

# Contrasting maternal and paternal genetic variation of hunter-gatherer groups in Thailand

Wibhu Kutanan<sup>1,2\*</sup>, Jatupol Kampuansai<sup>3</sup>, Piya Changmai<sup>4</sup>, Pavel Flegontov<sup>4,5</sup>, Roland Schröder<sup>2</sup>, Enrico Macholdt<sup>2</sup>, Alexander Hübner<sup>2</sup>, Daoroong Kangwanpong<sup>3</sup> and Mark Stoneking<sup>2\*</sup>

<sup>1</sup>Department of Biology, Faculty of Science, Khon Kaen University, Khon Kaen, 40002, Thailand

<sup>2</sup>Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, 04103, Germany

<sup>3</sup>Department of Biology, Faculty of Science, Chiang Mai University, Chiang Mai, 50200, Thailand

<sup>4</sup> Department of Biology and Ecology, Faculty of Science, University of Ostrava, Ostrava, 70103, Czech Republic

<sup>5</sup> Institute of Parasitology, Biology Centre, Czech Academy of Sciences, České Budějovice, 37001, Czech Republic

## \*Corresponding authors:

1. Professor Mark Stoneking, Ph.D.

Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology  
Deutscher Platz 6, D04103 Leipzig, Germany

Tel: +49 341 3550 502; Fax: +49 341 3550 555; E-mail: [stoneking@eva.mpg.de](mailto:stoneking@eva.mpg.de)

2. Assistant Professor Wibhu Kutanan, Ph.D.

Department of Biology, Faculty of Science, Khon Kaen University, Mittapap Road, Khon Kaen, 40002, Thailand

Tel: +66 43 202 531; Fax: + 66 43 202 530; Email: [wibhu@kku.ac.th](mailto:wibhu@kku.ac.th)

## ABSTRACT

The Maniq and Mlabri are the only recorded nomadic hunter-gatherer groups in Thailand. Here, we sequenced complete mitochondrial (mt) DNA genomes and ~2.364 Mbp of non-recombining Y chromosome (NRY) to learn more about the origins of these two enigmatic populations. Both groups exhibited low genetic diversity compared to other Thai populations, and contrasting patterns of mtDNA and NRY diversity: there was greater mtDNA diversity in the Maniq than in the Mlabri, while the converse was true for the NRY. We found basal uniparental lineages in the Maniq, namely mtDNA haplogroups M21a, R21 and M17a, and NRY haplogroup K. Overall, the Maniq are genetically similar to other negrito groups in Southeast Asia. By contrast, the Mlabri haplogroups (B5a1b1 for mtDNA and O1b1a1a1b and O1b1a1a1b1a1 for the NRY) are common lineages in Southeast Asian non-negrito groups, and overall the Mlabri are genetically similar to their linguistic relatives (Htin and Khmu) and other groups from northeastern Thailand. In agreement with previous studies of the Mlabri, our results indicate that the Mlabri do not directly descend from the indigenous negritos. Instead, they likely have a recent origin (within the past 1,000 years) by an extreme founder event (involving just one maternal and two paternal lineages) from an agricultural group, most likely the Htin or a closely-related group.

## Introduction

The indigenous peoples of Southeast Asia (SEA) are often referred to as “Negritos” because of their differences in phenotypic appearances, i.e. on average shorter stature, frizzier hair, and darker skin color<sup>1-2</sup>. The phenotypic resemblances of SEA negritos and African pygmies initially suggested a separate origin in Africa, although convergent evolution now seems more likely (for reviews see Détroit et al.<sup>3</sup> and Stock<sup>4</sup>). Hunting-gathering was the mode of subsistence traditionally practiced by most SEA negritos, whose communities are distributed in three regions: Andaman Islands, Peninsular Thailand and Malaysia, and the Philippines<sup>5</sup>. Several genetic studies have been conducted in some of these relic populations and indicated deep ancestry<sup>6-13</sup> which could link these people with the ancestral groups who migrated from Africa into SEA during the late Pleistocene<sup>14</sup>. However, the hunter-gatherer groups of Thailand are relatively under-studied compared to those from other SEA regions.

There are two hunter-gatherer groups in Thailand, the Maniq and the Mlabri. The Maniq, also known as Sakai or Kensui, have a census size of ~250 individuals<sup>15-16</sup> and are considered to be related to the Semang of Malaysia<sup>17</sup>. For example, the Maniq language belongs to the northern Aslian group of Austroasiatic languages, as do the Semang languages<sup>18</sup>, although this most likely reflects a later introduction by agriculturalists<sup>19-20</sup>. A previous study of sequences of the first hypervariable region (HVR) of the mtDNA control region reported a relatively large divergence of the Sakai (Maniq) from southern Thailand from other Thai groups<sup>21</sup>.

The other hunter-gatherer group, the Mlabri, are characterized by a nomadic traditional lifestyle, in which they construct shelters from the leaves of banana trees that they move out when the banana leaves turn yellow. Their neighbors call them “Phi Tong Luang” which means the “Spirits of the Yellow Leaves”. Although they do not exhibit the typical “negrito”

phenotypes, their mode of subsistence is hunting and gathering<sup>22-23</sup>. This, however, is disputed by Waters (2005)<sup>24</sup> who believed that the Mlabri have had contact and trade with neighboring ethnic groups, and thus they are not hunter-gatherer in the strict sense. The conflicting theories as to how they sustain their livelihood make it interesting to study their origin and lineages. With a census size of only ~130 individuals<sup>16</sup>, the Mlabri live in a mountainous area of northern Thailand and speak a language belonging to the Khmuic branch of Austroasiatic languages. Khmuic languages are also spoken by their neighbors, Htin (of Mal and Pray subgroups) and Khmu, who are also the other hilltribes in northern Thailand. Preceding genetic studies reported extremely reduced genetic diversity in the Mlabri and suggested a cultural reversion from agriculture to hunting and gathering<sup>25</sup> and genetic affinities between the Mlabri and the Mal, a Htin subgroup<sup>26</sup>.

Comparisons of variation in maternally-inherited mtDNA and paternally-inherited, non-recombining portions of the Y chromosome (NRY) have been useful in illuminating differences in the female vs. male demographic history of human populations<sup>27-28</sup>. However, such comparisons are complicated by the different molecular methods that have been typically employed to assay variation (e.g., HVR sequencing for mtDNA vs. genotyping of SNPs and/or STR loci for the NRY). Advances in next-generation sequencing make it feasible to obtain mtDNA and NRY sequences that allow directly comparable, unbiased inferences about female vs. male demographic history<sup>29-30</sup>. Here, we present and analyze complete mtDNA genome sequences and ~2.364 Mbp of NRY sequence from the Maniq and the Mlabri (Fig. 1). These two groups have very distinct mtDNA and NRY lineages and contrasting patterns of variation. Overall, our results support an affiliation of the Maniq with other SEA negrito groups, and a recent origin of the Mlabri lineages.

## Results

### *Mitochondrial DNA*

We generated 29 complete mtDNA sequences (18 from Mlabri and 11 from Maniq) with mean coverages ranging from 277X to 1342X. Extraordinarily low mtDNA diversity was observed in the Mlabri ( $h = 0.29 \pm 0.12$ ; MPD =  $0.29 \pm 0.32$ ;  $\pi = 0.00002 \pm 0.00002$ ;  $S = 1$ ) with just one polymorphic site (nt 43) defining two haplotypes. With 5 haplotypes defined, the Maniq showed higher diversity values ( $h = 0.82 \pm 0.08$ ; MPD =  $29.85 \pm 14.16$ ;  $\pi = 0.0018 \pm 0.00097$ ;  $S = 64$ ) than the Mlabri. However, both of them have lower genetic diversity values than other populations from Thailand (Figure S1)<sup>31-32</sup>.

The two different mtDNA sequences found in the Mlabri belong to haplogroup B5a1b1 and have a coalescent age of  $\sim 735.48$  ya (95% HPD: 1.71 – 2321.88; Figure S2). The overall coalescent age of B5a1b1 is  $\sim 13.99$  kya, which was similar to our previous study<sup>32</sup>. The network analysis (Fig. 2) also supports a single maternal ancestor for the Mlabri. The average number of mutations among the 18 Mlabri sequences is 0.29; with a mutation rate of one mutation in every 3,624 years<sup>33</sup>, this suggests an age for the Mlabri mtDNA variation of  $\sim 1,000$  years, close to the coalescence age estimate. Overall, the mtDNA results are in keeping with previous analyses suggesting a single maternal ancestor for the Mlabri and a recent founding age within the past 1000 years<sup>25</sup>.

By contrast, there are three basal haplogroups (M17a, M21a and R21) observed in the Maniq. When combined with sequences from other populations<sup>11,31-32</sup> (Table S1), the respective estimated ages of M17a, M21a and R21 are  $\sim 23.52$ ,  $\sim 18.44$  and  $\sim 8.83$  kya (Figure S3). The ages of M17a and M21a are much older than our previous estimations in a Thai dataset<sup>32</sup> while R21 is

newly-reported in populations from Thailand. The networks and MCC trees of M21a and R21 indicate a close relatedness of the Maniq sequences with those from the Jehai (an aboriginal Malaysian negrito group), while the Maniq M17a sequences cluster with sequences from the Laotian and Austroasiatic-speaking Blang (Fig. 2; Figure S3).

The MDS plot based on phiST distances indicates that the studied groups have large genetic distances from each other and from most other populations (Fig. 3a). In the NJ tree (Fig. 3b), the Mlabri do not show any close relationship with other SEA hunter-gatherer or negrito groups, but instead are closest to other northeastern Thai populations (Seak, Soa and Bru). In contrast, the Maniq cluster with populations from West Malaysia, and are closest to the Jehai, one of the Semang negrito groups of Malaysia (Fig. 3b).

### *Y chromosome*

We generated 14 sequences (10 from the Mlabri and 4 from the Maniq) for ~2.364 Mb of the NRY, with mean coverages ranging from 10X to 25X. In contrast to the mtDNA results, the Mlabri have higher NRY diversity (six different sequences;  $h = 0.7778 \pm 0.1374$ ; MPD =  $17.46 \pm 8.94$ ;  $\pi = 7.39 \times 10^{-6}$ ;  $S = 37$ ) than the Maniq, which show just one haplotype (i.e., all four sequences were identical). It should be noted that these four unrelated sequences were collected from three different locations (sites 1, 3 and 4) (Fig. 1).

Among six Y chromosomal haplotypes in ten Mlabri males, an MCC tree shows two clades: one comprises five haplotypes of haplogroup O1b1a1a1b which diverged ~1.43 kya and the other is a single haplotype, found in five individuals, classified to haplogroup O1b1a1a1b1a1 (Figure S4), suggesting two paternal lineages for the Mlabri. When we combine the Y sequences of haplogroup O1b1a1a1b and its sublineages retrieved from previous studies (Table S2), the

newly reported age of haplogroup O1b1a1a1b is ~6.89 kya with a signal of population expansion. The MCC tree shows that the O1b1a1a1b sequences of the Mlabri are clustered with sequences from the Htin subgroup Mal, while the O1b1a1a1b1a1 sequences are clustered with sequences from the Khmu and the Htin subgroup Pray (Figure S4). The network of haplogroup O1b1a1a1b shows an overall star-like structure, suggesting lineage expansion, and further supports two paternal lineages for the Mlabri (Fig. 4). The NRY sequence of the four unrelated Maniq males belongs to haplogroup K, and when compared to the two haplogroup K2b1 sequences<sup>30</sup>, the estimated coalescent age is ~41.10 kya (Figure S5). The Maniq sequence thus adds to a small set of basal Y chromosomes sequences.

The Mlabri NRY sequences cluster with other northern Thai populations in the MDS plot and NJ tree (Fig. 5a and 5b), particularly with their linguistic relatives, Khmu and the Htin subgroup Pray. Although the Maniq are highly diverged from other populations due to their very low diversity, they show closest affinity with Papuan, island SEA (ISEA), and South Asian populations (Fig. 5b).

## Discussion

We here report and analyze sequences of mtDNA genomes and ~2.364 Mbp of the NRY from the Mlabri and Maniq, two enigmatic hunter-gatherer groups from Thailand. The large genetic divergence of the Mlabri and the Maniq from each other and from the other populations probably reflects genetic drift associated with isolation and very small population sizes. Notably, we observed contrasting patterns of variation in mtDNA vs. the NRY: mtDNA diversity is higher in the Maniq and lower in the Mlabri, while the reverse is true for the NRY.

The SEA negritos are of interest because they are characterized by basal genetic lineages that likely reflect the earliest dispersal of modern humans from Africa<sup>6,34</sup>. However, there is also evidence that at least some negrito groups have interacted and mixed with non-negrito groups<sup>10-12,35</sup>. Among the three basal mtDNA haplogroups (M21a, M17a and R21) observed in the Maniq (Fig. 2), two (M21a and R21) have been previously reported in negrito groups from Malaysia<sup>9,13</sup>. MtDNA lineages of the Maniq (also known as the Kensui) were mentioned as unpublished data in a previous study<sup>13</sup>, which also found M21a and R21 to be common lineages, along with several minor SEA-specific haplogroups (B5a, F1a1a, M7c3c and N9a6a), indicating some SEA influences. In contrast, we did not detect any SEA-specific mtDNA haplogroups (B, F and M7) in the Maniq, but instead we found M17a, which has been reported at low frequency in MSEA<sup>32,36</sup> and ISEA<sup>37</sup>. Moreover, we detected Y chromosomal haplogroup K in the Maniq, which occurs elsewhere in populations of Papua New Guinea<sup>38-39</sup>, the Philippines<sup>28</sup> and Australia<sup>43</sup>. Haplogroup K is an ancient lineages that is dated here to ~41.10 kya, which is younger than previous estimates in Aboriginal Australians (~48.40 kya)<sup>40</sup>. Overall, the Maniq display basal genetic lineages and closest genetic affinities with indigenous groups in Malaysia or ISEA, supporting a common ancestry of the Maniq with other SEA negrito communities. The reduced genetic diversity (and population size) of the Maniq probably occurred as a consequence of agricultural expansions to this region. The archaeological evidence in southern Thailand suggests that the Neolithic period of the Malay Peninsula started around 4 kya<sup>19</sup> and after the expansion of agriculturalists who also brought Austroasiatic languages to this region, the Proto-Aslian groups were fragmented<sup>20</sup>, probably leading to gradual isolation.

The Mlabri present a different picture. Their mtDNA and NRY haplotypes are closely related to those of neighboring agricultural groups, rather than belonging to basal haplogroups



that are characteristic of typical hunter-gatherer groups. Overall, the mtDNA and NRY results suggest a severe founder event in the Mlabri ancestors, involving just one mtDNA lineage and two NRY lineages that occurred ~1000 ya. These results are in good agreement with previous studies that suggested that the Mlabri originated from an agricultural group and reverted to hunting-gathering<sup>25</sup>. We further attempted to identify the closest living relatives of the Mlabri, as these might be their source population. We found close genetic relationships in the NRY lineages between the Mlabri and various Htin subgroups as well as the Khmu, in agreement with an oral tradition, which indicates that the Htin are the ancestors of the Mlabri<sup>25</sup>. A shared common ancestor of these Khmuic-speaking groups might be another explanation. However, the Mlabri mtDNA lineage is most closely related to the Seak, Soa and Bru groups from northeastern Thailand (Fig. 3). Historical evidence indicates that the Mlabri, Soa, Seak and Bru in Thailand migrated from the present-day area of Laos about 100-200 ya<sup>23</sup>. Although each group migrated independently, their close relatedness might reflect maternal gene flow among various groups in Laos. However, the Mlabri stand out among these groups in apparently having just one maternal ancestor (Fig. 2).

A limitation of our study is that the sample sizes are quite small, which reflects the difficulty in contacting these elusive and enigmatic groups. Nonetheless, a rigorous participant recruitment procedure, and overall agreement with previous studies, suggests that our results are indeed representative of these small populations. Notably, the Mlabri are probably not indigenous SEA hunter-gatherers, but rather originated recently from an agricultural group via an extreme founder event, reflected in their one maternal and two paternal ancestral lineages. The Htin and Khmu are their potential ancestors. By contrast, the Maniq are likely to be indigenous SEA hunter-gatherers whose ancient ancestry is not mixed with non-negrito populations; further

study of additional Maniq samples from the other bands as well as autosomal data would provide more information on the history and relationships of this group.

## **Methods**

### **Samples**

Genomic DNA samples from 18 Mlabri who live in a village located in Wieng Sa District, Nan Province, Northern Thailand were obtained from a previous study<sup>41</sup>. Because the census size of the Mlabri is about 130 individuals, we collected all possible unrelated samples which could be assumed to be representative of this group. The Maniq still lead a highly mobile way of life in the forested Bantad Mountain of three southern Thai Provinces, i.e. Patthalung, Trang and Satun. They live in about 9 small bands or groups, with 8-50 individuals per group, and a total census size of about 250 individuals. All groups are quite closely related, and there is constant movement of family members between groups<sup>15,17</sup>. We obtained saliva samples from 11 Maniq from four different bands in three provinces, i.e. Pabon District (site 1: 3 samples) and Kong-Ra District (site 2: 1 sample), Phatthalung Province; Palien District, Trang Province (site 3: 2 samples); and La-Ngu District, Satun Province (site 4: 5 samples) (Fig. 1). Informed consent was obtained prior to the collection of Maniq saliva samples, and all volunteers were interviewed by interpreters and coordinators to screen for subjects unrelated for at least two generations. We extracted DNA using the QIAamp DNA Midi Kit (Qiagen, Germany) according to manufacturer's directions. Research on human subjects was approved by Khon Kaen University, and the Ethics Commission of the University of Leipzig Medical Faculty.

### **Sequencing**

## *mtDNA*

Twenty-nine complete mtDNA sequences (18 Mlabri and 11 Maniq) were generated from genomic libraries prepared with double indices and enriched for mtDNA as described previously<sup>42-43</sup>; the libraries were sequenced on the Illumina Hiseq 2500. MtDNA consensus sequences were obtained as described by Arias-Alvis et al.<sup>44</sup> with minor modifications. We used Bustard for Illumina standard base calling with a read length of 76 bp. We manually checked sequences with Bioedit ([www.mbio.ncsu.edu/BioEdit/bioedit.html](http://www.mbio.ncsu.edu/BioEdit/bioedit.html)) and aligned the sequences to the Reconstructed Sapiens Reference Sequence (RSRS)<sup>45</sup> using MAFFT 7.271<sup>46</sup>.

## *Y chromosome*

We designed a capture probe set for the NRY following a previous study<sup>47</sup> and using genome sequences included in SGDP Panel A<sup>48</sup> and Panel B<sup>49</sup>. Sites were kept that were included in the UCSC table browser *CRG Align* 100 mappability track of the human reference genome *hg19* (9,655,499 bp). In order to obtain confident Y haplogroup calls later, we selected regions with at least one ISOGG SNP (version 8.78) and then applied the following filter criterion: a region was removed when its length was lower than 500 bp and its repeat count of 24-mers was higher than 10 in an average window of 120 bp using the *CRG Align* 24 mappability track from the UCSC table browser. A total of 2,264 regions with a combined length of 2,364,049 bp were included into the capture probe set, which included 2,071 ISOGG SNPs (version 11.144).

Genomic libraries were prepared for each sample using a double indices scheme<sup>48</sup>. We enriched the libraries for the NRY regions via in-solution hybridization-capture<sup>49</sup>

using our probe set and the Agilent Sure Select system (Agilent, CA, USA) and sequenced on the Illumina HiSeq 2500 platform, generating paired-end reads of 125 bp length. Standard Illumina base-calling was performed using Bustard. We trimmed Illumina adapters and merged completely overlapping paired sequences using leeHOM<sup>50</sup> and de-multiplexed the pooled sequencing data using deML<sup>51</sup>.

The sequencing data were aligned against the human reference genome *hg19* using BWA's *aln* algorithm<sup>52</sup> and all sequence pairs that aligned to the NRY region which are previously defined were retained<sup>47</sup>. Duplicate reads were removed using PicardTools *MarkDuplicates*, indel realignment was performed using GATK *IndelRealigner*<sup>53</sup> and base quality were re-calibrated using GATK *BaseRecalibrator*. We identified single nucleotide variants (SNVs) using GATK *UnifiedGenotyper* v3.3-0 across all samples simultaneously setting the parameter *ploidy* to 1 and using DBSNP build 138 as a prior position list. The identified SNVs were further filtered as previously described<sup>54</sup>. We imputed all samples with missing genotype information at any of these variant sites using BEAGLE<sup>55</sup>.

## Statistical analyses

Haplogroup assignment was performed by Haplogrep<sup>56</sup> for mtDNA and yHaplo<sup>57</sup> for the NRY. Summary statistics for the studied populations and genetic distances ( $\Phi_{st}$ ) among the studied and compared populations were computed by Arlequin 3.5.1.3<sup>58</sup>. Comparative data for the mtDNA genome sequences were selected from the literature encompassing East Asia<sup>31-32,36,59-64</sup>, Island Southeast Asia<sup>12-13,34,65-67</sup>, Taiwan<sup>68</sup>, India<sup>69</sup>, Andaman Islands<sup>6,70</sup>, Australia<sup>71-73</sup> and New Zealand<sup>74</sup>. Further information on the populations used for mtDNA comparison are listed in Table S1. Overlapping NRY sequences of relevant populations from Thailand (Htin, Khmu, Soa, Bru and Seak) (unpublished data), East Asia, Oceania, South Asia and also Indian

groups from the United Kingdom and the United States of America<sup>30,75-76</sup> were included for comparison; additional details are provided in Table S2.

The  $\Phi_{st}$  distance matrices of both mtDNA and the NRY were used to construct multi-dimensional scaling (MDS) plots and Neighbor Joining (NJ) trees<sup>77</sup> by STATISTICA 13.0 (StatSoft, Inc., USA) and MEGA 7.0<sup>78</sup>, respectively. Median-joining networks<sup>79</sup> of haplogroups without pre- and post-processing steps were performed by Network ([www.fluxus-engineering.com](http://www.fluxus-engineering.com)) and visualized in Network publisher 1.3.0.0.

BEAST 1.8.0 was used to construct maximum clade credibility (MCC) trees by haplogroup, based on Bayesian Markov Chain Monte Carlo (MCMC) analyses<sup>80</sup>. We ran jModel test 2.1.7<sup>81</sup> for selecting the suitable models in each run during the creation of the input file of BEAST via BEAUTi v1.8.0<sup>80</sup>. For mtDNA, the BEAST calculation was done with the data partitioned between coding and noncoding regions with mutation rates of  $1.708 \times 10^{-8}$  and  $9.883 \times 10^{-8}$ , respectively<sup>33</sup>. For the NRY, we used a mutation rate of  $8.71 \times 10^{-10}$ ,<sup>82</sup> and the BEAST input files were modified by an in-house script to add in the invariant sites found in our dataset. After BEAST runs, we used Tracer 1.6 to check the results and TreeAnnotator v1.6.2 and FigTree v 1.4.0. to assemble and draw the Bayesian MCC trees, respectively. The Bayesian MCMC estimates (BE) and credible intervals (CI) of haplogroup coalescent times were calculated using the RSRS for rooting the mtDNA tree and African sequences belonging to haplogroup B2 (GS000035246-ASM) for rooting the NRY tree<sup>30</sup>.

## References

1. Coon, C.S. *The Origin of Races*. 112 (Knopf, 1962).
2. Bulbeck, D. Craniodental Affinities of Southeast Asia's "Negritos" and the Concordance with Their Genetic Affinities. *Hum Biol* **85**, 95–134 (2013).
3. Détroit, F., Corny, J., Dizon, E.Z. & Mijares, A.S. "Small size" in the Philippine human fossil record: is it meaningful for a better understanding of the evolutionary history of the negritos? *Hum Biol* **85**, 45–65 (2013).
4. Stock, J.T. The Skeletal Phenotype of "Negritos" from the Andaman Islands and Philippines Relative to Global Variation among Hunter-Gatherers. *Hum Biol* **85**, 67–94 (2013).
5. Higham, C. Hunter-gatherers in Southeast Asia: From prehistory to the present. *Hum Biol* **85**, 21–44 (2013).
6. Thangaraj, K. et al. Reconstructing the origin of Andaman Islanders. *Science* **308**, 996 (2005)
7. Hill, C. et al. Phylogeography and ethnogenesis of aboriginal Southeast Asians. *Mol Biol Evol* **23**: 2480–2491 (2006).
8. Delfin, F. et al. The Y-chromosome landscape of the Philippines: extensive heterogeneity and varying genetic affinities of Negrito and non-Negrito groups. *Eur J Hum Genet* **19**, 224–230 (2011).
9. Oppenheimer, S. mtDNA variation and southward Holocene human dispersals within mainland Southeast Asia. In: Enfield, N.J., editor. *Dynamics of Human Diversity: The Case of Mainland Southeast Asia*, 81–108 (Canberra Pacific Linguistics, 2011)
10. Reich, D. et al. 2011. Denisova Admixture and the First Modern Human Dispersals into Southeast Asia and Oceania. *Am J Hum Genet* **89**, 516–528 (2011).
11. Jinam, T.A. et al. Evolutionary history of continental southeast Asians: "early train" hypothesis based on genetic analysis of mitochondrial and autosomal DNA data. *Mol Biol Evol* **29**, 3513–3527 (2012).
12. Delfin, F. 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* **22**, 228–237 (2014).

13. Jinam, T.A. et al. Discerning the Origins of the Negritos, First Sundaland People: Deep Divergence and Archaic Admixture. *Genome Biol Evol* **9**, 2013–2022 (2017).
14. Endicott, P. Introduction: revisiting the "negrito" hypothesis: a transdisciplinary approach to human prehistory in southeast Asia. *Hum Biol* **85**, 7–20 (2013).
15. Nagata, S. Subgroup ‘names’ of the Sakai (Thailand) and the Semang (Malaysia): a literature survey. *Anthropol Sci* **114**, 45–57 (2006).
16. Bradley, D. East and Southeast Asia. In: Moseley, C., editor. *Encyclopedia of the world's endangered languages*. 349-424 (Routledge, 2007).
17. Rattanakrangsri, K., Maneerat, T. & Colchester, M. The Mani people of Thailand on the agricultural frontier. In: Colchester, M. & Chao, S., editor. *Conflict or Consent? The oil palm sector at a crossroads*. 282–301 (Forest Peoples Programme, 2013).
18. Simons, G.F. & Fennig, C.D. *Ethnologue: Languages of the World, Twentieth edition* (SIL International, 2017). Online: <http://www.ethnologue.com>.
19. Bellwood, P. Southeast Asia before history. In: Tarling, N. editor. *The Cambridge History of Southeast Asia, Vol. 1, From Early Times to c.1800*. 55–136 (Cambridge University Press, 1992).
20. Dunn, M., Kruspe, N. & Burenhult, N. Time and Place in the Prehistory of the Aslian Languages. *Hum Biol* **85**. 383–400 (2013).
21. Fucharoen, G., Fucharoen, S. & Horai S. Mitochondrial DNA polymorphisms in Thailand. *J Hum Genet* **46**, 115–125 (2001).
22. Pookajorn, S. *The Phi Tong Luang (Mlabri): A hunter–gatherer group in Thailand* (Odeon Store Printing House, 1992)
23. Schliesinger, J. *Ethnic groups of Thailand: non-Tai-speaking peoples* (White Lotus Press, 2000).
24. Waters T. Comment on "Recent origin and cultural reversion of a hunter-gatherer group". *PLoS Biol* **3**, e269 (2005).

25. Oota, H. et al. Recent origin and cultural reversion of a hunter–gatherer group. *PLoS Biol* **3**, e71 (2005).
26. Xu, S. et al. Genetic evidence supports linguistic affinity of Mlabri-a hunter-gatherer group in Thailand. *BMC Genet* **11**,18 (2010)
27. Kayser, M. et al. The impact of the Austronesian expansion: evidence from mtDNA and Y chromosome diversity in the Admiralty Islands of Melanesia. *Mol Biol Evol* **25**, 1362–1374 (2008).
28. Delfin, F. et al. Bridging near and remote Oceania: mtDNA and NRY variation in the Solomon Islands. *Mol Biol Evol* **29**, 545–564 (2012).
29. Lippold, S. et al. Human paternal and maternal demographic histories: insights from high-resolution Y chromosome and mtDNA sequences. *Investigative Genet* **5**, 13 (2014).
30. Karmin, M. et al. 2015. A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Res* **25**, 459–466 (2015).
31. Kutanan, W. 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* **136**, 85–98 (2017a).
32. Kutanan, W. et al. New insights from Thailand into the maternal genetic history of Mainland Southeast Asia. Preprint at <https://www.biorxiv.org/content/early/2017/07/12/162610> (2017b).
33. Soares, P. et al. Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet* **84**, 740–759 (2009).
34. Macaulay, V. et al. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* **308**, 1034–1036 (2005).
35. Aghakhanian, F. et al. Unravelling the Genetic History of Negritos and Indigenous Populations of Southeast Asia. *Genome Biol Evol* **7**, 1206–1215 (2015).



36. Peng, M.S. et al. Tracing the Austronesian footprint in mainland Southeast Asia: a perspective from mitochondrial DNA. *Mol Biol Evol* **27**, 2417–2430 (2010).
37. Tumonggor, M.K. et al. The Indonesian archipelago: An ancient genetic highway linking Asia and the Pacific. *J Hum Genet* **58**, 165–173 (2013).
38. Kayser, M. et al. Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea. *Am J Hum Genet* **72**, 281–302 (2003).
39. Van Oven, M. et al. Human genetics of the Kula Ring: Y-chromosome and mitochondrial DNA variation in the Massim of Papua New Guinea. *Eur J Hum Genet* **22**, 1393–1403 (2014).
40. Bergström, A. et al. Deep Roots for Aboriginal Australian Y Chromosomes. *Curr Biol* **26**, 809–813 (2016).
41. HUGO Pan-Asian SNP Consortium. Mapping human genetic diversity in Asia. *Science* **326**, 1541–1545 (2009).
42. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protoc* **6**, 1–10 (2010)
43. Maricic, T., Whitten, M. & Pääbo, S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One* **5**, e14004 (2010).
44. Arias-Alvis, 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. Preprint at <https://www.biorxiv.org/content/early/2017/07/06/160218> (2017).
45. Behar, D.M. et al. A “Copernican” reassessment of the human mitochondrial DNA tree from its root. *Am J Hum Genet* **90**, 675–684 (2012).
46. Katoh, K. & Standley, D.M. MAFFT Multiple Sequence Alignment Software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772–780 (2013).
47. Poznik, G.D. et al. Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females. *Science* **341**, 562–565 (2013).

48. Meyer, M. et al. A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222–226 (2012).
49. Prüfer, K. et al. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014).
50. Renaud, G., Stenzel, U. & Kelso, J. LeeHom: Adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Res* **42**, e141 (2014).
51. Renaud, G., Stenzel, U., Maricic, T., Wiebe, V., & Kelso, J. deML: Robust demultiplexing of Illumina sequences using a likelihood-based approach. *Bioinformatics* **31**, 770–772 (2015).
52. Li, H., & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
53. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297–1303 (2010).
54. Barbieri, C. et al. Refining the Y chromosome phylogeny with southern African sequences. *Hum Genet* **135**, 541–553 (2016).
55. Browning, B.L. & Browning, S.R. Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics* **194**, 459–471 (2013).
56. Kloss-Brandstätter A. et al. HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Hum Mutat* **32**, 25–32 (2010).
57. Poznik, G.D. Identifying Y-chromosome haplogroups in arbitrarily large samples of sequenced or genotyped men. Preprint at <https://www.biorxiv.org/content/early/2016/11/19/088716> (2016).
58. Excoffier, L. & Lischer, H.E.L. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* **10**, 564–567 (2010).
59. Zhao, M. et al. Mitochondrial genome evidence reveals successful Late Paleolithic settlement on the Tibetan Plateau. *Proc Natl Acad Sci USA* **106**, 21230–21235 (2009).

60. Zheng, H.X. et al. Major population expansion of East Asians began before Neolithic time: evidence of mtDNA genomes. *PLoS One* **6**: e25835 (2011).
61. Zhang, X. et al. Analysis of mitochondrial genome diversity identifies new and ancient maternal lineages in Cambodian aborigines. *Nat Commun* **4**, 2599 (2013).
62. Summerer, M. 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* **14**, 17 (2014).
63. Li, Y.C. et al. Ancient inland human dispersals from Myanmar into interior East Asia since the Late Pleistocene. *Sci Rep* **5**, 9473 (2015).
64. Diroma, M.A. et al. Extraction and annotation of human mitochondrial genomes from 1000 Genomes Whole Exome Sequencing data. *BMC Genom* **15**, S2 (2014).
65. Dancause, K.N., Chan, C.W., Arunotai, N.H. & Lum, J.K. Origins of the Moken Sea Gypsies inferred from mitochondrial hypervariable region and whole genome sequences. *J Hum Genet* **54**, 86–93 (2009).
66. Gunnarsdottir, E.D., Li, M., Bauchet, M., Finstermeier, K. & Stoneking, M. High-throughput sequencing of complete human mtDNA genomes from the Philippines. *Genome Res* **21** 1-11 (2011a).
67. Gunnarsdóttir, E.D. et al. Larger mitochondrial DNA than Y-chromosome differences between matrilineal and patrilineal groups from Sumatra. *Nat Commun* **2**, 228 (2011b).
68. Ko, A.M.S. et al. Early Austronesians: into and out of Taiwan. *Am J Hum Genet* **94**, 426–436 (2014).
69. Chandrasekar, A. et al. Updating phylogeny of mitochondrial DNA macrohaplogroup M in India: dispersal of modern human in South Asian corridor. *PLoS One* **4**, e7447 (2009).
70. Barik, S.S. et al. Detailed mtDNA genotypes permit a reassessment of the settlement and population structure of the Andaman Islands. *Am J Phys Anthropol* **136**, 19–27 (2008).

71. Ingman, M., Kaessmann, H., Pääbo, S. & Gyllensten, U. Mitochondrial genome variation and the origin of modern humans. *Nature* **408**, 708–713 (2000).
72. van Holst Pellekaan S.M., Ingman, M., Roberts-Thomson, J. & Harding, R.M. Mitochondrial genomics identifies major haplogroups in Aboriginal Australians. *Am J Phys Anthropol* **131**, 282–294 (2006).
73. Nagle, N. et al. Aboriginal Australian mitochondrial genome variation – an increased understanding of population antiquity and diversity. *Sci Rep* **7**, 43041 (2017).
74. Benton, M. et al. Correction: Complete Mitochondrial Genome Sequencing Reveals Novel Haplotypes in a Polynesian Population. *PLoS One* **7**, 10 (2012)
75. Mallick, S. et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016).
76. Poznik, G.D. et al. Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. *Nat Genet* **48**, 593–599 (2016).
77. Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406–425 (1987).
78. Kumar, S., Stecher, G. & Tamura, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* **33**, 1870–1874 (2016).
79. Bandelt, H.J., Forster, P. & Röhl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* **16** 37–48 (1999).
80. Drummond, A.J., Suchard, M.A., Xie, D. & Rambaut, A. A Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* **29**, 1969–1973 (2012).
81. Darriba, D., Taboada, G.L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* **9**, 772 (2012).
82. Helgason, A. et al. The Y-chromosome point mutation rate in humans. *Nat Genet* **47**, 453–457 (2015).

## **Acknowledgements**

We would like to thank the participants and the people involved with participant recruitment. We thank Prof. Murray Cox for discussion and valuable suggestions. W.K. and M.S. were supported by the Max Planck Society. WK was also funded by the Thailand Research Fund (Grant No. MRG5980146). P.F. was supported by the Institution Development Program of the University of Ostrava and by the Rector's award of the University of Ostrava. P.C. was supported by the grant no. 0924/2016/ŠaS from the Statutory City of Ostrava and by the grant no. 01211/2016/RRC “Strengthening international cooperation in science, research and education” from the Moravian-Silesian Region.

## **Author Contributions**

W.K. and M.S. conceived and designed the experiment, provided funding, and drafted the manuscript; W.K. and R.S. conducted sequencing; J.K., P.C., P.F. and D.K. collected the samples and extracted DNA; W.K., A.H. and E.M. analysed the results; All authors reviewed the manuscript.

## **Additional information**

### **Data availability**

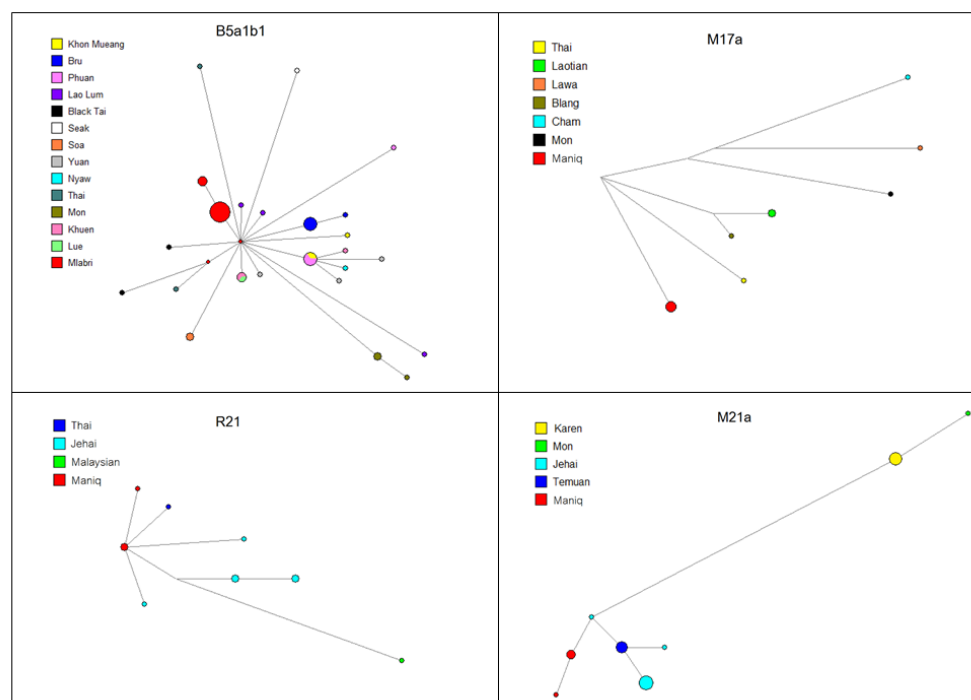
GenBank accession numbers will be provided upon acceptance for publication.

### **Competing financial interests**

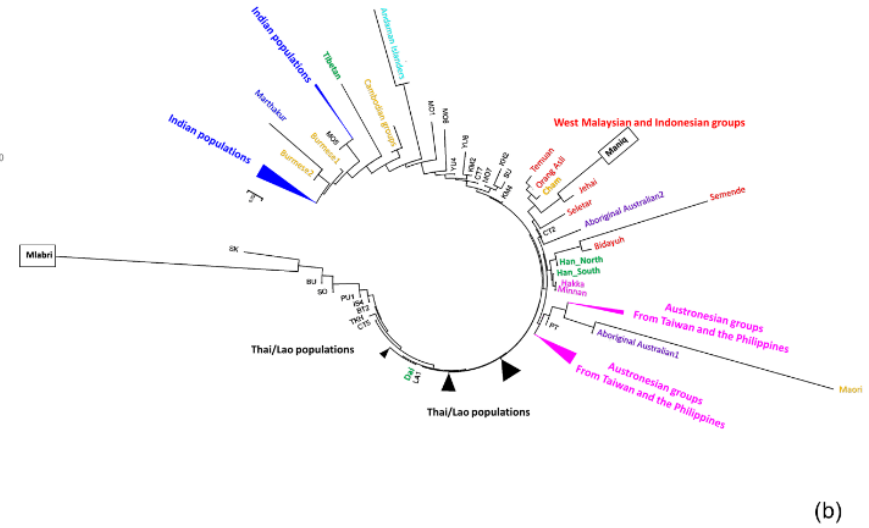
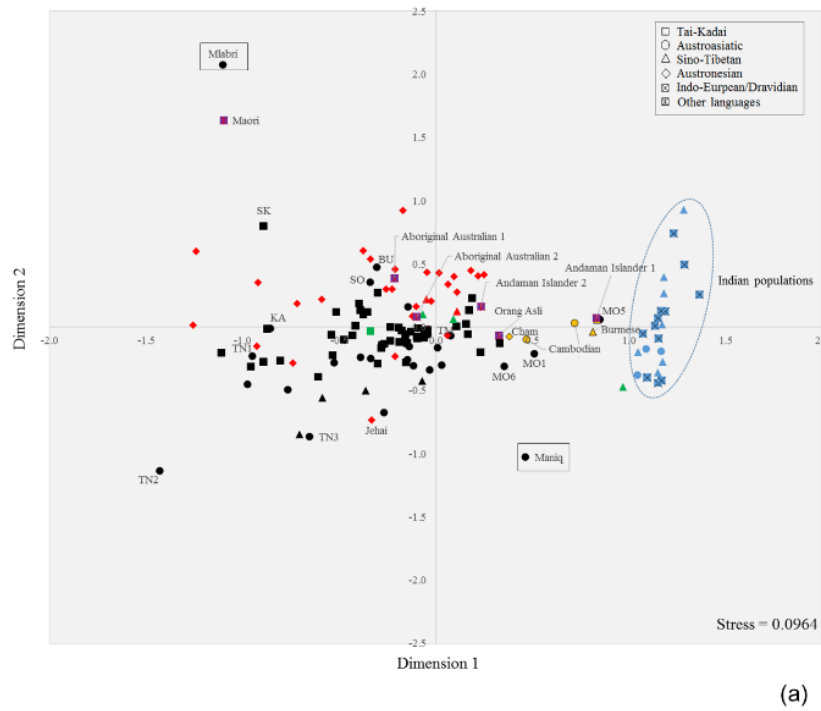
The authors declares no competing interests



**Figure 1** Map of the Mlabri and Maniq sampling locations (stars) and the Thai populations selected for NRY comparisons (comparative mtDNA data are from previous studies<sup>31,32</sup>). Further details concerning the populations are provided in Table S1 and S2. There are four Maniq collection sites; Pabon District (site 1) and Kong-Ra District (site 2), Phatthalung Province; Palien District, Trang Province (site 3); and La-Ngu District, Satun Province (site 4).

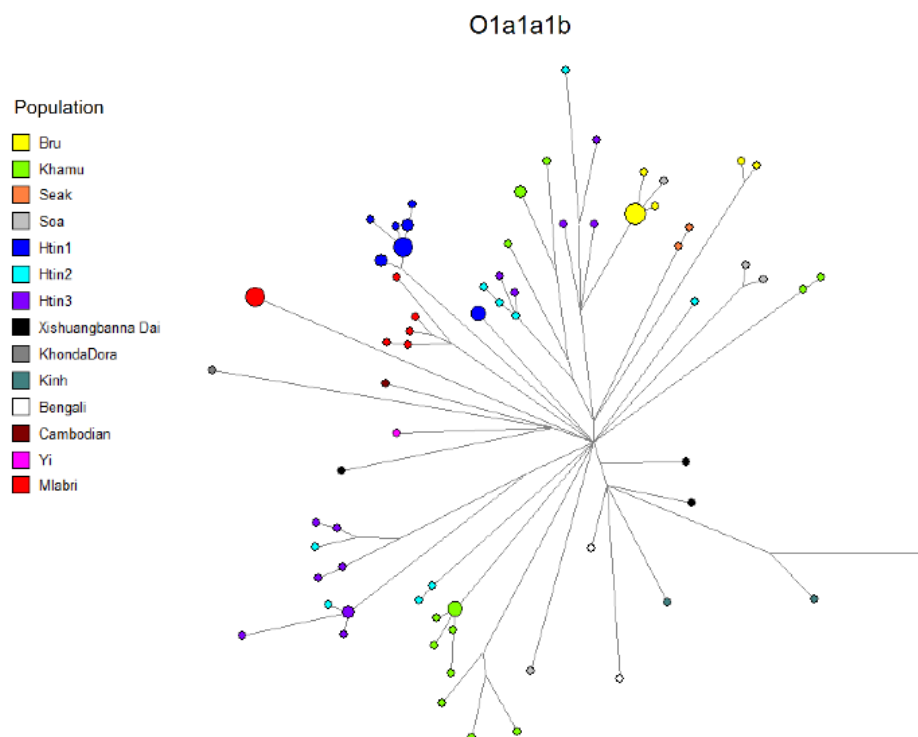


**Figure 2** Networks of mtDNA haplogroups: B5a1b1 is found in Mlabri and R21, M21a and M17a are found in Maniq.

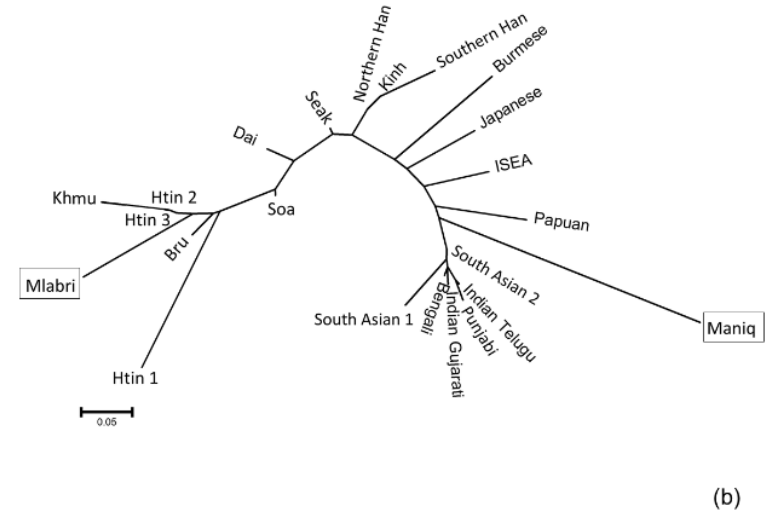
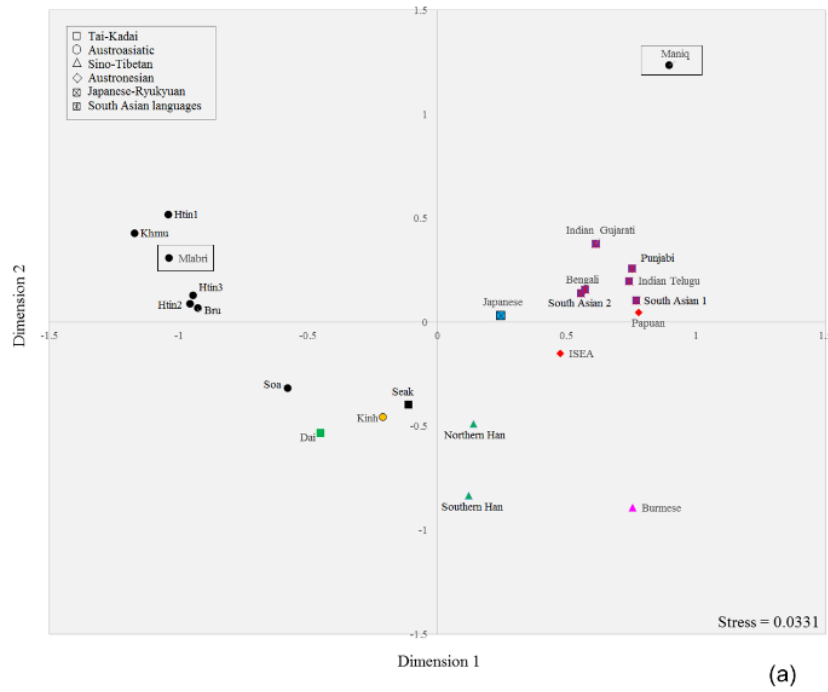


**Figure 3** The MDS plot (a) and NJ tree (b) based on the mtDNA  $\Phi_{st}$  distance matrix for 142 populations from Asia, New Zealand and Australia. Population details are provided in Table S1. TN1 is Htin subgroup Mal and TN2 and TN3 are Htin subgroup Pray. KA, SO, SK and BU are Khmu, So, Seak and Bru, respectively. Black indicates Thai/Lao populations; red indicates populations from Indonesia and Malaysia; green indicates populations from China.





**Figure 4** Network of NRY haplogroup O1a1a1b sequences.



**Figure 5** The MDS plot (a) and NJ tree (b) based on the Y chromosomal  $\Phi_{st}$  distance matrix for 23 populations from Asia and Oceania. Black circles indicates Thai populations. Population details are provided in Table S2.