Supplementary Figure legends

Fig. S1. Schematic representation of the four models used in the regional ABC analysis. Thicker arrows indicate a colonization event. t_0, time to the last colonization event; t_1, bottleneck duration; t_c, time between the colonization event and the population recovery (to the current population size, N) after the bottleneck phase.

Fig. S2. Potential distribution of *Cephaloflexa bergi.* (A) Prediction under present day climatic conditions; the colour gradient represents the probability that environmental conditions are suitable for this species. Distribution prediction under the CCSM (B) and MIROC (C) paleo- reconstructed scenarios during the LGM (21 kya); the red colour shows the suitable areas for the species under the 10 percentile training presence criterion (used as threshold). Black dots indicate localities used for the predictions.

Fig. S3. Detailed Bayesian Inference tree for the COI (A) and the ITS-1 (B) data sets, the numbers at the nodes indicate the posterior probability. (C) Neighbor-joining tree inferred from the D_{XY} values of the ITS-1 data. The colour codes correspond to sampling localities as shown in figure 1

Fig. S4. COI Bayesian Inference trees built after removing the most variable positions (5^{th} category). (A) Analyses conducted after removing from the alignment of all sequences those sites that were most variable (5^{th} category) when comparing only the 01-AR population individuals. (B) Analyses conducted after removing from the alignment of all sequences the most variable positions estimated comparing all sequences in the alignment. The sequences of outgroup specimens

F4063, F3952, F2518, F2516 and F2391 are deposited in GenBank under accession numbers KF971679 to KF971683.

Fig. S5: Inference tree inferred to estimate *C. be*rgi diversification times. Black spots indicate ancestral nodes for populations (the code is indicated) with two deep lineages (those with descendants in two major clades are represented by big spots). The blue horizontal bar indicates the 95% highest posterior density interval.

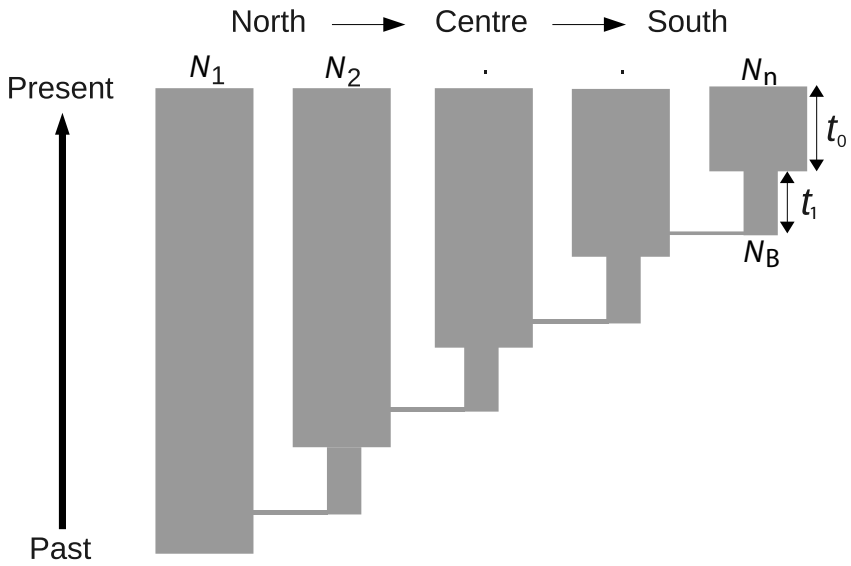
Fig. S6. Confusion matrices for the model choice. Bars represent the number pseudo-observed data sets (PODS) generated under the models indicated in the X-axis (true models) and assigned to the LSM (light grey) and LRM (dark grey) models by our model choice procedure.

Fig. S7. Plots comparing the ABC posterior probabilities of the best-supported model in localities 01-AR_1, 04-ED, 07-SO, 10-SB, 13-EC, 19-PC_1, 19-PC_2 and 28-SI_1 with those estimated in sets of 100 discrete bins of probability as in Peter *et al.* (2010) (Empirical probabilities). The probabilities were calibrated as indicated by the dashed lines.

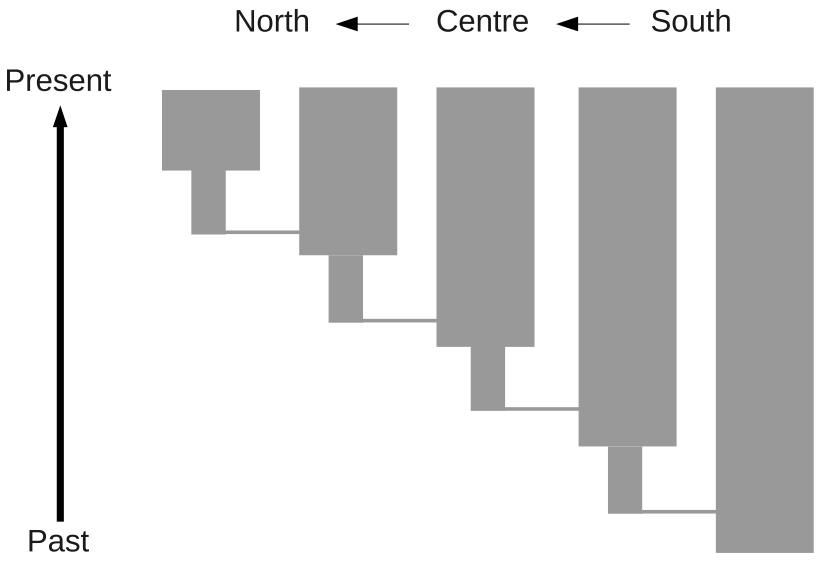
Fig. S8. Truncated prior (black), retained simulations (blue) and posterior density (red) distributions of model parameters under the best-supported model in localities 01-AR_1, 07-SO, 10-SB, 13-EC, 19-PC_1, 19-PC_2. θ, the product of the effective size and the mutation rate ($\theta = 4N_e u$) for COI (θ_{COI}) and ITS (θ_{ITS}). c and b, indicate the proportion of the source population that participates in the founding of current sampled SAF populations in the LSM and the LRM models, respectively. *T*c and *T*b, denote the times of the founding events.

Fig. S9. Distribution of posterior quantiles of true model parameter under the best-supported model in localities 01-AR_1, 07-SO, 10-SB, 13-EC, 19-PC_1, 19-PC_2. *p*-values of the Kolmogorov-Smirnov test for the uniform distribution are shown on each graph. θ, the product of the effective size and the mutation rate ($\theta = 4N_e u$) for COI (θ_{COI}) and ITS (θ_{ITS}). c and b, indicate the proportion of the source population that participates in the founding of current sampled SAF populations in the LSM and the LRM models, respectively. *T*c and *T*b, denote the times of the founding events.

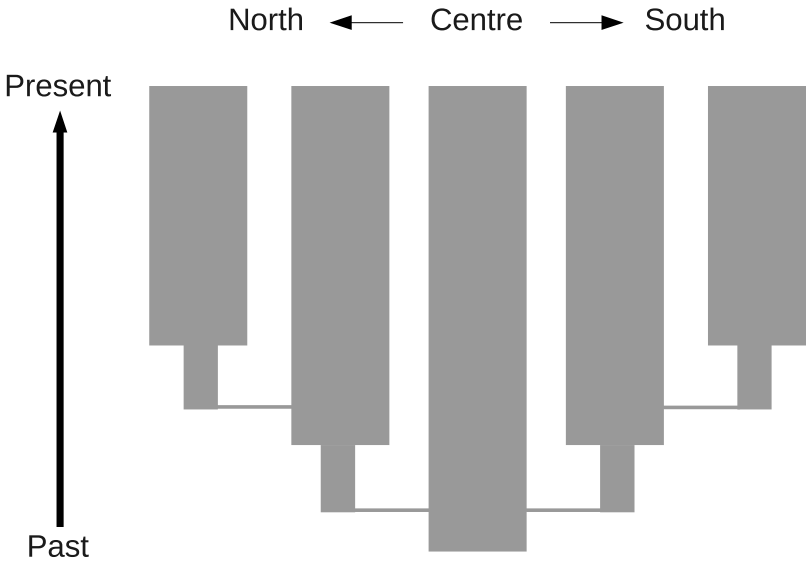
Fig. S1



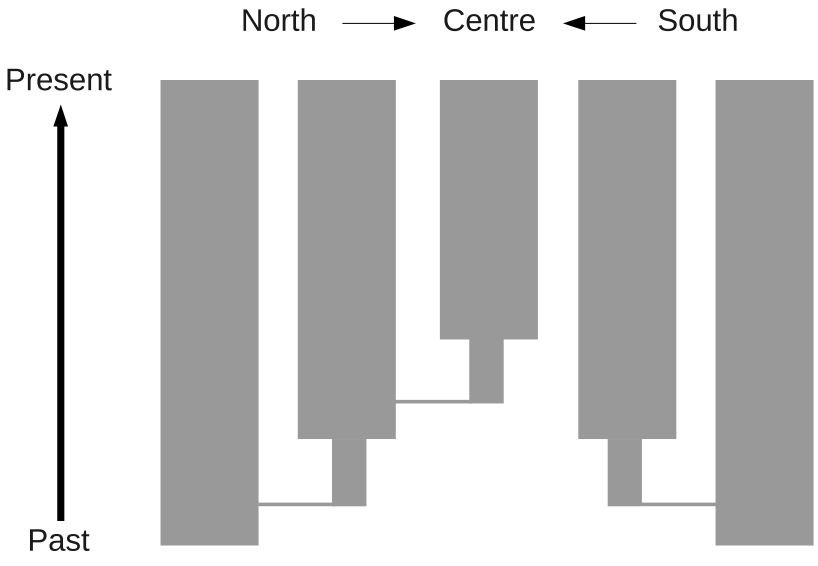
Model 1



Model 2



Model 3



Model 4

Fig. S2

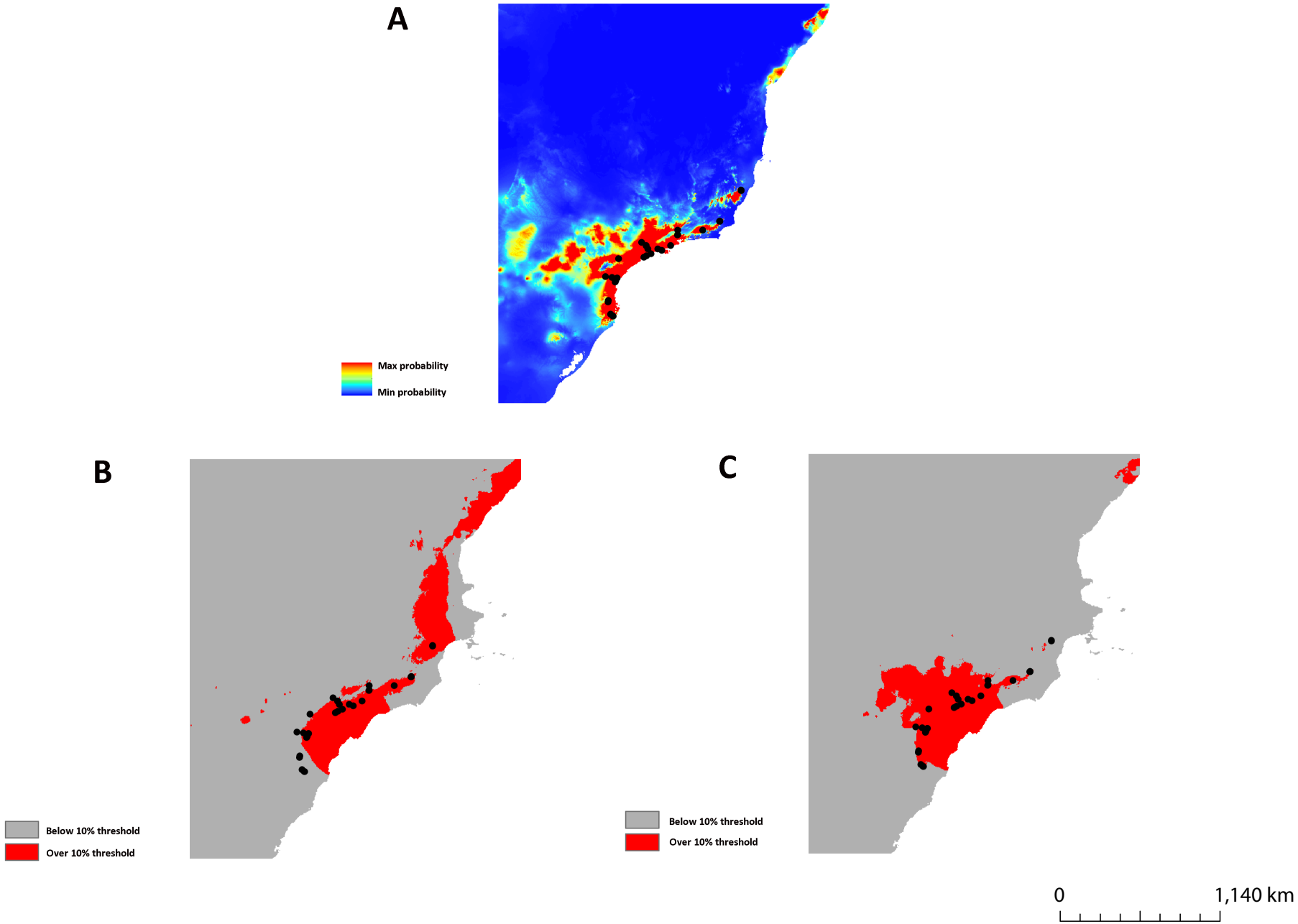

A
Max probability
Min probability
B
Below 10% threshold
Over 10% threshold
C
Below 10% threshold
Over 10% threshold
0
1,140 km

Fig. S3

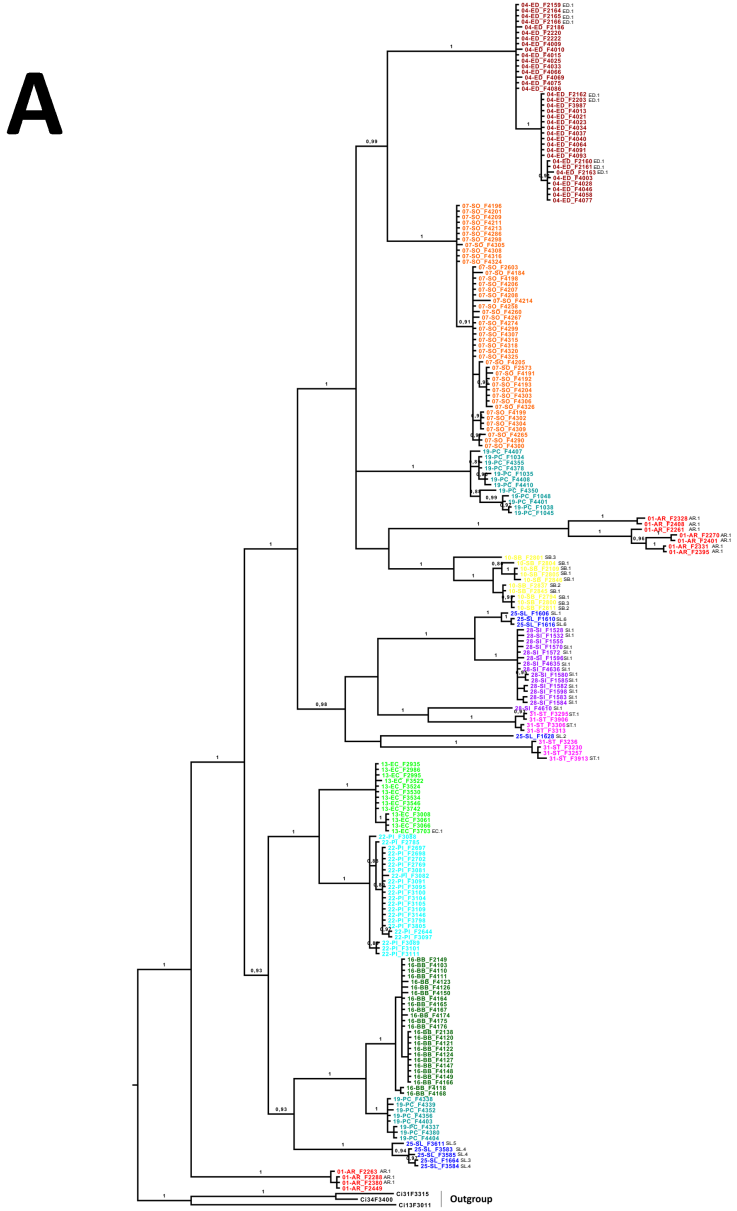


Fig. S3

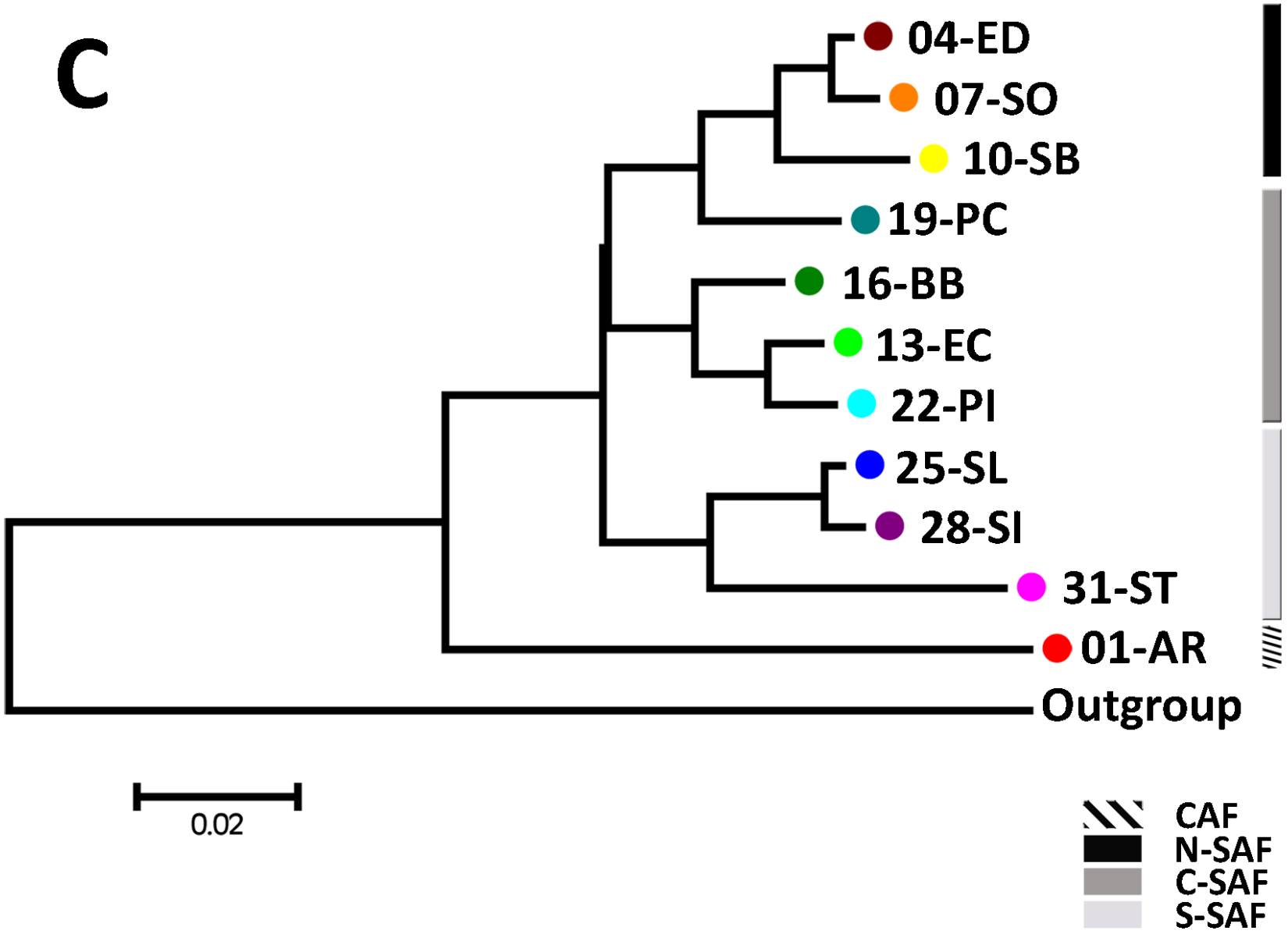

C
04-ED
07-SO
10-SB
19-PC
16-BB
13-EC
22-PI
25-SL
28-SI
31-ST
01-AR
Outgroup
0.02
CAF
N-SAF
C-SAF
S-SAF

Fig. S4

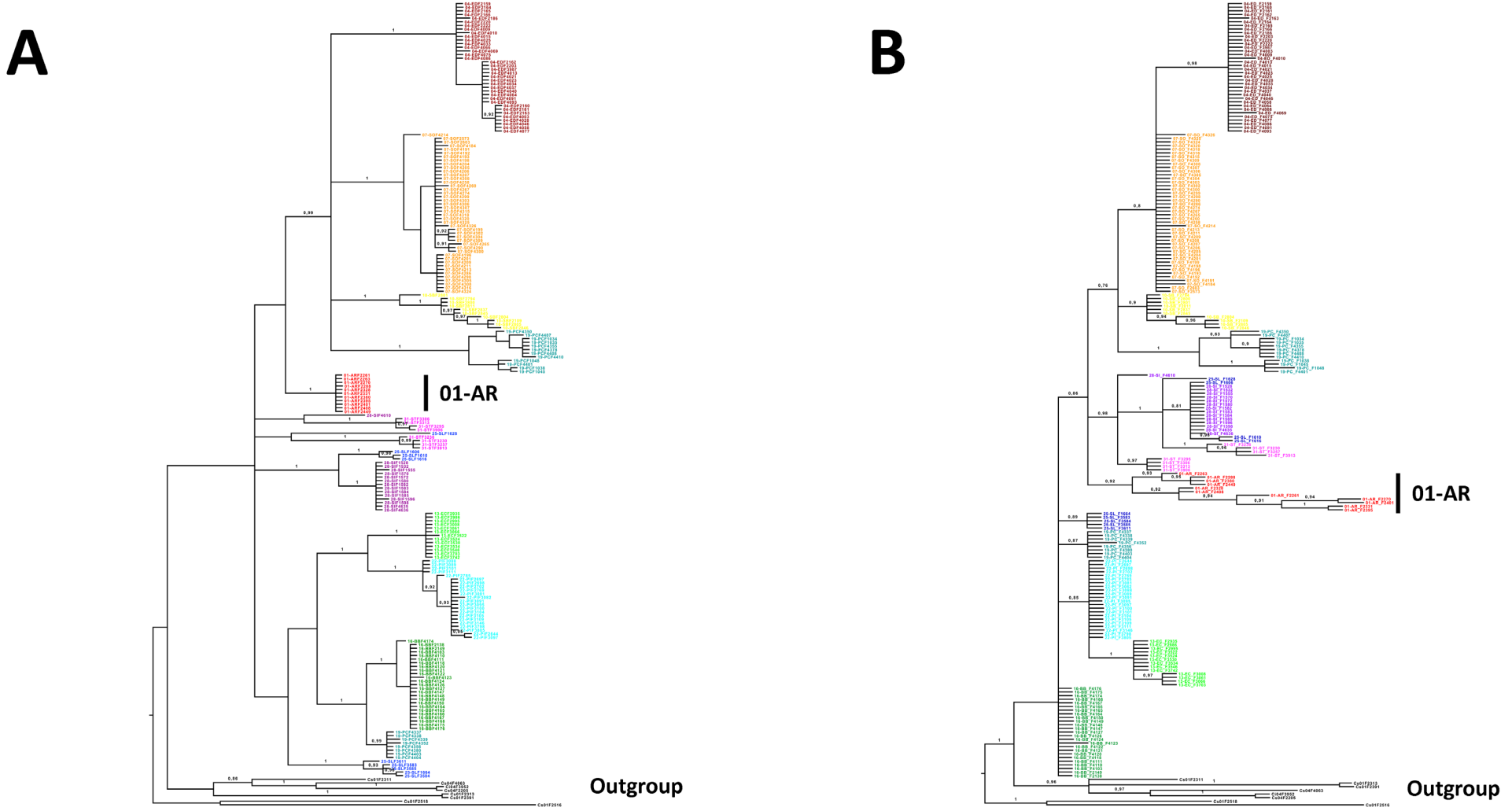


Fig. S5

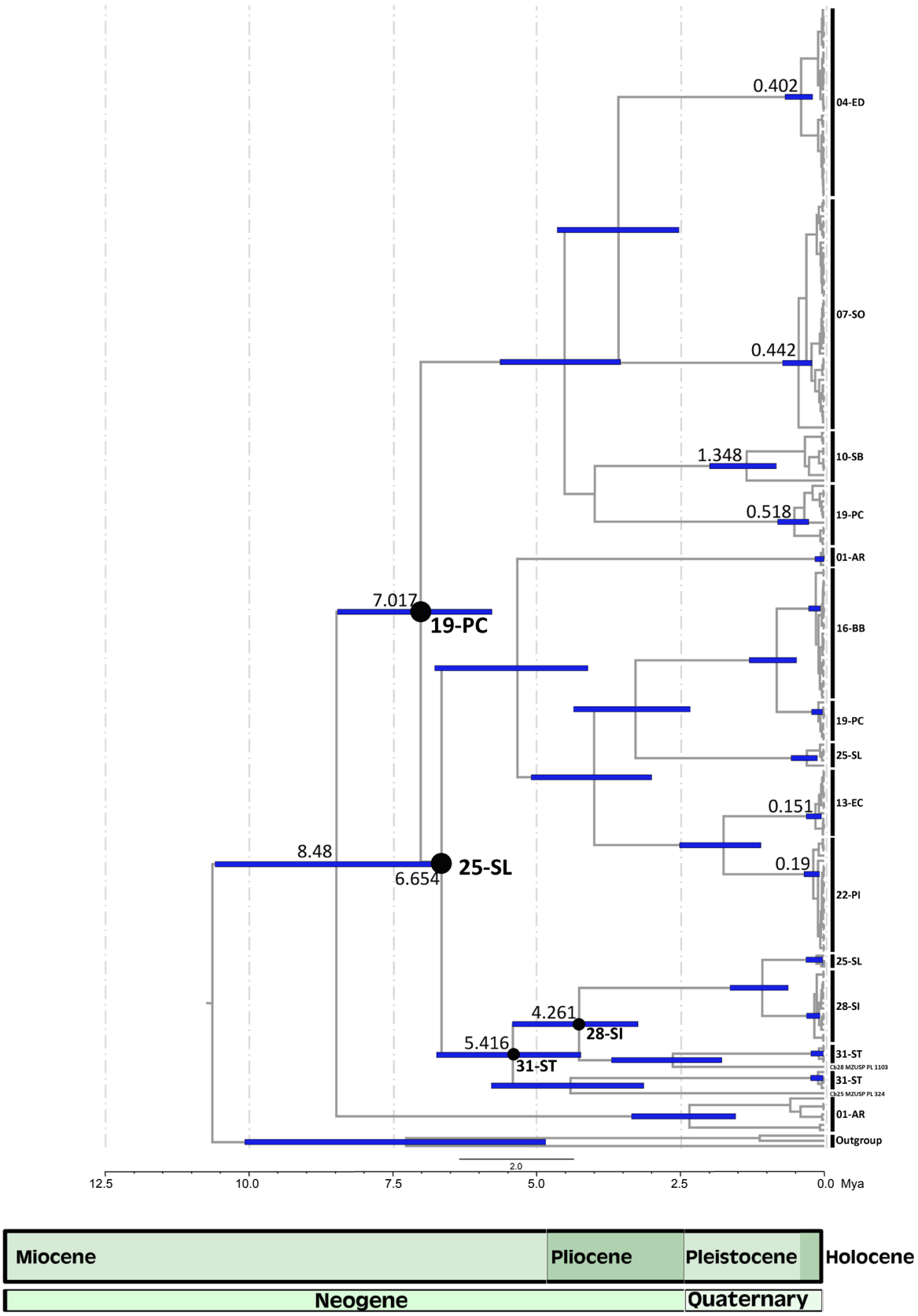

0.402
04-ED
07-SO
0.442
1.348
10-SB
19-PC
0.518
01-AR
7.017
19-PC
16-BB
19-PC
25-SL
13-EC
0.151
8.48
25-SL
6.654
0.19
22-PI
25-SL
28-SI
4.261
28-SI
5.416
31-ST
31-ST
Cb28 MZUSP PL 1103
31-ST
Cb25 MZUSP PL 324
01-AR
Outgroup
2.0
12.5
10.0
7.5
5.0
2.5
0.0 Mya
Miocene
Pliocene
Pleistocene
Holocene
Neogene
Quaternary

Fig. S6

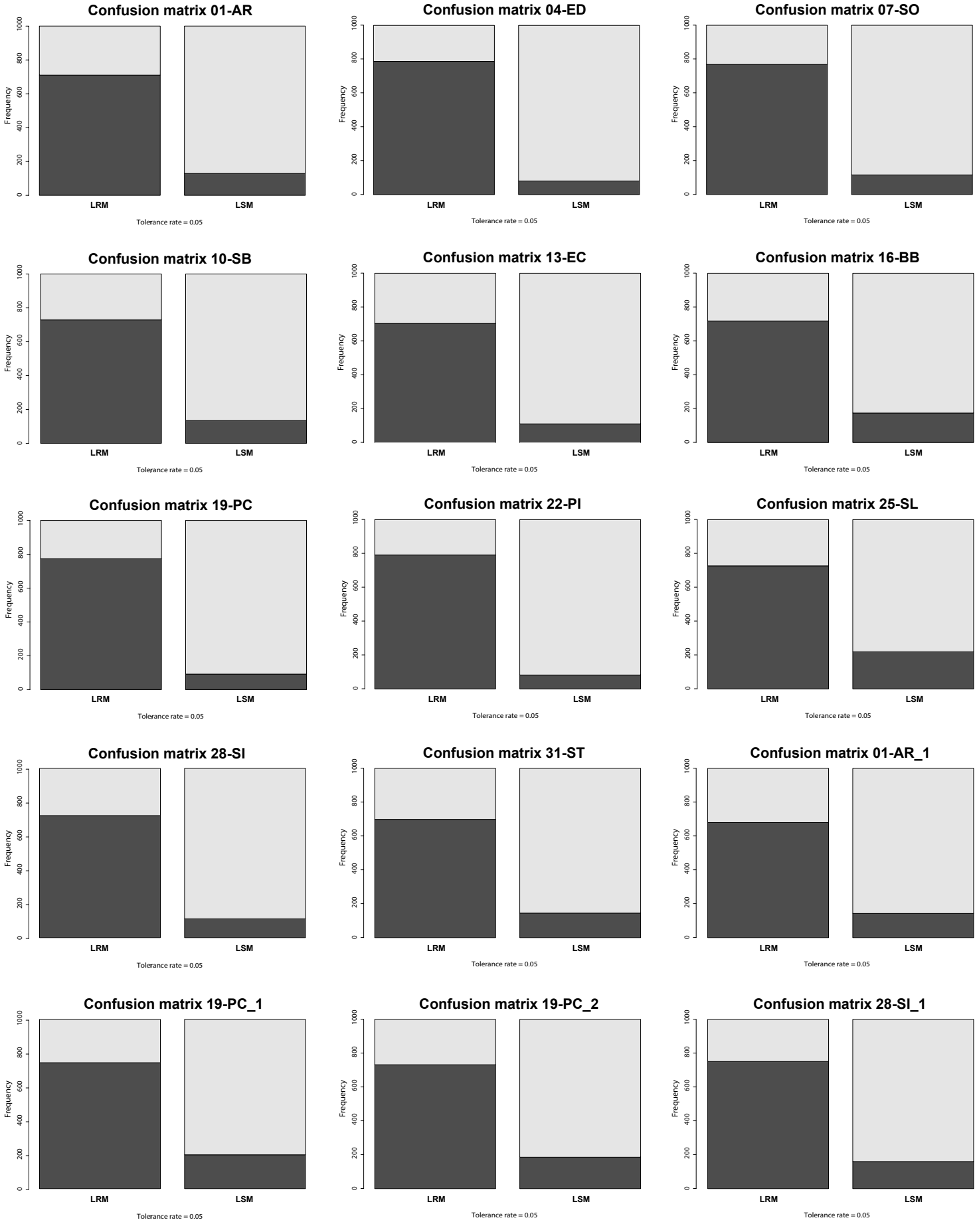

Confusion matrix 01-AR
Confusion matrix 04-ED
Confusion matrix 07-SO
Confusion matrix 10-SB
Confusion matrix 13-EC
Confusion matrix 16-BB
Confusion matrix 19-PC
Confusion matrix 22-PI
Confusion matrix 25-SL
Confusion matrix 28-SI
Confusion matrix 31-ST
Confusion matrix 01-AR_1
Confusion matrix 19-PC_1
Confusion matrix 19-PC_2
Confusion matrix 28-SI_1
Frequency
0
200
400
600
800
1000
LRM
LSM
Tolerance rate = 0.05

Fig. S7

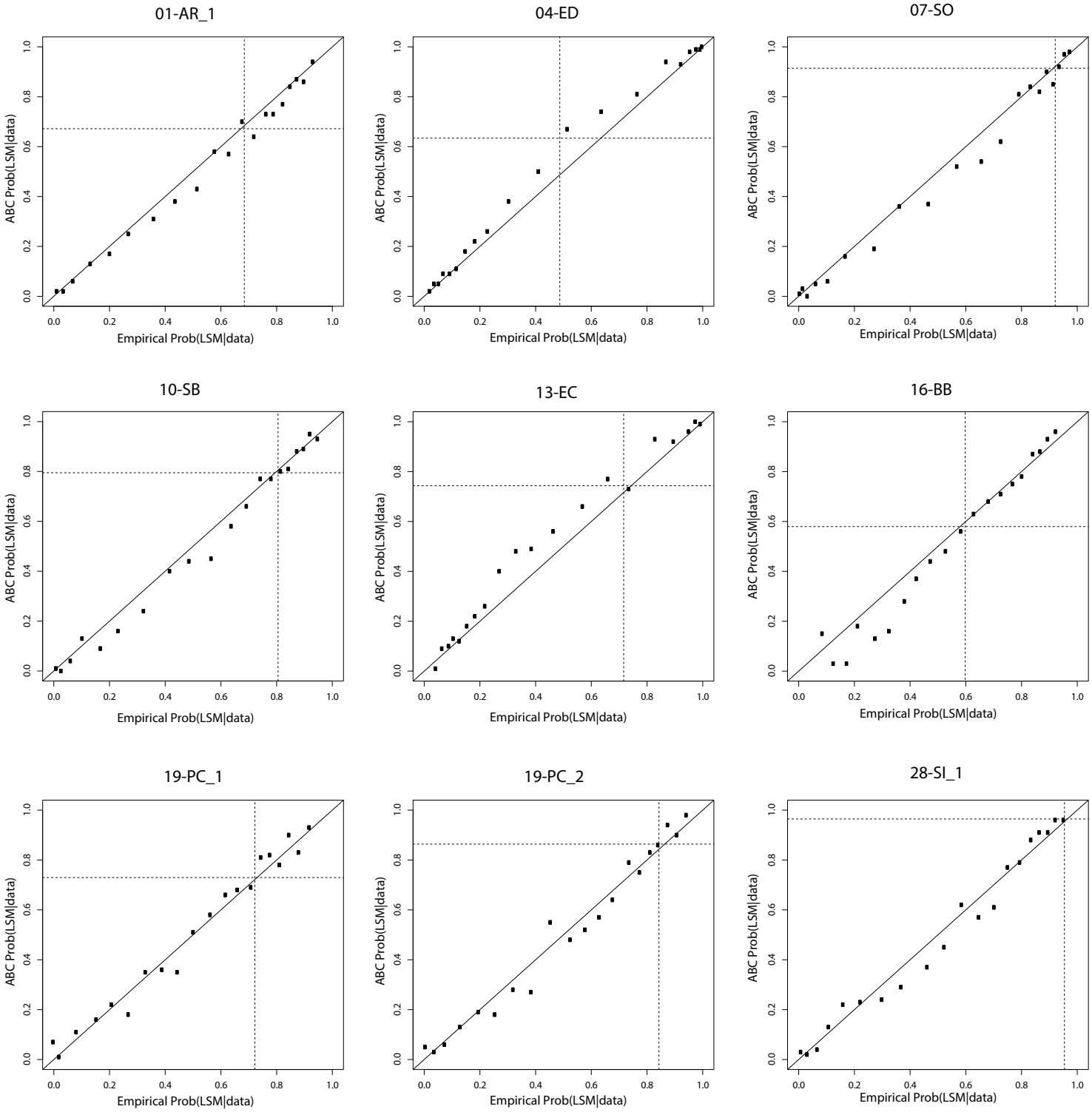

01-AR_1
04-ED
07-SO
10-SB
13-EC
16-BB
19-PC_1
19-PC_2
28-SI_1
ABC Prob(LSM|data)
Empirical Prob(LSM|data)
0.0
0.2
0.4
0.6
0.8
1.0

Fig. S8

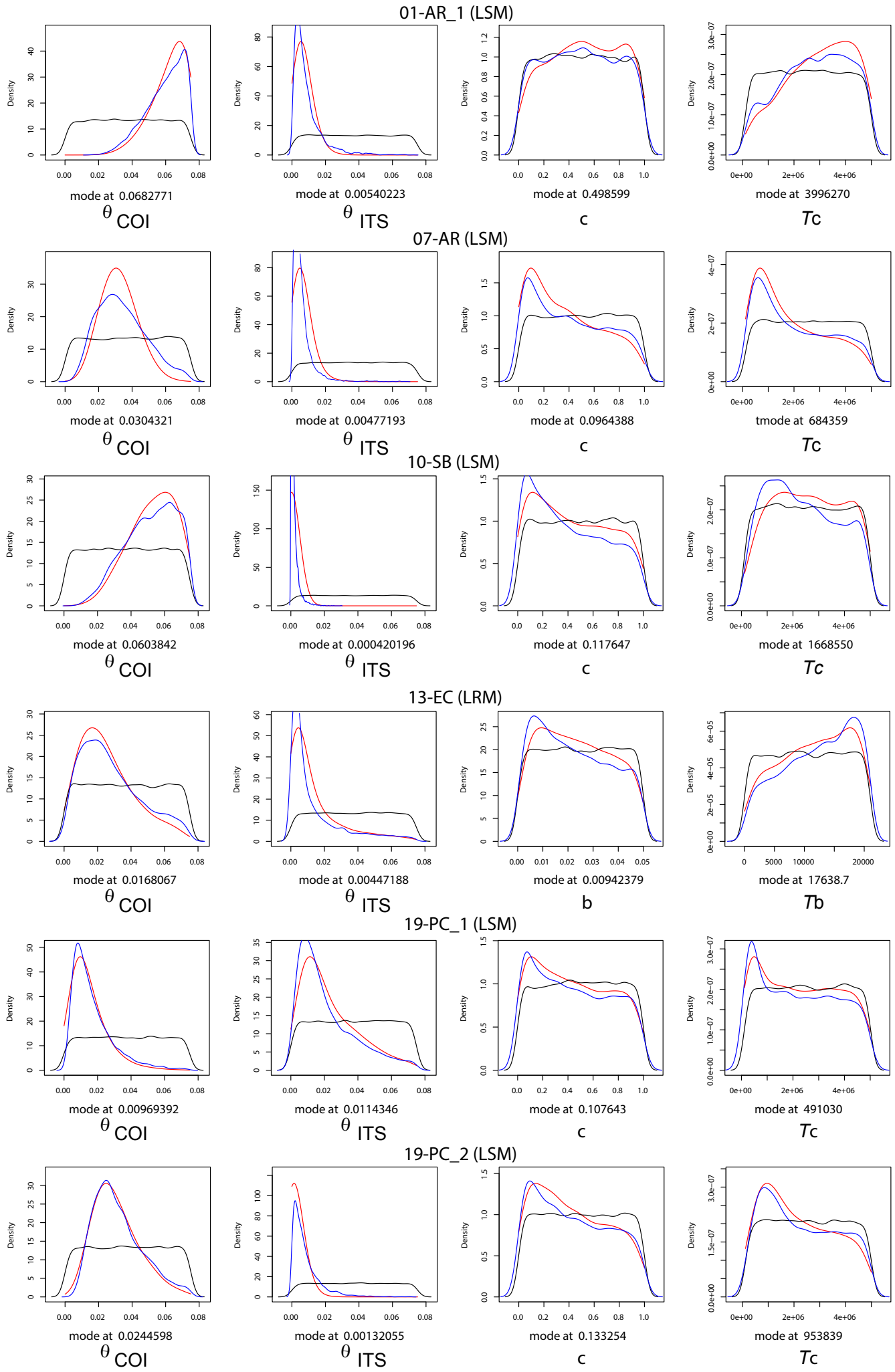
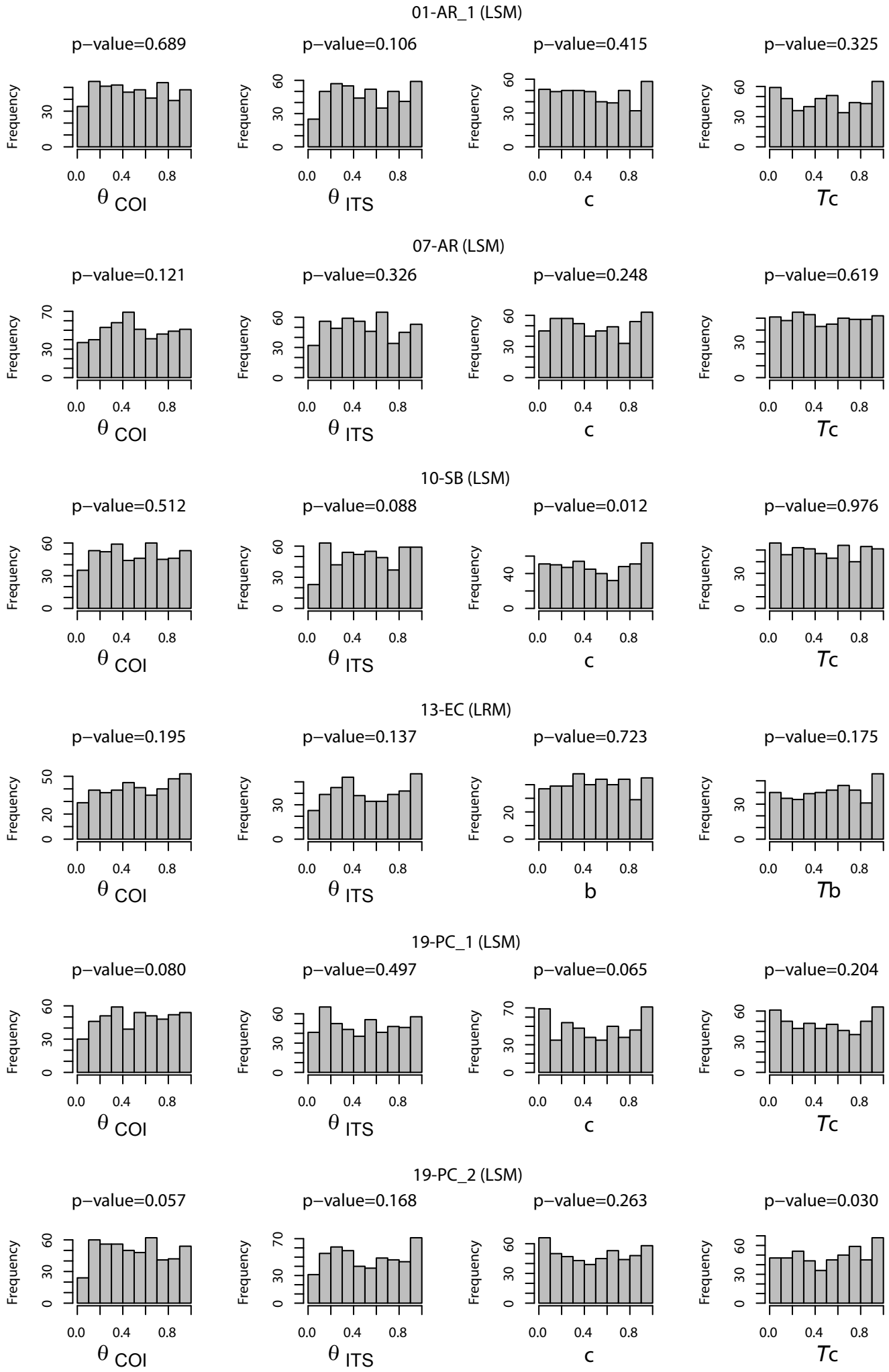

01-AR_1 (LSM)
mode at 0.0682771
θ COI
mode at 0.00540223
θ ITS
mode at 0.498599
c
mode at 3996270
Tc
07-AR (LSM)
mode at 0.0304321
θ COI
mode at 0.00477193
θ ITS
mode at 0.0964388
c
tmode at 684359
Tc
10-SB (LSM)
mode at 0.0603842
θ COI
mode at 0.000420196
θ ITS
mode at 0.117647
c
mode at 1668550
Tc
13-EC (LRM)
mode at 0.0168067
θ COI
mode at 0.00447188
θ ITS
mode at 0.00942379
b
mode at 17638.7
Tb
19-PC_1 (LSM)
mode at 0.00969392
θ COI
mode at 0.0114346
θ ITS
mode at 0.107643
c
mode at 491030
Tc
19-PC_2 (LSM)
mode at 0.0244598
θ COI
mode at 0.00132055
θ ITS
mode at 0.133254
c
mode at 953839
Tc

Fig. S9


01-AR_1 (LSM)
p−value=0.689
p−value=0.106
p−value=0.415
p−value=0.325
07-AR (LSM)
p−value=0.121
p−value=0.326
p−value=0.248
p−value=0.619
10-SB (LSM)
p−value=0.512
p−value=0.088
p−value=0.012
p−value=0.976
13-EC (LRM)
p−value=0.195
p−value=0.137
p−value=0.723
p−value=0.175
19-PC_1 (LSM)
p−value=0.080
p−value=0.497
p−value=0.065
p−value=0.204
19-PC_2 (LSM)
p−value=0.057
p−value=0.168
p−value=0.263
p−value=0.030
Frequency
θ COI
θ ITS
c
b
Tc
Tb