

Supplementary Materials

Cultural variation impacts paternal and maternal genetic lineages of the Hmong-Mien and Sino-Tibetan groups from Thailand

Wibhu Kutanan^{1,2,*}, Rasmi Shoocongdej³, Metawee Srikummool⁴, Alexander Hübner², Thanatip Suttipai¹, Suparat Srithawong¹, Jatupol Kampuansai^{5,6}, and Mark Stoneking^{2,*}

There are 10 Supplementary Figures and 10 Supplementary Tables in this manuscript.

Supplementary Figures

Fig. S1 Bar plots of the distribution of major haplogroups for (A) MSY and (B) mtDNA. The new populations are placed at the left of the figure.

Fig. S2 Bayesian maximum clade credibility (MCC) tree of MSY sequences belonging to: (A) haplogroup C*; (B) haplogroup F; (C) haplogroup O2a2a*.

Fig. S3 Bayesian maximum clade credibility (MCC) trees of mtDNA sequences belonging to: (A) haplogroup A; (B) haplogroup B4; (C) haplogroup B5a1c1a; (D) haplogroup B6; (E) haplogroup C; (F) haplogroup D; (G) haplogroup F1g; (H) haplogroup G; and (I) haplogroup R9.

Fig. S4 The MDS plot for 73 Thai/Lao populations based on the MSY Φ_{st} distances for (A) dimension 1 vs. 2; (B) dimension 1 vs. 3; (C) dimension 2 vs. 3 and based on mtDNA Φ_{st} distances for (D) dimension 1 vs. 2; (E) dimension 1 vs. 3; (F) dimension 2 vs. 3.

Fig. S5 The heat plot of the five-dimensional MDS for the 73 Thai/Lao populations. (A) MSY; (B) mtDNA.

Fig. S6 Correspondence Analysis based on MSY haplogroup frequencies for (A) dimension 1 vs. 2; (B) dimension 1 vs. 3; (C) dimension 1 vs. 4; (D) dimension 2 vs. 3; (E) dimension 2 vs. 4; (F) dimension 3 vs. 4.

Fig. S7 Correspondence Analysis based on mtDNA haplogroup frequencies for (A) dimension 1 vs. 2; (B) dimension 1 vs. 3; (C) dimension 1 vs. 4; (D) dimension 2 vs. 3; (E) dimension 2 vs. 4; (F) dimension 3 vs. 4.

Fig. S8 The BSPs based on the MSY and mtDNA for 14 populations. Solid lines are the median estimated effective population size (y axis) through time from the present in years (x axis). The 95% highest posterior density limits are indicated by dotted lines.

Fig. S9 Bar plots of (A) haplotype diversity and (B) mean number of pairwise differences for patrilocal (blue) and matrilocal (orange) populations. Population abbreviations are in Table S1.

Fig. S10 Genetic variation among populations within groups, defined by ethnicity, language and cultural practice.

Supplementary Tables

Table S1 General information concerning the studied populations and genetic diversity values.

Table S2 Sequencing coverage and MSY haplogroup of each individual.

Table S3 Sequencing coverage and mtDNA haplogroup of each individual.

Table S4 Details for populations included in the MSY and mtDNA comparisons.

Table S5 Frequency of MSY/mtDNA lineages of Northeast Asian (NEA) and Southeast Asian (SEA) origin in Hmong-Mien and Sino-Tibetan speaking populations from China and Vietnam.

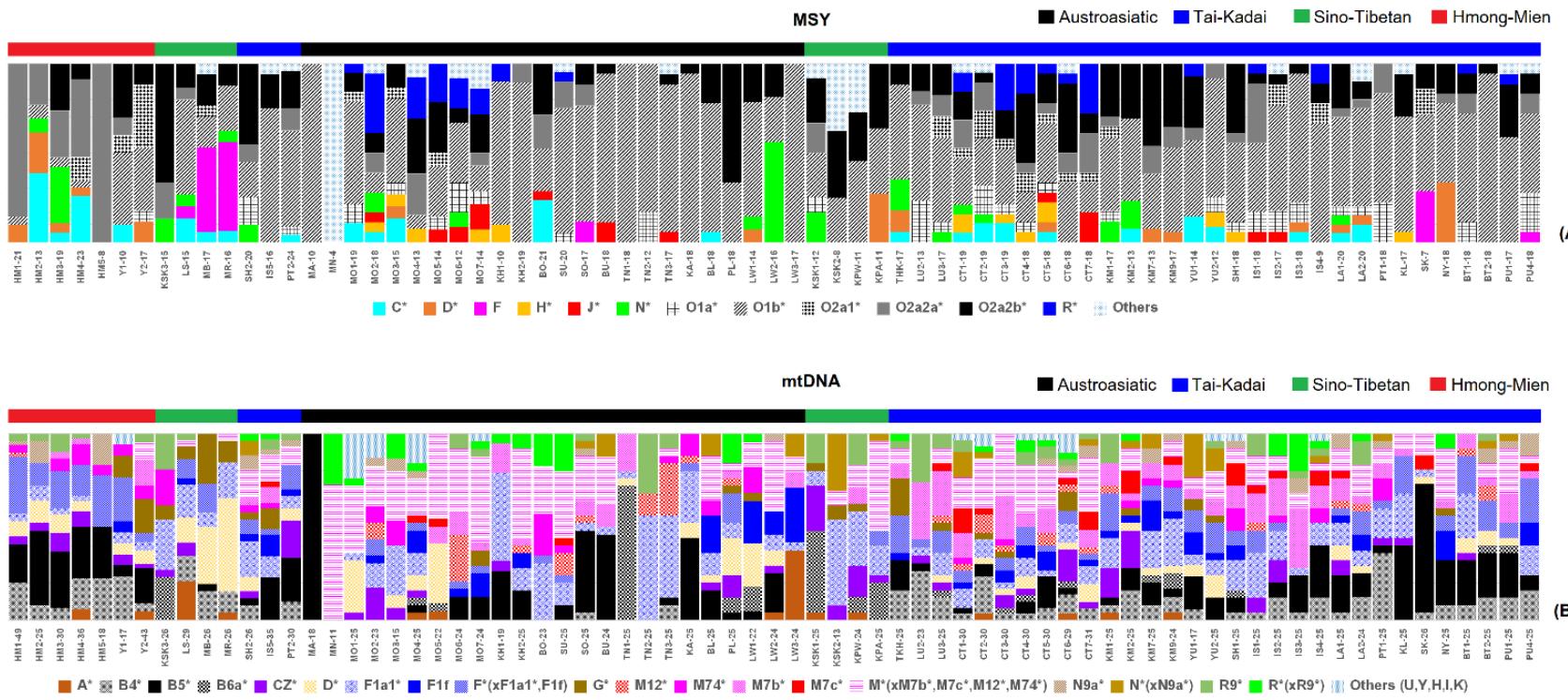
Table S6 Details on BEAST analyses.

Table S7 MSY haplogroup frequencies. The populations sequenced in this study are on the left side of the table, and newly-reported haplogroups are in red font.

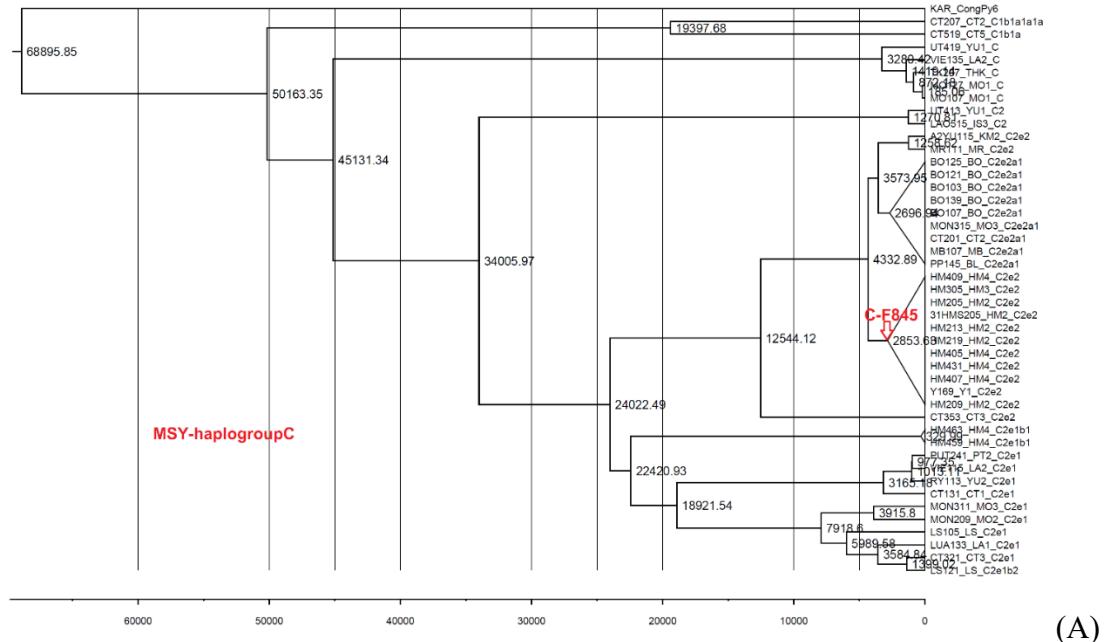
Table S8 MtDNA haplogroup frequencies. The populations sequenced in this study are on the left side of the table, and newly-reported haplogroups are in red font.

Table S9 Mann-Whitney U test results for comparing genetic diversity values between different language families and residence patterns.

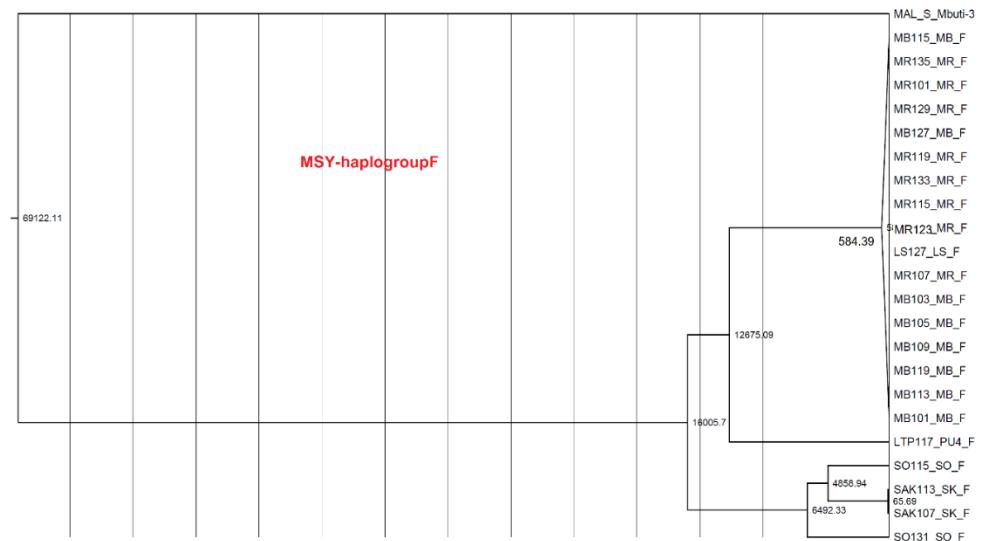
Table S10 Haplotype sharing between populations of the same patrilocal or matrilocal group, for groups with at least two different populations sampled.



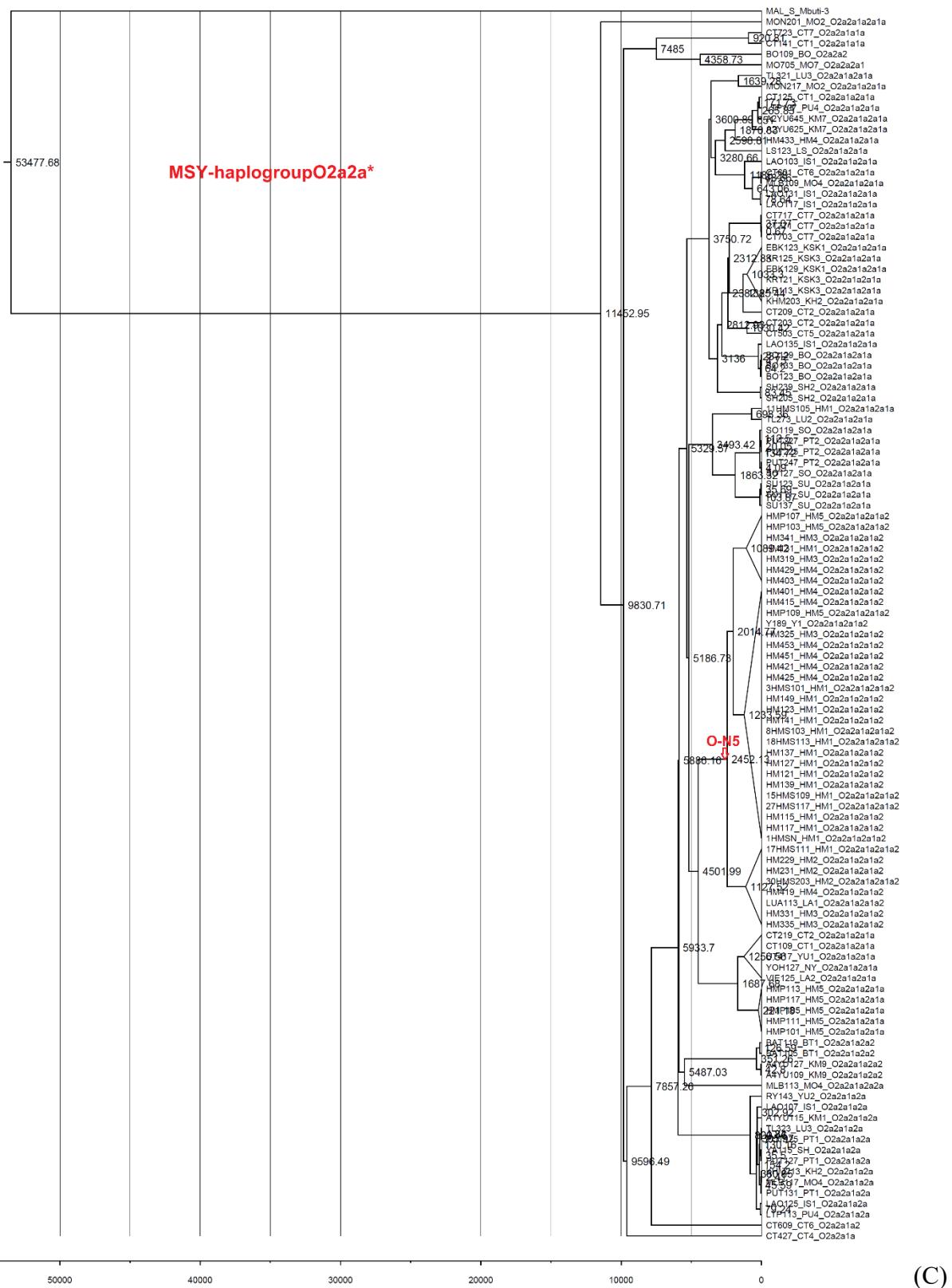
Supplementary Fig. S1 Bar plots of the distribution of major haplogroups for (A) MSY and (B) mtDNA. The new populations are placed at the left of the figure.



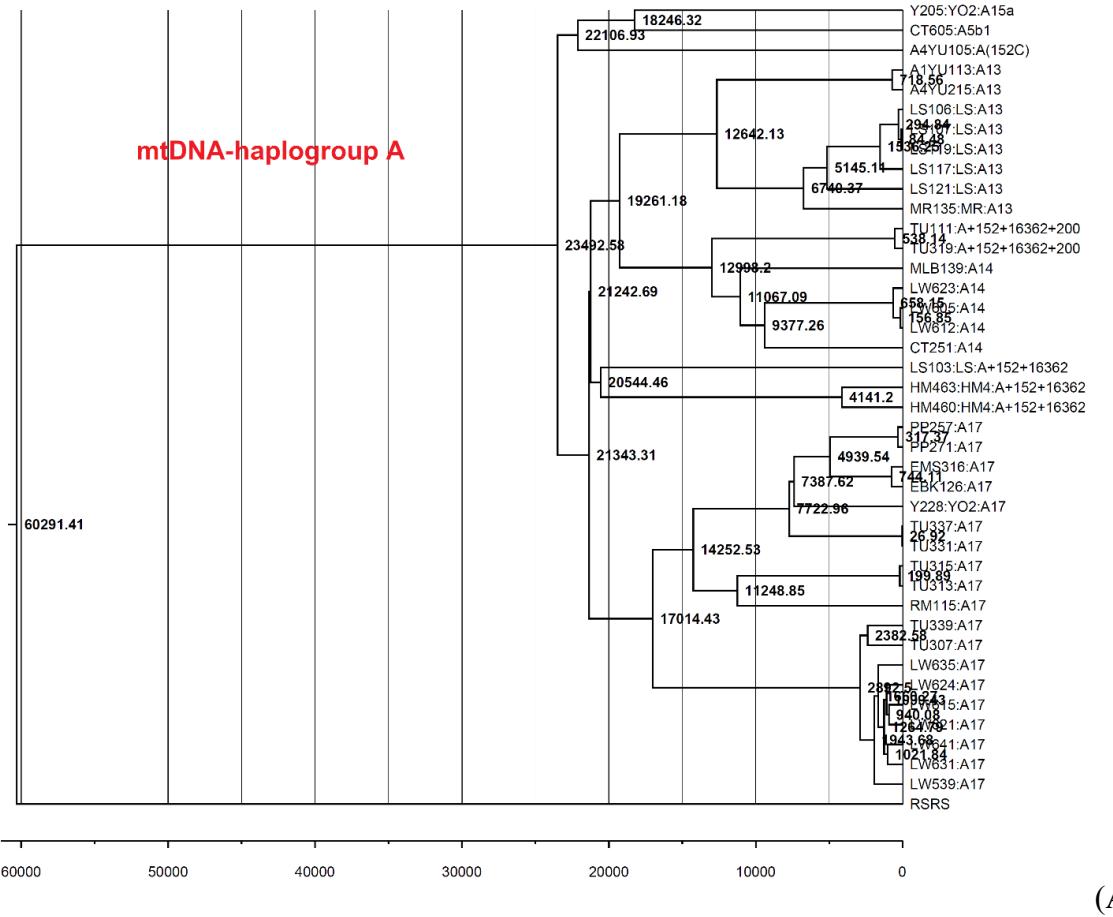
(A)

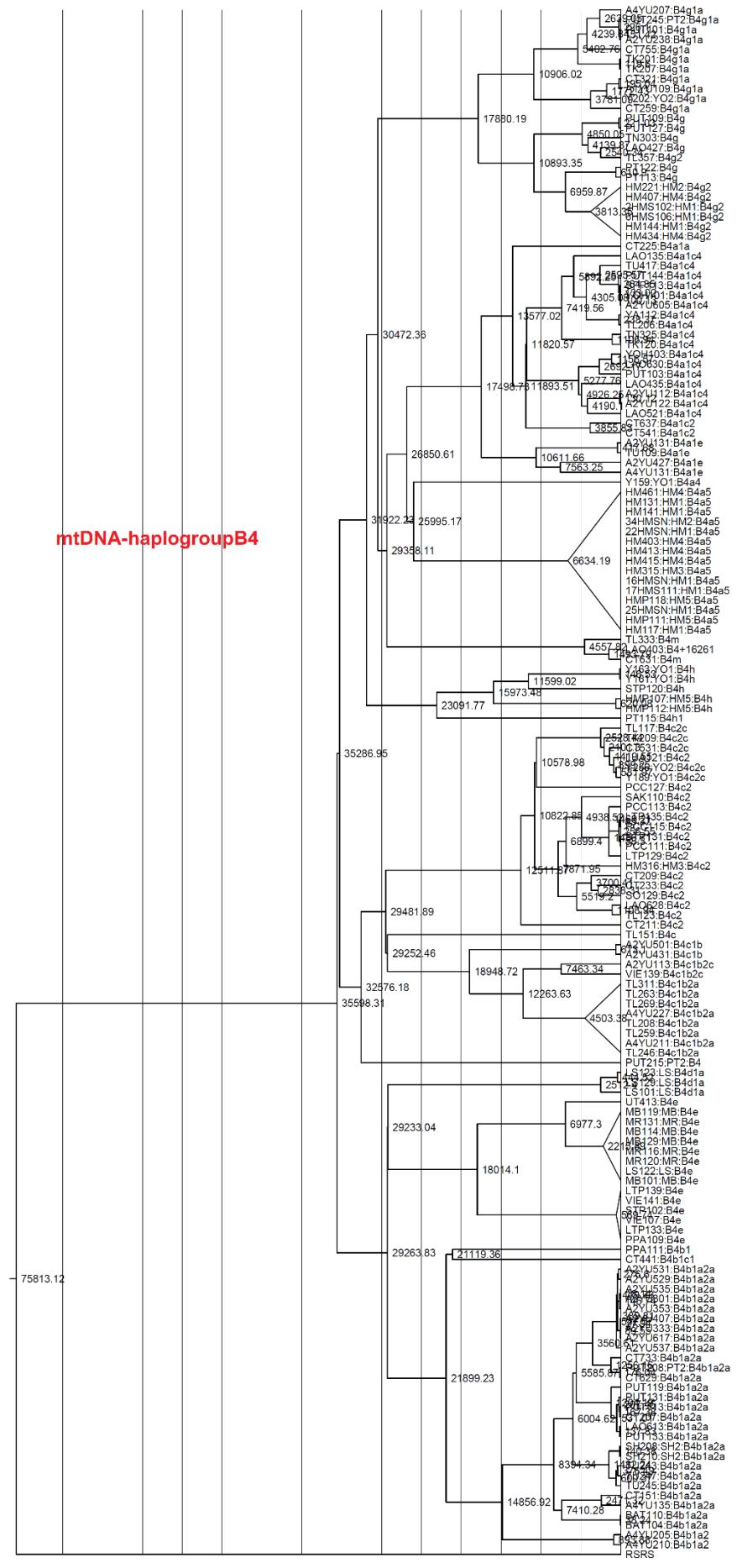


(B)

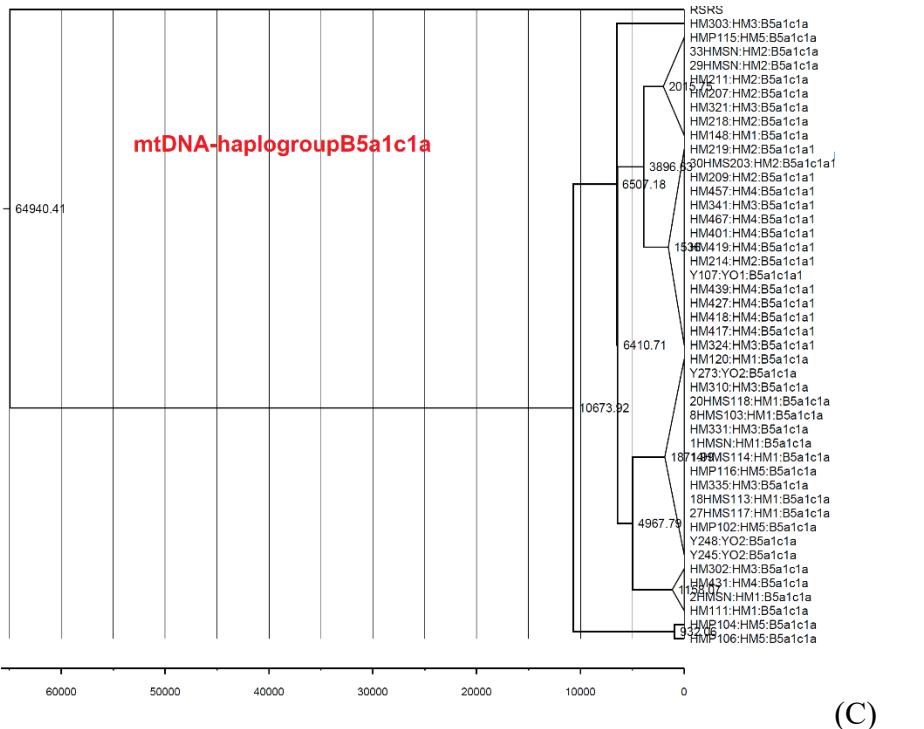


Supplementary Fig. S2 Bayesian maximum clade credibility (MCC) tree of MSY sequences belonging to: (A) haplogroup C*; (B) haplogroup F; (C) haplogroup O2a2a*.

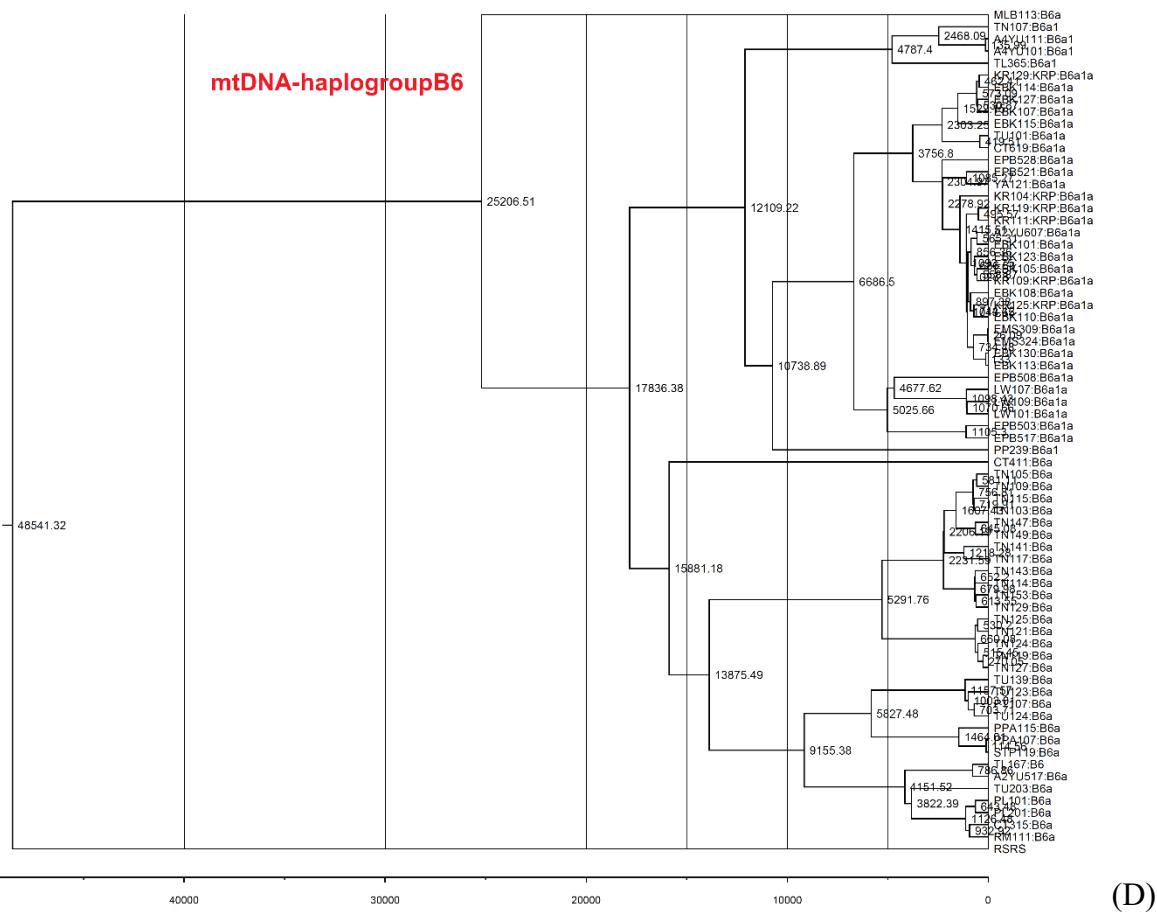




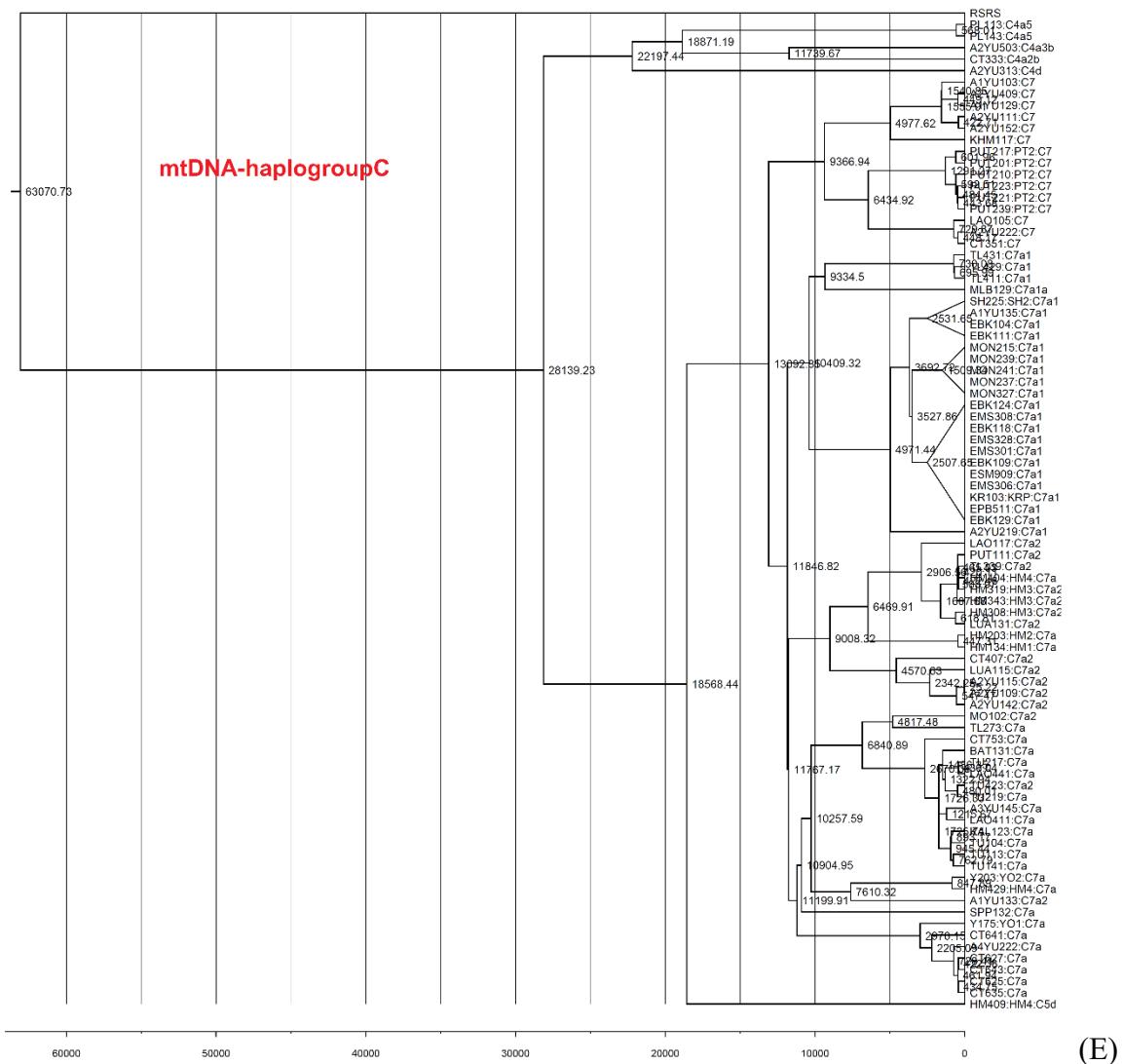
(B)

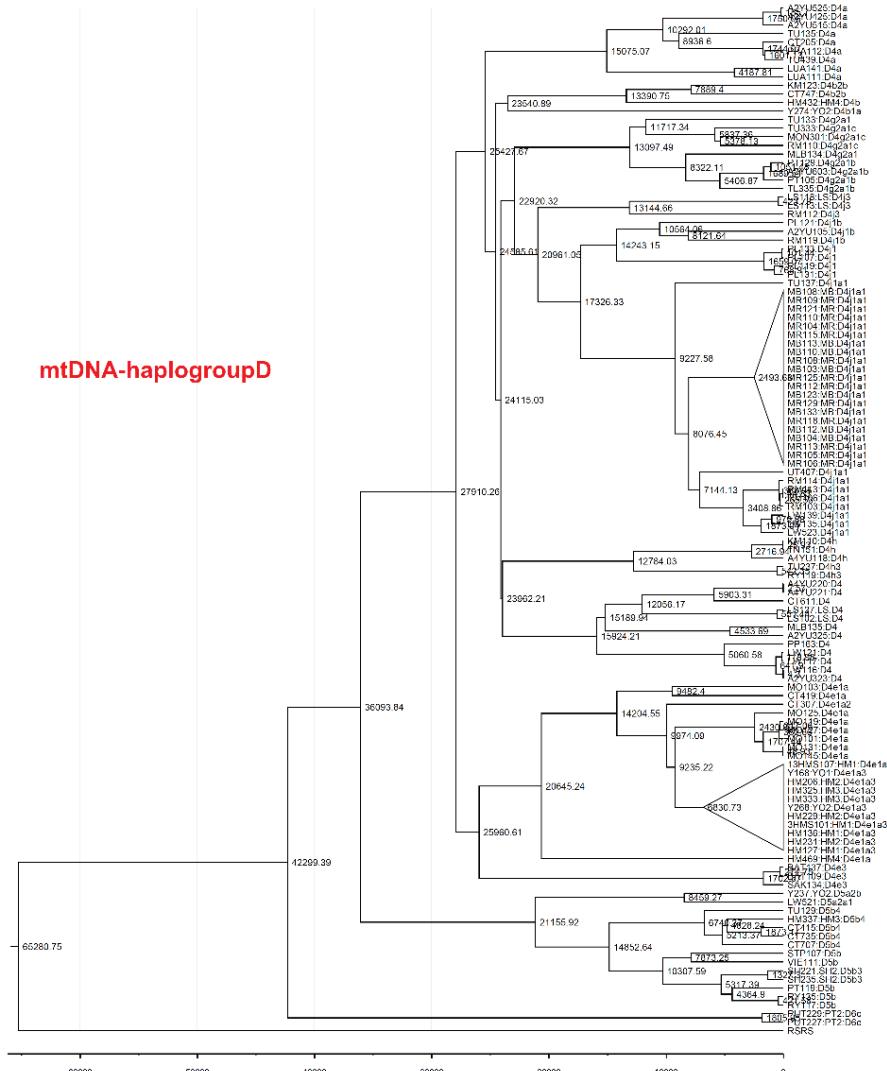


(C)

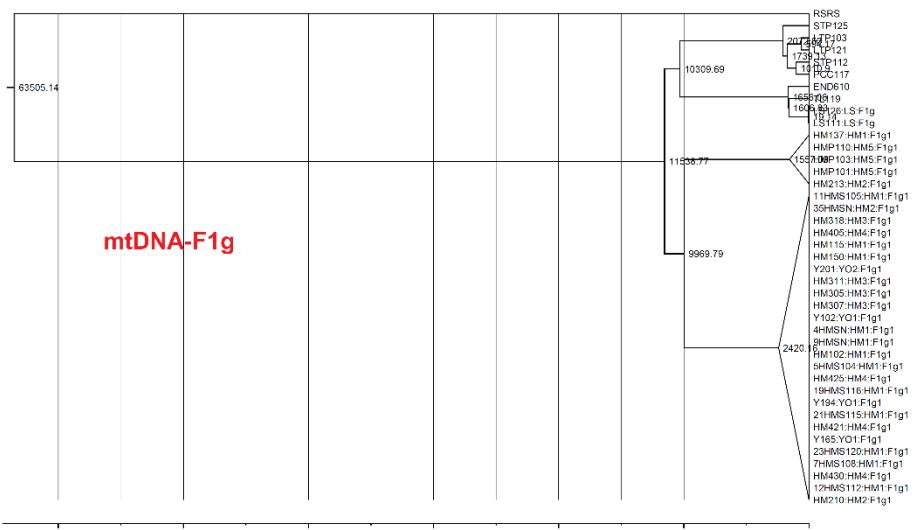


(D)

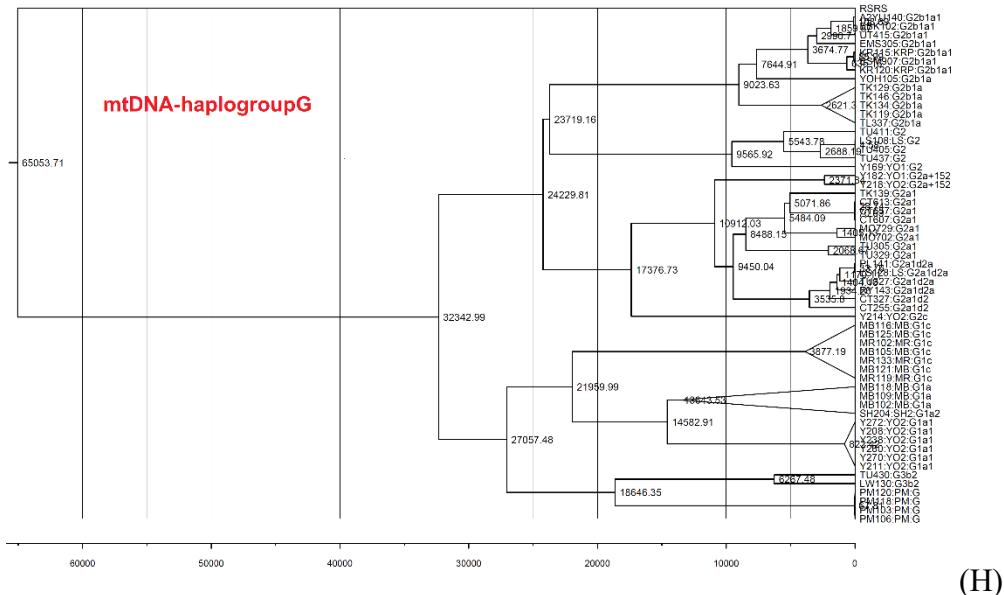




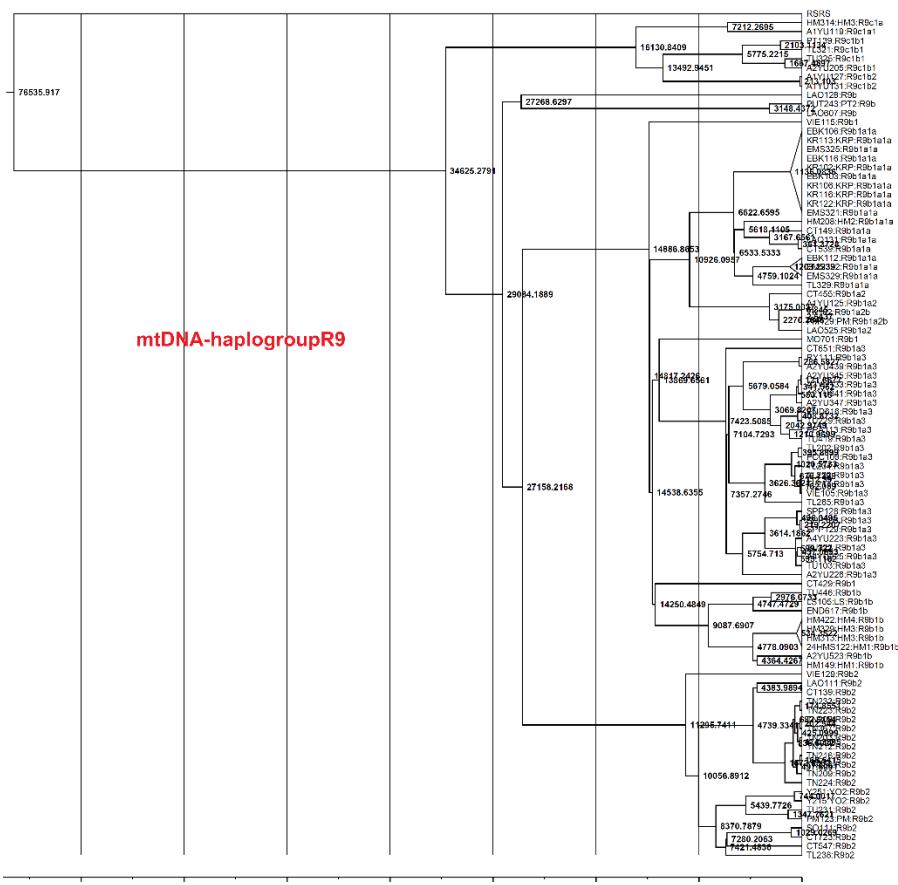
(F)



(G)



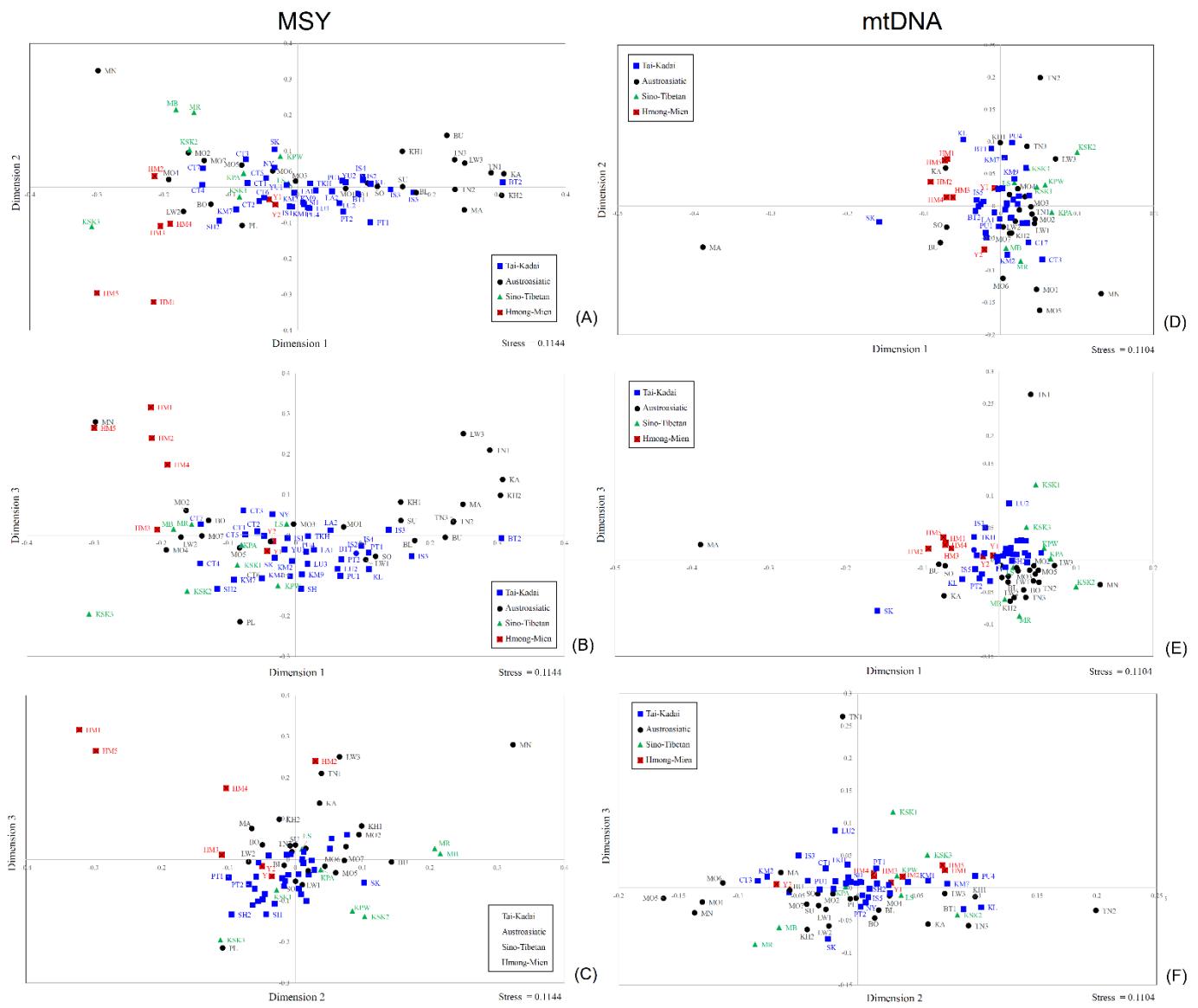
(H)



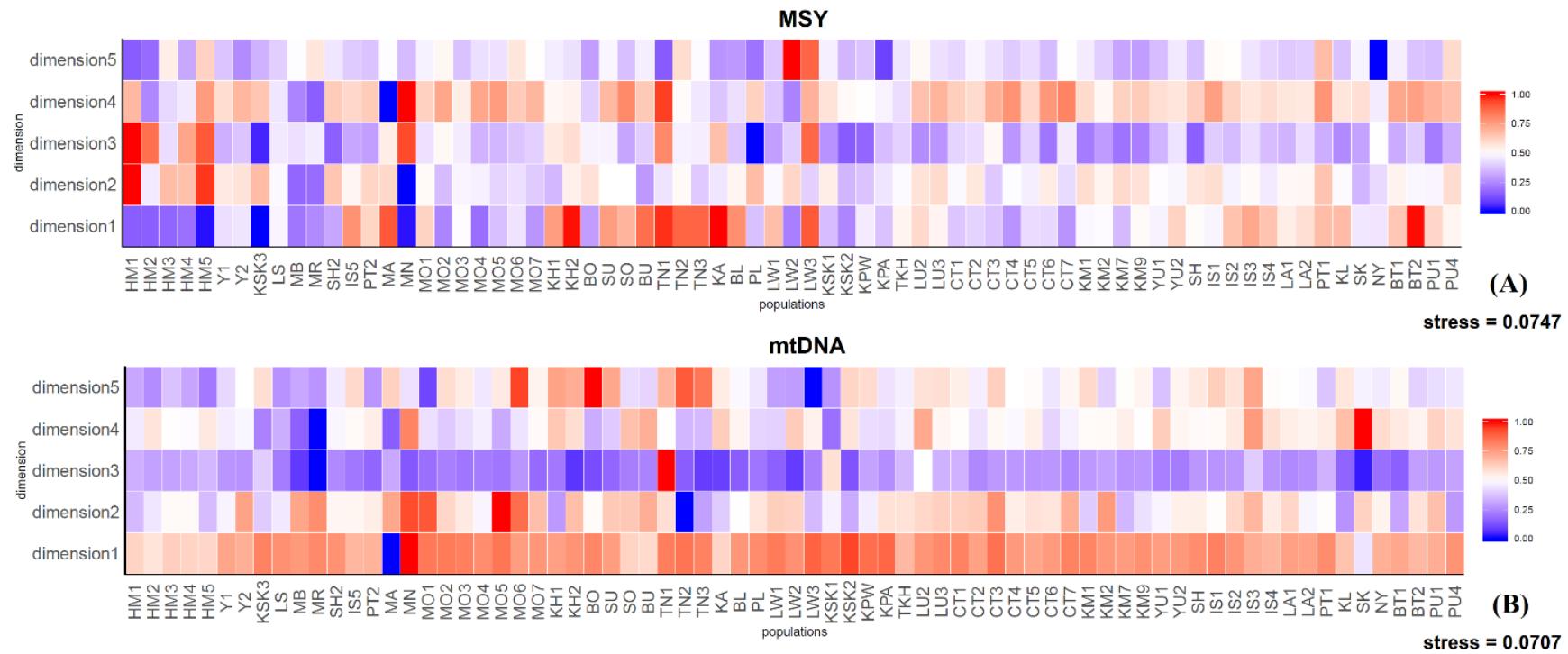
(I)

Supplementary Fig. S3 Bayesian maximum clade credibility (MCC) trees of mtDNA sequences

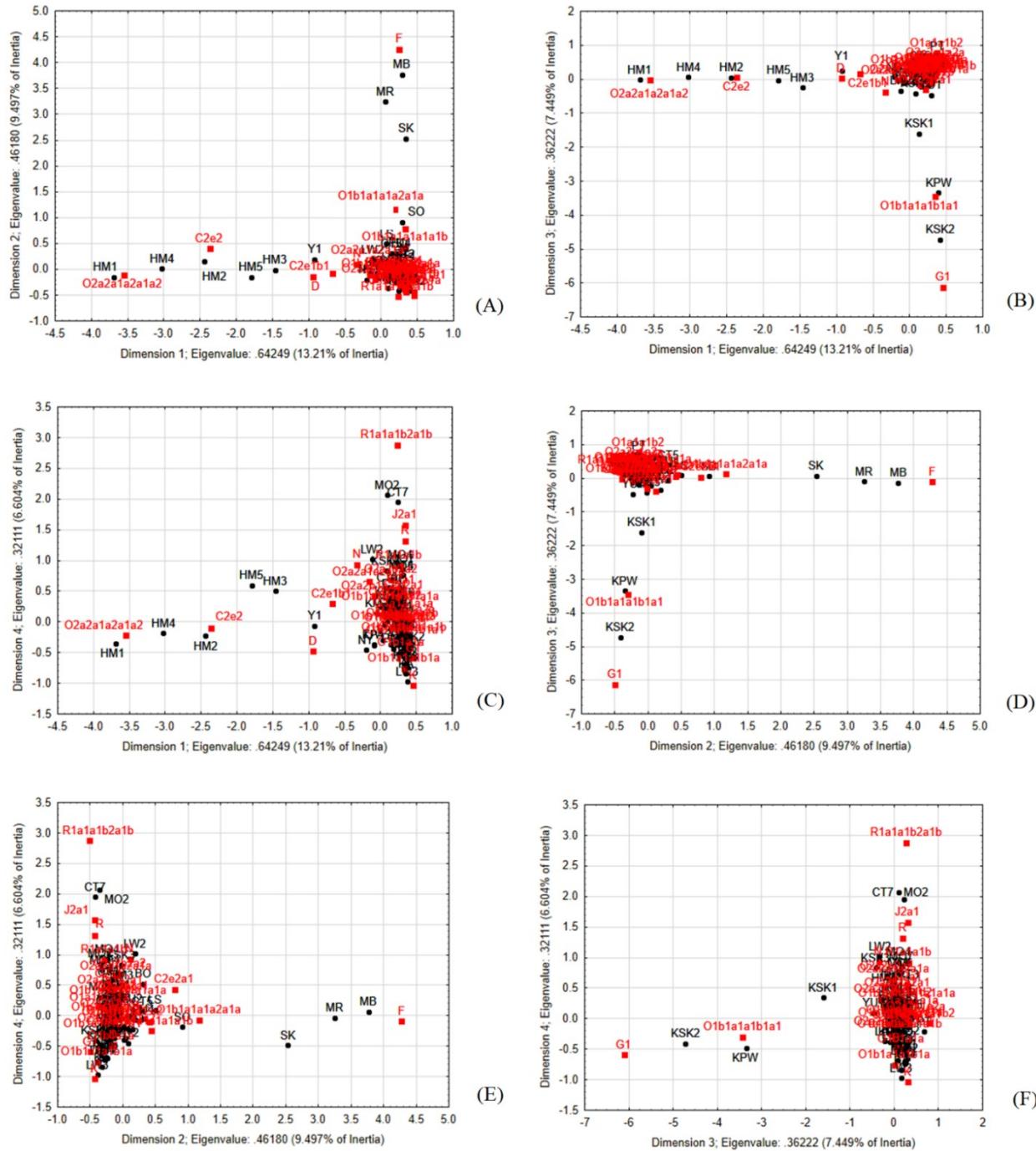
belonging to: (A) haplogroup A; (B) haplogroup B4; (C) haplogroup B5a1c1a; (D) haplogroup B6; (E) haplogroup C; (F) haplogroup D; (G) haplogroup F1g; (H) haplogroup G; and (I) haplogroup R9.



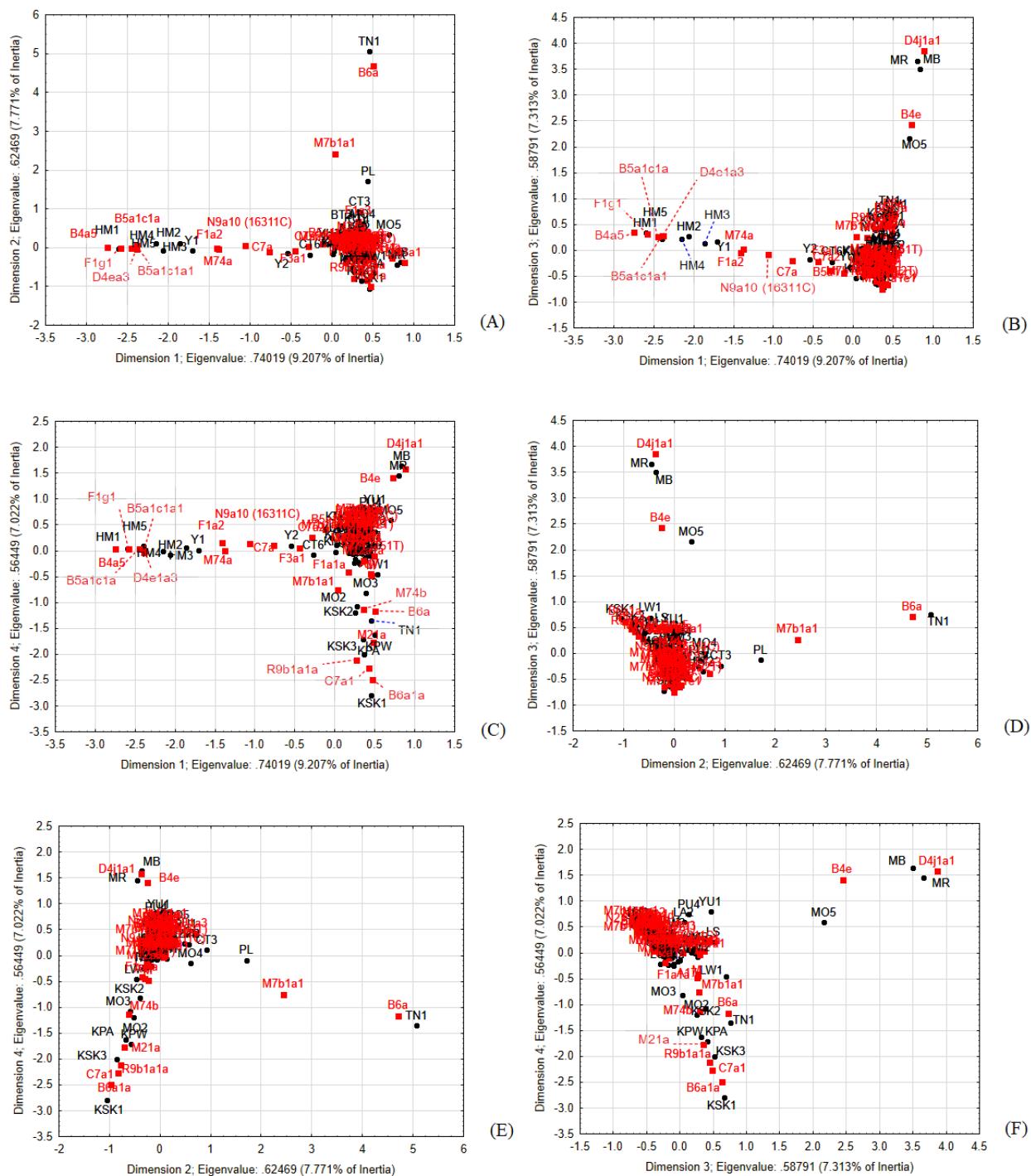
Supplementary Fig. S4 The MDS plot for 73 Thai/Lao populations based on the MSY Φ_{st} distances for (A) dimension 1 vs. 2; (B) dimension 1 vs. 3; (C) dimension 2 vs. 3 and based on mtDNA Φ_{st} distances for (D) dimension 1 vs. 2; (E) dimension 1 vs. 3; (F) dimension 2 vs. 3.



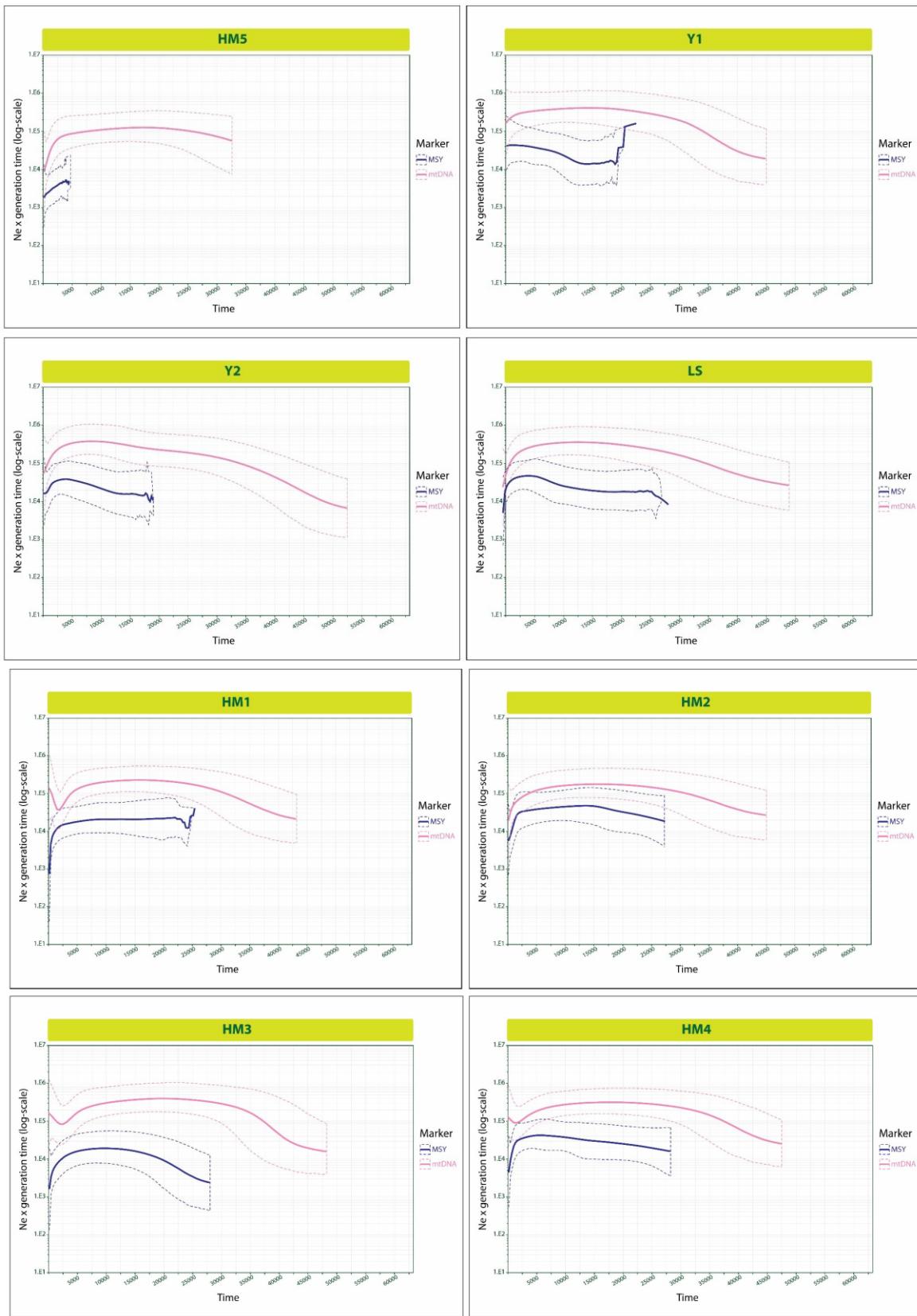
Supplementary Fig. S5 The heat plot of the five-dimensional MDS for the 73 Thai/Lao populations. (A) MSY; (B) mtDNA.

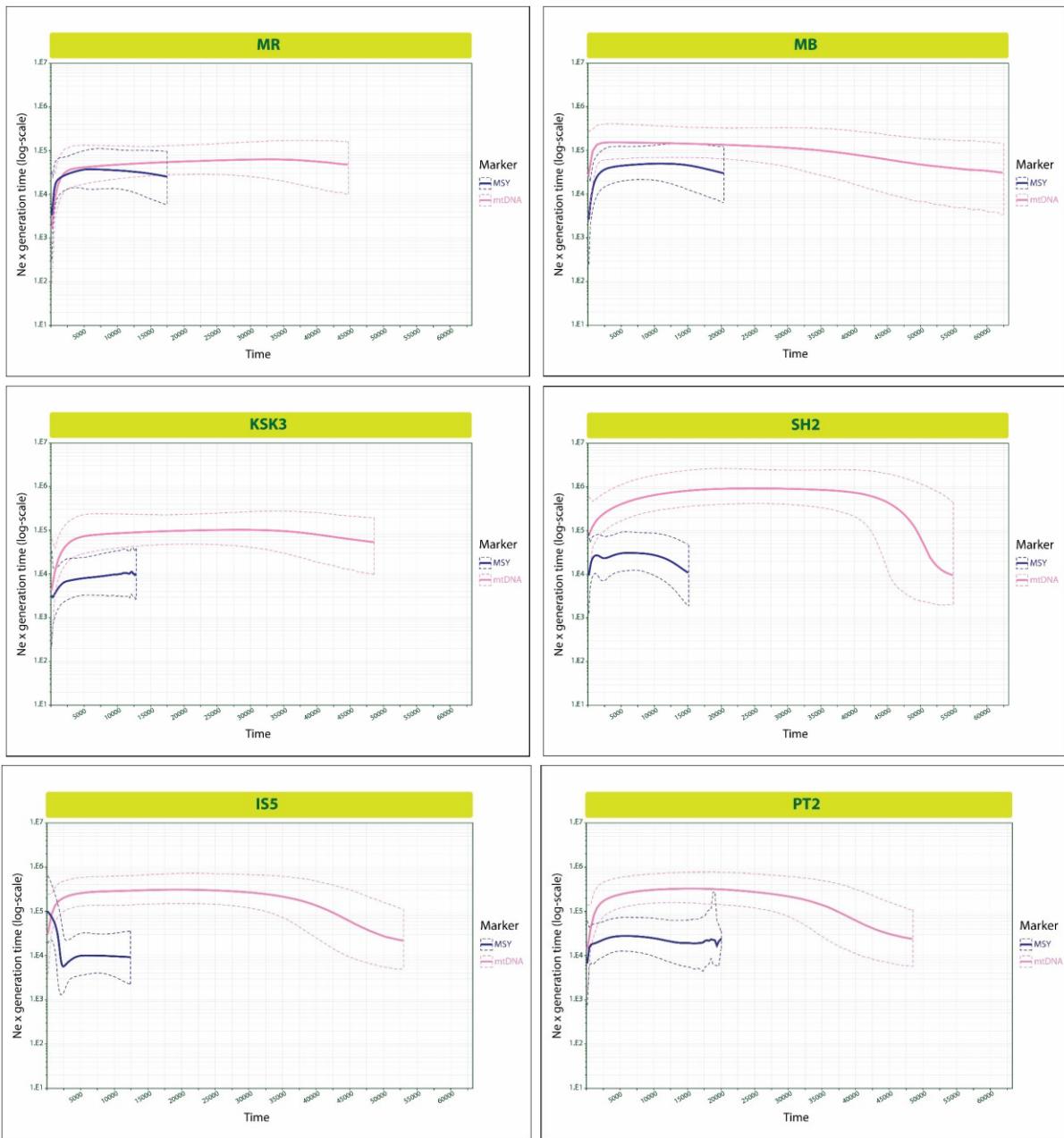


Supplementary Fig. S6 Correspondence Analysis based on MSY haplogroup frequencies for (A) dimension 1 vs. 2; (B) dimension 1 vs. 3; (C) dimension 1 vs. 4; (D) dimension 2 vs. 3; (E) dimension 2 vs. 4; (F) dimension 3 vs. 4.

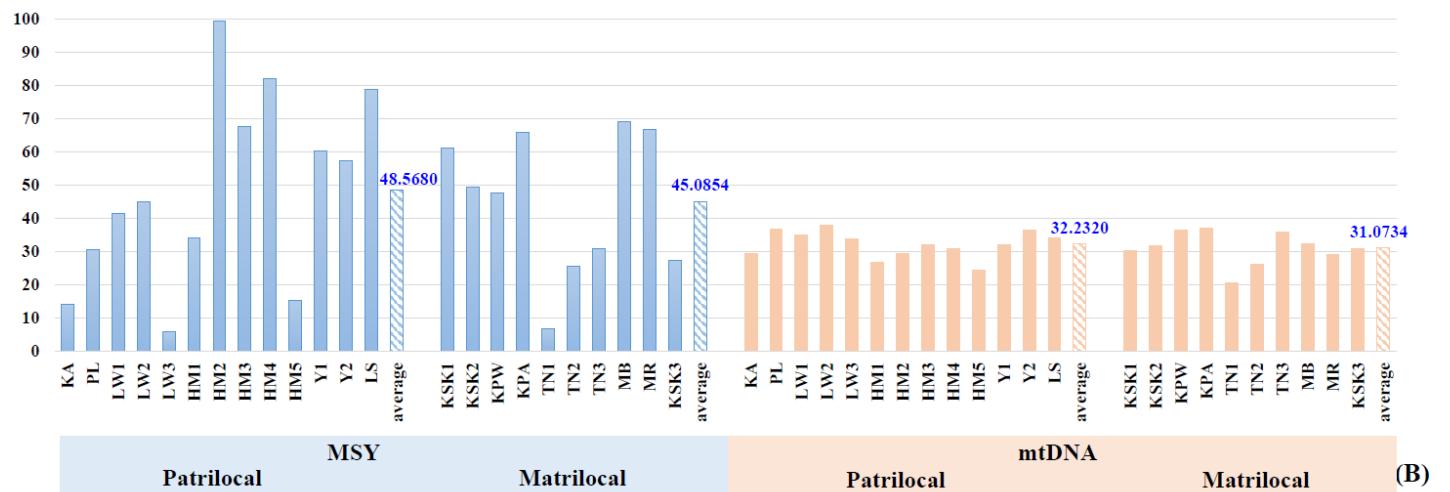
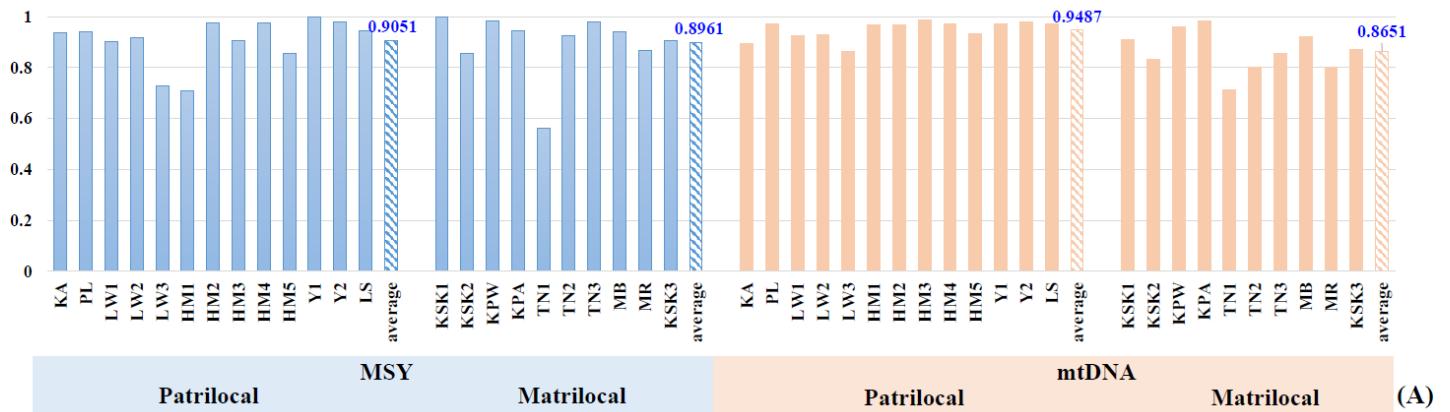


Supplementary Fig. S7 Correspondence Analysis based on mtDNA haplogroup frequencies for (A) dimension 1 vs. 2; (B) dimension 1 vs. 3; (C) dimension 1 vs. 4; (D) dimension 2 vs. 3; (E) dimension 2 vs. 4; (F) dimension 3 vs. 4.

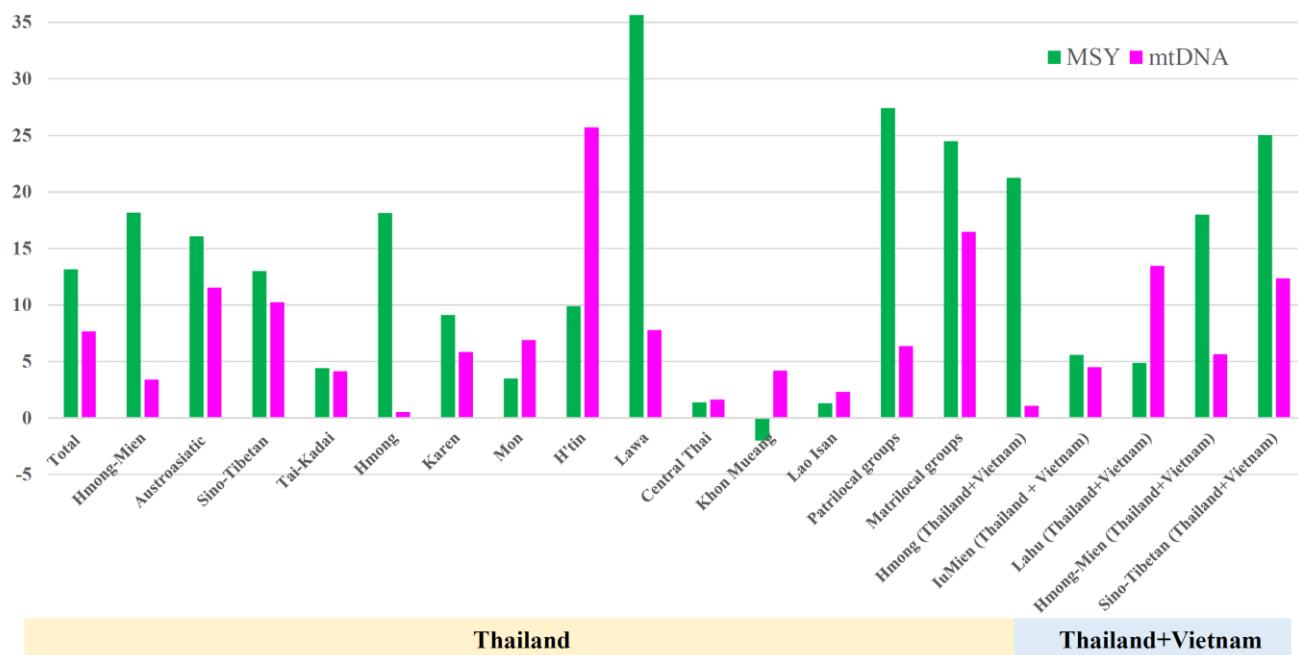




Supplementary Fig. S8 The BSPs based on the MSY and mtDNA for 14 populations. Solid lines are the median estimated effective population size (y axis) through time from the present in years (x axis). The 95% highest posterior density limits are indicated by dotted lines.



Supplementary Fig. S9 Bar plots of (A) haplotype diversity and (B) mean number of pairwise differences of patrilocal (blue) and matrilocal (orange). Population abbreviations are in Table S1.



Supplementary Fig. S10. Genetic variation among populations within groups, defined by ethnicity, language and cultural practice.