



**HAL**  
open science

## South Asian maternal and paternal lineages in southern Thailand and the role of sex-biased admixture

Wipada Woravatin, Mark Stoneking, Metawee Srikumool, Jatupol Kampaunsai, Leonardo Arias, Wibhu Kutanan

### ► To cite this version:

Wipada Woravatin, Mark Stoneking, Metawee Srikumool, Jatupol Kampaunsai, Leonardo Arias, et al.. South Asian maternal and paternal lineages in southern Thailand and the role of sex-biased admixture. PLoS ONE, 2023, 18 (9), pp.e0291547. 10.1371/journal.pone.0291547 . hal-04367499

**HAL Id: hal-04367499**

**<https://hal.science/hal-04367499>**

Submitted on 30 Dec 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## RESEARCH ARTICLE

# South Asian maternal and paternal lineages in southern Thailand and the role of sex-biased admixture

Wipada Woravatin<sup>1</sup>, Mark Stoneking<sup>2,3</sup>, Metawee Srikumool<sup>4</sup>, Jatupol Kampuansai<sup>5,6</sup>, Leonardo Arias<sup>2,7</sup>, Wibhu Kutanana<sup>1,8\*</sup>

**1** Department of Biology, Faculty of Science, Khon Kaen University, Khon Kaen, Thailand, **2** Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany, **3** Biométrie et Biologie Évolutive, UMR 5558, CNRS & Université de Lyon, Lyon, France, **4** Department of Biochemistry, Faculty of Medical Science, Naresuan University, Phitsanulok, Thailand, **5** Department of Biology, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand, **6** Research Center in Bioresources for Agriculture, Industry and Medicine, Chiang Mai University, Chiang Mai, Thailand, **7** Centre for Linguistics, Faculty of Humanities, Leiden University, Leiden, The Netherlands, **8** Department of Biology, Faculty of Science, Naresuan University, Phitsanulok, Thailand

\* [wibhuk@nu.ac.th](mailto:wibhuk@nu.ac.th)



## OPEN ACCESS

**Citation:** Woravatin W, Stoneking M, Srikumool M, Kampuansai J, Arias L, Kutanana W (2023) South Asian maternal and paternal lineages in southern Thailand and the role of sex-biased admixture. PLoS ONE 18(9): e0291547. <https://doi.org/10.1371/journal.pone.0291547>

**Editor:** Guanglin He, Xiamen University, CHINA

**Received:** June 22, 2023

**Accepted:** August 31, 2023

**Published:** September 14, 2023

**Copyright:** © 2023 Woravatin et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The mtDNA dataset generated during the current study are available in GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) with accession numbers OQ731946–OQ731991 and the Y chromosomal dataset are available in European Nucleotide Archive (project code: PRJEB61777).

**Funding:** W.W. and W.K. were supported by National Research Council of Thailand (NRCT): (N41A640275). W.K., J.K. and M.Sr. were funded by the Institute of Suvarnabhumi Studies, Thailand Academy of Social Sciences, Humanities and Arts:

## Abstract

Previous genome-wide studies have reported South Asian (SA) ancestry in several Mainland Southeast Asian (MSEA) populations; however, additional details concerning population history, in particular the role of sex-specific aspects of the SA admixture in MSEA populations can be addressed with uniparental markers. Here, we generated ~2.3 mB sequences of the male-specific portions of the Y chromosome (MSY) of a Tai-Kadai (TK)-speaking Southern Thai group (SouthernThai\_TK), and complete mitochondrial (mtDNA) genomes of the SouthernThai\_TK and an Austronesian (AN)-speaking Southern Thai (SouthernThai\_AN) group. We identified new mtDNA haplogroups, e.g. Q3, E1a1a1, B4a1a and M7c1c3 that have not previously reported in Thai populations, but are frequent in Island Southeast Asia and Oceania, suggesting interactions between MSEA and these regions. SA prevalent mtDNA haplogroups were observed at frequencies of ~35–45% in the Southern Thai groups; both of them showed more genetic relatedness to Austroasiatic (AA) speaking Mon than to any other group. For MSY, SouthernThai\_TK had ~35% SA prevalent haplogroups and exhibited closer genetic affinity to Central Thais. We also analyzed published data from other MSEA populations and observed SA ancestry in some additional MSEA populations that also reflects sex-biased admixture; in general, most AA- and AN-speaking groups in MSEA were closer to SA than to TK groups based on mtDNA, but the opposite pattern was observed for the MSY. Overall, our results of new genetic lineages and sex-biased admixture from SA to MSEA groups attest to the additional value that uniparental markers can add to studies of genome-wide variation.

TASSHA, Ministry of Higher Education, Science, Research and Innovation. This study was also partially supported by the Global and Frontier Research University Fund, Naresuan University (Grant number: R2566C051).

**Competing interests:** NO authors have competing interests.

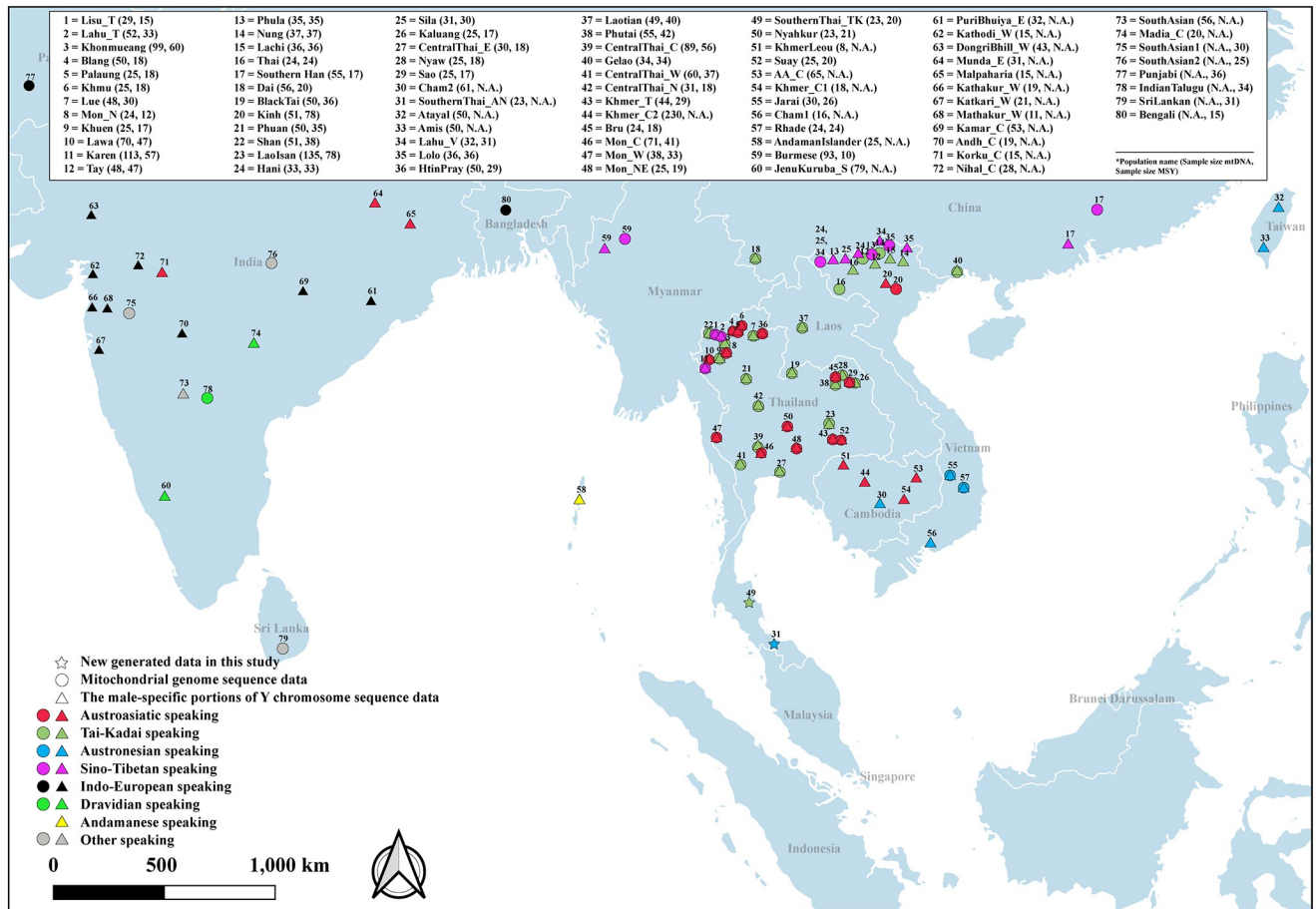
## Introduction

Located on the Malay Peninsula, Southern Thailand shares borders with Central Thailand to the North, the Andaman Sea to the West, the Gulf of Thailand to the East and Malaysia to the South. Most southern Thai people speak languages belonging to the Tai-Kadai (TK) and Austronesian (AN) language families. The KhonTai (TK) and ThaiMalay (AN) account for ~66% and 33% of the population, respectively, whereas the Maniq, one of the minor groups in the region, speak an Austroasiatic (AA) language. Maniq is one of the indigenous peoples of Southeast Asia (SEA) and probably descend from mid-Holocene hunter-gatherers [1–3], since they carry ancient genetic lineages [4, 5]. Overall, there is evidence of a long history of human occupation of Southern Thailand [6], as attested by the diversification of population subsistence patterns, cultural artifacts of Paleolithic hunter-gatherers associated with the Hòabinhian archaeological tradition dated to ~27–38 thousand years ago (kya) [7], and human remains at Moh Khiew Cave in the Krabi Province (dated to ~26–8 kya) [8–10].

As in other areas of SEA, the shift from hunter-gatherer to farming lifestyles in the Malay Peninsula probably happened during the Neolithic Period ~5–4 kya [11, 12]. An admixed genetic portrait between the deeply diverged East Eurasian lineage and East Asians was proposed for Neolithic people who had genetic connections with some contemporary SEA AA-speaking populations [2, 13]. Therefore, AA languages probably were predominant in Mainland Southeast Asia (MSEA) during the Neolithic Period, and subsequently fragmented due to the influence of later expansions during the Bronze/Iron Age ~2.5–2 kya that are associated with present-day TK and AN speakers. In addition to migrations from East Asia (EA) to MSEA, vestiges of interactions between South Asia (SA) and MSEA are also found during the late Bronze Age to Iron Age [11]. Maritime trade networks stretched from China through SEA to India and the Mediterranean region, which allowed SEA people to adopt technological knowledge and culture from both China and India [11, 14].

Previous genome-wide studies found evidence of SA genetic ancestry in several present-day MSEA population [15, 16]. The SA ancestry in these populations ranged from 25% to 30% and dated to ~500 to 750 ya in one study [15], while another study estimated SA admixture proportions ranging from ~4% to 12% and dates between ~450 to 1,600 [16]. Additional insights into SA ancestry in SEA could come from analyses of the paternally-inherited male-specific portions of the Y chromosome (MSY) and the maternally-inherited mitochondrial (mt) DNA genome. These markers are able to reconstruct different aspects of the genetic history than obtained from genome-wide data, especially sex-biased admixture, which frequently occurs in human populations [17–20]. Contrasting patterns of MSY and mtDNA variation were reported previously in some MSEA populations, especially the highlanders in Thailand and Vietnam [21–25], and shown to be influenced mainly by postmarital residence patterns (i.e., matrilocality vs. patrilocality), genetic drift, and admixture with lowlanders.

Previous MSY and mtDNA studies reported the presence of haplogroups frequent in SA in some Thai populations, e.g., mtDNA haplogroups W3a1b, M30 and M45 in central Thais and Mon [26, 27] and MSY haplogroups R1a1 in Khmer, J2a in Nyahkur and J and R in Mon [23]. However, the role of sex-biased admixture related to SA ancestry in MSEA has not been studied, and moreover only a few groups from Southern Thailand have been investigated for mtDNA and MSY variation [5, 28, 29]. Here, we generated new data, consisting of complete mtDNA genome sequences from TK- and AN-speaking southern Thai populations, and ~2.3 mB of MSY sequences from the southern Thai\_TK. We explored maternally and paternally genetic lineages and genetic relationships between southern Thais and other SA/SEA populations (Fig 1). We investigated SA genetic contribution in our newly-generated southern Thai



**Fig 1. Map showing the locations of 88 populations from Southeast and South Asia that were analyzed for mtDNA (74 populations) and MSY (55 populations) variation.** The map was generated using the Quantum GIS, QGIS Development Team (2023), QGIS Geographic Information System. Open Source Geospatial Foundation (<https://www.qgis.org>) and boundaries were adapted and modified from <https://public.opendatasoft.com/explore/dataset/worldadministrative-boundaries/export/>.

<https://doi.org/10.1371/journal.pone.0291547.g001>

populations as well as in published data from MSEA, and we find heterogeneity in SA admixture and sex-biased admixture in MSEA populations.

## Materials and methods

### Samples

The 46 genomic DNA samples (23 for southern Thai\_TK and 23 for southern Thai\_AN) were from previous studies [15, 30]. For southern Thai\_TK, there are 20 males and 3 females but all 23 southern Thai\_AN donors are females. Samples were recruited during November 2019. Sample donors were unrelated for at least two generations and provided buccal samples with written informed consent, after the goals of the study were explained and all questions answered. The rights and identity of all donors have been protected during the entire process of this research, according to guidelines and regulations based on the protocol on human subjects which was approved by the Khon Kaen University Ethic Committee (Protocol No. HE622223). Genomic DNAs were extracted using the Gentra Puregene Buccal Cell Kit (Qiagen, Germany) according to the manufacturer’s instructions.

## Sequencing

On December 2019, genomic libraries with double indices were prepared and enriched for mtDNA as described previously [31, 32]. The libraries were sequenced on an Illumina HiSeq 2500 and mtDNA consensus sequences obtained as described by previous study [19] with minor modifications. Bustard was used for Illumina standard base calling with 76 bp for read length. Sequences were then manually checked and manipulated with Bioedit ([www.mbio.ncsu.edu/BioEdit/bioedit.html](http://www.mbio.ncsu.edu/BioEdit/bioedit.html)). The sequence alignment to the Reconstructed Sapiens Reference Sequence (RSRS) [33] was done by MAFFT 7.271 [34].

The same genomic libraries for 20 male samples of southernThai\_TK were enriched for ~2.3 mB of the MSY via in-solution hybridization-capture using a previously designed probe set [5, 23] and the Agilent Sure Select system (Agilent, CA). Sequencing was carried out on the Illumina HiSeq 2500 platform with paired-end reads of 125-bp length. Bustard was used for Illumina standard base calling, and leeHOM was used to trim Illumina adapters and merge completely overlapping paired sequences [35]. Then, deML was used to demultiplex the pooled sequencing data [36]. The alignment and post-processing pipeline of the sequencing data was carried out as previously described [5].

## Statistical analysis

The newly-generated 46 mtDNA sequences (S1 Table) were combined with 3,169 sequences from 72 populations from SA and SEA to obtain a broader picture of population relationships [24, 26, 37–47] for a total of 3,215 sequences belonging to 74 populations (S2 Table). Apart from these 74 populations, another published 31 sequences of haplogroup Q3 and 71 sequences of haplogroup E1a1a1 from SEA and Near Oceania were included in the median-joining network [48] of haplogroup Q3 [49–54] (S3 Table) and E1a1a1 [37, 55–57] (S4 Table). The network was constructed with Network 10.2 ([www.fluxus-engineering.com](http://www.fluxus-engineering.com)) without pre- and post-processing steps, and all nucleotide position were weighted equally. For MSY, combining the 20 new sequences of southern Thai\_TK from this study (S5 Table) with 1,674 sequences from previous studies [23–25, 58–60] brings the total to 1,694 sequences from 55 populations (S6 Table). For mtDNA, haplogroup assignment was performed by Haplogrep [61] with PhyloTree mtDNA tree Build 17 (<http://www.phylotree.org>) [62]. The mixture proportions of mtDNA samples and most likely detected mitochondrial haplogroups for each sample were estimated and identified by mixEMT software [63].

Using yHaplo [64], MSY haplogroups were assigned to the maximum depth possible given the phylogeny of the ISOGG Y-DNA Haplogroup Tree 2015 (<http://www.isogg.org/>) and the available genetic markers in our target region. Summary statistics of the genetic diversity within populations and the matrix of pairwise genetic distances ( $\Phi_{st}$ ) were obtained with Arlequin 3.5.1.3 [65]. To visualize population relatedness, the R package (R Development Core Team 2016) [66] was used to carry out nonparametric MDS analysis (based on the  $\Phi_{st}$  distance matrices for the MSY and mtDNA and using R function: isoMDS package: MASS) and to construct heat plots of the  $\Phi_{st}$  distance matrix and the matrix of shared haplotypes (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 MSY and mtDNA haplogroup frequencies.

Admix 2.0 was used to estimate admixture proportions [67, 68]. We used a simple admixture model consisting of two parental groups: East and Southeast Asian (ESEA) and South Asian (SA) (S2 and S6 Tables). Several analyses using 12 and 8 different parameters for mtDNA and MSY, respectively, i.e., admixture times (2000, 1000, 500, and 0 ya) and mutation

rates ( $0$ ,  $1.67 \times 10^{-8}$  and  $2.67 \times 10^{-8}$  for mtDNA and  $0$  and  $8.71 \times 10^{-10}$  for MSY) were performed (S7 Table) [69–71], with 10,000 bootstrap replicates.

To construct Bayesian skyline plots (BSP) per population, based on Bayesian Markov Chain Monte Carlo (MCMC) analyses, we used BEAST 1.8.4 [72]. BEAST input files were created with BEAUTi v1.8.2 after first running jModel test 2.1.7 in order to choose the most suitable model of sequence evolution [73]. For mtDNA, we executed BSP analyses per population with mutation rates of  $1.708 \times 10^{-8}$  and  $9.883 \times 10^{-8}$  for data partitioned between the coding and noncoding regions, respectively [69]. For the MSY, we used an MSY mutation rate of  $8.71 \times 10^{-10}$  substitutions/bp/year [71], and the BEAST input files were modified by an in-house script to add in the invariant sites found in our data set. Both strict and log normal relaxed clock models were run, with marginal likelihood estimation [74, 75]. After each BEAST run, the Bayes factor was computed from the log marginal likelihood of both models to choose the best-fitting BSP tree. Tracer 1.6 was used to generate the BSP plot from the BEAST results.

## Results

### Genetic diversity

We generated 46 complete mtDNA sequences from two populations, southernThai\_TK and southernThai\_AN, with mean coverage ranging from 788x to 4022x (overall average coverage 2346x) (S1 Table). There were no missing sites in any of the sequences, and an analysis for sample mixtures did not detect any evidence of multiple sequences in any sample. Compared to other populations in Thailand and elsewhere in Asia (S2 Table), high genetic diversity values for both populations were observed: haplotype diversity ( $h$ ) = 1.00 for both southernThai\_TK and southern Thai\_AN, mean number of pairwise difference (MPD) = 36.9 for southernThai\_TK and 34.7 for southern Thai\_AN, Tajima's  $D$  values =  $-2.18$  ( $P < 0.01$ ) for southernThai\_TK and  $-1.99$  ( $P < 0.01$ ) for southern Thai\_AN. Both the high genetic diversities and the significantly negative Tajima's  $D$  values suggest recent maternal expansions in these two groups.

The mean coverage for the 20 new MSY sequences, for the SouthernThai\_TK, ranges from 11.42x to 22.62x (overall average coverage 15.87x) (S5 Table). Genetic diversity values ( $h$  = 1.00, MPD = 98.8) indicate high genetic variation within this population compared to other Thai and SA/SEA populations (S6 Table). The Tajima's  $D$  value of  $-2.04$  ( $P < 0.01$ ) also indicates a recent history of paternal population expansion.

### Genetic lineages

We assigned haplogroups as prevalent in SEA/EA (found at high frequency in SEA/EA and absent or nearly so in SA, e.g. B, F, M7, R\*, and R9), prevalent in SA (e.g. M\*(xM7), M8, M9 and M12), or other (found at low frequency in either or both regions, e.g. G, M12, and N\*) (S1 Fig). Among a total of 22 mtDNA haplogroups detected in southernThai\_TK, there were 39.13% SEA/EA prevalent haplogroups (B4a1a, F1a1a, F1f, M7b1a1 (+16192), M7c1a3, R9c1b1) and 43.48% prevalent in SA (M3c1a, M17, M21b, M21b2, M22a, M24a, M26, M38, M50, M77); other haplogroups, accounted for 17.39% (N, N8, M12a2, R11b1 and R11'B6) (S1 Fig, S8 Table). For the 18 haplogroups in southernThai\_AN, SEA/EA prevalent haplogroups (B4c1b2a2, B4c2, B5a1a, F1a1a1, F1f, M7c1c3) had a frequency of 39.13%; SA prevalent haplogroups (M17c, M21a, M24a, M50a, M51b, M72a and M74b2) accounted for 34.78%, and other haplogroups (M12a, N9a6a, N22, Q3, E1a1a1) occurred with frequency 26.09% (S1 Fig, S8 Table).

Interestingly, there is one SouthernThai\_AN individual with mtDNA haplogroup Q3, which has never been reported in MSEA populations. We retrieved mtDNA genomes of haplogroup Q3 from previous studies [49–54] and constructed a median-joining network to investigate the mtDNA relationships within Q3. The SouthernThai\_AN Q3 sequence differed from all others but was closest to a sequence from East Timor (differing by 3 substitutions) and both sequences were more distantly related to a sequence from an individual belonging to the non-Austronesian-speaking Oro group from Papua New Guinea (S2 Fig). In addition, haplogroup E1a1a1, which has not been reported in any other Thai populations, was observed in SouthernThai\_AN. The network constructed from mtDNA genomes of haplogroup E1a1a1 from previous studies [37, 55–57, 76] showed a star-like structure that reflected lineage expansion of this haplogroup. All E1a1a1 sequences branch off from the central haplotype, which is present in individuals from the Philippines, and carry only a few distinct mutations (1 to 2), with the exception of some haplotypes from Taiwan and one from Indonesia that are separated by four to five mutations from the central node (S3 Fig).

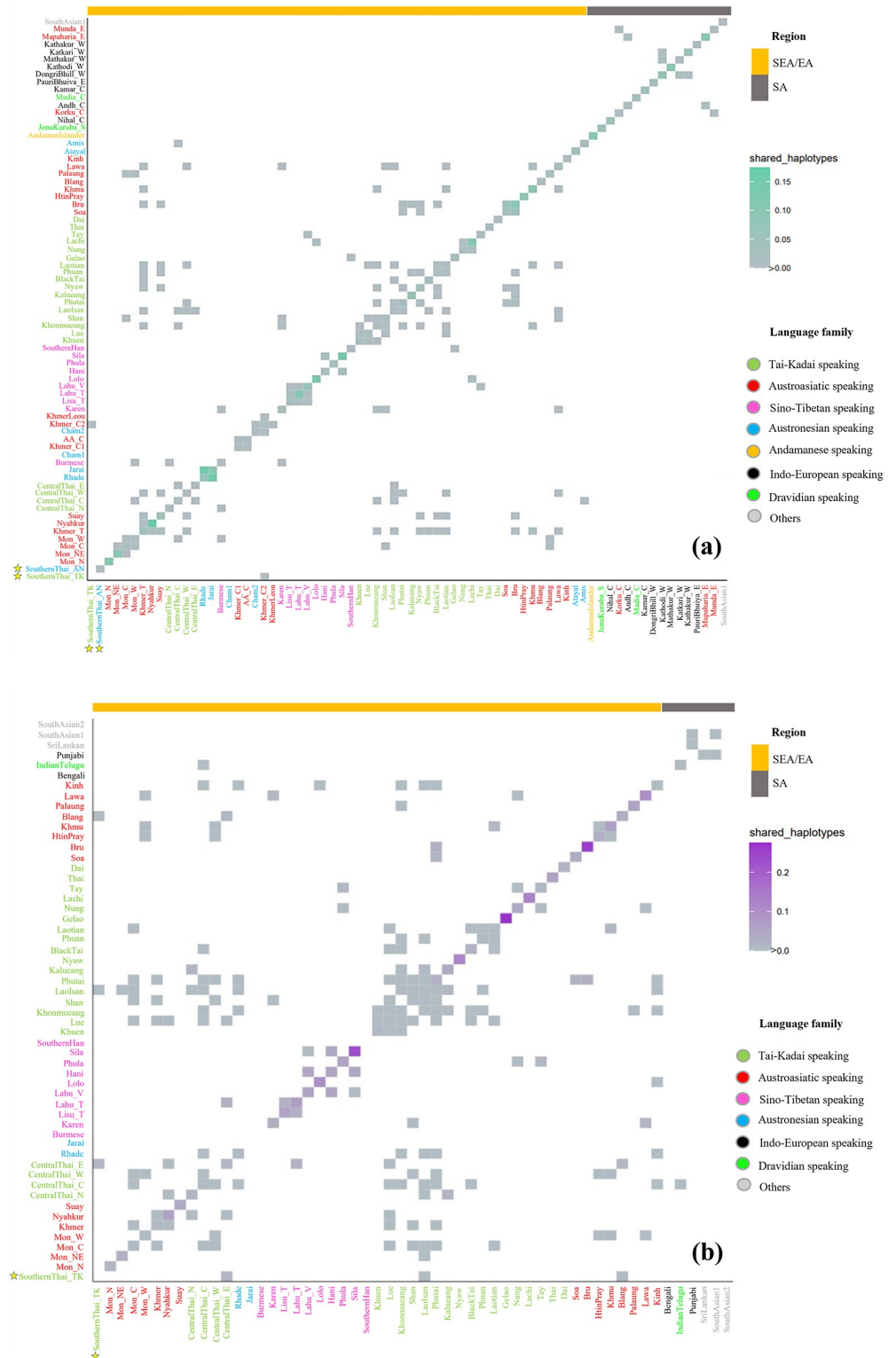
In other MSEA groups, SA prevalent haplogroups were in very high frequency (greater than 60%) in Burmese, three AN populations from Vietnam (Rhade, Jarai, Cham1) and Cambodian Khmer\_C1, while all southern Thais, almost all Mon groups (except Mon\_N), Nyahkur, Suay, and two populations from Cambodia (KhmerLeou and AA\_C) exhibited moderately high levels of SA prevalent haplogroups (30–60%) (S4 Fig). All central Thai groups, Mon\_N, Thai Khmer (Khmer\_T), and Cambodian Khmer (Khmer\_C2) had percentages of SA prevalent haplogroups ranging from 15–30%. The Correspondence Analysis (CA) based on haplogroup frequencies further illustrates these results, as the Burmese, Khmer\_C1, Rhade, Jarai, and Cham1 are positioned closest to SA populations, and in general the MSEA AA-speaking populations were closer to SA populations than other SEA/EA language groups; the southern Thai\_TK, differ notably from other TK groups in being closer to SA groups (S5 Fig).

There were 19 MSY haplogroups among the 20 newly generated sequences from the SouthernThai\_TK; the SEA/EA prevalent haplogroups O1 and O2 (O1a1a1a1a, O1b1a1a, O1b1a2a1, O1b1a2b1, O2a1c1a1a1, O2a1c1a7, O2a2a1a2a1a, O2a2b1a1a3a, O2b1a) were the most frequent at 40% (S6 Fig, S5 Table). SA prevalent haplogroups, identified as belonging to H\*, J\*, L\*, and R\* (H1a1d2, L1a, R, R1a1a1b and R1a1a1b2a1b) had a frequency of 35%, and another 25% were other haplogroups (C, C2e2a1, Q1a2, K and N1c2) (S6 Fig, S9 Table). When other MSEA were included, in contrast to mtDNA, the CA analysis based on MSY haplogroup frequencies indicated that TK groups (including the SouthernThai\_TK) were closer to the SA populations than AA groups (other than the Mon) (S7 Fig). This contrasting pattern was also observed in the AN-speaking Vietnamese groups. Specifically, SouthernThai\_TK, the CentralThai\_N, Mon\_C and Mon\_W had high percentages of SA prevalent haplogroups (greater than 30%) while Rhade, Jarai, Suay, Khmer, Mon\_NE and Nyahkur exhibited frequencies lower than 15% (S8 Fig).

## Genetic relationships among populations

Shared haplotypes within populations indicate smaller population size due to relatedness among individuals, whereas shared haplotypes between populations could reflect recent shared ancestry or contact. We observe some shared mtDNA haplotypes within the southernThai\_AN but not in southernThai\_TK, but some sharing between southernThai\_TK and Khmer from Cambodia (Khmer\_C2) (Fig 2a). No SA group shared mtDNA haplotypes with SEA/EA groups (Fig 2a).

Genetic relationships between populations can be inferred by measuring genetic distances. Pairwise  $\Phi_{st}$  genetic distances based on mtDNA showed no significant differences between the



**Fig 2. Frequency of shared mtDNA (a) and MSY (b) haplotypes within and between populations.** The font color of the population name indicates language family and the colored bar at the top indicates geographic region; the heat plot shows the frequency of shared types according to the key on the right. The new populations analyzed in this study are indicated by stars.

<https://doi.org/10.1371/journal.pone.0291547.g002>



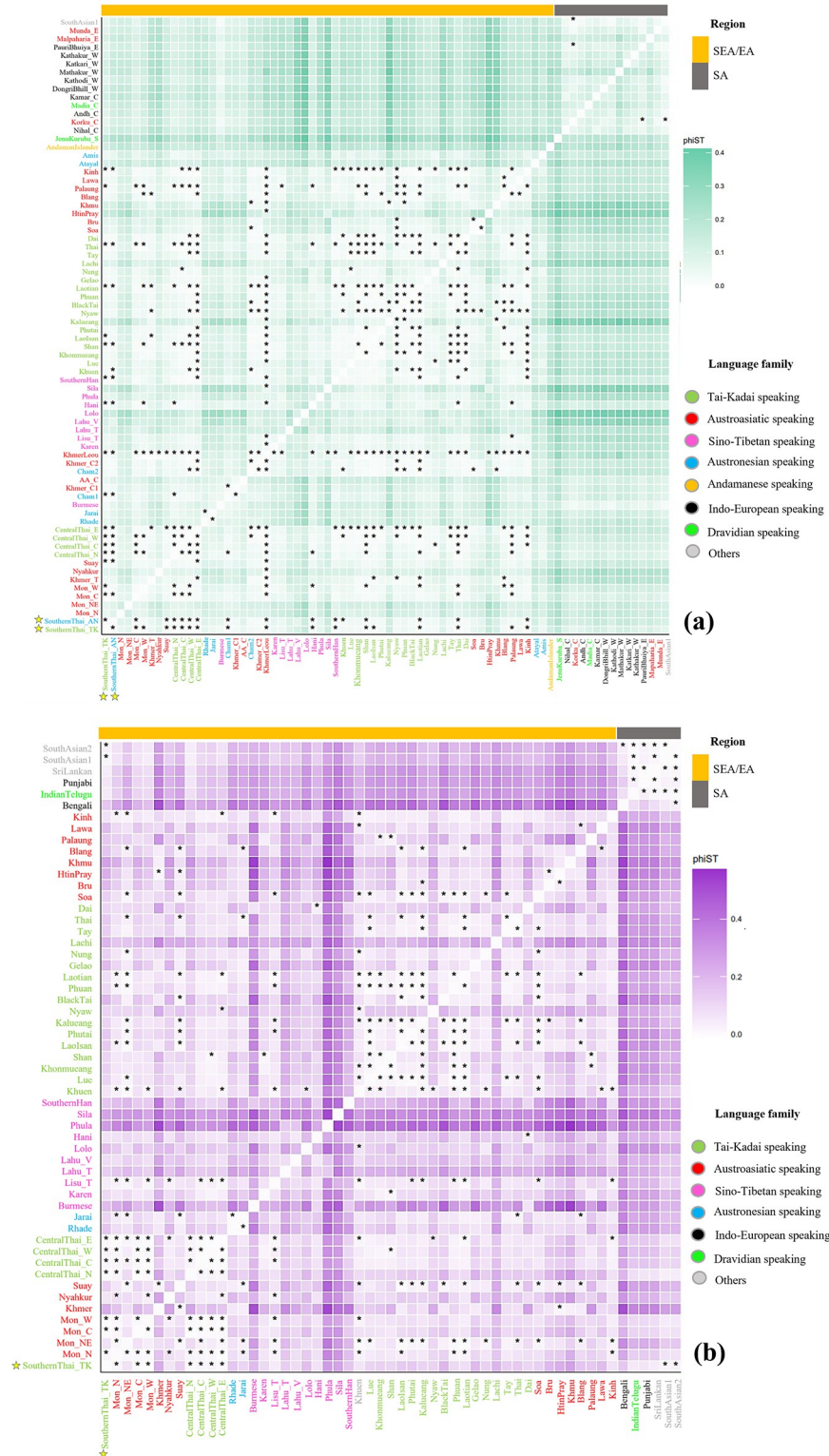
two southern Thai populations, and between these and several other populations, such as Mon\_C, all central Thais and Shan from Thailand, Cham1, Hani, Thai and Kinh from Vietnam, Cambodian KhmerLeou, southern Chinese Han, and Laotian populations, as shown in Fig 3a. In general, SA populations exhibited smaller genetic distances among them and larger genetic distances to most populations from SEA. However, southern Thais, central Thais, Mon, Burmese, Cambodian, and AN-speaking Rhade, Jarai, and Cham from Vietnam showed smaller distances to SA groups than to other SEA populations (Fig 3a).

Interestingly, AN-speaking Amis and Atayal were genetically different from AN-speaking groups from Vietnam and also from other SEA populations (Fig 3a). A multidimensional scaling (MDS) plot was employed to visualize further genetic relationships based on the  $\Phi_{st}$  distance matrix. The MDS plots based on three dimensions indicates genetic separation between SA and SEA/EA populations with some SEA populations, i.e., southern Thais, some central Thais and Mon groups in Thailand, Burmese, Cambodians and Cham1, Jarai and Rhade from Vietnam between these two clusters, suggesting an admixed genetic structure (Fig 4a and 4b).

For the MSY, southernThai\_TK shared haplotypes with centralThai\_E, Lao Isan and AA-speaking Blang from northern Thailand (Fig 2b). In general, we observed more shared haplotypes between AA- and TK-speaking groups in Thailand than between any other groups. SA populations shared more haplotypes among themselves than with SEA groups; only IndianTegu shared haplotypes with centralThai\_C. The heat plot of  $\Phi_{st}$  genetic distances also supports non-significant differences between southernThai\_TK and all central Thai populations (Fig 3b). In contrast to mtDNA, the southernThai\_TK also exhibit genetic similarity with almost all Mon populations, except Northeastern Thai Mon (Mon\_NE), and two South Asian populations. The SA populations showed genetic similarities among themselves and all of them differed significantly from SEA populations (except the two SA populations with the southernThai\_TK, as noted above). The MDS plots based on  $\Phi_{st}$  values show genetic heterogeneity of SEA Sino-Tibetan-speaking populations; they are also differentiated from a cloud of SA populations and a cloud of other SEA populations. The southernThai\_TK, some Mon and central Thai populations, Burmese, Phula, Vietnamese Lahu and Thai Lahu are positioned closer to the SA cloud than are other SEA populations (Fig 5a and 5b). In contrast to the mtDNA results, Jarai and Rhade had more affinity with SEA AA-speaking groups than with SA groups. The overall results from the genetic distance analyses indicate a contrasting pattern between maternal and paternal lineages in some SEA populations. For example, SouthernThai\_TK, CentralThai\_N, Nyahkur, and Mon\_C tend to show paternal genetic relationships closer to SA than do the maternal lineages in these groups (Fig 5), while Rhade, Jarai, and Mon\_NE showed the opposite pattern (Fig 4).

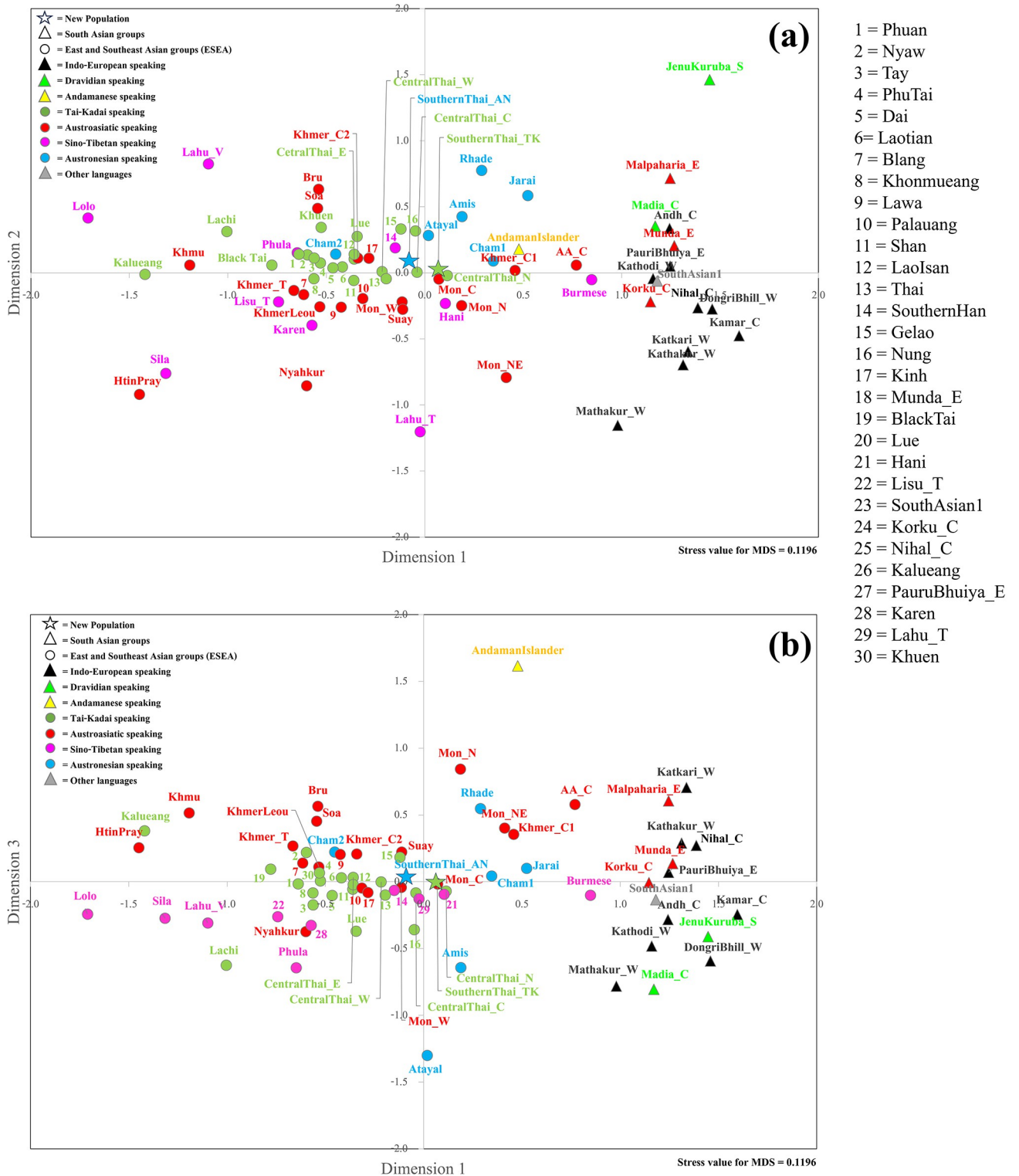
### Genetic contribution of South Asian to Southeast Asian populations

The results based on genetic lineages and genetic distance analyses indicate that southern Thais and several MSEA populations exhibit genetic relatedness to both SEA and SA populations, with some exhibiting contrasting mtDNA vs. MSY patterns, e.g. Mon, Nyahkur, Central Thai, Khmer, Burmese, Cham, Jarai and Rhade. All results support SA genetic influences in those groups that are consistent with genome-wide studies [15, 16]. Assuming that these results are indeed indicative of admixture, we then used a model-based approach to estimate the proportion of SA and SEA ancestry in these admixed groups. We used a simple admixture model consisting of two parental groups, East and Southeast Asian (ESEA) and South Asian (SA) (S2 and S6 Tables), using the coalescent-based method implemented in Admix 2.0 [67, 68]. The ESEA parental groups were chosen based on the SA admixture results from previous studies, which reported no significant SA-related components [15, 16] and based on



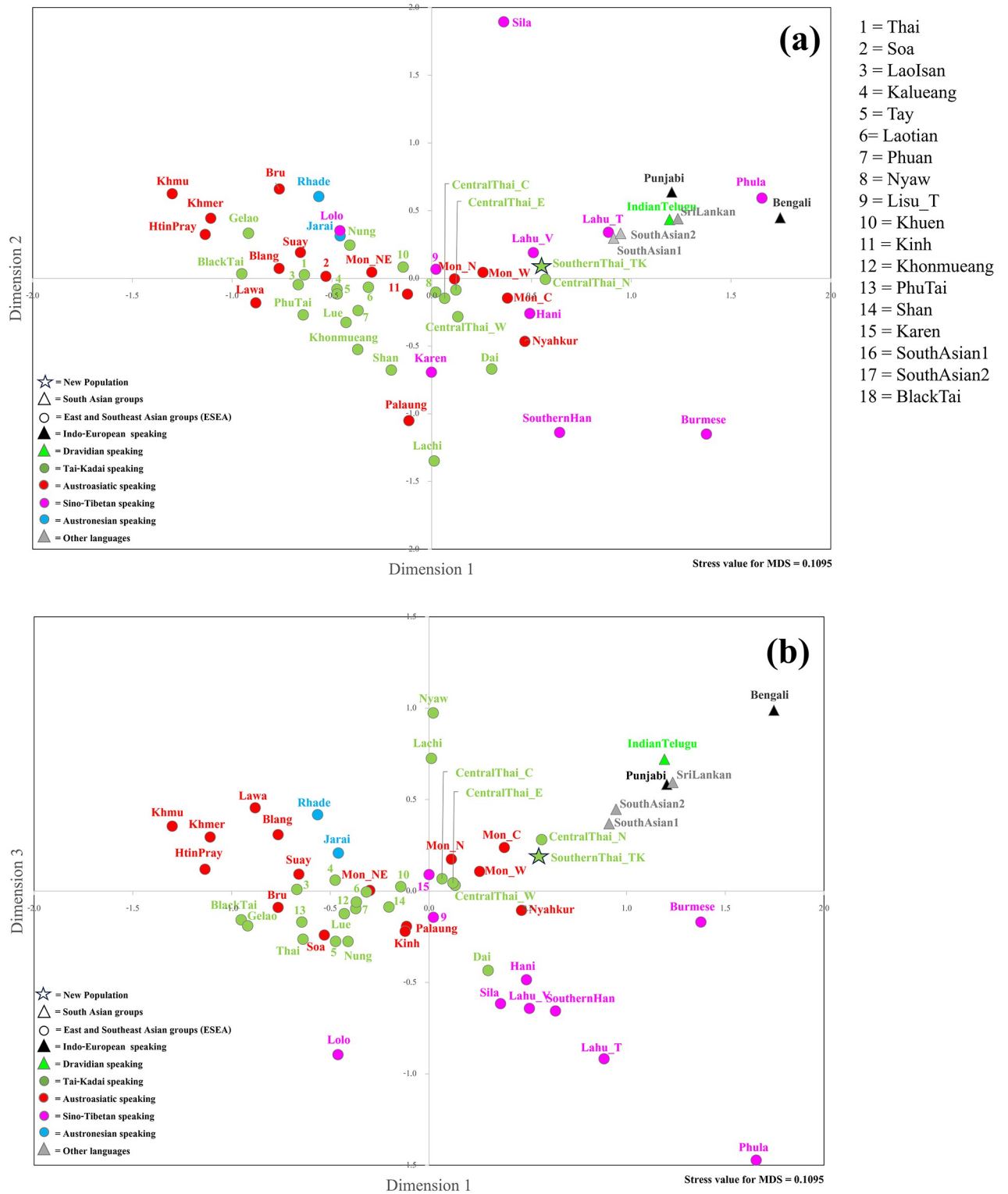
**Fig 3. Heat plot of  $\Phi_{st}$  values based on mtDNA (a) and MSY (b) haplotypes. The “\*” symbol indicates  $\Phi_{st}$  values that are not significantly different from zero ( $P > 0.05$ ). See legend of Fig 2 for other details.**

<https://doi.org/10.1371/journal.pone.0291547.g003>



**Fig 4. The three-dimensional MDS plot based on the  $\Phi_{st}$  distance matrix of 74 populations for mtDNA. The stress value is 0.1196.**

<https://doi.org/10.1371/journal.pone.0291547.g004>

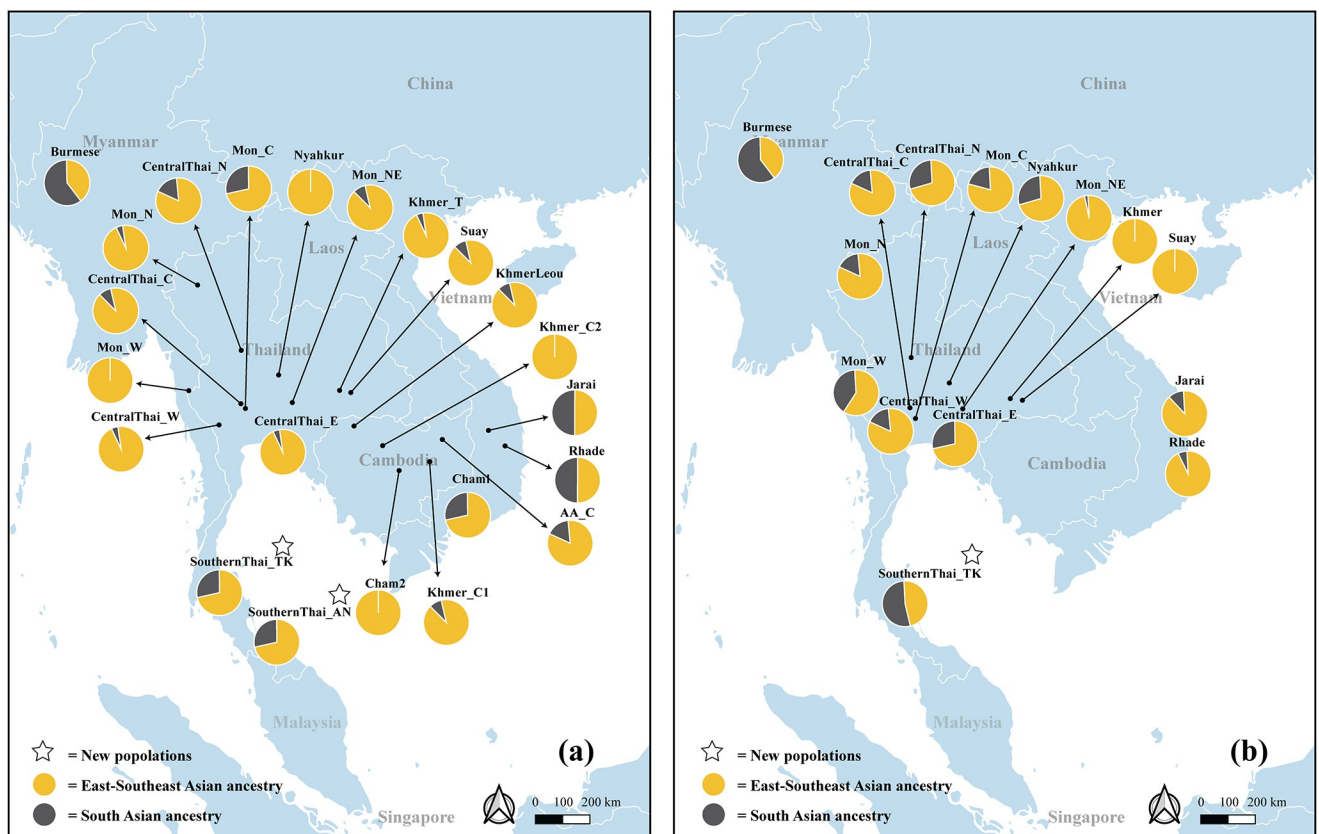


**Fig 5. The three-dimensional MDS plot based on the  $\Phi_{st}$  distance matrix of 55 populations for MSY sequences. The stress value is 0.1095.**

<https://doi.org/10.1371/journal.pone.0291547.g005>

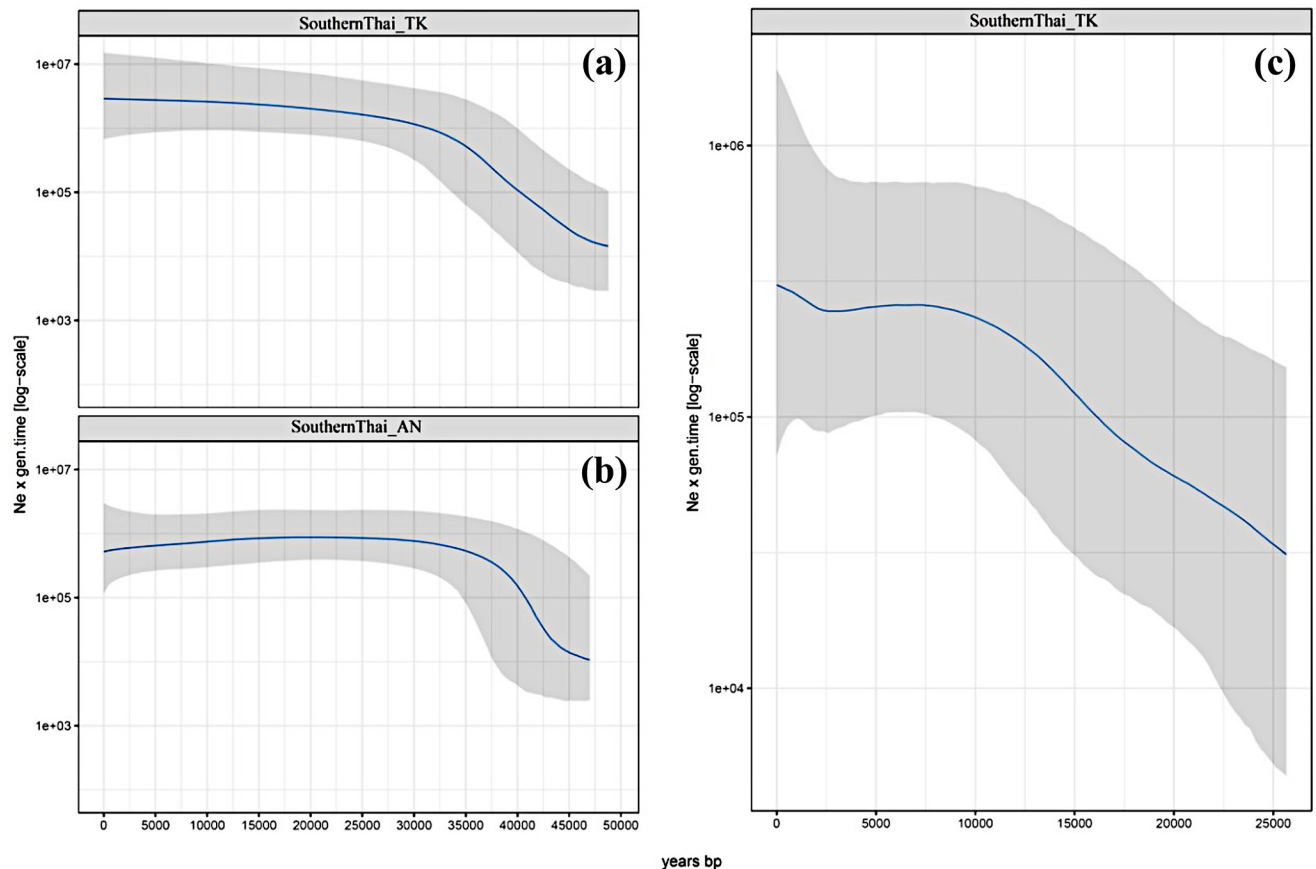
haplogroup distributions ESEA surrogates harbor very few SA lineages and SA source has very few (if any) ESEA lineages (S4 and S8 Figs). According to historical records and archaeological evidence as well as recent ancient DNA studies [77], the interaction between SA and MSEA started at most 2 kya; we therefore considered admixture times starting 2 kya. We performed several analyses using different parameters, i.e., admixture times (2000, 1000, 500, and 0 ya) and mutation rates (0,  $1.67 \times 10^{-8}$  and  $2.67 \times 10^{-8}$  for mtDNA and 0 and  $8.71 \times 10^{-10}$  for MSY) (S7 Table) [69–71]. Because admixture results were consistent in all analyses (S10 and S11 Tables), we consider admixture results at 2,000 years since the admixture event occurred with mutation rates of  $1.67 \times 10^{-8}$  for mtDNA and  $8.71 \times 10^{-10}$  for the MSY.

The maternal contribution from SA to SEA was highest in Burmese (61.4%), Rhade (50.0%), and Jarai (49.3%), while levels of SA ancestry in SouthernThai\_TK, SouthernThai\_AN, Mon\_C, CentralThai\_N, AA\_C and Vietnamese Cham1 ranged from ~15–30%. Other populations showed SA contributions lower than 15%, and no SA contribution was detected in Khmer\_C2, Cham2, Nyahkur and Mon\_W (Fig 6a). For the MSY (Fig 6b), Burmese and SouthernThai\_TK had the highest estimated SA ancestry (59.4% and 54.32%) while Mon\_W had 38.83% SA-related ancestry. Mon\_N, Mon\_C, Nyahkur, and all Central Thai populations exhibited SA admixture ranging from ~15% to ~30%. The Jarai, Rhade and Mon\_NE had low levels of SA admixture (~2–10%) while Suay and Khmer showed no detectable SA admixture. As discussed in more detail below, we thus find contrasting patterns of



**Fig 6. Pie charts of South Asian ancestry in Southeast Asian populations.** (a) mtDNA, 22 populations; (b) MSY, 15 populations. The map was generated using the Quantum GIS, QGIS Development Team (2023), QGIS Geographic Information System. Open Source Geospatial Foundation (<https://www.qgis.org>) and was not taken from another source. and boundaries were adapted and modified from <https://public.opendatasoft.com/explore/dataset/worldadministrative-boundaries/export/>.

<https://doi.org/10.1371/journal.pone.0291547.g006>



**Fig 7. The BSPs based on mtDNA for the SouthernThai\_TK (a) and SouthernThai\_AN (b), and based on MSY for the SouthernThai\_TK (c).** Blue lines are the median estimated effective population size (y axis) through time from the present in years (x axis). The 95% highest posterior density limits are indicated by light grey shading.

<https://doi.org/10.1371/journal.pone.0291547.g007>

mtDNA vs MSY SA admixture in practically all populations, except for Burmese, who have the highest level of SA ancestry for both markers.

### The effective population size change over time

The Bayesian skyline plots (BSPs) of population size change ( $N_e$ ) over time based on mtDNA exhibit similar patterns for both Southern Thai populations newly sequenced here:  $N_e$  increased from 45 kya to 30 kya and then was stable until the present day (Fig 7a and 7b). For the MSY, the  $N_e$  of SouthernThai\_TK increased from 25 kya until ~7.5 kya, and then remained constant until a slight increase is observed at ~2.5 kya (Fig 7c). The pattern of mtDNA  $N_e$  of Southern Thai groups is similar to other Central Thai populations [27], whereas the MSY  $N_e$  of southernThai\_TK was unique in that no reduction in population size was detected throughout their demographic history, while for all TK groups a population size decline was observed between ~5–2.5 kya [23]. However, the size increase in SouthernThai\_TK ~3–2 kya, is characteristic of other TK populations [23].

### Discussion

Previous studies based on autosomal markers revealed unique aspects of southern Thai populations compared to other Thai populations; in particular, the presence of South

Asian-associated ancestry [15]. Although sex-biased admixture of the human populations can be revealed by genomic data based on comparing X chromosomal and autosomal variation [78–80], there are some limitations of using genome-wide data to infer sex-biased admixture, e.g. when there is an extremely low ancestry proportion, low number of X chromosomal SNPs, and potential differences in ascertainment bias between the X chromosomal and autosomal SNPs on a genotyping array [78]. Studies of mtDNA and MSY variation provide a direct assessment of sex-biased admixture, and moreover these uniparental markers can also provide additional information on specific or new genetic lineages. Here, we analyze new complete mtDNA genomes and ~2.3 mB of the MSY in two southern Thai populations to reveal genetic lineages, demographic changes and their genetic structure based on sequencing data, and we additionally analyze sex-biased admixture involving South Asian ancestry in several other MSEA populations, including two southern Thai groups, that have evidence of South Asian ancestry in genome-wide data [15, 16].

While all of the MSY lineages found in our study have been found previously in MSEA, we did find two novel mtDNA haplogroups, E1a1a1 and Q3, in SouthernThai\_AN; these two haplogroups have not been reported in MSEA. Haplogroup Q3 is particularly interesting, as the origin of haplogroup Q and sublineages has been placed in Northern Sahul (highland and coastal New Guinea and Near Oceania) [54], and it has not been reported previously from MSEA or western ISEA. The network of haplogroup Q3 sequences showed that the most closely related sequence to the southernThai\_AN Q3 sequence comes from East Timor (S2 Fig). A recent study of ancient DNA from Wallacea found evidence for contact between MSEA and the Nusa Tenggara Islands of southern Wallacea (which include East Timor) [81]; while it is tempting to speculate that the Q3 sequence may also reflect such contact, it could also reflect other episodes of contact, as the Timor area was suggested to be a secondary pathway that modern humans used to travel between Sundaland and Sahul after initial colonization [82].

Haplogroup E1a1a1 that is abundant in Taiwan and ISEA [37, 83]. The network of haplogroup E1a1a1 sequences showed a star-like structure (S3 Fig), suggesting lineage expansion of E1a1a1 in the ISEA that is probably associated with the spread of Austronesian languages. In addition, some of the SEA/EA mtDNA haplogroups in Southern Thais, e.g., B4a1a and M7c1c3 are present elsewhere mainly in aboriginal Taiwanese and ISEA populations [84]. The sharing of these recent sublineages indicates contact between MSEA and ISEA, which has not been observed previously from Thai/Lao mtDNA data [23, 24, 26]. However, the AN-speaking southern Thai from this study exhibited closer genetic relatedness to MSEA populations than to the AN-speaking Amis and Atayal from Taiwan (Fig 3a). While the AN-speaking Cham populations from Vietnam and Cambodia are similar to SouthernThai\_AN in their relationships with other populations (Fig 3a), the AN-speaking Jarai and Rhade from Vietnam exhibit a distinct maternal genetic structure. The ancestors of AN and TK groups were thought to originate in Southeast China ~6–5 kya and then they split. TK ancestors moved southward to MSEA whereas AN ancestors entered Taiwan and spread out of Taiwan to ISEA ~4–3 kya, thus Taiwan is likely the original home for AN groups [37, 85, 86]. Previous mtDNA studies of AN speaking Cham, Jarai and Rhade suggested that cultural diffusion played an important role in shaping the genetic structure of those groups [25, 38, 39], while a genome-wide study supported both migration and cultural diffusion [87]. Here, we suggest that the Southern Thai\_AN language was introduced largely by cultural diffusion, but the mtDNA haplogroup profile does indicate some links with AN-speaking people from ISEA or alternatively, the southernThai\_AN might reflect a migration of AN speakers to southern Thailand, followed by extensive admixture with local populations.

Interactions between SA and MSEA probably occurred since the late Bronze Age to Iron Age and influenced the formation of early states or civilizations, e.g. Pyu, Funan, Champa, Dvaravati, and Langkasuka [11, 14]; a recent ancient DNA study documented a high genetic contribution (~40–50%) from South Asia in prehistoric people who lived during the early period of Funan, one of the earliest states in MSEA in present-day Cambodia [77]. The later expansion of Indian culture in SEA might further be continued through the historical period by the activity of additional Indian immigrants [88], which could affect the overall demographic structure of MSEA populations. Previous genetic dating of the SA admixture in most present-day MSEA populations ranged during the historical period from ~400 to ~800 ya [15, 16] with the older dates in Nyahkur, Khmer, Jarai and Rhade [16]; these results support significant historical South Asian interaction in MSEA. Although both autosomal DNA studies of ancient samples and present-day people provided new information on the level of SA admixture in MSEA population and admixture dates that started since the Iron Age through the historical period, sex-specific demographic histories are more difficult to explore with autosomal SNPs.

Several analyses based on haplogroup frequency (S4–S8 Figs), and genetic distance (Figs 3–5) indicated the SA ancestry in several Thai and MSEA populations for both mtDNA and MSY. Admixture analysis based on model-based estimates also supported SA proportions in these MSEA populations (S10 and S11 Tables) with heterogeneous patterns of sex-biased admixtures that were detected among groups (Fig 6). We find less sex-biased admixture in Burmese and they have high SA proportions (~60%) for both mtDNA and MSY (Fig 6). Myanmar shares a border with India to the West, which would allow multiple migrations and promote substantial gene flow. The other MSEA groups exhibited stronger sex-biased admixtures. In general, based on mtDNA, almost all of the AA- and AN- speaking groups are genetically closer to SA than TK groups, and the SouthernThai\_TK was notably different from CentralThai groups in being closer to SA groups (Fig 4, S4 and S5 Figs). The AA-speaking Mon\_W is an exception since they have no mtDNA SA ancestry but elevated MSY SA ancestry (~40%) (S10 Table). SA admixture in Mon would have resulted from Mon-Burmese interactions. Mon, Burmese and Karen have been reported to share a mutation in the G6PD gene [89], which supports their close genetic affinities.

However, there are heterogeneous sex-biased admixture patterns among Mon populations in Thailand. Historically, multiple (at least nine) migrations of the Mon groups from Myanmar to Thailand occurred during the 16<sup>th</sup> to 19<sup>th</sup> centuries A.D. [90, 91]; different spatial and temporal migrations could promote various interactions between either Mon and TK groups or new Mon and old settled Mon groups, which then led to heterogeneous sex-biased admixture patterns and contrasting maternal and paternal genetic variation [17, 24, 25].

Interestingly, the AN-speaking groups in Vietnam (e.g. Jarai and Rhade) also exhibited a strong sex-biased SA admixture pattern. The Jarai and Rhade showed almost 50% SA maternal ancestry, but less than 10% SA paternal ancestry (Fig 6 and S10 and S11 Tables). Jarai and Rhade live in the highland area and are matrilocal, but some sources indicate a previous connection with Cham, the neighboring lowland population in Vietnam [92]. This connection is also supported by the results from a recent genome-wide data analysis [87]. There is evidence that the ancestors of Vietnamese AN groups might have had matrilineal societies [93] and might have interacted with each other and also with other MSEA groups, e.g., Khmer, Jarai and Rhade probably received SA components from other lowland populations and subsequently other factors, e.g., residence pattern (matrilocal in this case) could have influenced patterns of genetic variation.

In conclusion, we have expanded the study of high-resolution mtDNA and MSY sequences in MSEA with new data from Southern Thai populations, although a limitation



of our study is that the sample sizes are relatively small. However, the overall agreement with previous studies suggests that our results are indeed representative of Southern Thais. We explored additional insights into the genetic histories of MSEA populations; the mtDNA haplogroups E1a1a1, Q3, B4a1a and M7c1c3 that were found in Southern Thai populations reflect MSEA and ISEA connections. Both southern Thai populations exhibited close genetic relatedness with AA-speaking Mon and Central Thais, and SA ancestry was one of the shared characteristics. Consistent with previous autosomal DNA studies, evidence of SA ancestry was found in some Thai and other MSEA groups based on analyses of specific lineages, haplogroup frequencies, and genetic distance analyses. The model-based estimates of admixture proportions confirm these results for both mtDNA and MSY. However, patterns of sex-biased admixture vary greatly among groups. Burmese show little sex bias and have elevated SA ancestries for both mtDNA and MSY and this probably reflects their close position to India. The other groups exhibit strong sex bias—for example, Jarai and Rhade in Vietnam, which exhibit more SA mtDNA than MSY ancestry, probably influenced by population interactions and cultural practices. In sum, new mtDNA genetic lineages (e.g., Q3 and E1a1a1) and evidence for sex-biased South Asian admixture in several MSEA populations from our study support the usefulness of high-resolution uni-parental markers to reconstruct aspects of the genetic history of populations that have not been previously revealed by autosomal markers.

## Supporting information

**S1 Fig. Bar plot of percentages of each major mtDNA haplogroup in four groups of populations: SouthernThai\_TK, SouthernThai\_AN, Southeast Asian/East Asian (SEA/EA) and South Asian (SA).** Data from 56 SEA/EA populations, and 16 SA populations.

(TIF)

**S2 Fig. Median-joining network based on 32 mtDNA genomes of haplogroup Q3 from Southern Thailand, East Timor, and Papua New Guinea.** Number of substitutions are shown on the lines connecting the nodes.

(TIF)

**S3 Fig. Median-joining network based on 72 mtDNA genomes of haplogroup E1a1a1 from Southern Thailand, Taiwan, the Philippines, Indonesia and Saudi Arabia.** Number of substitutions are shown on the lines connecting the nodes.

(TIF)

**S4 Fig. Bar plot of percentages of three categories of mtDNA haplogroups in 74 populations: Southeast Asian/East Asian (SEA/EA) prevalent, South Asian (SA) prevalent and other.**

(TIF)

**S5 Fig. Correspondence analysis (CA) plot based on mtDNA haplogroup frequencies in 74 populations.**

(PDF)

**S6 Fig. Bar plot of percentages of each major MSY haplogroup in three groups of populations: SouthernThai\_TK, Southeast Asian/East Asian (SEA/EA) and South Asian (SA).** Data from 55 SEA/EA populations and 16 SA populations.

(TIF)

**S7 Fig. Correspondence analysis (CA) plot based on MSY haplogroup frequency of 55 populations.**

(PDF)

**S8 Fig. Bar plot of percentages of three categories of MSY haplogroups: Southeast Asian/East Asian (SEA/EA) prevalent, South Asian (SA) prevalent, and other haplogroups, in 55 populations.**

(TIF)

**S1 Table. Sequencing coverage and mtDNA haplogroup of each individual.**

(XLSX)

**S2 Table. General information on the studied populations and mtDNA genetic diversity values.**

(XLSX)

**S3 Table. General information on additional mtDNA sequences of haplogroup Q3, used in the network analysis.**

(XLSX)

**S4 Table. General information on additional mtDNA sequences of haplogroup E1a1a1, used in the network analysis.**

(XLSX)

**S5 Table. Sequencing coverage and MSY haplogroup of each SouthernThai\_TK individual.**

(XLSX)

**S6 Table. General information on the studied populations and MSY genetic diversity values.**

(XLSX)

**S7 Table. Parameters in the admixture analysis.**

(XLSX)

**S8 Table. MtDNA haplogroup frequencies.**

(XLSX)

**S9 Table. MSY haplogroup frequencies.**

(XLSX)

**S10 Table. Estimated admixture proportions in admixed populations based on mtDNA sequences.**

(XLSX)

**S11 Table. Estimated admixture proportions in admixed populations based on MSY sequences.**

(XLSX)

## Acknowledgments

We would like to thank all participants who donated biological samples and Sukrit Sangkhano and Suparat Srithawong for coordinating sample collections. We greatly appreciate Roland Schröder and Alexander Hübner for technical assistance. We thank the Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany, for hosting laboratory facilities and the Max Planck Society for research support.

## Author Contributions

**Conceptualization:** Wipada Woravatin, Mark Stoneking, Wibhu Kutanan.

**Formal analysis:** Wipada Woravatin, Leonardo Arias.

**Funding acquisition:** Wibhu Kutanan.

**Investigation:** Wipada Woravatin.

**Methodology:** Wipada Woravatin, Metawee Srikumool, Jatupol Kampuansai, Leonardo Arias.

**Supervision:** Mark Stoneking, Wibhu Kutanan.

**Writing – original draft:** Wipada Woravatin, Wibhu Kutanan.

**Writing – review & editing:** Mark Stoneking, Metawee Srikumool, Jatupol Kampuansai, Leonardo Arias, Wibhu Kutanan.

## References

1. Aghakhanian F, Yunus Y, Naidu R, et al. Unravelling the genetic history of Negritos and indigenous populations of Southeast Asia. *Genome Biol Evol.* 2015; 7:1206–1215. <https://doi.org/10.1093/gbe/evv065> PMID: 25877615
2. McColl H, Racimo F, Vinner L, et al. The prehistoric peopling of Southeast Asia. *Science.* 2018; 361:88–92. <https://doi.org/10.1126/science.aat3628> PMID: 29976827
3. Göllner T, Larena M, Kutanan W, et al. Unveiling the Genetic History of the Maniq, a Primary Hunter-Gatherer Society. *Genome Biology and Evolution.* 2022; 14:evac021. <https://doi.org/10.1093/gbe/evac021> PMID: 35143674
4. Jinam TA, Hong L-C, Phipps ME, 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; 29:3513–3527. <https://doi.org/10.1093/molbev/mss169> PMID: 22729749
5. Kutanan W, Kampuansai J, Changmai P, et al. Contrasting maternal and paternal genetic variation of hunter-gatherer groups in Thailand. *Sci Rep.* 2018; 8:1536. <https://doi.org/10.1038/s41598-018-20020-0> PMID: 29367746
6. Oota H, Kurosaki K, Pookajorn S, et al. Genetic Study of the Paleolithic and Neolithic Southeast Asians. *Human Biology.* 2001; 73:225–231. <https://doi.org/10.1353/hub.2001.0023> PMID: 11446426
7. Anderson DD. Lang Rongrien Rockshelter: A Pleistocene, Early Holocene Archaeological Site from Krabi, Southwestern Thailand. Philadelphia: University of Pennsylvania Museum of Archaeology and Anthropology; 1990.
8. Pookajorn S. Preliminary report of excavation at Moh Khiew cave, Krabi province, Sakai cave, Trang province and ethnoarchaeological research of hunter-gatherer group so-called “Sakai or Semang” at Trang province: the Hoabinhian Research Project in Thailand. Bangkok: Silpakorn University; 1991.
9. Pookajorn S. Final report of excavations at Moh-Khiew Cave, Krabi Province, Sakai Cave, Trang Province, and ethnoarchaeological research of hunter-gatherer group so-called Mani or Sakai or Orang Asli at Trang Province: the Hoabinhian Research Project in Thailand. Bangkok: Silpakorn University; 1994.
10. Auetrakulvit P, Forestier H, Khaokhiew C, et al. New Excavation at Moh Khiew Site (southern Thailand). *Crossing Borders in Southeast Asian Archaeology.* 2012;62–74.
11. Higham C. Early Mainland Southeast Asia: From First Humans to Angkor. Bangkok: River Books Press; 2014.
12. Enfield NJ, Comrie B, editors. Languages of Mainland Southeast Asia: The State of the Art. 1st ed. De Gruyter Mouton; 2015.
13. Lipson M, Cheronet O, Mallick S, et al. Ancient genomes document multiple waves of migration in Southeast Asian prehistory. *Science.* 2018; 361:92–95. <https://doi.org/10.1126/science.aat3188> PMID: 29773666
14. Coedes G. The Indianized states of south-east Asia. East-West Center Press; 1968.
15. Kutanan W, Liu D, Kampuansai J, et al. Reconstructing the Human Genetic History of Mainland Southeast Asia: Insights from Genome-Wide Data from Thailand and Laos. *Molecular Biology and Evolution.* 2021; 38:3459–3477. <https://doi.org/10.1093/molbev/msab124> PMID: 33905512

16. Changmai P, Jaisamut K, Kampuansai J, et al. Indian genetic heritage in Southeast Asian populations. *PLoS Genet.* 2022; 18:e1010036. <https://doi.org/10.1371/journal.pgen.1010036> PMID: 35176016
17. Wen B, Xie X, Gao S, et al. Analyses of Genetic Structure of Tibeto-Burman Populations Reveals Sex-Biased Admixture in Southern Tibeto-Burmans. *Am J Hum Genet.* 2004; 74:856–865. <https://doi.org/10.1086/386292> PMID: 15042512
18. Wilkins JF. Unraveling male and female histories from human genetic data. *Curr Opin Genet Dev.* 2006; 16:611–617. <https://doi.org/10.1016/j.gde.2006.10.004> PMID: 17067791
19. Arias L, Schröder R, Hübner A, et al. Cultural Innovations Influence Patterns of Genetic Diversity in Northwestern Amazonia. *Molecular Biology and Evolution.* 2018; 35:2719–2735. <https://doi.org/10.1093/molbev/msy169> PMID: 30169717
20. Bajić V, Barbieri C, Hübner A, et al. Genetic structure and sex-biased gene flow in the history of southern African populations. *Am J Phys Anthropol.* 2018; 167:656–671. <https://doi.org/10.1002/ajpa.23694> PMID: 30192370
21. Oota H, Settheetham-Ishida W, Tiwawech D, et al. Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence. *Nat Genet.* 2001; 29:20–21. <https://doi.org/10.1038/ng711> PMID: 11528385
22. Besaggio D, Fuselli S, Srikumool M, et al. Genetic variation in Northern Thailand Hill Tribes: origins and relationships with social structure and linguistic differences. *BMC Evol Biol.* 2007; 7:S12. <https://doi.org/10.1186/1471-2148-7-S2-S12> PMID: 17767728
23. Kutanan W, Kampuansai J, Srikumool M, et al. Contrasting Paternal and Maternal Genetic Histories of Thai and Lao Populations. *Molecular Biology and Evolution.* 2019; 36:1490–1506. <https://doi.org/10.1093/molbev/msz083> PMID: 30980085
24. Kutanan W, Shoocongdej R, Srikumool M, 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; 28:1563–1579. <https://doi.org/10.1038/s41431-020-0693-x> PMID: 32690935
25. Macholdt E, Arias L, Duong NT, et al. The paternal and maternal genetic history of Vietnamese populations. *Eur J Hum Genet.* 2020; 28:636–645. <https://doi.org/10.1038/s41431-019-0557-4> PMID: 31827276
26. Kutanan W, Kampuansai J, Srikumool M, 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. <https://doi.org/10.1007/s00439-016-1742-y> PMID: 27837350
27. Kutanan W, Kampuansai J, Brunelli A, et al. New insights from Thailand into the maternal genetic history of Mainland Southeast Asia. *Eur J Hum Genet.* 2018; 26:898–911. <https://doi.org/10.1038/s41431-018-0113-7> PMID: 29483671
28. Dancause KN, Chan CW, Arunotai NH, et al. Origins of the Moken Sea Gypsies inferred from mitochondrial hypervariable region and whole genome sequences. *J Hum Genet.* 2009; 54:86–93. <https://doi.org/10.1038/jhg.2008.12> PMID: 19158811
29. Seetaraso T, Kutanan W, Kampuansai J, et al. Unique Genetic Structure of Y-chromosomal Lineage of the Moken from the Andaman Sea of Thailand. 2020; 19:1066–1079. <http://cmuir.cmu.ac.th/jspui/handle/6653943832/77733>
30. Srikumool M, Srithawong S, Muisuk K, et al. Forensic and genetic characterizations of diverse southern Thai populations based on 15 autosomal STRs. *Sci Rep.* 2022; 12:655. <https://doi.org/10.1038/s41598-021-04646-1> PMID: 35027632
31. Meyer M, Kircher M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc.* 2010; 6:1–10. <https://doi.org/10.1101/pdb.prot5448> PMID: 20516186
32. Maricic T, Whitten M, Pääbo S. Multiplexed DNA Sequence Capture of Mitochondrial Genomes Using PCR Products. *PLOS ONE.* 2010; 5:e14004. <https://doi.org/10.1371/journal.pone.0014004> PMID: 21103372
33. Behar DM, van Oven M, Rosset S, et al. A “Copernican” Reassessment of the Human Mitochondrial DNA Tree from its Root. *Am J Hum Genet.* 2012; 90:675–684. <https://doi.org/10.1016/j.ajhg.2012.03.002> PMID: 22482806
34. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013; 30:772–780. <https://doi.org/10.1093/molbev/mst010> PMID: 23329690
35. Renaud G, Stenzel U, Kelso J. LeeHom: Adaptor trimming and merging for Illumina sequencing reads. *Nucleic acids research.* 2014;42. <https://doi.org/10.1093/nar/gku699> PMID: 25100869

36. Renaud G, Stenzel U, Maricic T, et al. deML: robust demultiplexing of Illumina sequences using a likelihood-based approach. *Bioinformatics*. 2015; 31:770–772. <https://doi.org/10.1093/bioinformatics/btu719> PMID: 25359895
37. Ko AM-S, Chen C-Y, Fu Q, et al. Early Austronesians: Into and Out Of Taiwan. *The American Journal of Human Genetics*. 2014; 94:426–436. <https://doi.org/10.1016/j.ajhg.2014.02.003> PMID: 24607387
38. Peng M-S, Quang HH, Dang KP, et al. Tracing the Austronesian Footprint in Mainland Southeast Asia: A Perspective from Mitochondrial DNA. *Molecular Biology and Evolution*. 2010; 27:2417–2430. <https://doi.org/10.1093/molbev/msq131> PMID: 20513740
39. Duong NT, Macholdt E, Ton ND, et al. Complete human mtDNA genome sequences from Vietnam and the phylogeography of Mainland Southeast Asia. *Sci Rep*. 2018; 8:11651. <https://doi.org/10.1038/s41598-018-29989-0> PMID: 30076323
40. Sun C, Kong Q-P, Zhang Y-P. The role of climate in human mitochondrial DNA evolution: A reappraisal. *Genomics*. 2007; 89:338–342. <https://doi.org/10.1016/j.ygeno.2006.11.005> PMID: 17188837
41. Chandrasekar A, Kumar S, Sreenath J, et al. Updating Phylogeny of Mitochondrial DNA Macrohaplogroup M in India: Dispersal of Modern Human in South Asian Corridor. *PLoS One*. 2009; 4:e7447. <https://doi.org/10.1371/journal.pone.0007447> PMID: 19823670
42. Zhang X, Qi X, Yang Z, et al. Analysis of mitochondrial genome diversity identifies new and ancient maternal lineages in Cambodian aborigines. *Nat Commun*. 2013; 4:2599. <https://doi.org/10.1038/ncomms3599> PMID: 24121720
43. Diroma MA, Calabrese C, Simone D, et al. Extraction and annotation of human mitochondrial genomes from 1000 Genomes Whole Exome Sequencing data. *BMC Genomics*. 2014; 15:S2. <https://doi.org/10.1186/1471-2164-15-S3-S2> PMID: 25077682
44. Summerer M, Horst J, Erhart G, et al. Large-scale mitochondrial DNA analysis in Southeast Asia reveals evolutionary effects of cultural isolation in the multi-ethnic population of Myanmar. *BMC Evolutionary Biology*. 2014; 14:17. <https://doi.org/10.1186/1471-2148-14-17> PMID: 24467713
45. Li Y-C, Wang H-W, Tian J-Y, et al. Ancient inland human dispersals from Myanmar into interior East Asia since the Late Pleistocene. *Sci Rep*. 2015; 5:9473. <https://doi.org/10.1038/srep09473> PMID: 25826227
46. Kloss-Brandstätter A, Summerer M, Horst D, et al. An in-depth analysis of the mitochondrial phylogenetic landscape of Cambodia. *Sci Rep*. 2021; 11:10816. <https://doi.org/10.1038/s41598-021-90145-2> PMID: 34031453
47. Thangaraj K, Chaubey G, Kivisild T, et al. Reconstructing the origin of Andaman Islanders. *Science*. 2005; 308:996. <https://doi.org/10.1126/science.1109987> PMID: 15890876
48. Bandelt HJ, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol*. 1999; 16:37–48. <https://doi.org/10.1093/oxfordjournals.molbev.a026036> PMID: 10331250
49. Ingman M, Gyllenstein U. Mitochondrial genome variation and evolutionary history of Australian and New Guinean aborigines. *Genome Res*. 2003; 13:1600–1606. <https://doi.org/10.1101/gr.686603> PMID: 12840039
50. Friedlaender JS, Friedlaender FR, Hodgson JA, et al. Melanesian mtDNA Complexity. *PLOS ONE*. 2007; 2:e248. <https://doi.org/10.1371/journal.pone.0000248> PMID: 17327912
51. Hartmann A, Thieme M, Nanduri LK, et al. Validation of microarray-based resequencing of 93 worldwide mitochondrial genomes. *Hum Mutat*. 2009; 30:115–122. <https://doi.org/10.1002/humu.20816> PMID: 18623076
52. Lippold S, Xu H, Ko A, et al. Human paternal and maternal demographic histories: insights from high-resolution Y chromosome and mtDNA sequences. *Investigative Genetics*. 2014; 5:13. <https://doi.org/10.1186/2041-2223-5-13> PMID: 25254093
53. Gomes SM, Bodner M, Souto L, et al. Human settlement history between Sunda and Sahul: a focus on East Timor (Timor-Leste) and the Pleistocene mtDNA diversity. *BMC Genomics*. 2015; 16:70. <https://doi.org/10.1186/s12864-014-1201-x> PMID: 25757516
54. Pedro N, Brucato N, Fernandes V, et al. Papuan mitochondrial genomes and the settlement of Sahul. *J Hum Genet*. 2020; 65:875–887. <https://doi.org/10.1038/s10038-020-0781-3> PMID: 32483274
55. Gunnarsdóttir ED, Li M, Bauchet M, et al. High-throughput sequencing of complete human mtDNA genomes from the Philippines. *Genome Res*. 2011; 21:1–11. <https://doi.org/10.1101/gr.107615.110> PMID: 21147912
56. Gunnarsdóttir ED, Nandineni MR, Li M, et al. Larger mitochondrial DNA than Y-chromosome differences between matrilineal and patrilineal groups from Sumatra. *Nat Commun*. 2011; 2:228. <https://doi.org/10.1038/ncomms1235> PMID: 21407194

57. Delfin F, Min-Shan Ko A, Li M, 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; 22:228–237. <https://doi.org/10.1038/ejhg.2013.122> PMID: 23756438
58. Karmin M, Saag L, Vicente M, et al. A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Res.* 2015; 25:459–466. <https://doi.org/10.1101/gr.186684.114> PMID: 25770088
59. Mallick S, Li H, Lipson M, et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature.* 2016; 538:201–206. <https://doi.org/10.1038/nature18964> PMID: 27654912
60. Poznik GD, Xue Y, Mendez FL, et al. Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. *Nat Genet.* 2016; 48:593–599. <https://doi.org/10.1038/ng.3559> PMID: 27111036
61. Kloss-Brandstätter A, Pacher D, Schönherr S, et al. HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Hum Mutat.* 2011; 32:25–32. <https://doi.org/10.1002/humu.21382> PMID: 20960467
62. van Oven M, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat.* 2009; 30:E386–394. <https://doi.org/10.1002/humu.20921> PMID: 18853457
63. Vohr SH, Gordon R, Eizenga JM, et al. A phylogenetic approach for haplotype analysis of sequence data from complex mitochondrial mixtures. *Forensic Sci Int Genet.* 2017; 30:93–105. <https://doi.org/10.1016/j.fsigen.2017.05.007> PMID: 28667863
64. Poznik GD. 2016. Identifying Y-chromosome haplogroups in arbitrarily large samples of sequenced or genotyped men, unpublished data. Available from: <https://www.biorxiv.org/content/early/2016/11/19/088716> (last accessed May 8, 2022).
65. 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; 10:564–567. <https://doi.org/10.1111/j.1755-0998.2010.02847.x> PMID: 21565059
66. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. 2016.
67. Bertorelle G, Excoffier L. Inferring admixture proportions from molecular data. *Molecular Biology and Evolution.* 1998; 15:1298–1311. <https://doi.org/10.1093/oxfordjournals.molbev.a025858> PMID: 9787436
68. Dupanloup I, Bertorelle G. Inferring Admixture Proportions from Molecular Data: Extension to Any Number of Parental Populations. *Molecular Biology and Evolution.* 2001; 18:672–675. <https://doi.org/10.1093/oxfordjournals.molbev.a003847> PMID: 11264419
69. Soares P, Ermini L, Thomson N, et al. Correcting for Purifying Selection: An Improved Human Mitochondrial Molecular Clock. *Am J Hum Genet.* 2009; 84:740–759. <https://doi.org/10.1016/j.ajhg.2009.05.001> PMID: 19500773
70. Fu Q, Mitnik A, Johnson PLF, et al. A Revised Timescale for Human Evolution Based on Ancient Mitochondrial Genomes. *Current Biology.* 2013; 23:553–559. <https://doi.org/10.1016/j.cub.2013.02.044> PMID: 23523248
71. Helgason A, Einarsson AW, Guðmundsdóttir VB, et al. The Y-chromosome point mutation rate in humans. *Nat Genet.* 2015; 47:453–457. <https://doi.org/10.1038/ng.3171> PMID: 25807285
72. Drummond AJ, Suchard MA, Xie D, et al. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution.* 2012; 29:1969–1973. <https://doi.org/10.1093/molbev/mss075> PMID: 22367748
73. Darriba D, Taboada GL, Doallo R, et al. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods.* 2012; 9:772–772. <https://doi.org/10.1038/nmeth.2109> PMID: 22847109
74. Baele G, Lemey P, Bedford T, et al. Improving the Accuracy of Demographic and Molecular Clock Model Comparison While Accommodating Phylogenetic Uncertainty. *Molecular Biology and Evolution.* 2012; 29:2157–2167. <https://doi.org/10.1093/molbev/mss084> PMID: 22403239
75. Baele G, Lemey P. Bayesian evolutionary model testing in the phylogenomics era: matching model complexity with computational efficiency. *Bioinformatics.* 2013; 29:1970–1979. <https://doi.org/10.1093/bioinformatics/btt340> PMID: 23766415
76. Marrero P, Abu-Amero KK, Larruga JM, et al. Carriers of human mitochondrial DNA macrohaplogroup M colonized India from southeastern Asia. *BMC Evol Biol.* 2016; 16:246. <https://doi.org/10.1186/s12862-016-0816-8> PMID: 27832758
77. Changmai P, Pinhasi R, Pietrusewsky M, et al. Ancient DNA from Protohistoric Period Cambodia indicates that South Asians admixed with local populations as early as 1st-3rd centuries CE. *Sci Rep.* 2022; 12:22507. <https://doi.org/10.1038/s41598-022-26799-3> PMID: 36581666

78. Tätte K, Pagani L, Pathak AK, et al. The genetic legacy of continental scale admixture in Indian Austro-asiatic speakers. *Sci Rep*. 2019; 9:3818. <https://doi.org/10.1038/s41598-019-40399-8> PMID: 30846778
79. Ongaro L., Molinaro L., Flores, et al. Evaluating the Impact of Sex-Biased Genetic Admixture in the Americas through the Analysis of Haplotype Data. *Genes*. 2021; 12(10):1580. <https://doi.org/10.3390/genes12101580> PMID: 34680976
80. Wu D., Li P. Y., Pan B, et al. Genetic Admixture in the Culturally Unique Peranakan Chinese Population in Southeast Asia. *Mol Biol Evol*. 2021; 38: 4463–4474. <https://doi.org/10.1093/molbev/msab187> PMID: 34152401
81. Oliveira S, Nägele K, Carlhoff S, et al. Ancient genomes from the last three millennia support multiple human dispersals into Wallacea. *Nat Ecol Evol*. 2022; 6:1024–1034. <https://doi.org/10.1038/s41559-022-01775-2> PMID: 35681000
82. Kealy S, Louys J, O'Connor S. Least-cost pathway models indicate northern human dispersal from Sunda to Sahul. *Journal of Human Evolution*. 2018; 125:59–70. <https://doi.org/10.1016/j.jhevol.2018.10.003> PMID: 30502898
83. Soares P, Trejaut JA, Loo J-H, et al. Climate Change and Postglacial Human Dispersals in Southeast Asia. *Molecular Biology and Evolution*. 2008; 25:1209–1218. <https://doi.org/10.1093/molbev/msn068> PMID: 18359946
84. Brandão A, Eng KK, Rito T, et al. Quantifying the legacy of the Chinese Neolithic on the maternal genetic heritage of Taiwan and Island Southeast Asia. *Hum Genet*. 2016; 135:363–376. <https://doi.org/10.1007/s00439-016-1640-3> PMID: 26875094
85. Gray RD, Drummond AJ, Greenhill SJ. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science*. 2009; 323:479–483. <https://doi.org/10.1126/science.1166858> PMID: 19164742
86. Sidwell P. Southeast Asian mainland: linguistic history. In: *The Global Prehistory of Human Migration*, Bellwood P, editors. John Wiley & Sons; 2014. pp. 259–268.
87. Liu D, Duong NT, Ton ND, et al. Extensive Ethnolinguistic Diversity in Vietnam Reflects Multiple Sources of Genetic Diversity. *Molecular Biology and Evolution*. 2020; 37:2503–2519. <https://doi.org/10.1093/molbev/msaa099> PMID: 32344428
88. Mabbett IW. The “Indianization” of Southeast Asia: reflections on the prehistoric sources. *Journal of Southeast Asian studies*. 1977; 8.
89. Nuchprayoon I, Louicharoen C, Charoenvej W. Glucose-6-phosphate dehydrogenase mutations in Mon and Burmese of southern Myanmar. *J Hum Genet*. 2008; 53:48–54. <https://doi.org/10.1007/s10038-007-0217-3> PMID: 18046504
90. Ocharoen S. Mons in Thailand. Thailand Research Fund; 1998.
91. Roy EV. Safe haven: Mon refugees at the capitals of Siam from the 1500s to the 1800s. *The Journal of the Siam Society*. 2010; 98:151–184.
92. Hickey GC. *The Major Ethnic Groups of The South Vietnamese Highlands*. RAND corporation; 1964.
93. Bentley RA, Pradier B, Kyaw AA, et al. Kinship and migration in prehistoric mainland Southeast Asia: An overview of isotopic evidence. *Archaeological Research in Asia*. 2021; 25:100260.