

**Supplementary Information for:**

**Genomic portrait and relatedness patterns of the Iron Age Log Coffin culture in northwestern Thailand**

Selina Carlhoff<sup>1\*</sup>, Wibhu Kutanan<sup>2</sup>, Adam B Rohrlach<sup>1,3</sup>, Cosimo Posth<sup>4,5</sup>, Mark Stoneking<sup>6,7</sup>, Kathrin Nägele<sup>1</sup>, Rasmi Shoocongdej<sup>8,9\*</sup>, Johannes Krause<sup>1\*</sup>

<sup>1</sup>Department of Archaeogenetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.

<sup>2</sup>Department of Biology, Faculty of Science, Naresuan University, Phitsanulok, Thailand.

<sup>3</sup>School of Computer and Mathematical Sciences, University of Adelaide, Adelaide, Australia.

<sup>4</sup>Institute for Archaeological Sciences, Archaeo- and Palaeogenetics, University of Tübingen, Tübingen, Germany.

<sup>5</sup>Senckenberg Centre for Human Evolution and Palaeoenvironment, University of Tübingen, Tübingen, Germany.

<sup>6</sup>Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.

<sup>7</sup>Université Lyon 1, CNRS, Laboratoire de Biométrie et Biologie Evolutive, UMR 5558, Villeurbanne, France.

<sup>8</sup>Department of Archaeology, Silpakorn University, Bangkok, Thailand.

<sup>9</sup>The Prehistoric Population and Cultural Dynamics in Highland Pang Mapha Project, Princess Maha Chakri Sirindhorn Anthropology Centre, Bangkok, Thailand.

\*Corresponding authors: selina\_carlhoff@eva.mpg.de, SHOOCONGDEJ\_R@su.ac.th, krause@eva.mpg.de

## Supplementary Notes

### Site descriptions

#### *Ban Rai (BRB)*

Ban Rai (19° 31'18" N 98° 10'31" E) is a rock shelter on top of the southern edge of the Lang River valley near Ban Huai Rai (Ban Hai) village in the Pang Mapha District of Mae Hong Son Province. Excavations under the Highland Archaeology Project led by Rasmi Shoocongdej yielded a rich, well-dated stratigraphic sequence indicating two occupation layers and prehistoric rock paintings of human figures and animals<sup>1-9</sup>. The pre-Log Coffin culture layers have been dated to 12.5 to 8 k BP and were characterized by a wide range of lithics<sup>10,11</sup>, an abundance of faunal remains<sup>12</sup>, and a primary flexed burial<sup>1</sup>.

In its second phase, the rock shelter was used exclusively as a burial site assigned to the Log Coffin culture (2040±30-1770±30 BP)<sup>13</sup>. This layer yielded human remains, pottery sherds, iron tools (chisel, hoop), and 15 log coffins and supporting posts dug into older layers<sup>1,9</sup>.

Dendrochronological and <sup>14</sup>C studies of log coffin heads showed the simultaneous use of different carving styles<sup>14,15</sup>, while oxygen chronology indicated that precipitation was higher during the Log Coffin period<sup>16</sup>.

#### *Lahu Pot (LHP)/Moa Mussur (MMS)*

Lahu Pot (19° 35' N, 98° 15' E) is a stalagmite cave located in the Lang River basin near Wa-Na Luang village in the Pang Mapha District. The lower cave yielded six log coffins, human bones, and teeth, while four log coffins and human bones were found in the upper cave. Dates from Lahu Pot Cave range from 1,540 ± 120 to 2,080 ± 60 BP<sup>17</sup>.

#### *Pang Kham 1 (PKL)*

Pang Kham 1 Cave (19° 40' N, 98° 12' E) is located 1,034 m above sea level in the Huay Pong Saen Pig basin in the Pang Mapha District. It can be divided into three areas of which only two housed log coffins. The second area yielded several log coffins, iron tools, bronze decorations, earthenware, and human bones. The third area housed many coffins, earthenware, and a limestone-embedded skull. A total of 10 log coffins were recovered and dated to 1,343 ± 59 BP and 1,698 ± 59 BP<sup>18</sup>.

### *Tham Lod rock shelter (THL)*

The Tham Lod rock shelter (19° 34' N 98° 16' E) is located at the base of a limestone cliff overhang around 250 m from a Nam Lang River tributary in the vicinity of the Tham Lod Natural and Wildlife Education Center. It was excavated from 2002 to 2006 under the direction of Rasmi Shoocondej (Highland Archeological Project in Pang Mapha) supported by Thailand Research Funds (TRF) and Silpakorn University<sup>2-9</sup>.

Tham Lod yielded the longest stratigraphic sequence of the region (10 layers over 4.5 m). More than 100,000 archaeological remains date from the late Pleistocene into the Holocene, including lithic artefacts, faunal and floral remains, pottery sherds, and metal objects from upper layers<sup>7,12</sup>. Dates produced from various materials indicate use as a temporary settlement for lithic production<sup>10,11,19</sup>, including Hòabìnhan-typical stone artefacts (sumatraliths) in layers 4 to 6<sup>20</sup>, food preparation, and as a cemetery<sup>8,9,21</sup> from 35,782 ± 266 BP (layer 10) over 27,620 ± 170 BP (layer 8) to ~3 kaa ago (layer 2), albeit the absence of the Middle Holocene<sup>20</sup>. Recent isotopic studies of human and faunal teeth enamel tracked the expansion of forest-grassland ecosystems in the region. They suggested a northern expansion of the savanna corridor during the terminal Pleistocene providing favourable conditions for hunter-gatherers across mountainous areas<sup>22</sup>.

Two burials and two fragmentary skeletons were discovered in the youngest Pleistocene layers associated with faunal remains, shells, flakes, and used cobble tools<sup>8,23,24</sup>.

### *Tham Lod cave (THL)*

Tham Lod cave (19° 34' N, 98° 17' E) is situated near the Tham Lod Natural and Wildlife Education Center, about 300 m from the Tham Lod rockshelter. The cave is a multi-chamber cave complex with the Nam Lang river passing through the cave. The burial site is located in the upper chambers, where three chambers contain fragments of ten log coffins, earthenware sherds, iron rings, and a decorated tooth, but the site has been badly disturbed by humans and animals. Pieces of log coffin from the site date to 1,450±110 - 1240±90 BP<sup>17</sup>.

### *Ban Wang Hai (WAH)*

Ban Wang Hai (18° 32' N, 99° 00' E) is situated in Kuang river valley, Mueang Lamphun district, Lamphun province. The site was first excavated by the 4th Regional Office of Fine Arts Department in 1987. Seven skeletal remains were discovered along with grave goods including bronze and glass bracelets, iron tools, stone tools, spindle whorls and earthenware pottery as well as stone and glass beads (the 4th Regional Office of Fine Arts Department

1987). The excavation was continued as part of the Thai-French Prehistoric Project between 1996-1998 and dated to 429-657 CE<sup>25</sup>.

*Yappa Nhae 1 (YPN)*

Yappa Nhae 1 Cave (19° 36' N, 98° 14' E) is located 900 m above sea level in the Mae La Na Basin near Ya-Pa Nae village in the Pang Mapha District. Indirectly dated between 2,200 and 1,200 BP, four log coffins were found in Chamber 3 and 4 in addition to human remains of at least nine individuals (five adults, two juveniles, one baby, one unknown), a decorated tooth, iron tools, bronze coins, and earthenware.

*Yappa Nhae 2 (YPN)*

Yappa Nhae 2 Cave (19° 36' N, 98° 14' E) is located 964 m above sea level in the Mae La Na basin near Ya-Pa Nae village in the Pang Mapha District. Four log coffins were discovered in the middle, upper and lower chamber indirectly dated to 2,020 to 1,200 BP from the log coffin heads<sup>26</sup>. These were associated with human remains of at least 23 individuals (19 adults and four children), earthenware, chisels, animal bones, and iron tools.

## Supplementary Methods

### Shotgun screening

After demultiplexing, the shotgun-sequencing data was processed with nf-core/eager v2.4.0 (<https://nf-co.re/eager>)<sup>27</sup> using Nextflow v21.04.0<sup>28</sup>. General sequencing quality control was conducted with FastQC v0.11.9<sup>29</sup> and fastP v0.20.1<sup>30</sup>. Forward and reverse adapters were removed using AdapterRemoval v2.3.2<sup>31</sup> with one overlapping position trimming reads based on 30 bp read length and 20 base quality and skipping base quality trimming on the 5' end. Using BWA v0.7.17-r1188<sup>32,33</sup>, the reads were aligned to the human reference genome *hg19* allowing for 1 mismatch per read, maximum edit distance of 2, disabling the seeding and two gaps. The resulting bam files were then filtered with SAMtools v1.12<sup>34</sup> for a minimum mapping quality and minimum read length of 30, and unmapped reads were discarded. Mapping statistics were generated with Qualimap v2.2.2-dev<sup>35</sup> and bedtools v2.30.0<sup>36</sup>. Duplicates were removed using Picard MarkDuplicates v2.26.0 (<http://broadinstitute.github.io/picard/>) and C->T misincorporation frequency assessed with DamageProfiler v0.4.9<sup>37</sup> of all reads up to 100 bp, considering 15 bp of each read and setting the maximum misincorporation frequency to be displayed. Endogenous DNA was estimated using endorS.py v0.4 (<https://github.com/aidaanva/endorS.py>), and the final report generated with MultiQC v1.11<sup>38</sup>.

```
nextflow run nf-core/eager
    -profile eva,archgen,big_data
    -r 2.4.0
    --input Thailand.tsv
    --min_adap_overlap 1
    --clip_readlength 30
    --clip_min_read_quality 20
    --preserve5p
    --mapper bwaaln
    --bwaalnno 0.01
    --bwaalno 2
    --run_bam_filtering true
    --bam_mapping_quality_threshold 30
    --bam_filter_minreadlength 30
    --bam_unmapped_type discard
    --dedupper markduplicates
    --damageprofiler_length 100
    --damageprofiler_threshold 15
    --damageprofiler_yaxis 0.3
```

### In-solution enrichment

The 1240K-enriched sequences were processed with nf-core/eager v2.4.0 as described above with the addition of providing the 1240K positions in a BED file, mitochondrial to nuclear ratio calculation using MTNucRatioCalculator v0.7, genetic sex estimation with sexdetermine v1.1.2<sup>39</sup> and nuclear contamination estimation with ANGSD v0.935<sup>40</sup>. After trimming off 2 bp at both ends of the reads with bamUtil v1.0.15<sup>41</sup>, the resulting bam files were genotyped with pileupCaller from sequenceTools v1.4.0.5 (<https://github.com/stschiff/sequenceTools>) on the 1240K SNP panel calling random haploids and utilizing all sites.

```
nextflow run nf-core/eager
    -profile eva,archgen,big_data
    -r 2.4.0
    --input Thailand.tsv
    --min_adap_overlap 1
    --clip_readlength 0
    --clip_min_read_quality 20
    --preserve5p
    --mapper bwaaln
    --bwaaln 0.01
    --bwaalno 2
    --run_bam_filtering true
    --bam_mapping_quality_threshold 30
    --bam_filter_minreadlength 30
    --bam_unmapped_type discard
    --dedupper markduplicates
    --damageprofiler_length 100
    --damageprofiler_threshold 15
    --damageprofiler_yaxis 0.3
    --snpcature_bed
    --run_mtnucratio true
    --mtnucratio_header MT
    --run_sexedetermine true
    --run_nuclear_contamination true
    --contamination_chrom_name chrX
    --run_trim_bam true
    --bamutils_clip_half_udg_left 2
    --bamutils_clip_half_udg_right 2
    --run_genotyping true
    --genotyping_tool pileupcaller
    --genotyping_source trimmed
    --pileupcaller_method randomHaploid
    --pileupcaller_transition AllSites
```

## Supplementary Discussion

### Ancient DNA authentication

After shotgun sequencing, the level of endogenous human DNA content ranged from <0.1% to 57.92%, with a median of 0.285% (Supplementary Data 1). Overall, samples from Yappa Nhae were better preserved compared to the other sites, and petrous bones yielded much more endogenous DNA compared to teeth.

We assessed if the sequenced reads were ancient rather than present-day human DNA by the rate of cytosine to thymine mis-incorporations at the 5' end of reads ('ancient DNA damage', Supplementary Figure 2A, Supplementary Data 1). Most samples showed an elevated damage rate (7-26%) at the first base, which drops off further into the read due to UDG-half treatment. Since ancient DNA reads are known to be fragmented, we also evaluated the median read length, which ranged between 30 and 55 bp (Supplementary Figure 2B, Supplementary Data 1). Based on previous comparisons of ancient and present-day DNA<sup>42</sup>, 33 sequences showed the damage pattern and read length characteristic of ancient DNA and an endogenous DNA content suitable for in-solution capture (>0.1%).

### Runs of homozygosity

We also used hapROH v0.6<sup>43</sup> to estimate effective population size ( $N_e$ ), i.e., the size of an idealized population that would give rise to the same variance of allele frequencies, with a maximum likelihood approach (MLE\_ROH\_  $N_e$ ) to 977 (845-1,130) after removing the four individuals with ROH >20 cM. For the individual sites, we calculated the following  $N_e$  values: 50,000 (2,610-50,000) for Ban Rai, 1,047 (483-2,936) for Lahu Pot, 1,316 (784-2,453) for Tham Lod, and 1,047 (881-1,254) for Yappa Nhae and 518 (382-716) for Long Long Rak.

### IBD calculation

To further investigate the genetic relationships between the studied individuals, we analysed long shared segments of the genome, also called identity-by-descent (IBD) blocks. Due to sequencing differences and low coverage, we were not able to include the published individuals from Long Long Rak in our analysis<sup>44</sup>. To assess imputation accuracy, we also plotted genotypes before and after imputation onto a PCA calculated on present-day East and Southeast Asian groups, as described previously. In PCA space, the imputed individuals are

more spread out and show a slight drift towards groups from East Asia, but generally overlap with the positions of the non-imputed individuals (Supplementary Figure 5A).

### Principal component analysis

We merged the newly generated genotype calls with published data from present-day and ancient individuals from northern to southern East Asia<sup>44-55</sup> and used smartpca v16000 (lsqmode: YES, shrinkmode: YES) from EIGENSOFT v7.2.1<sup>56</sup> to calculate several sets of principal components.

The initial PCA was calculated on present-day East and Southeast Asian individuals<sup>45,46,48,49,53,55</sup> and with all ancient individuals projected onto the scaffold from the present-day individuals (Supplementary Figure 6A). Along PC1 we can observe a north-south cline, while PC2 stretches between the Onge from the Andaman Islands and the Taiwanese Atayal. The newly genotyped ancient individuals cluster together with present-day individuals from Thailand (Padaung Karen, Eastern Lawa, Western Lawa, Mon) as well as the Log Coffin-associated individuals from Long Long Rak<sup>49</sup>. This position is markedly different from Bronze and Iron Age individuals from Ban Chiang, northeastern Thailand<sup>54</sup> and other contemporaneous individuals from Guangxi, China<sup>53</sup>.

The second PCA was calculated on the present-day individuals from Thailand only<sup>49</sup> which separated according to the genetic drift of specific groups (Supplementary Figure 6B). The ancient individuals clustered with present-day groups affiliated with all language families except Hmong-Mien (Blang, Padaung Karen, Laotian, Eastern Lawa, Western Lawa). However, the Ban Chiang individuals were only associated with Austroasiatic-speaking groups, such as the Bru, Khmu, and Soa, and the younger individuals from Guangxi clustered with Hmong-Mien speakers.

### *f*-statistics

To formally investigate the observed patterns in the PCA and ADMIXTURE analyses, we used qp3Pop v651 and qpDstat v980 from AdmixTools v7.0.2<sup>45</sup>.

First, we tested the different affinities of the Ban Chiang Bronze Age individuals<sup>49</sup> and the ancient individuals from northwestern Thailand to all published ancient groups from East Asia<sup>49</sup> with  $f_4(\text{Mbuti.DG}, \text{ancient East Asia}; \text{Thailand\_BA}, \text{Thailand\_1700BP})$ . While there



were no ancient groups that showed closer affinity to the Bronze Age individuals, ancient individuals from northern East Asia (e.g., Miaozigou\_MN, Upper\_YR\_IA, WLR\_LN, Mongolia\_EIA\_7\_Xiognu, Upper\_YR\_LN, YR\_LBIA, Mongolia\_EIA\_3, Mongolia\_LBA\_2\_Ulaanzukh) and Nepal (Mebrak\_2125BP.SG, Samdzong\_1500BP.SG) resulted in significantly positive values indicating extra genetic affinity present in the ancient individuals from northwestern Thailand (Supplementary Figure 7A). We also observed this pattern when comparing with the Iron Age individual from Ban Chiang<sup>55</sup>, albeit with a lower number of overlapping SNPs.

Tests with Long Long Rak in the third position showed significant attraction of the ancient individuals from Thailand to Neolithic individuals from Vietnam, Tanshishan, Mongolia\_EIA\_8, Bianbian, and Boisman\_MN, while Vietnam\_BA\_DongSonCulture.SG was significantly closer to Long Long Rak (Supplementary Figure 7B).

The statistic  $f_3(\text{Mbuti}; \text{present-day groups, Thailand 1700BP})$  indicated that the ancient individuals from northwestern Thailand shared most drift with the Austroasiatic-speaking Mlabri and Htin from Thailand<sup>53</sup> followed by the Kra-Dai-speaking Maonan and Gelao from southern China<sup>44</sup> (Supplementary Figure 8). The tested groups also included present-day groups from Thailand<sup>53</sup>, Malaysia<sup>49</sup> and other South and East Asian groups<sup>45,46</sup>.

The results of  $f_4(\text{Mbuti, present-day Thai; Bronze Age Ban Chiang, Thailand 1700BP})$  showed a closer affinity of all present-day groups from Thailand<sup>57</sup> to the newly genotyped ancient individuals compared to the Bronze Age individuals from northeastern Thailand<sup>49</sup>. The values were particularly high for the Palaung, Skaw Karen, Iu Mien, and Shan, whereas tests including the Khmu, Kalueang, Soa, Austronesian speakers from southern Thailand, Htin Pray and Bru were not significant (Supplementary Figure 9).

We also tested if any present-day groups from Thailand showed closer affinity to the Log Coffin-associated genomes compared to other present-day groups ( $f_4(\text{Mbuti, present-day Thailand; Thailand 1700BP, present-day groups})$ ). Here, we were particularly interested in the hill tribes living in the highlands (Karen, Lahu, Lisu, Akha, Lawa, Htin, Khmu, Hmong, and Iu Mien), who are culturally and genetically isolated from the lowlanders<sup>44</sup>, and chose the Sino-Tibetan-speaking Padaung Karen, the Austroasiatic-speaking Htin Pray and the Hmong Mien-speaking Hmong Njua for statistical testing based on close geographical or position in PCA space. Since Tai-Kadai languages have been hypothesized to have spread from southern

China in the Iron Age<sup>54</sup>, we repeated the same test with the Tai-Kadai speaking Shan in the second position, who live near the excavation site.

The results indicate that the Log Coffin-associated individuals are closer to the tested groups when compared with the Onge, Jehai, and Salar, while the tested groups are all genetically closer to other groups of the same language family than to the ancient genomes (Supplementary Figure 10). Since there is no present-day group or groups from one language family, which show a closer genetic affinity to the ancient genomes, we conclude that the present-day genetic diversity in Thailand is the result of different demographic processes than captured by the ancient genomes from northwestern Thailand.

### qpAdm

We then used the  $f_4$  results to inform ancestry modelling with qpAdm v1520 from AdmixTools v7.0.2<sup>45</sup>.

We first tried to model the ancient individuals from Thailand as an admixture between Bronze Age Ban Chiang and ancient individuals from northern East Asia, which showed the strongest affinity to the Log Coffin-associated individuals (Miaozigou\_MN, Upper\_YR\_IA, WLR\_LN, Mongolia\_EIA\_7\_Xiogou, Upper\_YR\_LN, YR\_LBIA, Mongolia\_EIA\_3, Mongolia\_LBA\_2\_Ulaanzukh, Mebrak\_2125BP.SG, Samdzong\_1500BP.SG, Fig. S8). The published ancient individuals were rotated from left (sources) to right (references). The tested groups from the Yellow River region resulted in well-fit models estimating around 50% northern East Asian ancestry in the ancient individuals from northwestern Thailand (Supplementary Figure 11).

A parallel setup modelling the ancestry of the Long Long Rak individuals resulted in very similar results. However, the amount of Bronze Age Ban Chiang ancestry was slightly higher in all models.

Setting the Iron Age individuals from Ban Chiang as the target, however, showed that no genetic input from for northern East Asian groups was required (nested  $p > 0.05$ ).

Initially it was suggested that all Neolithic Southeast Asian individuals, including the individuals from Ban Chiang, represent an admixture between deep Onge- (hunter-gatherer) and Ami-related (early farmer) groups<sup>44</sup>, where Hòabìnhian-associated individuals from Laos and Malaysia were the best proxies for the hunter-gatherer ancestry<sup>54</sup>. However, later studies showed that another hunter-gatherer source, Longlin from Guangxi Province<sup>52</sup>, produced

better fitting models for Neolithic Man Bac in Vietnam and some other ancient Southeast Asians, while others retained Hòabìnhian-associated influence<sup>44,49,52,54</sup>. To account for this diverse ancestry set present in the region and assess the predominant genetic influence for all ancient groups from Thailand, we tested models using both the Hòabìnhian-associated individual from Laos<sup>58</sup> and the Longlin individual from Guangxi<sup>44</sup> as the hunter-gatherer source or in the reference groups. As representatives for early farmer ancestry we rotated the Neolithic individuals from southern China<sup>49</sup> from sources to references.

For the newly genotyped individuals using the Hòabìnhian-associated individual as the hunter-gatherer source resulted in better fitting models, i.e., higher p-values, although most models with the Longlin individual did fit ( $p > 0.05$ ). The best-fitting model estimated  $36.0 \pm 3.1\%$  Hòabìnhian-related and  $64.0 \pm 3.1\%$  Suogang-related ancestry (Supplementary Figure 12).

The Hòabìnhian-associated individual was also the preferred source for Long Long Rak with the best-fitting model favouring Liangdao1 as the southern East Asian source.

The models for the Iron Age individuals from Ban Chiang resulted in larger hunter-gatherer ancestry components compared to the Iron Age groups from northwestern Thailand and the differentiation between models including the Hòabìnhian and Longlin as sources was less pronounced.

Bronze Age Ban Chiang was best modelled as an admixture between Longlin and Qihe.

Overall, the best-performing southern East Asian source varied, but most models involving Liangdao1 showed sufficient model-fit, and hence was selected for further modelling.

Combining the information from the previous two analyses, we simultaneously estimated ancestry proportions related to hunter-gatherers, early farmers, and northern East Asians. We set the Hòabìnhian-associated individual from Laos and Liangdao1 as sources, while the remaining southern East Asian individuals were added to the references, and all northern East Asians rotated from references to sources (Supplementary Figure 13, Supplementary Data 5). The model could not assign any ancestry related to northern East Asia in the Bronze and Iron Age individuals from Ban Chiang, whereas the new individuals from northwestern Thailand showed a large component related to Upper\_YR\_LN, Miaoziyou\_MN, and Upper\_YR\_IA. However, using this setup, we could not find a sufficiently well-fit model for Long Long Rak ( $p < 0.05$ ).

To investigate the geographical spread of the northern East Asian-related ancestry, we attempted to fit most previously published individuals as an admixture between a Longlin/Hòabìnhian-associated individual, Liangdao1 and northern East Asian<sup>50</sup>. While we could model Laos\_Unknown.SG, Thailand 1700BP, Vietnam\_Historical.SG and Vietnam\_LN.SG as Longlin, Liangdao1 and Upper Yellow River, the models with the Hòabìnhian-associated individual achieved higher p-values and fit for more northern East Asian groups. Our analyses indicate that hunter-gatherer-related ancestry is not necessary for the 1.5-0.5-ka-old individuals from Guangxi (BaBanQinCen, GaoHuaHua, LaCen, Shenxian and Yiyang), Myanmar\_LN\_BA\_lc from Oakaie and Vietnam\_BA from Nui Nap (nested model 100  $p > 0.05$  or negative admixture proportion). Conversely, northern East Asian ancestry is not required for modelling Baojianshan, Dushan, Thailand\_BA, Thailand\_IA (Ban Chiang) and Vietnam\_N (Man Bac) (nested model  $p > 0.05$ ).

We repeated the analysis described above for the newly sequenced genomes individually using the Hòabìnhian-associated individual, Liangdao1, and Upper YR LN as respective proxies with the remaining ancient genomes as references (Mbuti.DG, Longlin, Qihe, Xitoucun, Tanshishan, Suogang, Miaozigou\_MN, Mongolia\_EIA\_3, Mongolia\_EIA\_7\_Xiongnu, Mongolia\_LBA\_2\_Ulaanzukh, Nepal\_Mebrak\_2125BP.SG, Nepal\_Samdzong\_1500BP.SG, Upper\_YR\_IA, WLR\_LN, YR\_LBIA). We could show some variation in hunter-gatherer, early farmer, and northern East Asian ancestry, but overall, the proportions were similar, around 15%, 45%, and 40%, respectively (Supplementary Figure 14). The lowest northern East Asian component was estimated for THL003, while THL004 could also be modelled without any hunter-gatherer contribution (nested  $p > 0.05$ ). This ancestry model did not fit for YPN010, YPN017, and YPN030 ( $p < 0.05$ ).

### ADMIXTURE

We estimated proportions of ancestry using *ADMIXTURE* v.1.3.0<sup>59</sup> on a dataset containing present-day and ancient individuals from East and Southeast Asia<sup>44-46,49,50,52-54,59</sup>. For  $K=8$  the corresponding ancestry components are maximized in the following groups: Mbuti (pink), French (light blue), Onge (dark green), Mlabri (dark purple), Atayal (red), Lahu (light green), Htin Pray (yellow), Boisman\_MN (orange; Supplementary Figure 16). The newly generated genomes from northwestern Thailand are split into ~65% Boisman and ~35% Htin Pray ancestry component with most individuals additionally containing a few per cent Mlabri, and some individuals very little French or Atayal ancestry. This is similar to the individuals from

Long Long Rak (Thailand\_IA.SG/WGC)<sup>49</sup>, but differs from Ban Chiang (Thailand\_BA)<sup>60</sup>, which lack the Boisman and contain more Lahu component.

#### Sex-biased admixture

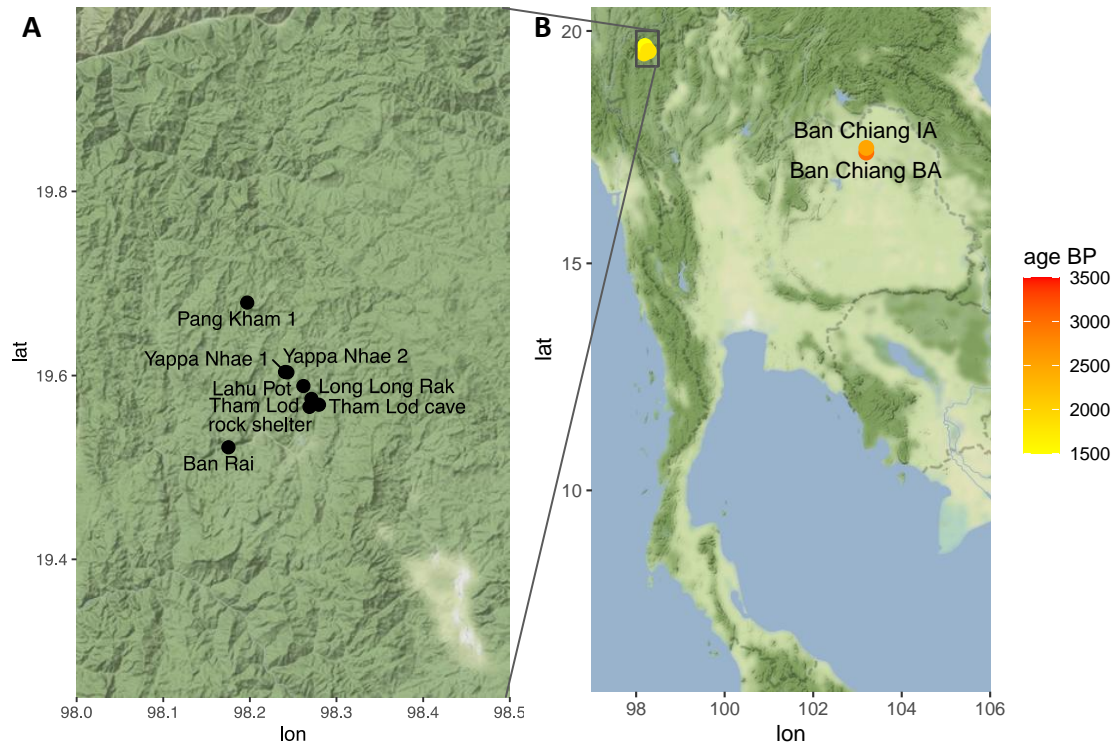
Due to the lack of X-chromosome coverage in individuals from northern East Asia<sup>43</sup>, we were not able to conduct the qpAdm analyses exclusively on X-chromosome data to assess differences in admixture proportions on the autosome and X-chromosome potentially related to sex-biased admixture.

An ADMIXTURE analysis including present-day individuals from Thailand was similarly inconclusive.

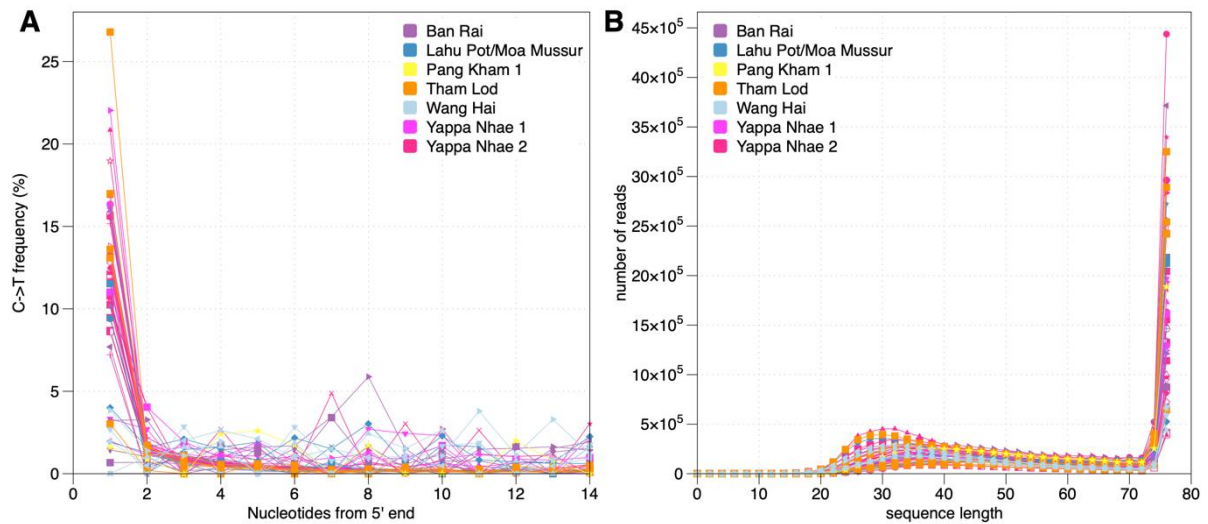
#### Admixture time dating

We used DATES v753<sup>60</sup> to estimate the time of admixture between the hunter-gatherer/Yangtze River, hunter-gatherer/Yellow River and Yangtze River/Yellow River components. However, the linkage disequilibrium decay curves were flat and the resulting standard errors very large, when using Laos\_Hoabinhian, Liangdao1 and Upper\_YR\_LN as representatives for these ancestry components.

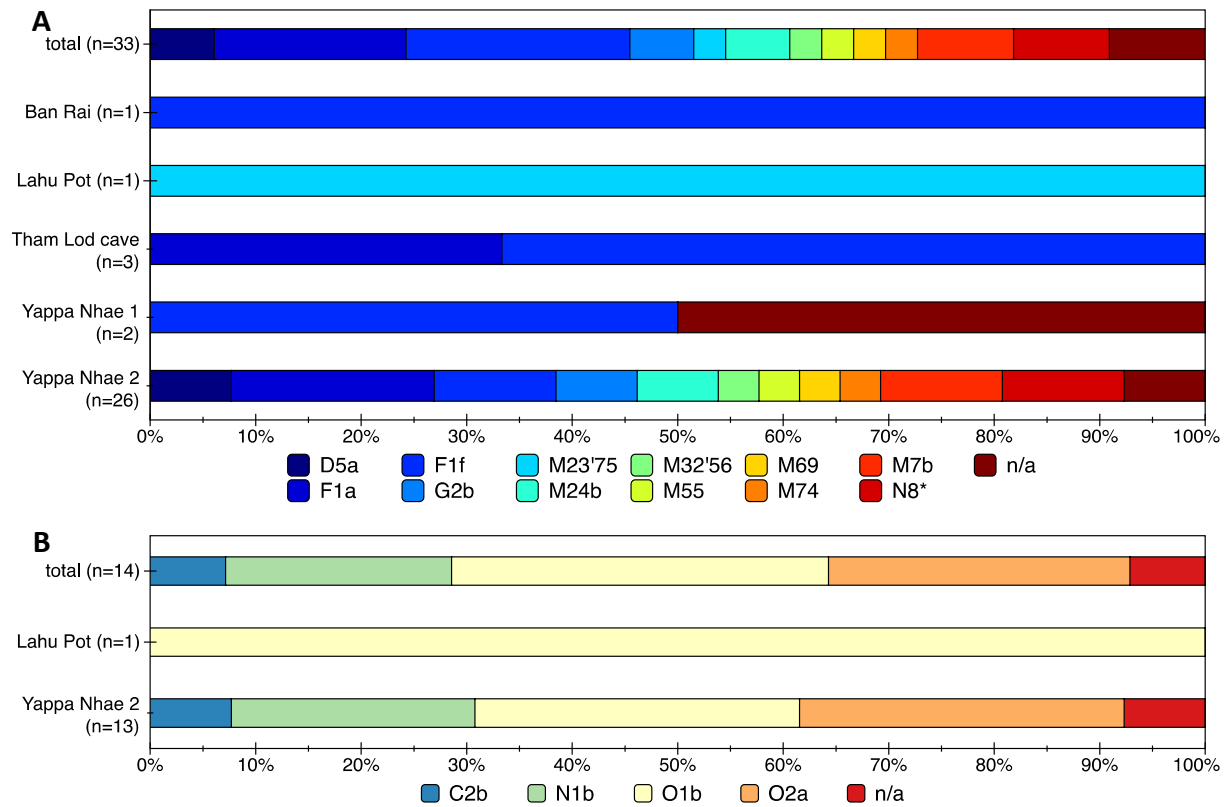
## Supplementary Figures



**Supplementary Figure 1.** Maps of archaeological sites used in this study; **(A)** the location of the archaeological sites from which data was produced for this study and Long Long Rak<sup>53</sup>; **(B)** the location of archaeological sites in Thailand for which ancient DNA data has been published<sup>49</sup>; map created with ggmap v3.0.0<sup>61</sup>, map tiles by Stamen Design (CC BY 3.0), data by OpenStreetMap (ODbL).

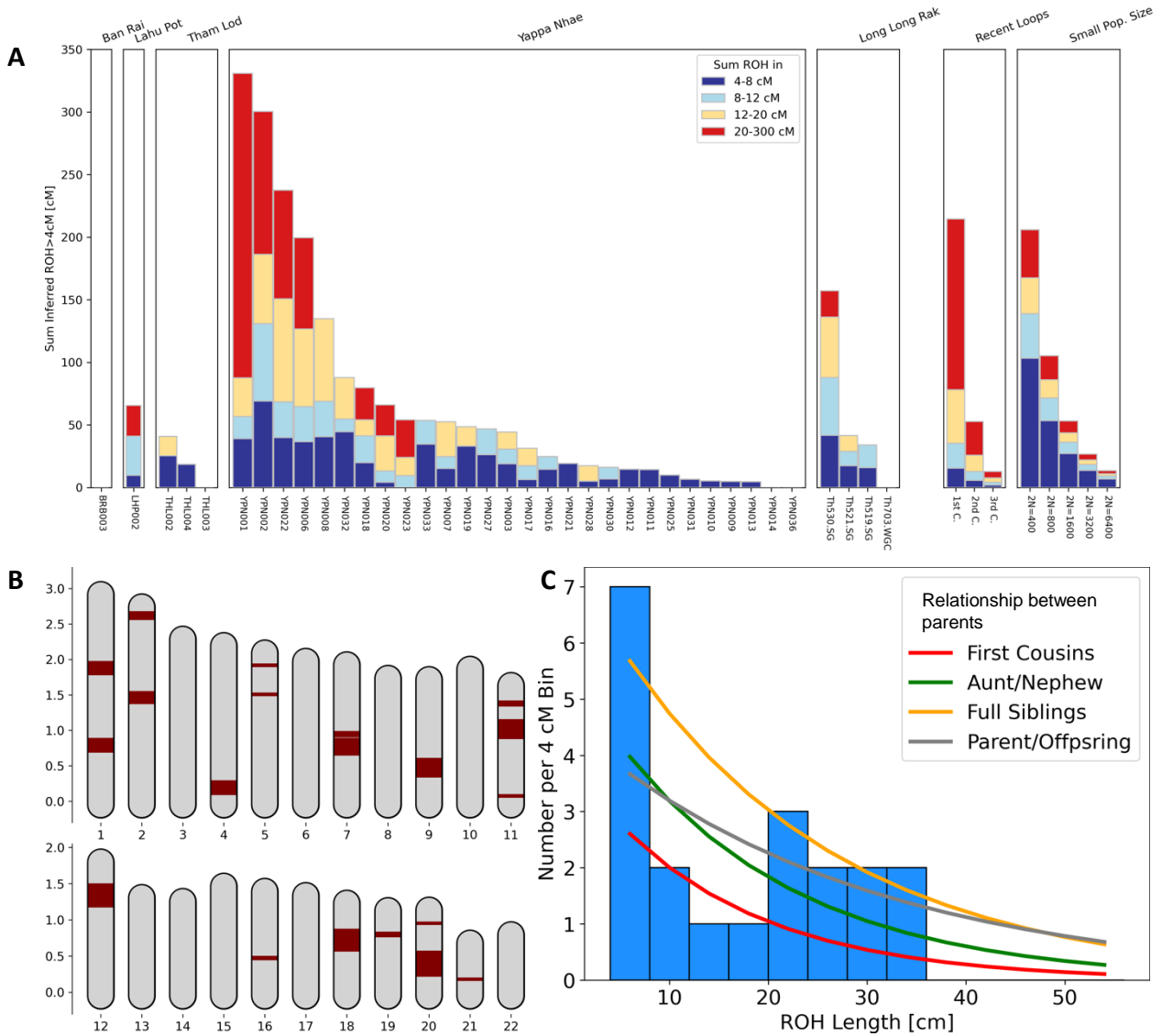


**Supplementary Figure 2.** Quality control plots of ancient genomes from seven archaeological sites in northwestern Thailand; **A**, frequency of C-T transitions at the 5' end of the DNA fragments; **B**, read length distribution; plotted with DataGraph v4.4 (Visual Data Tools, Inc., <http://www.visualdatatools.com/DataGraph/>). Source data are provided as a Source Data file.

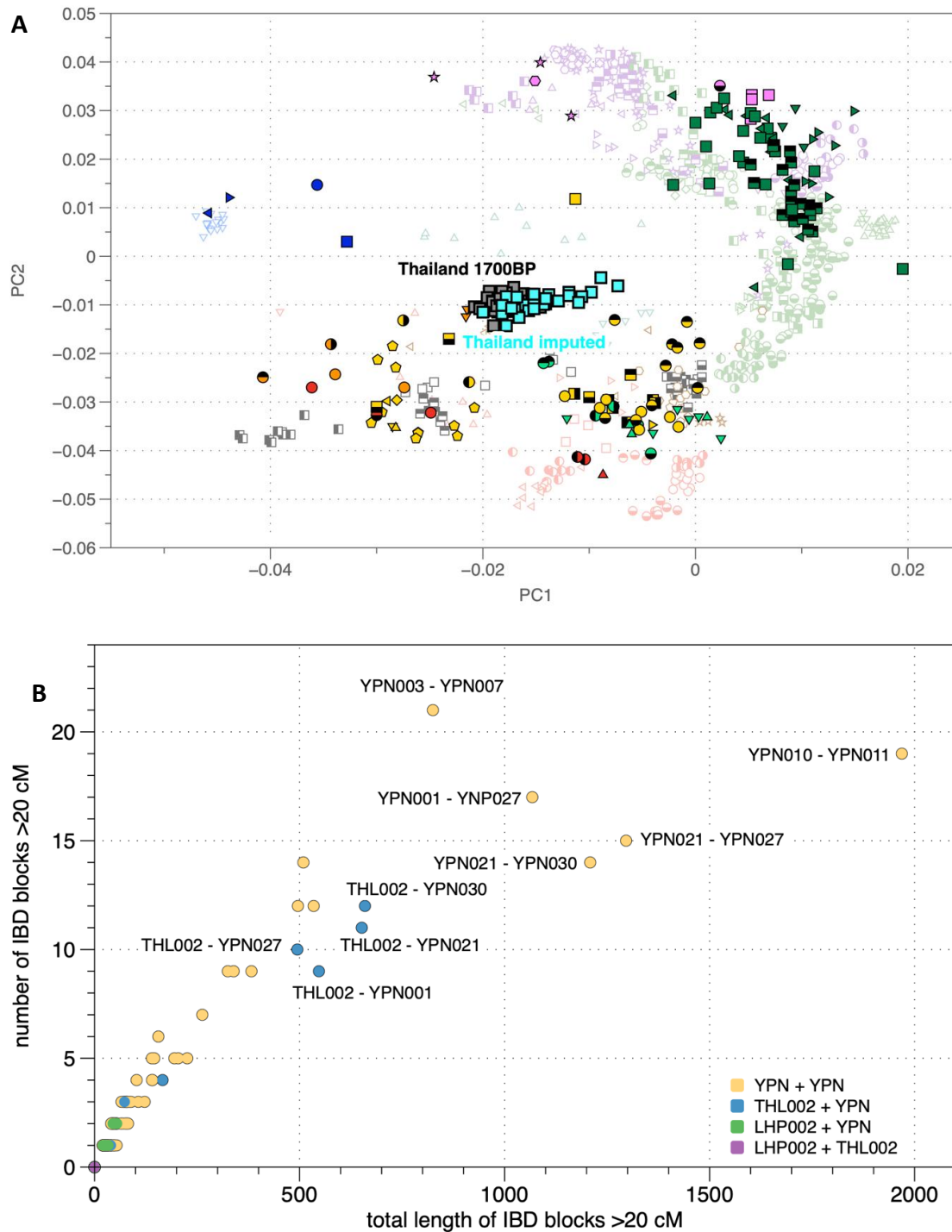


**Supplementary Figure 3.** The diversity of mitochondrial (A) and Y-haplogroups (B) of ancient Log Coffin-associated genomes from northwestern Thailand represented pooled across sites as well as for individual sites, plotted with DataGraph v4.4 (Visual Data Tools, Inc., <http://www.visualdatatools.com/DataGraph/>). Source data are provided as a Source Data file.

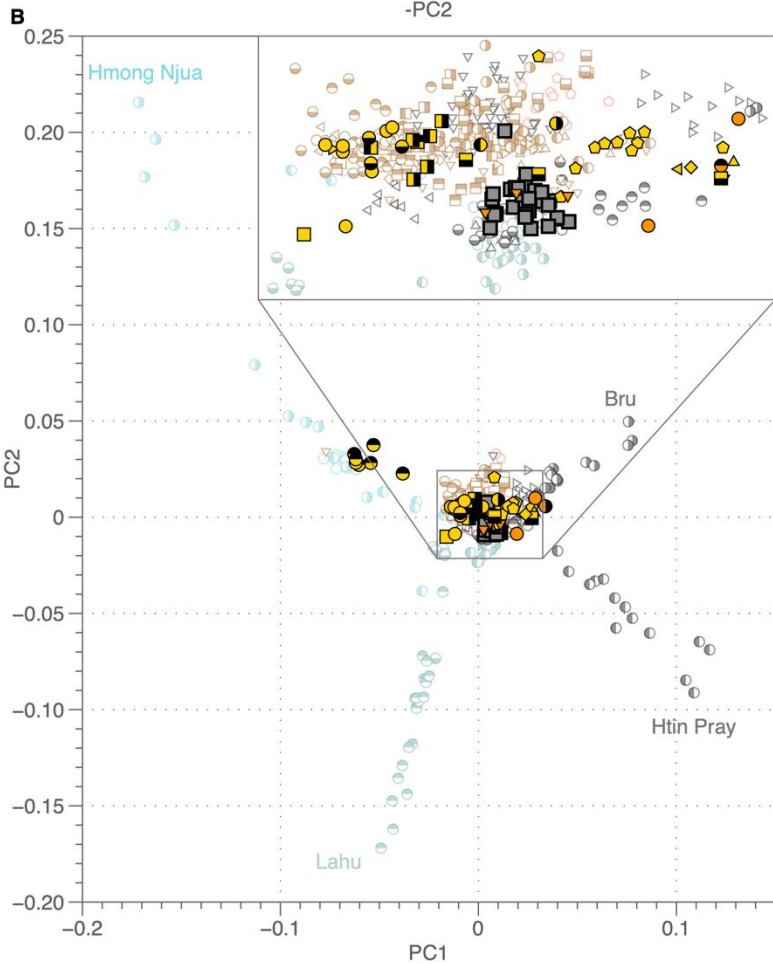
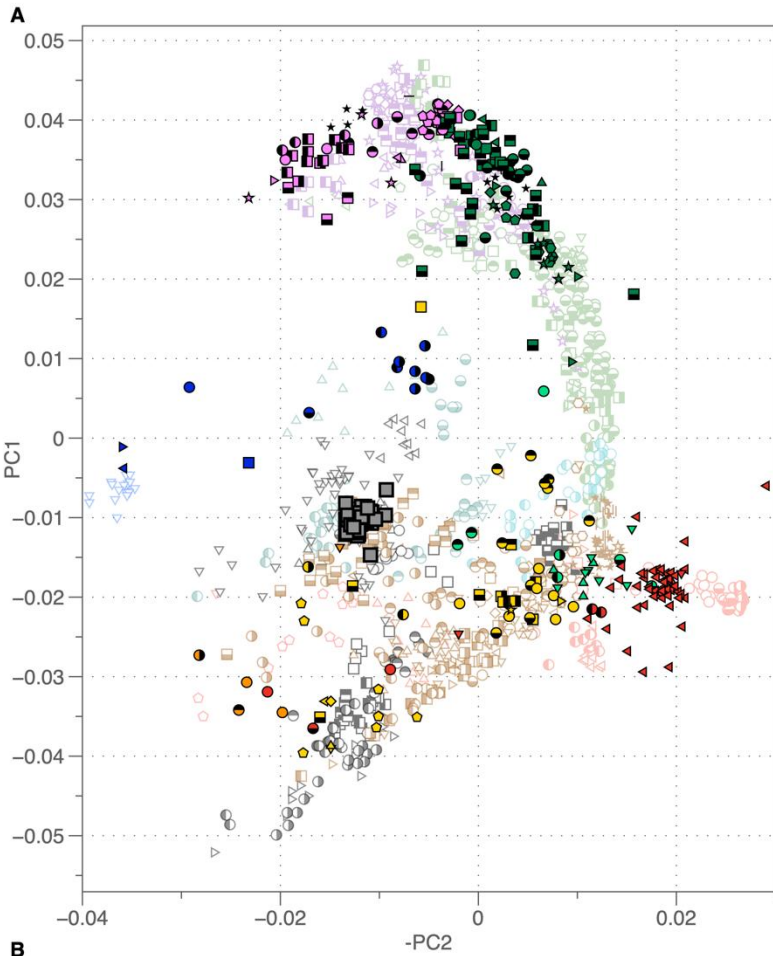




**Supplementary Figure 4.** Runs of homozyosity (ROH) of ancient individuals from northwestern Thailand: (A) lengths of ROH for each ancient individual, previously simulated data expected for different parental relationships and population sizes displayed on the right<sup>49</sup>, (B) location of ROH within the karyotype of YPN001 and (C) ROH length distribution of YPN001 (blue bars) with lines depicting the “expected” ROH trajectory of relationships between parents superimposed. Plotted with hapROH v0.6<sup>53</sup>, source data are provided as a Source Data file.

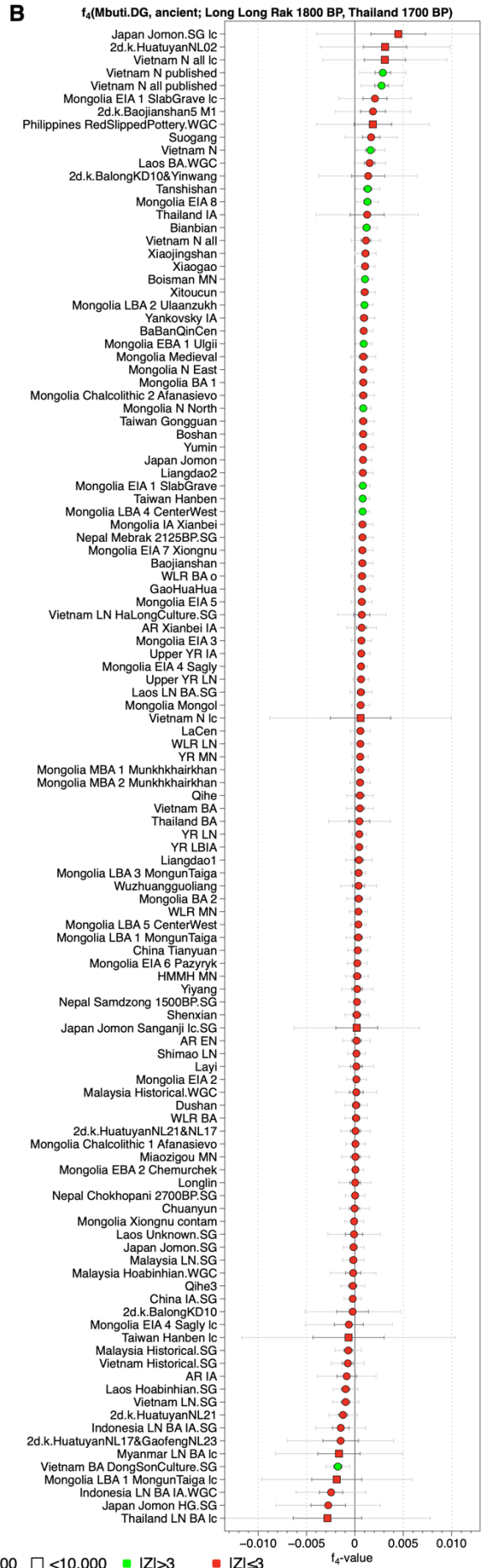
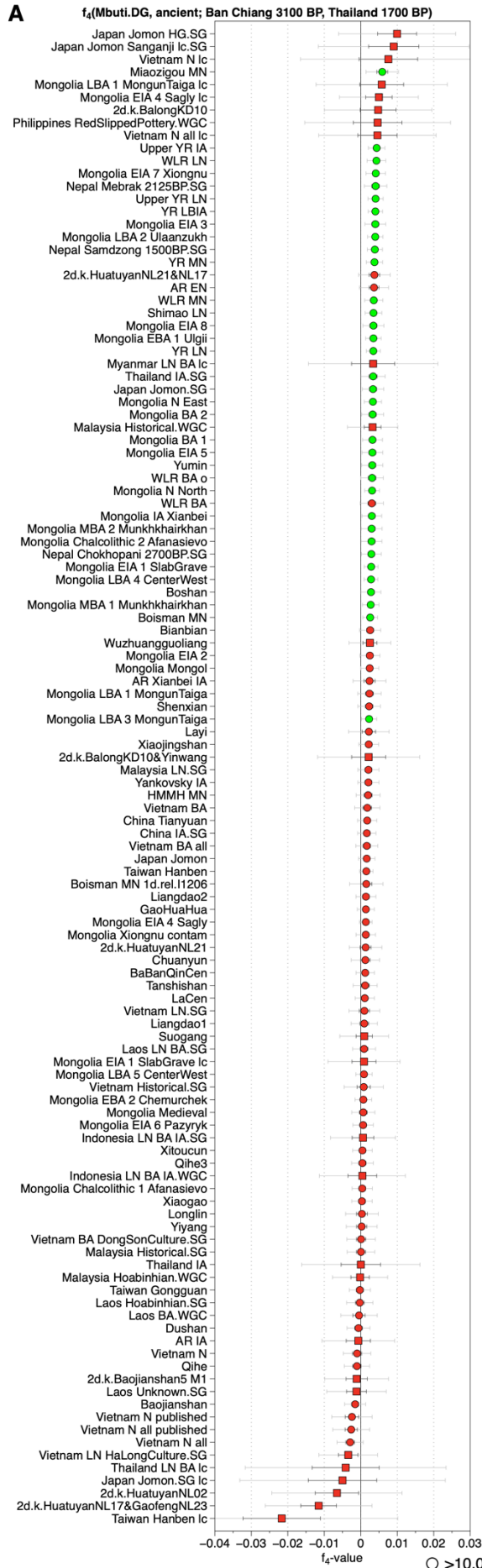


**Supplementary Figure 5.** (A) Principal component analyses calculated on present-day East and Southeast Asian individuals<sup>45,46,48,49,55</sup> showing the position of the Log Coffin-associated genomes before (grey squares) and after imputation (turquoise squares), full legend as in Supplementary Figure 6, source data are provided as a Source Data file; (B) Shared IBD blocks between newly generated Log Coffin-associated individuals from Thailand coloured by genetic relationships between individuals from different burial sites (Supplementary Data 4); plotted with DataGraph v4.4 (Visual Data Tools, Inc., <http://www.visualdatatools.com/DataGraph/>).

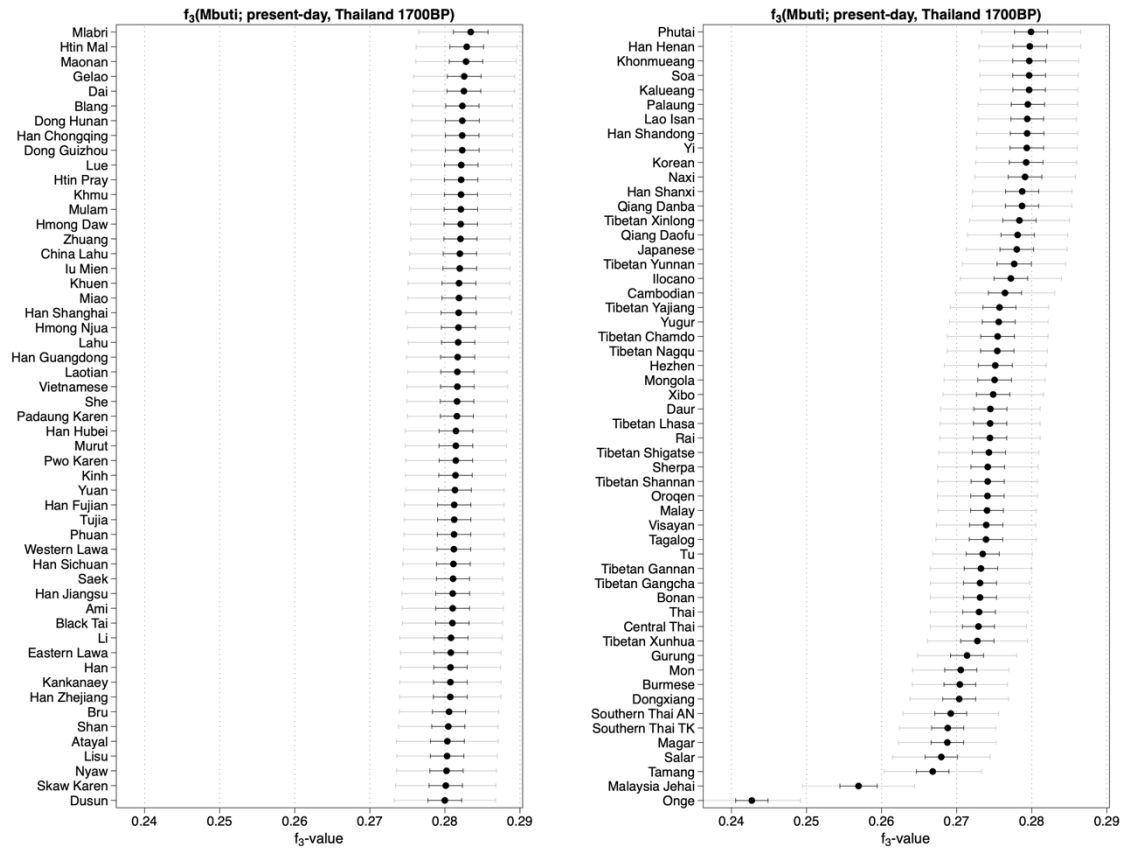


- Thailand 1700 BP
- China Tianyuan
- Japan Jomon
- Japan Jomon HG
- ▲ Laos Hoabinhian
- ▲ Malaysia Hoabinhian
- Longlin
- Amur River EN
- Amur River IA
- Amur River Xianbei IA
- Bianbian
- Boisman MN
- ▲ Boshan
- ▼ HMMH MN
- ▲ Miaozigou MN
- ▲ Upper YR IA
- Upper YR LN
- West Liao River BA
- West Liao River BA o
- West Liao River LN
- West Liao River MN
- Wuzhuangguoliang
- ◆ Xiaogao
- Xiaojingshan
- Yellow River LBIA
- ★ Yellow River LN
- ★ Yellow River MN
- Yiyang
- Yumin
- Shimao LN
- Mongolia BA 1
- Mongolia BA 2
- Mongolia Chalcolithic 2 Afanasievo
- Mongolia EBA 1 Ulgii
- Mongolia EIA 1 SlabGrave
- Mongolia EIA 3
- Mongolia EIA 7 Xiongnu
- Mongolia EIA 8
- Mongolia IA Xianbei
- Mongolia LBA 1 MongunTaiga
- Mongolia LBA 2 Ulaanzukh
- Mongolia LBA 4 CenterWest
- Mongolia MBA 1 Munkhkhairkhan
- Mongolia Medieval
- Mongolia Mongol
- Mongolia N East
- Mongolia N North
- Nepal Chokhopani 2700 BP
- ★ Nepal Mebrak 2125 BP
- ★ Nepal Samdzong 1500 BP
- Chuanyun
- Liangdao1
- Liangdao2
- Qihe
- Qihe3
- Suogang
- ▲ Tanshishan
- ▼ Xitoucun
- ★ Shenxian
- BaBanQinCen
- Baojianshan
- Dushan
- GaoHuaHua
- LaCen
- ▲ Laos BA
- ▼ Laos LN BA
- ▼ Laos Unknown
- ▼ Layi
- Myanmar LN BA Ic
- Thailand Ban Chiang 3100BP
- Thailand Ban Chiang 2500BP
- ▼ Thailand Long Long Rak 1800BP
- Thailand Ban Chiang LNBA
- Vietnam BA
- Vietnam BA DongSonCulture
- Vietnam Historical
- Vietnam LN
- Vietnam LN HaLongCulture
- Vietnam N
- Vietnam N all
- Vietnam N published
- Indonesia LN BA IA 1
- Indonesia LN BA IA 2
- Malaysia Historical 1
- Malaysia Historical 2
- Malaysia LN
- ▲ Philippines RedSlippedPottery
- ▼ Taiwan Gongguan
- ▼ Taiwan Hanben
- Ami
- Atayal
- Dusun
- Ilocano
- Kankanaey
- Malay
- Malaysia Jehai
- Murut
- Southern Thai AN
- Tagalog
- Visayan
- Burmese
- China Lahu
- Lahu
- Lisu
- Padaung Karen
- Pwo Karen
- Skaw Karen
- Blang
- Bru
- Cambodian
- Eastern Lawa
- Htin Mal
- HtinPray
- Khmu
- Kirih
- Miabri
- Mon
- Palaung
- Soa
- Vietnamese
- Western Lawa
- Black Tai
- Central Thai
- Dai
- Gelao
- Kaluengang
- Khonmueang
- Khuen
- Lao Isan
- Laotian
- Li
- Maonan
- Mulam
- Nyaw
- Phuan
- Phutai
- Saek
- Shan
- Southern Thai TK
- Thai
- Yuan
- Zhuang
- Onge
- Hmong Daw
- Hmong Njua
- Iu Mien
- Bonan
- Dong Guizhou
- Dong Hunan
- Dongxiang
- Han
- Han Chongqing
- Han Fujian
- Han Guangdong
- Han Henan
- Han Hubei
- Han Jiangsu
- Han Shandong
- Han Shanghai
- Han Shanxi
- Han Sichuan
- Han Zhejiang
- Japanese
- Korean
- Magar
- Miao
- Naxi
- Qiang Danba
- Qiang Daofu
- She
- Tu
- Tujia
- Yi
- Yugur
- Daur
- Gurgun
- Hezhen
- Mongola
- Oroqen
- Rai
- Salar
- Sherpa
- Tamang
- Tibetan Chamdo
- Tibetan Gangcha
- Tibetan Gannan
- Tibetan Lhasa
- Tibetan Nagqu
- Tibetan Shannan
- Tibetan Shigatse
- Tibetan Xinlun
- Tibetan Xunhua
- Tibetan Yajiang
- Tibetan Yunnan
- Xibo

**Supplementary Figure 6.** Principal component analyses calculated on (A) present-day East Asian individuals<sup>45,46,48,49,53,55</sup> and (B) present-day individuals from Thailand<sup>49</sup>; plotted with DataGraph v4.4 (Visual Data Tools, Inc., <http://www.visualdatatools.com/DataGraph/>). Source data are provided as a Source Data file.

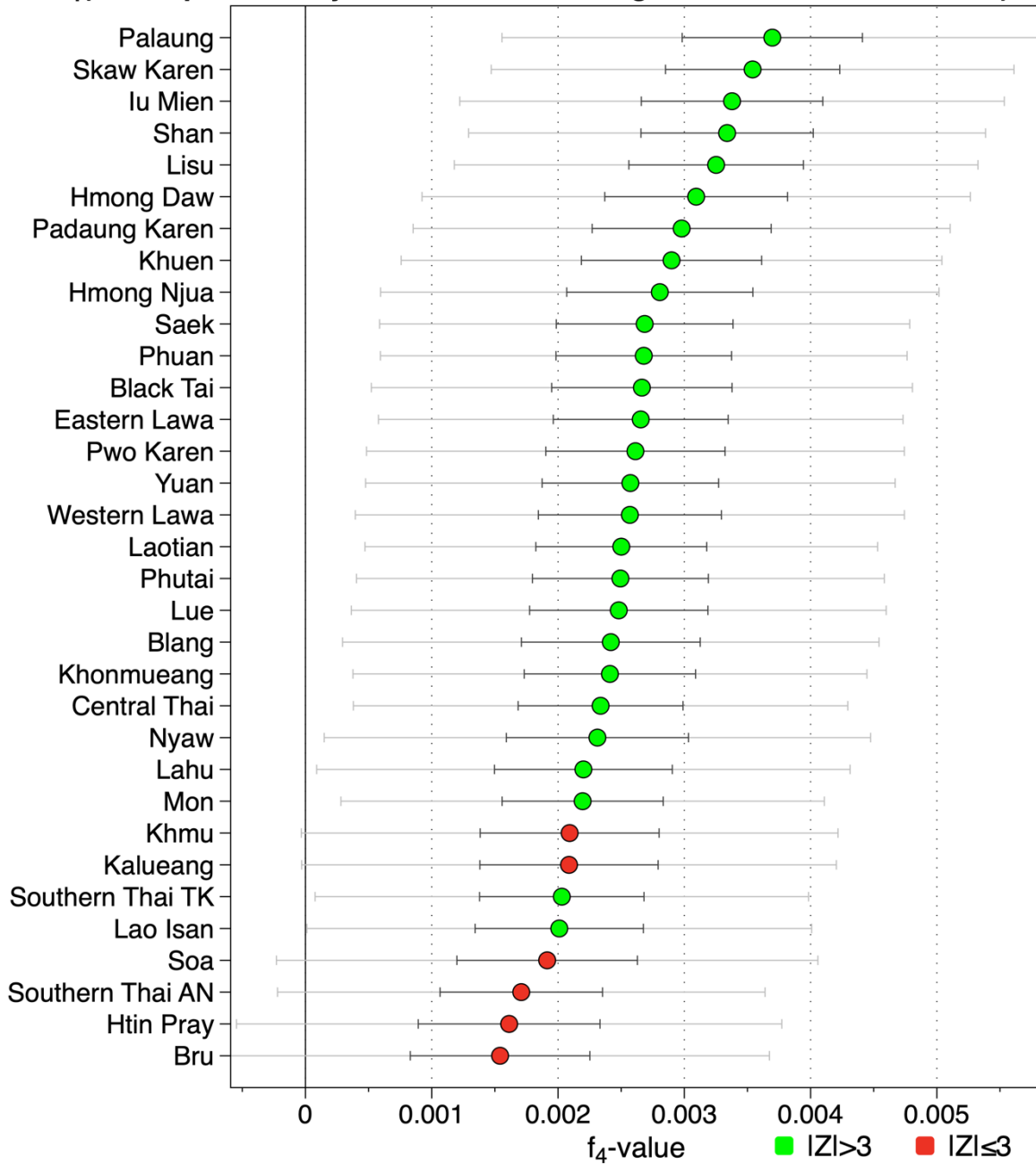


**Supplementary Figure 7.** Observed  $f_4$ -statistics comparing the genetic affinities of the newly genotyped individuals from northwestern Thailand with (A) Bronze Age northeastern Thailand<sup>50</sup> and (B) Iron Age northwestern Thailand<sup>44</sup>; red and green indicate z-scores such that  $|Z| \leq 3$  and  $|Z| > 3$  respectively, circles and squares denote  $f_4$ -statistics where there were more and less than 10,000 overlapping SNPs respectively, dark grey lines represent  $\pm 1$  standard error as calculated by the qpDstat<sup>45</sup>, light grey lines represent  $\pm 3$  standard errors; plotted with DataGraph v4.4 (Visual Data Tools, Inc., <http://www.visualdatatools.com/DataGraph/>). Source data are provided as a Source Data file.



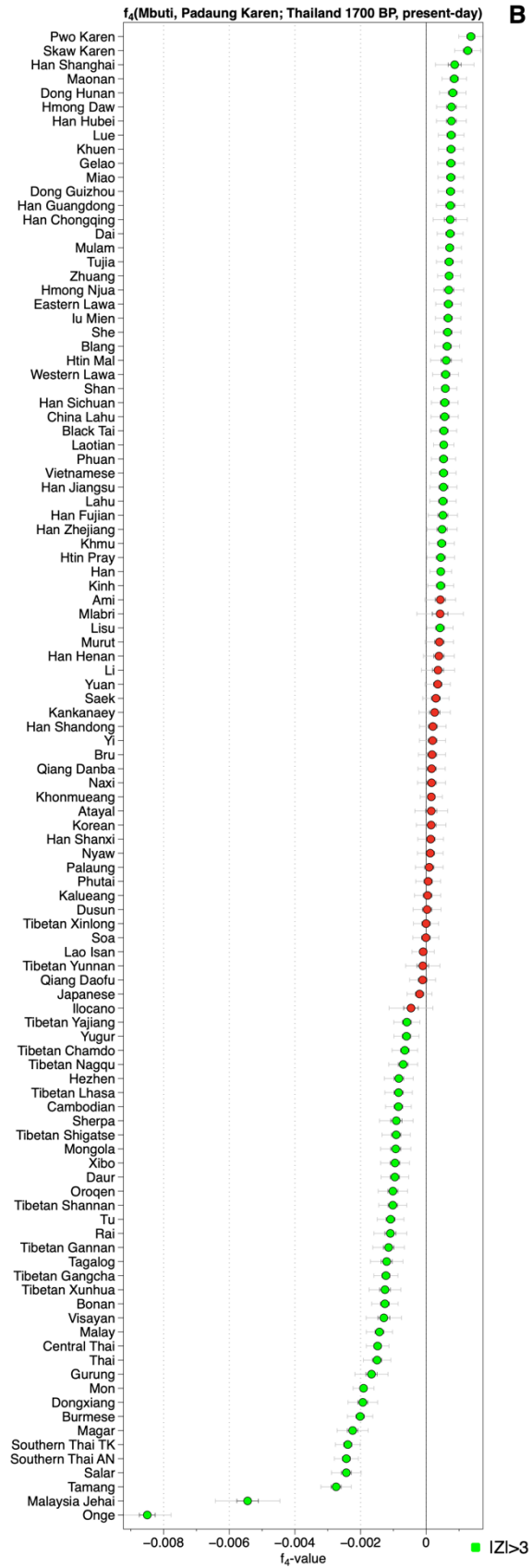
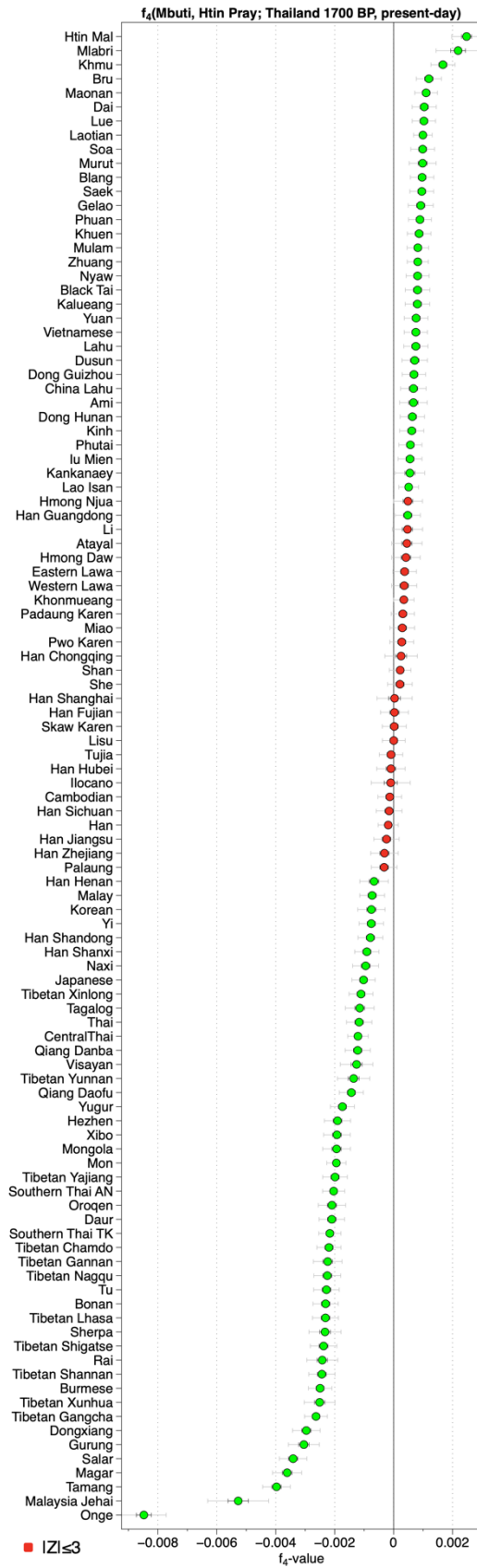
**Supplementary Figure 8.** Observed  $f_3$ -stastics showing shared genetic drift between ancient genomes from northwestern Thailand and present-day groups from the region<sup>44–46,49,53,55</sup>, dark grey lines represent  $\pm 1$  standard error as calculated by the qp3Pop<sup>45</sup>, light grey lines represent  $\pm 3$  standard errors, plotted with DataGraph v4.4 (Visual Data Tools, Inc., <http://www.visualdatatools.com/DataGraph/>). Source data are provided as a Source Data file.

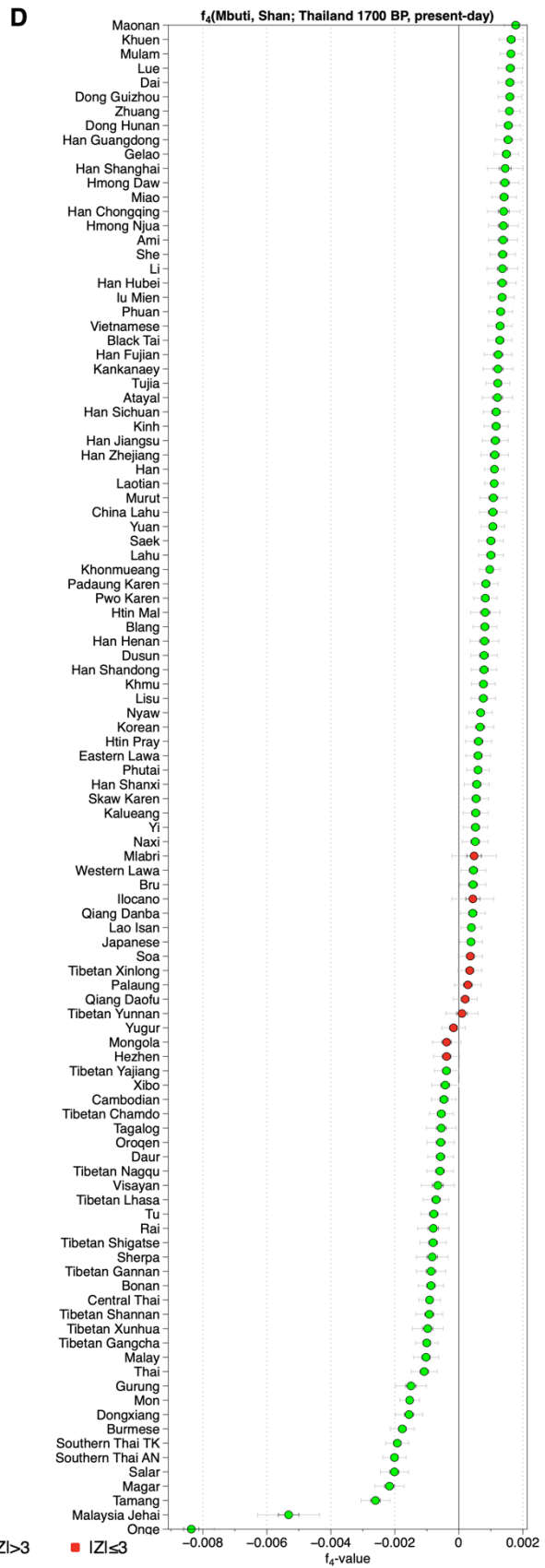
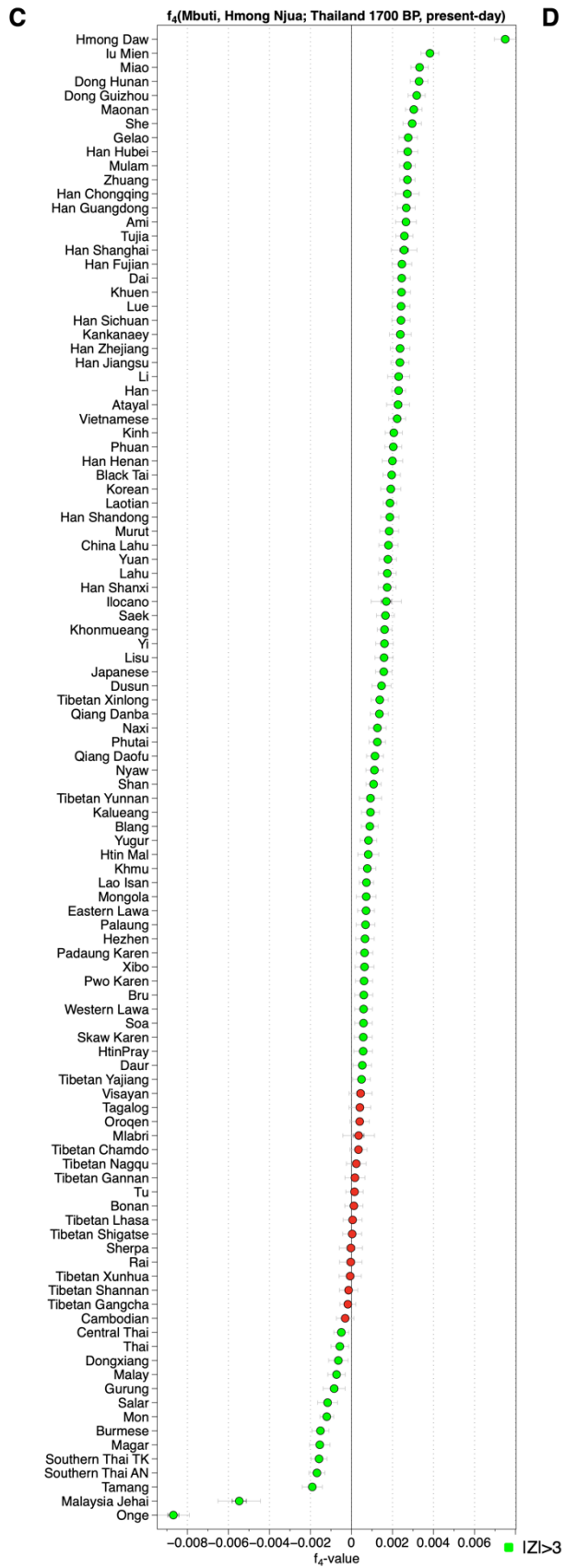
**$f_4$ (Mbuti, present-day Thailand; Ban Chiang 3100BP, Thailand 1700BP)**



**Supplementary Figure 9.** Observed  $f_4$ -statistics comparing the affinity of the newly generated genomes from ancient northwestern Thailand and Bronze Age northeastern Thailand<sup>52</sup> to present-day groups from Thailand<sup>50</sup>, red and green indicate z-scores such that  $|Z| \leq 3$  and  $|Z| > 3$  respectively, dark grey lines represent  $\pm 1$  standard error as calculated by the qpDstat<sup>45</sup>, light grey lines represent  $\pm 3$  standard errors, plotted with DataGraph v4.4 (Visual Data Tools, Inc., <http://www.visualdatatools.com/DataGraph/>). Source data are provided as a Source Data file.

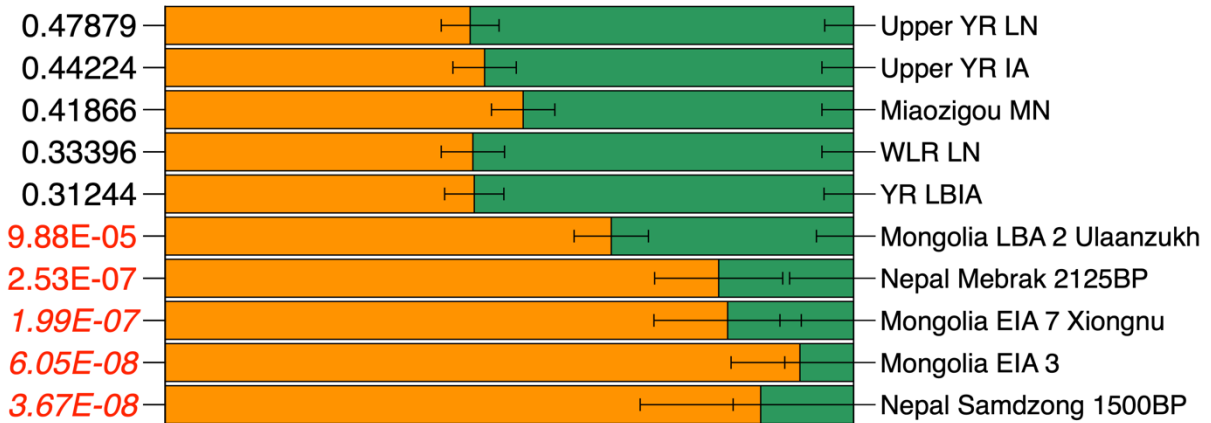


**A****B**

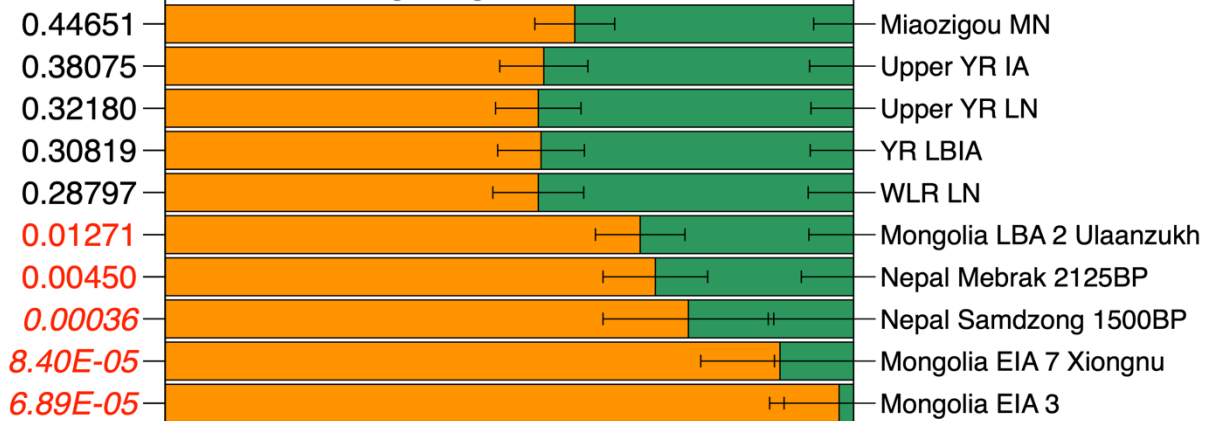


**Supplementary Figure 10.** Observed  $f_4$ -statistics comparing the affinity of the newly generated genomes from ancient northwestern Thailand to present-day groups<sup>44-46,49,53,55</sup> and hill tribes from Thailand<sup>53</sup> speaking Sino-Tibetan (**A**), Austroasiatic (**B**), Hmong Mien (**C**) and Tai-Kadai (**D**) languages, red and green indicate z-scores such that  $|Z| \leq 3$  and  $|Z| > 3$  respectively, dark grey lines represent  $\pm 1$  standard error as calculated by the qpDstat<sup>45</sup>, light grey lines represent  $\pm 3$  standard errors, plotted with DataGraph v4.4 (Visual Data Tools, Inc., <http://www.visualdatatools.com/DataGraph/>). Source data are provided as a Source Data file.

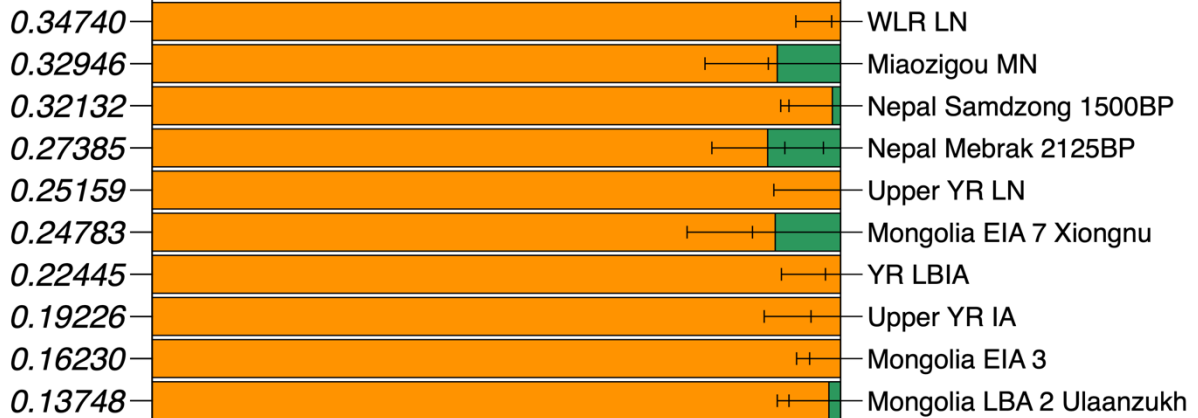
### Thailand 1700 BP



### Long Long Rak 1800 BP

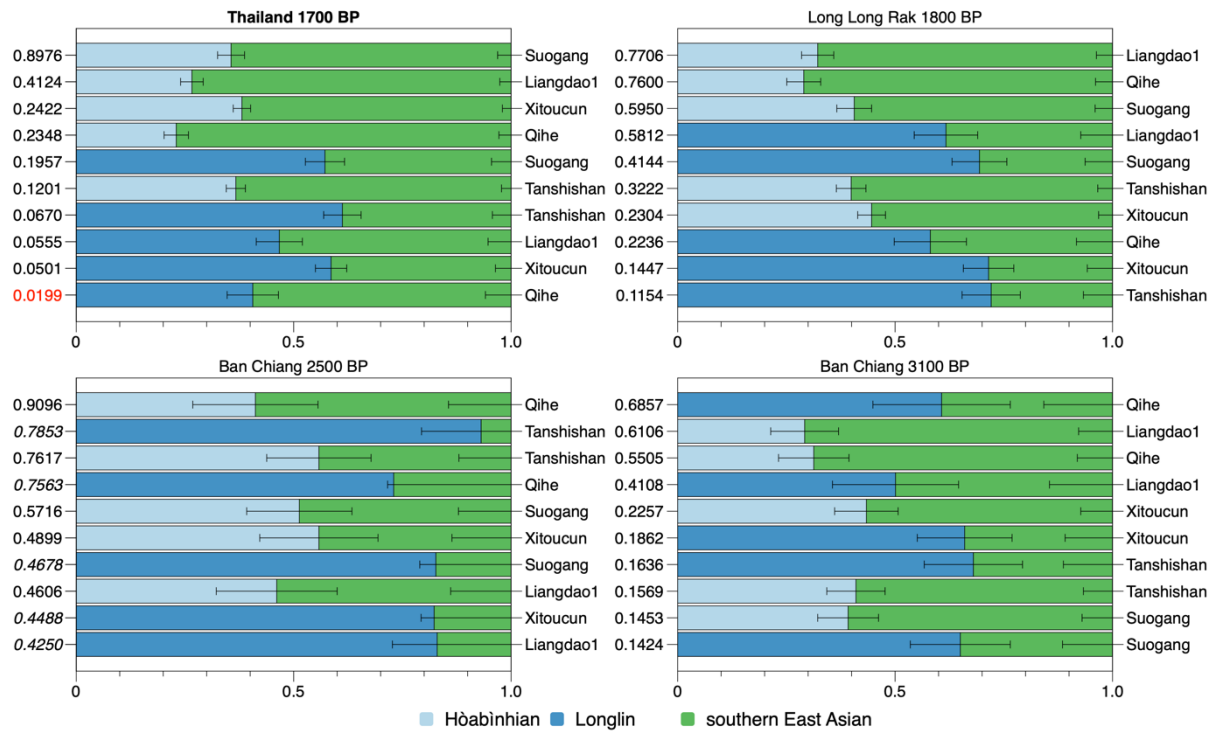


### Ban Chiang 2500 BP

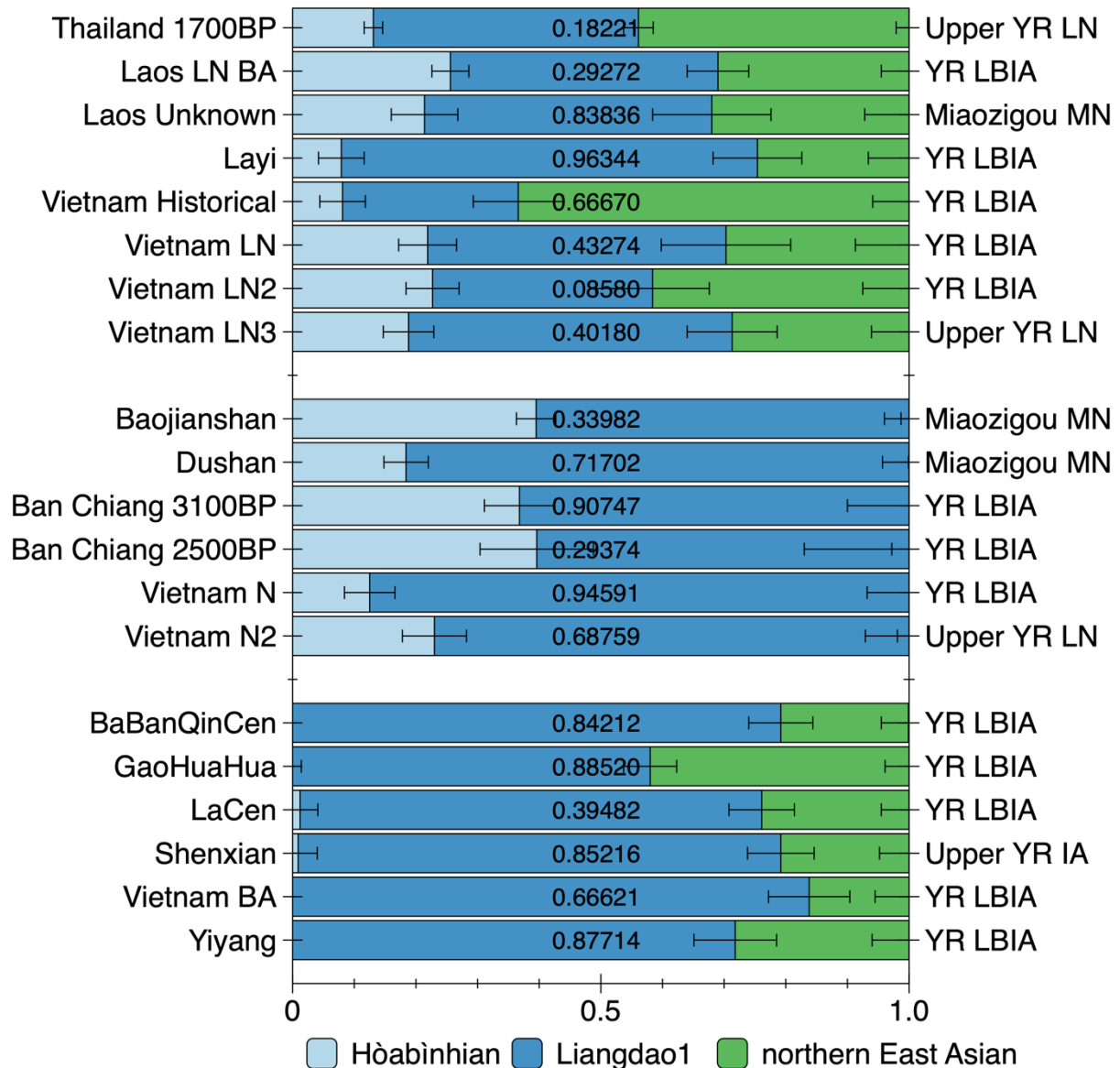


0 0.25 0.50 0.75 1.00  
■ Ban Chiang 3100 BP ■ northern East Asian

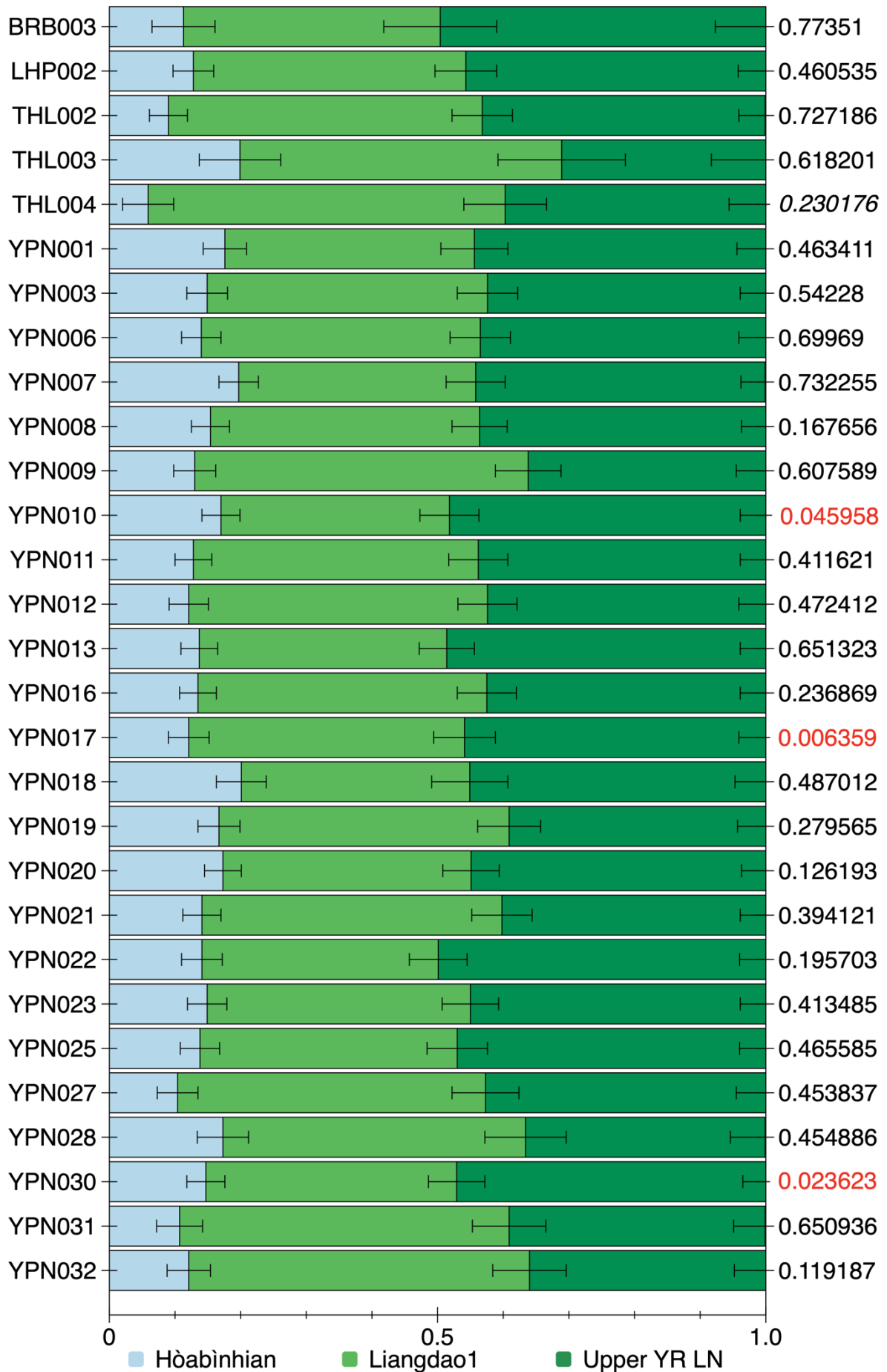
**Supplementary Figure 11.** Results of qpAdm analyses modelling ancient individuals from Iron Age northeastern and -western Thailand<sup>44,49</sup> as an admixture between Bronze Age northeastern Thailand<sup>62</sup> and ancient groups from northern East Asia<sup>50</sup>; p-values are given on the left where red indicates  $p < 0.05$  and italicized values indicate nested  $p > 0.05$ , northern East Asian groups are displayed on the left, error bars in grey show  $\pm 1$  standard error as calculated with block jackknife in qpAdm v.1520<sup>45</sup>; plotted with DataGraph v4.4 (Visual Data Tools, Inc., <http://www.visualdatatools.com/DataGraph/>). Source data are provided as a Source Data file.



**Supplementary Figure 12.** Results of qpAdm analyses modelling the ancestry of ancient genomes from Thailand as an admixture between the Hòabinhian-associated individual from Laos<sup>44</sup>, Longlin<sup>54</sup> and southern East Asian individuals<sup>52</sup>; p-values are displayed on the left, where red indicates  $p < 0.05$  and italicized values indicate nested  $p > 0.05$ , the southern East Asian sources are displayed on the right, error bars in grey show  $\pm 1$  standard error as calculated with block jackknife in qpAdm v.1520<sup>45</sup>; plotted with DataGraph v4.4 (Visual Data Tools, Inc., <http://www.visualdatatools.com/DataGraph/>). Source data are provided as a Source Data file.

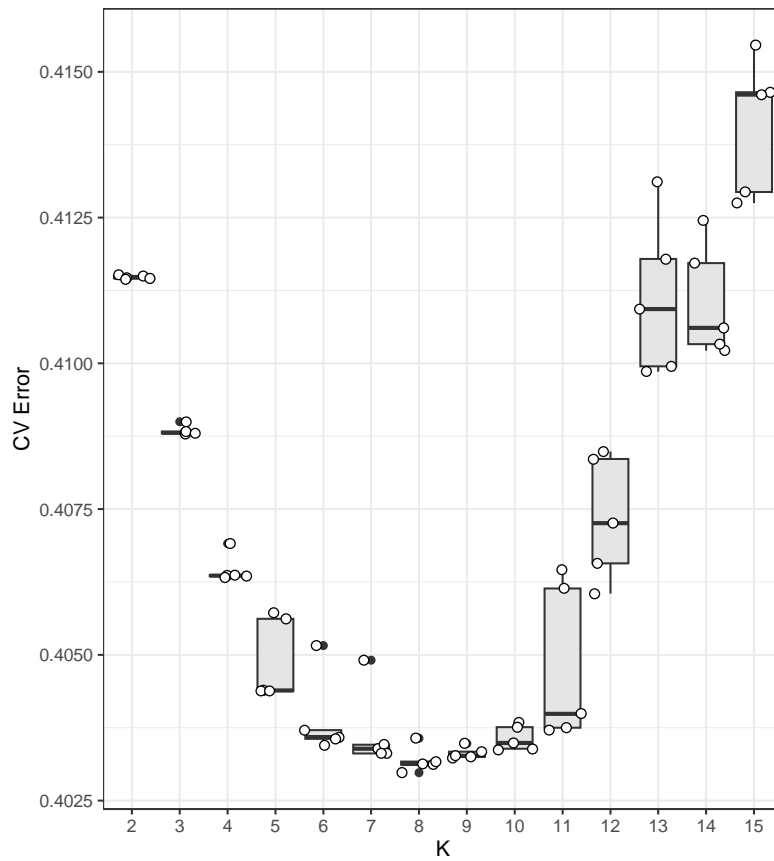


**Supplementary Figure 13.** Results of qpAdm analyses modelling ancient individuals from Southeast Asia as a three-way admixture between hunter-gatherers (light blue), ancient southern (light green), and ancient northern East Asian groups (dark green); p-values are displayed within the bars, northern East Asian sources on the left, error bars in grey show  $\pm 1$  standard error as calculated with block jackknife in qpAdm v.1520<sup>45</sup>; plotted with DataGraph v4.4 (Visual Data Tools, Inc., <http://www.visualdatatools.com/DataGraph/>). Source data are provided as a Source Data file.

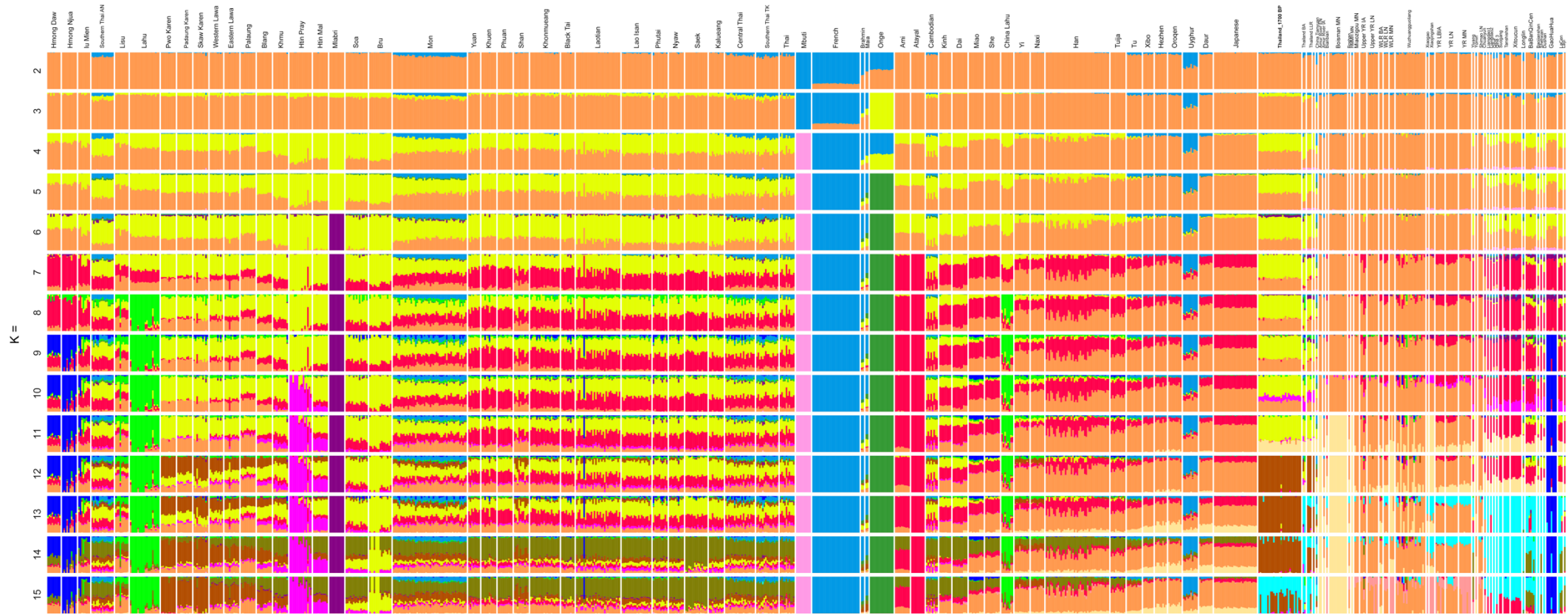




**Supplementary Figure 14.** Results of qpAdm analyses modelling the newly generated ancient genomes from northwestern Thailand as an admixture of Hòabìnhiàn-associated individual from Laos<sup>44</sup>, Liangdao1<sup>52</sup> and Upper Yellow River Late Neolithic (Upper YR LN)<sup>50</sup>; p-values are displayed on the right, where red numbers indicate  $p < 0.05$  and italic nested  $p > 0.05$ , error bars in grey show  $\pm 1$  standard error as calculated with block jackknife in qpAdm v.1520<sup>45</sup>; plotted with DataGraph v4.4 (Visual Data Tools, Inc., <http://www.visualdatatools.com/DataGraph/>). Source data are provided as a Source Data file.



**Supplementary Figure 15.** Cross-validation errors of ADMIXTURE analysis of present-day and ancient individuals from East and Southeast Asia<sup>44–46,49,50,52–54,59</sup>, showing the median of 5 replicates, the 25<sup>th</sup> and 75<sup>th</sup> percentiles, and all individual data points, plotted with ggplot2 v3.4.4<sup>62</sup>. Source data are provided as a Source Data file.



**Supplementary Figure 16.** ADMIXTURE results of present-day and ancient individuals from East and Southeast Asia<sup>44–46,49,50,52–54,59</sup>; plotted with AdmixturePlotter.R (<https://github.com/TCLamnidis/AdmixturePlotter>). Source data are provided as a Source Data file.

## Supplementary References

1. Treerayapiwat, C. Patterns of Habitation and Burial Activity in the Ban Rai Rock Shelter, Northwestern Thailand. *Asian Perspectives* **44**, 231–245 (2005).
2. Shoocongdej, R. *The Final Report of Survey and Database of Caves in Pang Mapha district, Mae Hong Son Province*. (2000).
3. Shoocongdej, R. *Progress Report 1. Highland Archaeology Project in Pang Mapha Phase I*. (2001).
4. Shoocongdej, R. *Progress Report 2. Highland Archaeology Project in Pang Mapha Phase I*. (2002).
5. Shoocongdej, R. *Progress Report 3. Highland Archaeology Project in Pang Mapha Phase I*. (2002).
6. Shoocongdej, R. *Progress Report 4. Highland Archaeology Project in Pang Mapha Phase I*. (2002).
7. Shoocongdej, R. *Final Report 1. Highland Archaeology Project in Pang Mapha Phase I*. (2003).
8. Shoocongdej, R. *Highland Archaeology in Pang Mapha*. (Silpakorn University Press, 2004).
9. Shoocongdej, R. Late pleistocene activities at the Tham Lod rockshelter in highland Pang Mapha, Mae hong Son province, northwestern Thailand. in *Uncovering Southeast Asia's Past: Selected Papers from the 10th International Conference of the European Association of Southeast Asian Archaeologists* 22–37 (NUS press Singapore, 2006).
10. Marwick, B. Human behavioural ecology and stone artefacts in northwest Thailand during the terminal Pleistocene and Holocene. *Southeast Asian Prehistory: The 11th EurASEAA Conference* 73–80 (2008).

11. Marwick, B. Multiple Optima in Hoabinhian flaked stone artefact palaeoeconomics and palaeoecology at two archaeological sites in Northwest Thailand. *Journal of Anthropological Archaeology* **32**, 553–564 (2013).
12. Shoocongdej, R. & Wattanapituksakul, A. Faunal assemblages and demography during the Late Pleistocene (MIS 2-1) to Early Holocene in Highland Pang Mapha, Northwest Thailand. *Quat. Int.* **563**, 51–63 (2020).
13. Shoocongdej, R. *et al.* *Final Report of Prehistoric Population and Cultural Dynamic in Highland Pang Mapha, Mae Hong Son Province. Volume II: Archaeology.* (2021).
14. Wannasri, S., Pumijumnong, N. & Shoocongdej, R. Teak log coffin head styles in northern Thailand: Time sequencing with dendrochronology. *Sci. Asia.* **33**, 047 (2007).
15. Pumijumnong, N. & Wannasri, S. Teak Log Coffins in Northwest Thailand: Dated by Dendrochronology and <sup>14</sup>Cwiggle Matching. *Applied Environmental Research* **37**, 1–16 (2015).
16. Preechamart, S. *et al.* Tree-ring oxygen isotope chronology of teak log coffins in northwestern Thailand and its relationship with Pacific Decadal Oscillation and El Niño-Southern Oscillation. *Quat. Int.* (2021) doi:10.1016/j.quaint.2021.04.007.
17. Grave, P. Dating of Log Coffins in Northwestern Thailand. (1994).
18. Dilokwanich, S. *An Exploration and Database System of the Caves, Mae Hong Son Province.* (2000).
19. Marwick, B. What attributes are important for the measurement of assemblage reduction intensity? Results from an experimental stone artefact assemblage with relevance to the Hoabinhian of mainland Southeast Asia. *J. Archaeol. Sci.* **35**, 1189–1200 (2008).
20. Chitkament, T., Gaillard, C. & Shoocongdej, R. Tham Lod rockshelter (Pang Mapha district, north-western Thailand): Evolution of the lithic assemblages during the late Pleistocene. *Quat. Int.* **416**, 151–161 (2016).

21. Shoocongdej, R. *Final Report 2. Highland Archaeological Project in Pang Mapha Phase II.* (2007).
22. Suraprasit, K., Shoocongdej, R., Chintakanon, K. & Bocherens, H. Late Pleistocene human paleoecology in the highland savanna ecosystem of mainland Southeast Asia. *Sci. Rep.* **11**, 16756 (2021).
23. Khaewkamput, V. Archaeology of Tham Lod: Late Pleistocene-Late Holocene life ways. in *A Conference Proceeding on Human, Culture, and Paleoenvironment on Highland Pang Pang Mapha, Mae Hong Song Province* (Silpakorn University, Bangkok, 2003).
24. Wattanapituksakul, S. Late Pleistocene Mammal Teeth from the Tham Lod Rockshelter, Amphoe Pang Mapha, Changwat Mae Hong Son. (Chulalongkorn University, Bangkok, 2006).
25. Pautreau, J.-P., Mornais, P. & Doy-Asa, T. *Ban Wang Hai: excavations of an Iron-Age cemetery in northern Thailand.* (Silkworm Books, 2003).
26. Wannasri, S. *A Dendroarchaeological Study of Log Coffins: Bo Krai Cave and Ban Rai Rockshelter in Pang Mapha District, Mae Hong Son Province.* (Mahidol University, 2004).
27. Fellows Yates, J. A. *et al.* Reproducible, portable, and efficient ancient genome reconstruction with nf-core/eager. *PeerJ* **9**, e10947 (2021).
28. Di Tommaso, P. *et al.* Nextflow enables reproducible computational workflows. *Nat. Biotechnol.* **35**, 316–319 (2017).
29. Andrews, S. FastQC A Quality Control tool for High Throughput Sequence Data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
30. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
31. Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 88 (2016).

32. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
33. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-bio.GN]* (2013).
34. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
35. Okonechnikov, K., Conesa, A. & García-Alcalde, F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **32**, 292–294 (2016).
36. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
37. Neukamm, J., Peltzer, A. & Nieselt, K. DamageProfiler: Fast damage pattern calculation for ancient DNA. *Bioinformatics* (2021) doi:10.1093/bioinformatics/btab190.
38. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
39. Lamnidis, T. C. *et al.* Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe. *Nat. Commun.* **9**, 5018 (2018).
40. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 356 (2014).
41. Jun, G., Wing, M. K., Abecasis, G. R. & Kang, H. M. An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. *Genome Res.* **25**, 918–925 (2015).
42. Sawyer, S., Krause, J., Guschanski, K., Savolainen, V. & Pääbo, S. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS One* **7**, e34131 (2012).
43. Ringbauer, H., Novembre, J. & Steinrücken, M. Parental relatedness through time revealed by runs of homozygosity in ancient DNA. *Nat. Commun.* **12**, 5425 (2021).

44. McColl, H. *et al.* The prehistoric peopling of Southeast Asia. *Science* **361**, 88–92 (2018).
45. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
46. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014).
47. Jeong, C. *et al.* Long-term genetic stability and a high-altitude East Asian origin for the peoples of the high valleys of the Himalayan arc. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 7485–7490 (2016).
48. Skoglund, P. *et al.* Genomic insights into the peopling of the Southwest Pacific. *Nature* **538**, 510–513 (2016).
49. Lipson, M. *et al.* Ancient genomes document multiple waves of migration in Southeast Asian prehistory. *Science* **361**, 92–95 (2018).
50. Ning, C. *et al.* Ancient genomes from northern China suggest links between subsistence changes and human migration. *Nat. Commun.* **11**, 2700 (2020).
51. Yang, M. A. *et al.* 40,000-Year-Old Individual from Asia Provides Insight into Early Population Structure in Eurasia. *Curr. Biol.* **27**, 3202-3208.e9 (2017).
52. Yang, M. A. *et al.* Ancient DNA indicates human population shifts and admixture in northern and southern China. *Science* **369**, 282–288 (2020).
53. Kutanan, W. *et al.* Reconstructing the Human Genetic History of Mainland Southeast Asia: Insights from Genome-Wide Data from Thailand and Laos. *Mol. Biol. Evol.* **38**, 3459–3477 (2021).
54. Wang, T. *et al.* Human population history at the crossroads of East and Southeast Asia since 11,000 years ago. *Cell* (2021) doi:10.1016/j.cell.2021.05.018.
55. Wang, C.-C. *et al.* Genomic insights into the formation of human populations in East Asia. *Nature* **591**, 413–419 (2021).



56. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
57. Pittayaporn, P. Layers of Chinese Loanwords in Protosouthwestern Tai as Evidence for the Dating of the Spread of Southwestern Tai. *Manusya: Journal of Humanities* **17**, 47–68 (2014).
58. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
59. Mallick, S. *et al.* The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016).
60. Narasimhan, V. M. *et al.* The formation of human populations in South and Central Asia. *Science* **365**, (2019).
61. Kahle, D. & Wickham, H. ggmap: Spatial visualization with ggplot2. *The R Journal* **5**, 144–161 (2013).
62. Wickham, H. *Ggplot2*. (Springer International Publishing, 2016).