]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. [7]Department of Biology, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand. [8]Department of Archaeology, Silpakorn University, Bangkok, Thailand. [9]Department of Evolutionary Anthropology, Faculty of Life Sciences, University of Vienna, Vienna, Austria. [10]Transmission, Infection, Diversification and Evolution Group (Tide), Max Planck Institute of Geoanthropology, Jena, Germany. [11]Department of Archaeogenetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. [12]Institute of Biology, Vietnam Academy of Science and Technology, Hanoi, Vietnam. [13]Centre for Linguistics, Faculty of Humanities, Leiden University, Leiden 2300, The Netherlands. [14]Biométrie et Biologie Évolutive, UMR 5558, CNRS & Université de Lyon, Lyon, France.

Received: 25 January 2025 Accepted: 16 May 2025
Published online: 28 May 2025

References

1. Eberhard DM, Simons GF, Fennig CD. Ethnologue: languages of the World. 23rd eds. Texas: SIL International; 2020.
2. Radcliffe-Brown AR. The Andaman Islanders. Cambridge: Cambridge University Press; 2013.
3. Göllner T, Larena M, Kutanan W, Lukas H, Fieder M, Schaschl H. Unveiling the Genetic History of the Maniq, a Primary Hunter-Gatherer Society. Genome Biol Evol. 2022;14(4):evac021.
4. Hogan D. Men of the sea: coastal tribes of South Thailand's West Coast. J Siam Soc. 1972;60:205–35.
5. Attavanich M. A Study of living conditions in post-tsunami houses: the case of the moklen ethnic minority in Phang Nga Province, Southern Thailand. PhD dissertation, Kyoto University; 2016.
6. Bellina B, Blench R, Galipaud J-C. Sea Nomadism from the Past to the Present. In: Bellina B, Blench R, Galipaud J-C, editors. Sea Nomads of Southeast Asia: From the Past to the Present. Singapore: NUS Press; 2021. p. 1–27.
7. Chao C. Research Trends on Southeast Asian Sea-nomads. In: States, People, and Borders in Southeast Asia. Kyoto Review of Southeast Asia, issue 7. Kyoto University: Center for Southeast Asian Studies. 2006.
8. Bellina B. Khao Sam Kaeo: A Cosmopolitan Port-city or City-state in the South China Sea? In: Bellina B, editor. Khao Sam Kaeo: An Early Port-city Between the Indian Ocean and the South China Sea. Paris: Ecole française d'Extrême-Orient; 2017. p. 621–61.
9. Bellina B. Favereau A and Dussubieux L. Southeast Asian Early Maritime Silk Road Trading Polities Hinterland and the Sea Nomads of the Isthmus of Kra. In: Bellina B, Blench R, Galipaud J.-C, editors. Sea Nomads of Southeast Asia: From the Past to the Present. Singapore: NUS Press; 2021. pp. 102–41.
10. Andaya, Leonard Y. Leaves of the Same Tree: Trade and Ethnicity in the Straits of Melaka. In: The Orang Laut and the Malay, Honolulu. University of Hawai'I Press; 2008. pp. 173–201.
11. Chao C. The Orang Suku Laut of Riau, Indonesia: The Inalienable Gift of Territory. London; New York: Routledge; 2010.
12. Hall K. A History of Early Southeast Asia: Maritime Trade and Societal Development c. 100–1500. Lanham, MD: Rowman & Littlefield Publishers; 2011.
13. Blench R. The Linguistic Background to Southeast Asian Sea Nomadism. In: Bellina B, Blench R, Galipaud J-C, editors. Sea Nomads of Southeast Asia: From the Past to the Present. Singapore: NUS Press; 2021. p. 157–76.
14. Chao C. The Orang Suku Laut: Movement, Maps and Mapping. In: Bellina B, Blench R, Galipaud J-C, editors. Sea Nomads of Southeast Asia: From the Past to the Present. Singapore: NUS Press; 2021. p. 142–56.
15. Kusuma P, Cox MP, Pierron D, Razafindrazaka H, Brucato N, Tonasso L, et al. Mitochondrial DNA and the Y chromosome suggest the settlement of Madagascar by Indonesian sea nomad populations. BMC Genomics. 2015;16(1):191.
16. Kusuma P, Brucato N, Cox MP, Letellier T, Manan A, Nuraini C, et al. The last sea nomads of the Indonesian archipelago: genomic origins and dispersal. Eur J Hum Genet. 2017;25(8):1004–10.
17. Ilardo MA, Moltke I, Korneliussen TS, Cheng J, Stern AJ, Racimo F, et al. Physiological and genetic adaptations to diving in sea nomads. Cell. 2018;173(3):569-580.e15.
18. Dancause KN, Chan CW, Arunotai NH, Lum JK. Origins of the Moken Sea Gypsies inferred from mitochondrial hypervariable region and whole genome sequences. J Hum Genet. 2009;54(2):86–93.
19. Seetaraso T, Kutanan W, Kampuansai J, Muisuk K, Srikummool M. Unique Genetic Structure of Y-chromosomal Lineage of the Moken from the Andaman Sea of Thailand. CMUJ Nat Sci. 2020;19(3):1066–79.
20. Srikummool M, Srithawong S, Muisuk K, Sangkhano S, Suwannapoom C, Kampuansai J, et al. Forensic and genetic characterizations of diverse southern Thai populations based on 15 autosomal STRs. Sci Rep. 2022;12(1):655.
21. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989;123:585–95.
22. Zheng HX, Yan S, Qin ZD, Jin L. MtDNA analysis of global populations support that major population expansions began before Neolithic Time. Sci Rep. 2012;2:745.
23. Ko AM, Chen CY, Fu Q, Delfin F, Li M, Chiu HL, et al. Early Austronesians: into and out Of Taiwan. Am J Hum Genet. 2014;94:426–36.
24. McColl H, Racimo F, Vinner L, Demeter F, Gakuhari T, Moreno-Mayar JV, et al. The prehistoric peopling of Southeast Asia. Science. 2018;361(6397):88–92.
25. Sidwell P. 33 Southeast Asian mainland: linguistic history. In: Ness I, editor. The Encyclopedia of Global Human Migration. 1st ed: Wiley; 2013. pp. 1–10.
26. Grant WS, Liu M, Gao T, Yanagimoto T. Limits of Bayesian skyline plot analysis of mtDNA sequences to infer historical demographies in Pacific herring (and other species). Mol Phylogenet Evol. 2012;65(1):203–12.
27. Pittayaporn P. Layers of Chinese loanwords in proto-southwestern Tai as evidence for the dating of the spread of southwestern Tai. Manusya J Humanit. 2014;20:47–68.
28. Bettinger RL, Barton L, Morgan C. The origins of food production in North China: a different kind of agricultural revolution. Evol Anthropol. 2010;19:9–21.
29. Bellwood P. First farmers: The origins of agricultural societies. Oxford: Blackwell; 2005.
30. Delfin F, Min-Shan Ko A, Li M, Gunnarsdóttir ED, Tabbada KA, Salvador JM, et al. Complete mtDNA genomes of Filipino ethnolinguistic groups: a melting pot of recent and ancient lineages in the Asia-Pacific region. Eur J Hum Genet. 2014;22:228–37.
31. Tabbada KA, Trejaut J, Loo JH, Chen YM, Lin M, Mirazón-Lahr M, et al. Philippine mitochondrial DNA diversity: a populated viaduct between Taiwan and Indonesia? Mol Biol Evol. 2010;27(1):21–31.
32. Trejaut JA. Origin of the Bunun Indigenous People of Taiwan, a Review of Published Material Using Y-Chromosome and Mitochondrial DNA Gene Systems. DNA. 2022;2:185–201.
33. Larish MD. The Position of Moken and Moklen within the Austronesian Language Family (Thailand). Ann Arbor: University Microfilms; 2000.

34. Larish MD. Moken and Moklen. In: Adelaar S, Himmelmann N, editors. The Austronesian Languages of Asia and Madagascar. London: Routledge; 2005. p. 513–33.
35. Steinhauer H. On the Development of Urak Lawoi' Malay. Wacana, J Humanit Indonesia. 2008;10(1):117–43.
36. Robert CS. Changes in Patterns of Communication of the Urak Lawoi'. PhD dissertation. University of Wisconsin; 2010.
37. Thongtaweewiwat K. Community Rights and Interest of Thai Society: A Case Study of Moklen Land Ownership after the Tsunami in the Baan Tungwah, Khukkak Sub-district, Takuapa District, Phang Nga. Master thesis, Chulalongkorn University; 2006 (in Thai).
38. Hope S. Outcasts of the Islands: the Sea Gypsies of South East Asia. London: Harper Collins Publishers; 2000.
39. Nawichai P. Ethnic Group Livelihood Strategies and State Integration: Moken and the Hill People in Negotiation with the State. Doctoral dissertation, University of Passau; 2008.
40. SAC. Anthropology database of the Anthropology's Princess Maha Chakri Sirindhorn Anthropology Centre. 2021. Available online: https://ethnicity.sac.or.th/database-ethnic. Accessed 25 Feb 2025.
41. White WG. The Sea Gypsies of Malaya. Philadelphia: JB Lippincott Co.; 1922.
42. Hinshiranan N. The Analysis of Moken Opportunistic Foragers' Intragroup and Intergroup Relations. PhD dissertation, University of Hawai'i; 1996.
43. Ivanoff J. Sea Gypsies of Myanmar. Natl Geographic. April 2005: 36–55.
44. Sopher DE. The Sea Nomads: a Study Based on the Literature of the Maritime Boat People of Southeast Asia. Singapore: Lim Bian Han Government Printer; 1965.
45. Pattemore SW, Hogan DW. On the origin of the Urak Lawoi': a response to J. Ivanoff. J Siam Soc. 1989;77(2):75–8.
46. Kutanan W, Kampuansai J, Changmai P, Flegontov P, Schröder R, Macholdt E, et al. Contrasting maternal and paternal genetic variation of hunter-gatherer groups in Thailand. Sci Rep. 2018;8:1536.
47. Woravatin W, Stoneking M, Srikummool M, Kampuansai J, Arias L, Kutanan W. South Asian maternal and paternal lineages in southern Thailand and the role of sex-biased admixture. PLoS One. 2023;18(9):e0291547.
48. Kutanan W, Kitpipit T, Phetpeng S, Thanakiatkrai P. Forensic STR loci reveal common genetic ancestry of the Thai-Malay Muslims and Thai Buddhists in the deep Southern region of Thailand. J Hum Genet. 2014;59(12):675–81.
49. Arias L, Barbieri C, Barreto G, Stoneking M, Pakendorf B. High-resolution mitochondrial DNA analysis sheds light on human diversity, cultural interactions, and population mobility in Northwestern Amazonia. Am J Phys Anthropol. 2018;165(2):238–55.
50. Maricic T, Whitten M, Pääbo S. Multiplexed DNA Sequence Capture of Mitochondrial Genomes Using PCR Products. PLoS One. 2010;5:e14004.
51. Meyer M, Kircher M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb Protoc. 2010;6:1–10.
52. Renaud G, Stenzel U, Kelso J. LeeHom: adaptor trimming and merging for Illumina sequencing reads. Nucleic Acids Res. 2014;42:e141.
53. Renaud G, Stenzel U, Maricic T, Wiebe V, Kelso J. deML: robust demultiplexing of Illumina sequences using a likelihood-based approach. Bioinformatics. 2014;31(5):770–2.
54. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60.
55. Li M, Schroeder R, Ko A, Stoneking M. Fidelity of capture-enrichment for mtDNA genome sequencing: influence of NUMTs. Nucleic Acids Res. 2012;40(18):e137.
56. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM. Howell N Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet. 1999;23:147.
57. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv. 2013;3:13033997.
58. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297–303.
59. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011;27(21):2987–93.
60. Pradutkanchana S, Ishida T, Kimura R. Mitochondrial diversity of the sea nomads of Thailand. Genbank https://www.ncbi.nlm.nih.gov/nuccore/GU810005.1 – https://www.ncbi.nlm.nih.gov/nuccore/GU810079.1. 2010.
61. Kutanan W, Kampuansai J, Brunelli A, Ghirotto S, Pittayaporn P, Ruangchai S, et al. New insights from Thailand into the maternal genetic history of Mainland Southeast Asia. Eur J Hum Genet. 2018;26:898–911.
62. Kutanan W, Kampuansai J, Srikummool M, Kangwanpong D, Ghirotto S, Brunelli A, et al. Complete mitochondrial genomes of Thai and Lao populations indicate an ancient origin of Austroasiatic groups and demic diffusion in the spread of Tai-Kadai languages. Hum Genet. 2017;136:85–98.
63. Kutanan W, Shoocongdej R, Srikummool M, Hübner A, Suttipai T, Srithawong S, et al. Cultural variation impacts paternal and maternal genetic lineages of the Hmong-Mien and Sino-Tibetan groups from Thailand. Eur J Hum Genet. 2020;28:1563–79.
64. Thao DH, Dinh TH, Mitsunaga S, Duy D, Phuong NT, Anh NP, et al. Investigating demic versus cultural diffusion and sex bias in the spread of Austronesian languages in Vietnam. PLoS One. 2024;19(6):e0304964.
65. Duong NT, Macholdt E, Ton ND, Arias L, Schröder R, Van Phong N, et al. Complete human mtDNA genome sequences from Vietnam and the phylogeography of Mainland Southeast Asia. Sci Rep. 2018;8(1):11651.
66. Peng MS, Quang HH, Dang KP, Trieu AV, Wang HW, Yao YG, et al. Tracing the Austronesian footprint in Mainland Southeast Asia: a perspective from mitochondrial DNA. Mol Biol Evol. 2010;27(10):2417–30.
67. Kloss-Brandstätter A, Summerer M, Horst D, Horst B, Streiter G, Raschenberger J, et al. An in-depth analysis of the mitochondrial phylogenetic landscape of Cambodia. Sci Rep. 2021;11:10816.
68. Jinam TA, Hong LC, Phipps ME, Stoneking M, Ameen M, et al. Evolutionary history of continental southeast Asians: "early train" hypothesis based on genetic analysis of mitochondrial and autosomal DNA data. Mol Biol Evol. 2012;29:3513–27.
69. Gunnarsdóttir ED, Li M, Bauchet M, Finstermeier K, Stoneking M. High-throughput sequencing of complete human mtDNA genomes from the Philippines. Genome Res. 2011;21:1–11.
70. Gunnarsdóttir ED, Nandineni MR, Li M, Myles S, Gil D, Pakendorf B, et al. Larger mitochondrial DNA than Y-chromosome differences between matrilocal and patrilocal groups from Sumatra. Nat Commun. 2011;2:228.
71. Diroma MA, Calabrese C, Simone D, Santorsola M, Calabrese FM, Gasparre G, et al. Extraction and annotation of human mitochondrial genomes from 1000 genomes whole exome sequencing data. BMC Genomics. 2014;15:S2.
72. Zhang X, Qi X, Yang Z, Serey B, Sovannary T, Bunnath L, et al. Analysis of mitochondrial genome diversity identifies new and ancient maternal lineages in Cambodian aborigines. Nat Commun. 2013;4:2599.
73. Feng Y, Zhang H, Wang Q, Jin X, Le C, Liu Y, et al. Whole mitochondrial genome analysis of Tai-Kadai-speaking populations in Southwest China. Front Ecol Evol. 2022;10:1000493.
74. Zheng HX, Yan S, Qin ZD, Wang Y, Tan JZ, Li H, et al. Major population expansion of East Asians began before neolithic time: evidence of mtDNA genomes. PLoS One. 2011;6(10):e25835.
75. Summerer M, Horst J, Erhart G, Weißensteiner H, Schönherr S, Pacher D, et al. Large-scale mitochondrial DNA analysis in Southeast Asia reveals evolutionary effects of cultural isolation in the multi-ethnic population of Myanmar. BMC Evol Biol. 2014;14:17.
76. Li YC, Wang HW, Tian JY, Liu LN, Yang LQ, Zhu CL, et al. Ancient inland human dispersals from Myanmar into interior East Asia since the Late Pleistocene. Sci Rep. 2015;5:9473.
77. Chandrasekar A, Kumar S, Sreenath J, Sarkar BN, Urade BP, Mallick S, et al. Updating phylogeny of mitochondrial DNA macrohaplogroup m in India: dispersal of modern human in South Asian corridor. PLoS One. 2009;4(10):e7447.
78. Thangaraj K, Chaubey G, Kivisild T, Reddy AG, Singh VK, Rasalkar AA, et al. Reconstructing the origin of Andaman Islanders. Science. 2005;308:996.
79. Barik SS, Sahani R, Prasad BV, Endicott P, Metspalu M, Sarkar BN, et al. Detailed mtDNA genotypes permit a reassessment of the settlement

and population structure of the Andaman Islands. Am J Phys Anthropol. 2008;136(1):19–27.
80. Gomes SM, Bodner M, Souto L, Zimmermann B, Huber G, Strobl C, et al. Human settlement history between Sunda and Sahul: a focus on East Timor (Timor-Leste) and the Pleistocenic mtDNA diversity. BMC Genomics. 2015;16:70.
81. Pedro N, Brucato N, Fernandes V, André M, Saag L, Pomat W, et al. Papuan mitochondrial genomes and the settlement of Sahul. J Hum Genet. 2020;65:875–87.
82. Nagle N, van Oven M, Wilcox S, van Holst PS, Tyler-Smith C, Xue Y, et al. Aboriginal Australian mitochondrial genome variation - an increased understanding of population antiquity and diversity. Sci Rep. 2017;7:43041.
83. Behar DM, van Oven M, Rosset S, Metspalu M, Loogväli EL, Silva NM, et al. A "Copernican" Reassessment of the Human Mitochondrial DNA Tree from its Root. Am J Hum Genet. 2012;90:675–84.
84. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform. 2019;20(4):1160–6.
85. Schönherr S, Weissensteiner H, Kronenberg F, Forer L. Haplogrep 3 - an interactive haplogroup classification and analysis platform. Nucleic Acids Res. 2023;51(W1):W263–8.
86. van Oven M, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mut. 2009;30(2):E386–94.
87. Nei M. Molecular Evolutionary Genetics. New York: Columbia University Press; 1987.
88 Excoffier L, Lischer HEL. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour. 2010;10:564–7.
89. Lischer HE, Excoffier L. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics. 2012;28(2):298–9.
90. R Development Core Team: R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2011.
91. Excoffier L, Smouse P, Wuattro J. Analysis of molecular variance inferred from metric distance among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics. 1992;131:479–91.
92. Bertorelle G, Calafell F, Francalacci P, Bertranpetit J, Barbujani G. Geographic homogeneity and non-equilibrium patterns of mtDNA sequences in Tuscany. Italy Hum Genet. 1996;98:145–50.
93. Bandelt HJ, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 1999;16:37–48.
94. Ko AM, Chen CY, Fu Q, Delfin F, Li M, Chiu HL, Stoneking M, Ko YC. Early austronesians: into and out of Taiwan. GenBank https://www.ncbi.nlm.nih.gov/nuccore/KF540505. 2014.
95. Pugach I, Hübner A, Hung HC, Meyer M, Carson MT, Stoneking M. Ancient DNA from Guam and the peopling of the Pacific. Proc Natl Acad Sci USA. 2021;118(1):e2022112118.
96 Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 2019;15(4):e1006650.
97. Bouckaert RR, Drummond AJ. bModelTest: Bayesian phylogenetic site model averaging and model comparison. BMC Evol Biol. 2017;17(1):42.
98. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol. 2005;22:1185–92.
99. Soares P, Ermini L, Thomson N, Mormina M, Rito T, Röhl A, et al. Correcting for purifying selection: an improved human mitochondrial molecular clock. Am J Hum Genet. 2009;84(6):740–59.
100 Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using tracer 1.7. Syst Biol. 2018;67(5):901–4.

Publisher's Note