

A two-step DNA barcoding approach for delimiting moth species: moths of Dongling Mountain (Beijing, China) as a case study

Qian Jin^{1†}, Xi-Min Hu^{2†}, Hui-Lin Han³, Fen Chen¹, Wei-Jia Cai¹, Qian-Qian Ruan¹, Bo Liu¹, Gui-Jie Luo¹, Hao Wang¹, Xu Liu¹, Robert D. Ward⁴, Chun-Sheng Wu⁵, John-James Wilson^{6,7} and Ai-Bing Zhang^{2*}

¹*Suqian Institute of Agricultural Sciences, Jiangsu Academy of Agricultural Sciences, Suqian, Jiangsu 223800*

²*College of Life Sciences, Capital Normal University, Beijing 100048, China*

³*School of Forestry, Experiment Center, Northeast Forestry University, Haerbin 150040, China*

⁴*CSIRO National Research Collections Australia, GPO Box 1538, Hobart, Tasmania 7001, Australia*

⁵*Institute of Zoology, Chinese Academy of Sciences, 1 Beichen West Road, Chaoyang District, Beijing 100101, China*

⁶*School of Applied Sciences, Faculty of Computing, Engineering and Science, University of South Wales, Pontypridd, CF37 4AT, United Kingdom*

⁷*Department of Microbiology and Parasitology, Faculty of Medical Science, Naresuan University, Phitsanulok 65000, Thailand*

*Corresponding: E-mail: zhangab2008@gmail.com/zhangab2008@mail.cnu.edu.cn

†these authors contributed equally to this work

Supporting Information

FigureS1. Phylogenetic tree of Lepidoptera, inferred using neighbor-joining analysis of 351 COI sequences. Red colour highlights non-monophyletic at species level and blue colour highlights non-monophyletic at genus level.

FigureS2. Phylogenetic trees (ML) of 12 *Hypena* species with outgroups, constructed with different genes COI, COII, 28S, EF-1 α , and Wgl, and in combination, mtDNA (COI, COII), nDNA(28S, EF-1 α , Wgl) and mt+nDNA (COI, COII, 28S, EF-1 α , Wgl). Clades with different colors indicate different species. Numbers above or below branches indicate bootstrap values (less than 50 not shown).

FigureS3. Phylogenetic trees (NJ) of 12 *Hypena* species with outgroups, constructed with different genes COI, COII, 28S, EF-1 α , and Wgl, and in combination, mtDNA (COI, COII), nDNA(28S, EF-1 α , Wgl) and mt+nDNA (COI, COII, 28S, EF-1 α , Wgl). Clades with different colors indicate different species. Numbers above or below branches indicate bootstrap values (less than 50 not shown).

FigureS4. Phylogenetic trees (MP) of 12 *Hypena* species with outgroups constructed with different genes COI, COII, 28S, EF-1 α , and Wgl, and in combination, mtDNA (COI, COII), nDNA(28S, EF-1 α , Wgl) and mt+nDNA (COI, COII, 28S, EF-1 α , Wgl). Clades with different colors indicate different species. Numbers above or below branches indicate bootstrap values (less than 50 not shown).

FigureS5. Frequency distribution of intraspecific and interspecific genetic divergence in *Hypena* with different data sets COI, COII, 28S, EF-1 α , and Wgl, and in combination, mtDNA (COI, COII), nDNA(28S, EF-1 α , Wgl) and mt+nDNA (COI, COII, 28S, EF-1 α , Wgl). Total numbers of comparisons: 277 intraspecific and 1208 interspecific pairs across 12 *Hypena* species. Divergences calculated using Kimura's two parameter (K2P) model.

FigureS6. MOTU inferred in five loci and three combined gene categories of *Hypena* species using jMOTU method. a) Variation in the number of MOTUs inferred at different cut-off values (in the range 1-50 bp for COI, COII and combined gene nDNA, 1-30bp for 28S, EF-1 α and Wgl genes,

and 1–120 bp for combined genes mtDNA and mt+nDNA). Critical cutoff intervals for the different datasets are indicated with shaded red sections. b) Comparison of morphological species identified with MOTU compositions at the critical cutoff intervals. Different colors represent different MOTUs.

FigureS7. Automatic partition of the eight single and combined locus data sets using ABGD (Automatic Barcode Gap Discovery) method.

FigureS8. GMYC analysis of 12 *Hypena* species with five single loci COI, COII, 28S, EF-1 α , and Wgl. Ultrametric trees of *Hypena* species constructed with different loci (left) and lineage-through-time plots (right). The red vertical lines on the ultrametric trees indicate the maximum likelihood transition points of the switch in branching rates, as estimated by a general mixed Yule-coalescent (GMYC) model. The GMYC analysis was performed using a single threshold.

FigureS9. Scatterplot of a principal component analysis (PCA) in shape space of forewings. Different colour represents different species which are largely overlapping along PC1 and PC2 in Fig 5. *Red* - *sp1*, *yellow* - *sp3*, *green* - *sp5*, *blue* - *H. squalida*, *fuschia* - *H. stygiana*.

FigureS1



Figure

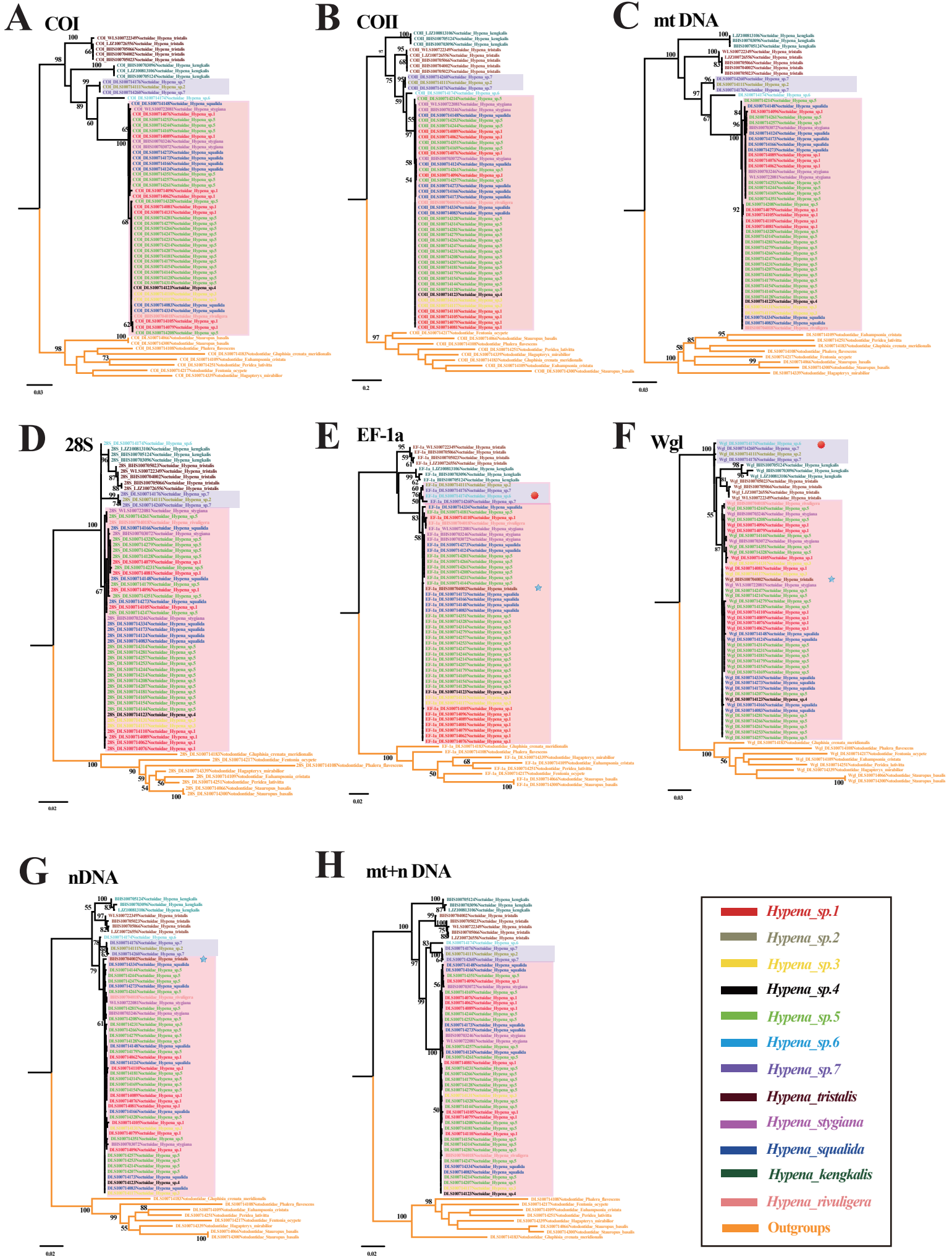
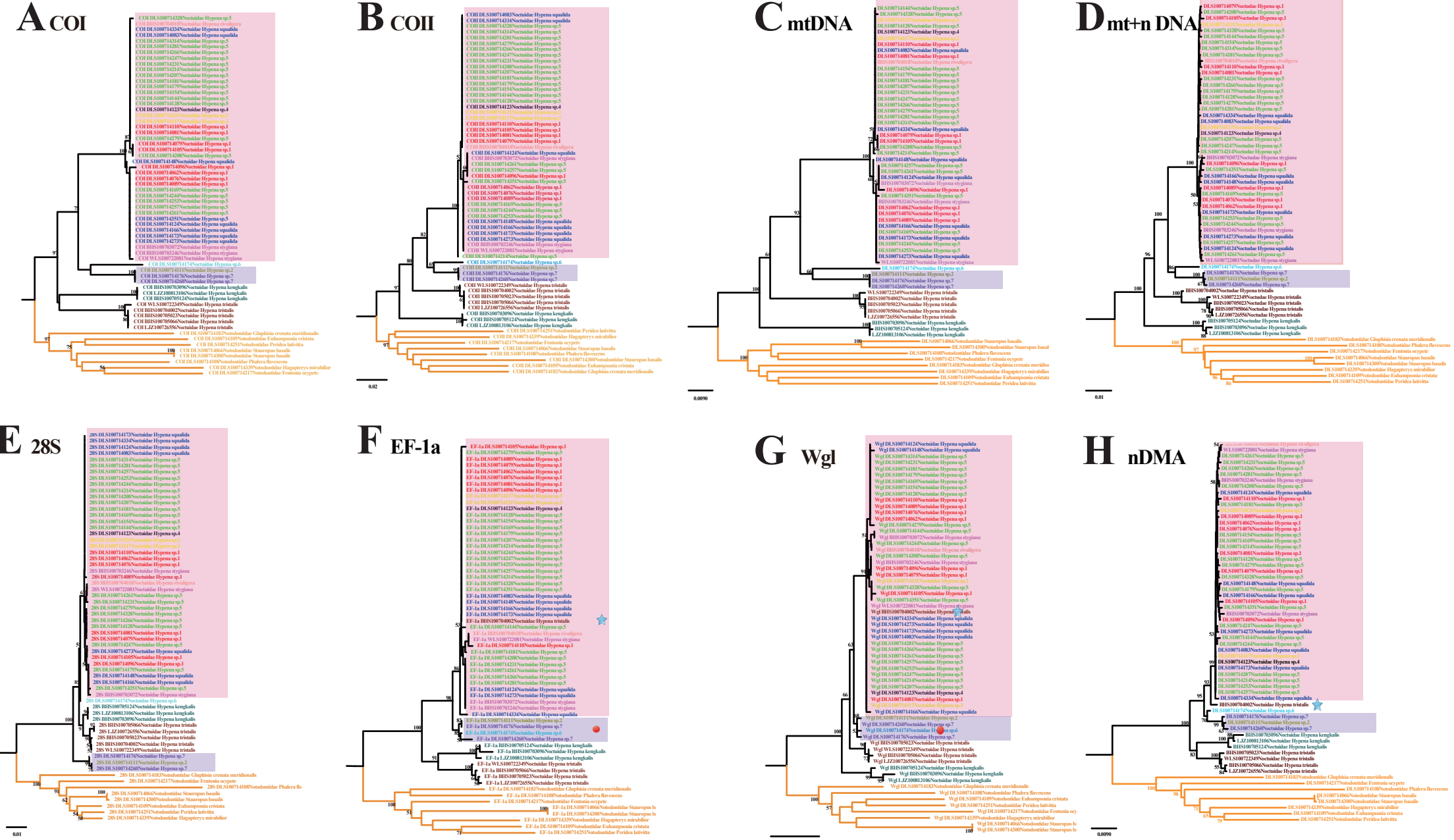
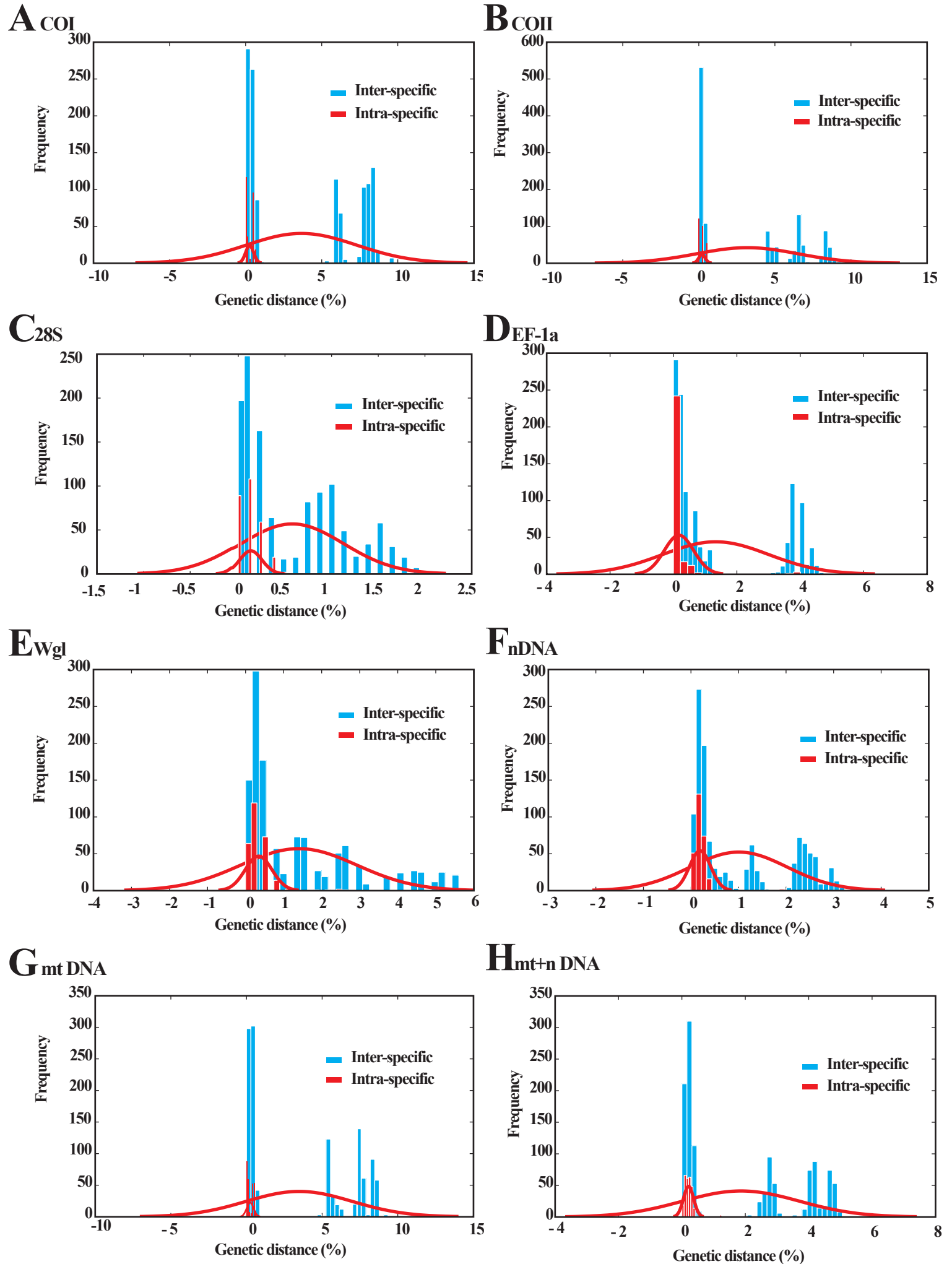


Figure S3

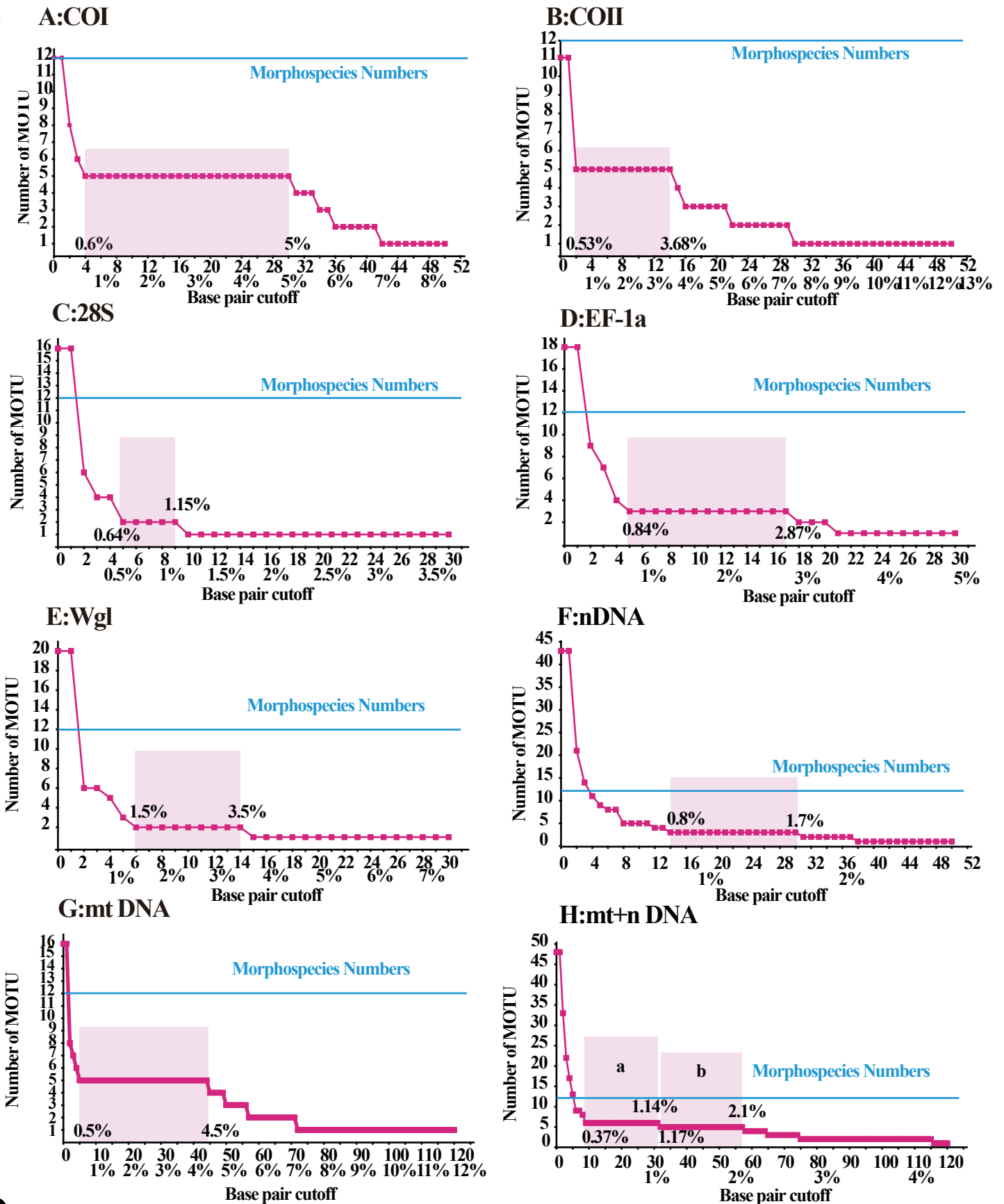


FigureS5



FigureS6

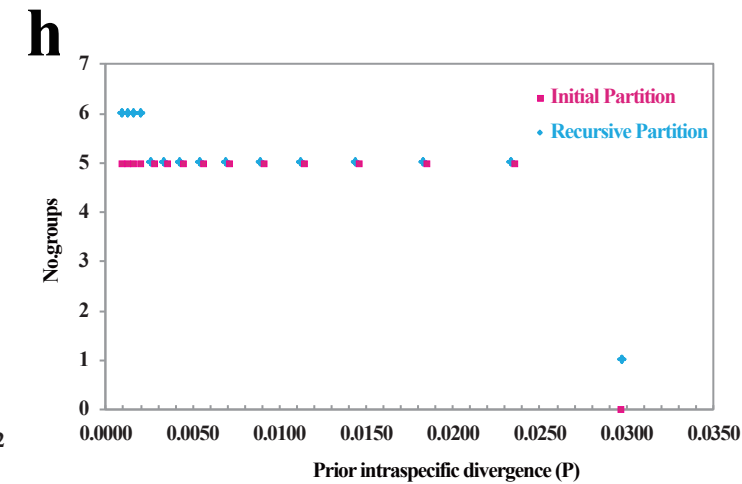
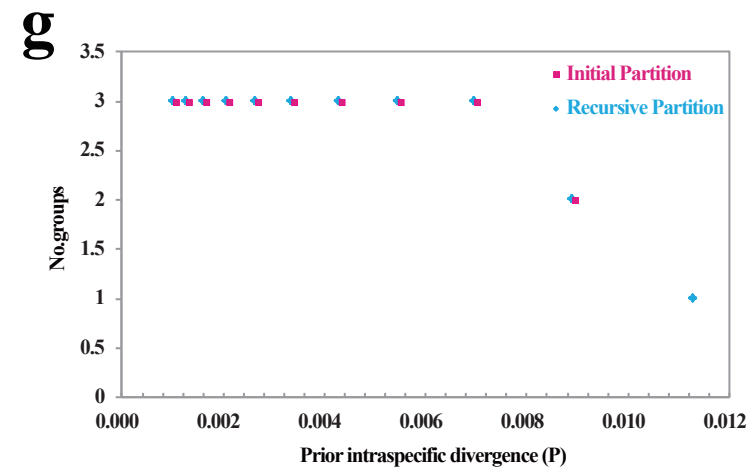
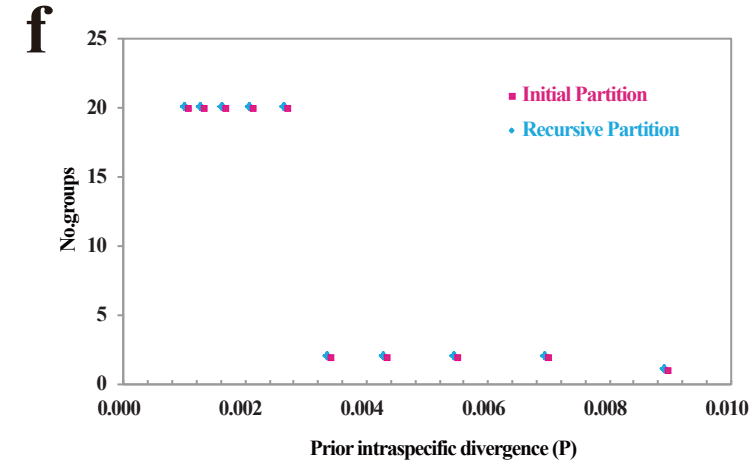
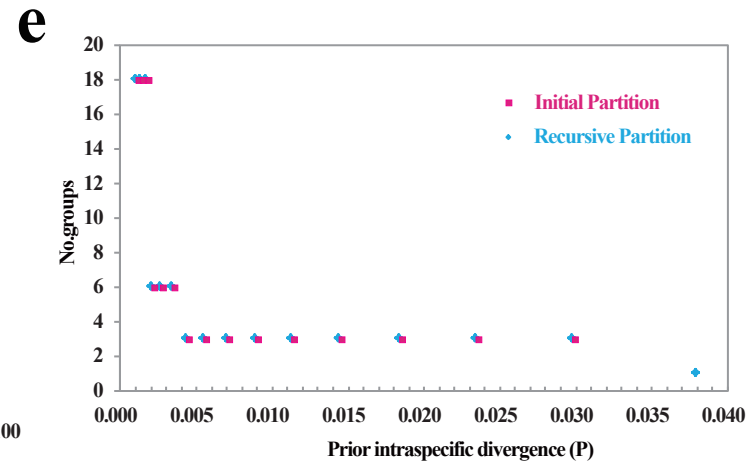
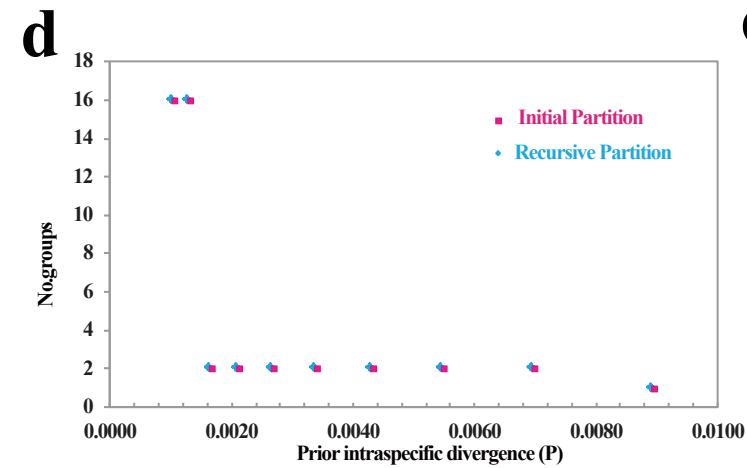
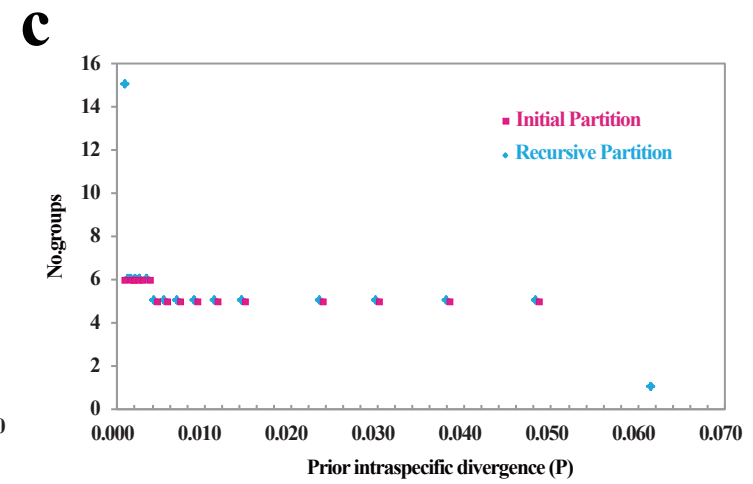
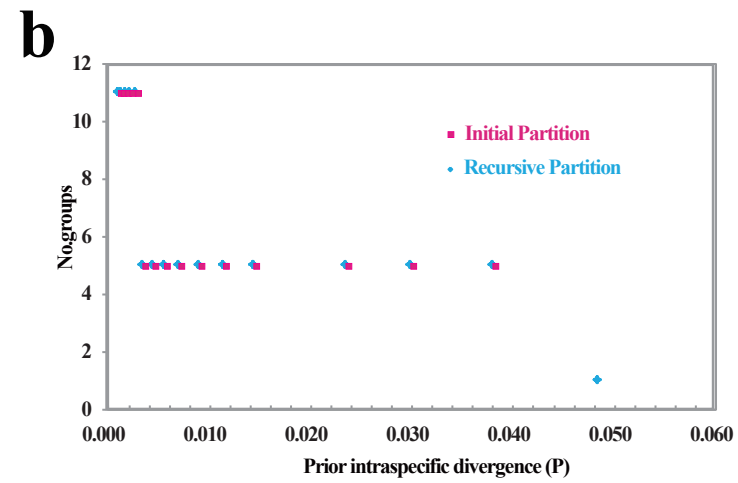
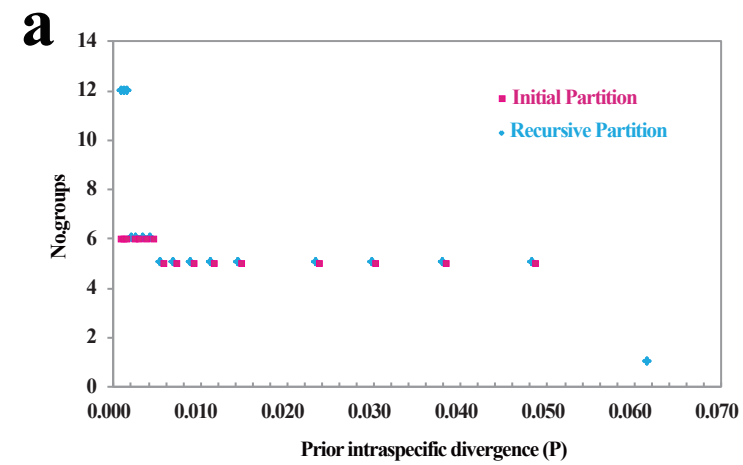
a



b



FigureS7



a : COI

e : EF-1a

b : COII

f : Wgl

c : COI+COII

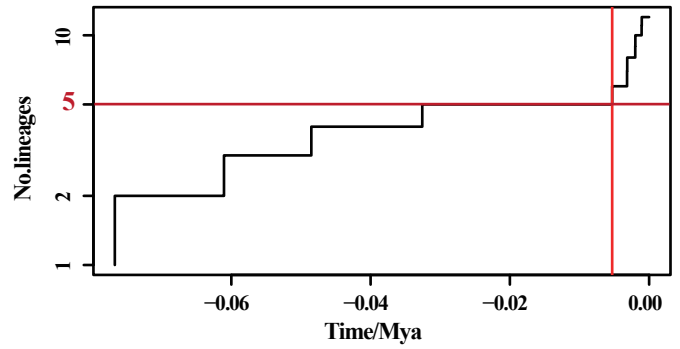
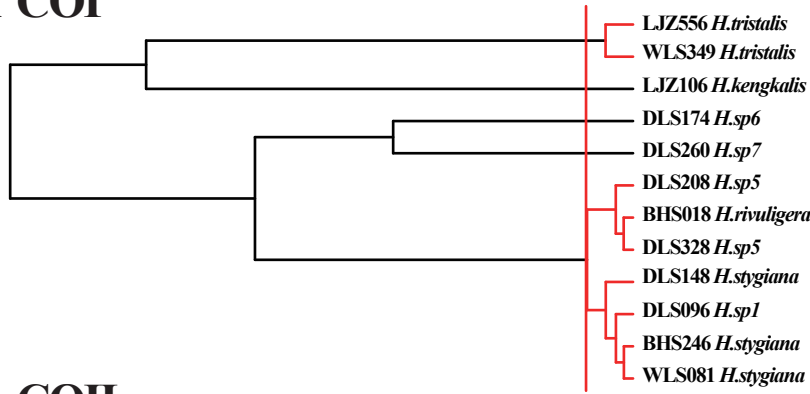
g : Nucleargenes

d : 28S

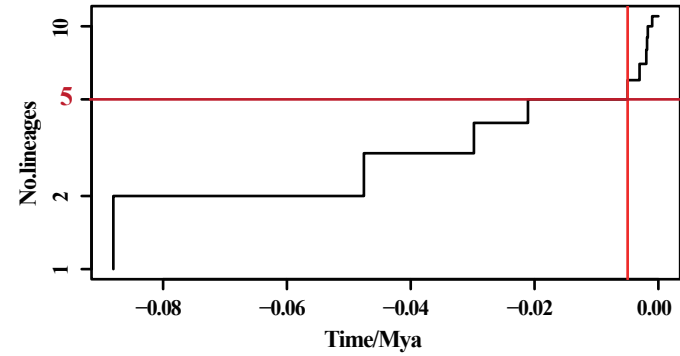
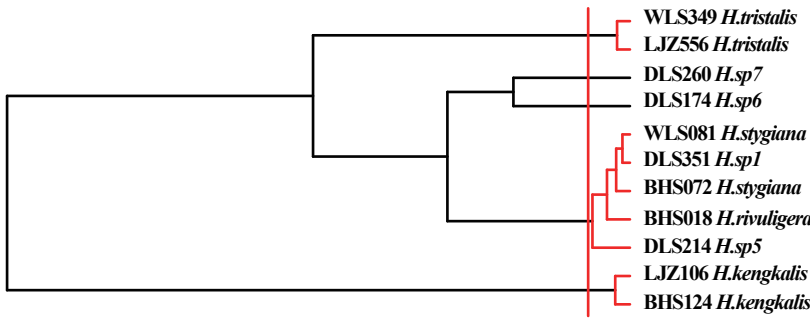
h : All 5 genes

FigureS8

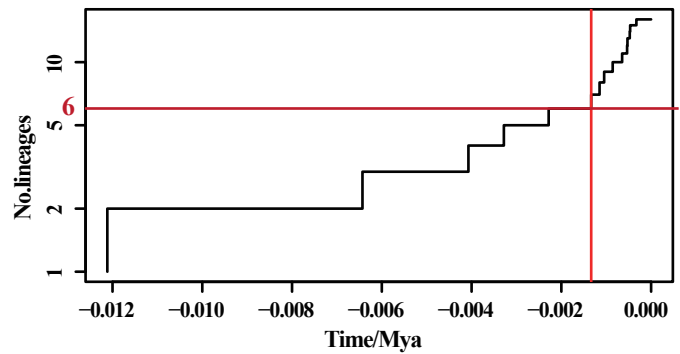
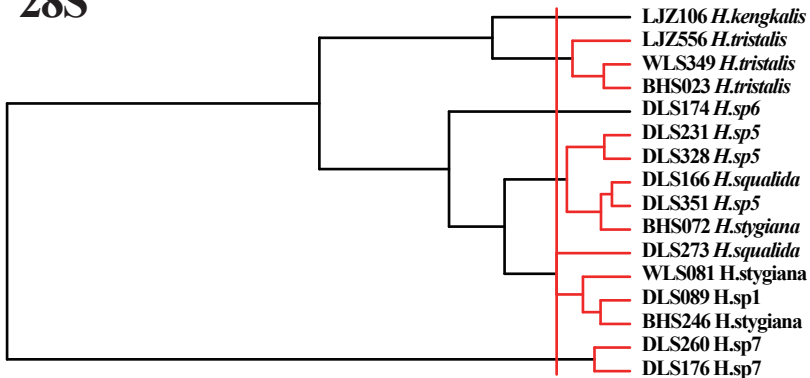
A COI



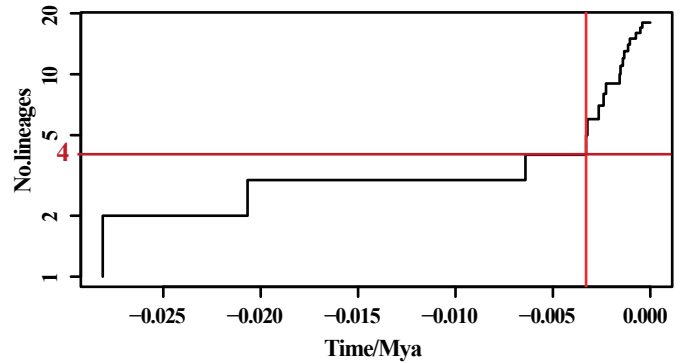
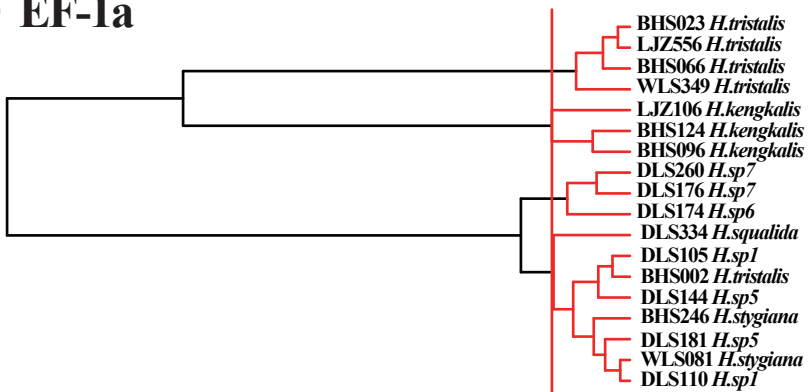
B COII



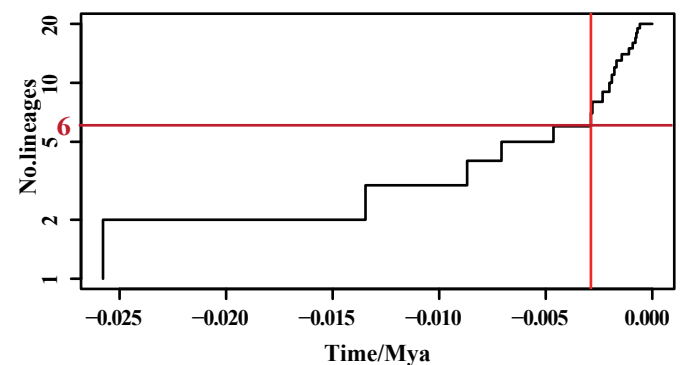
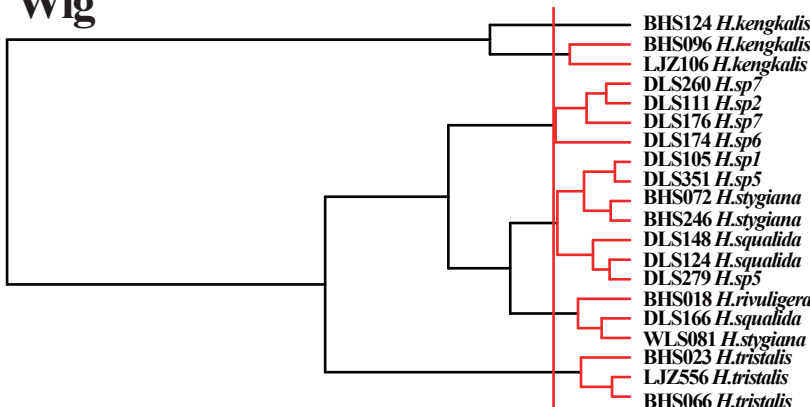
C 28S



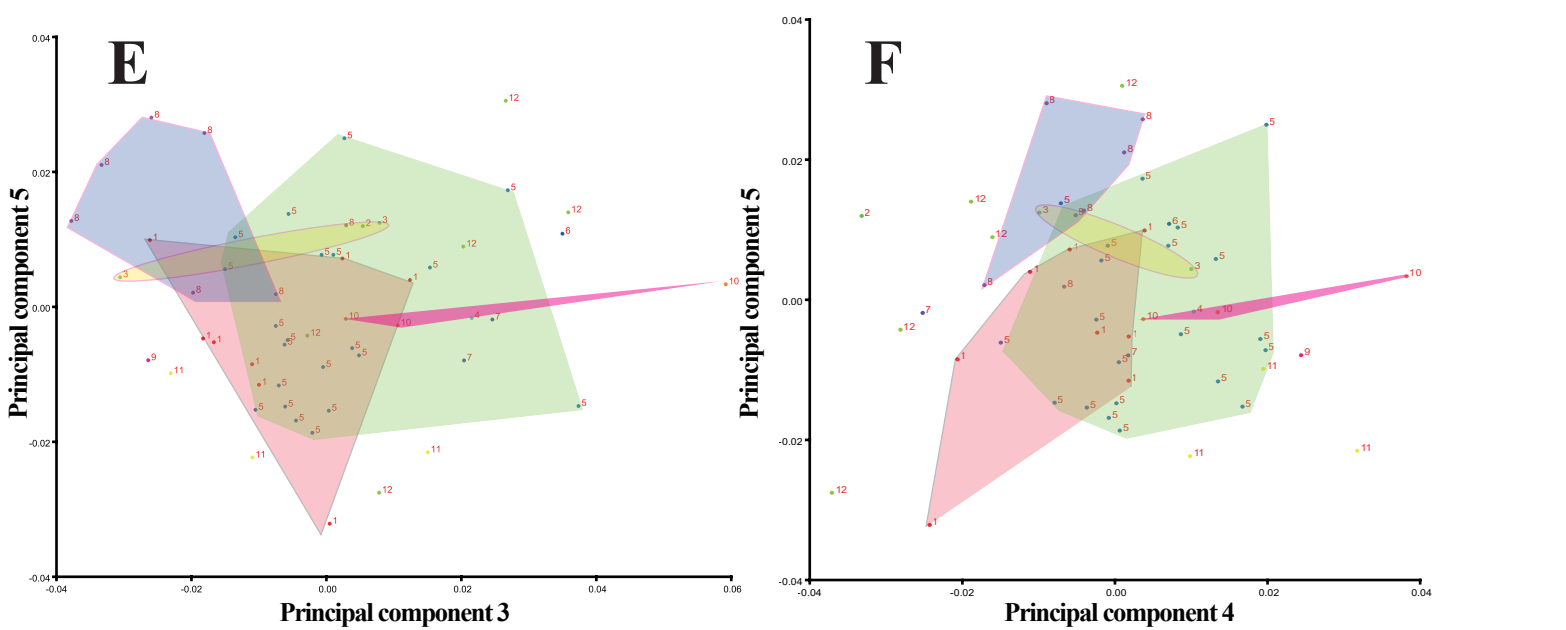
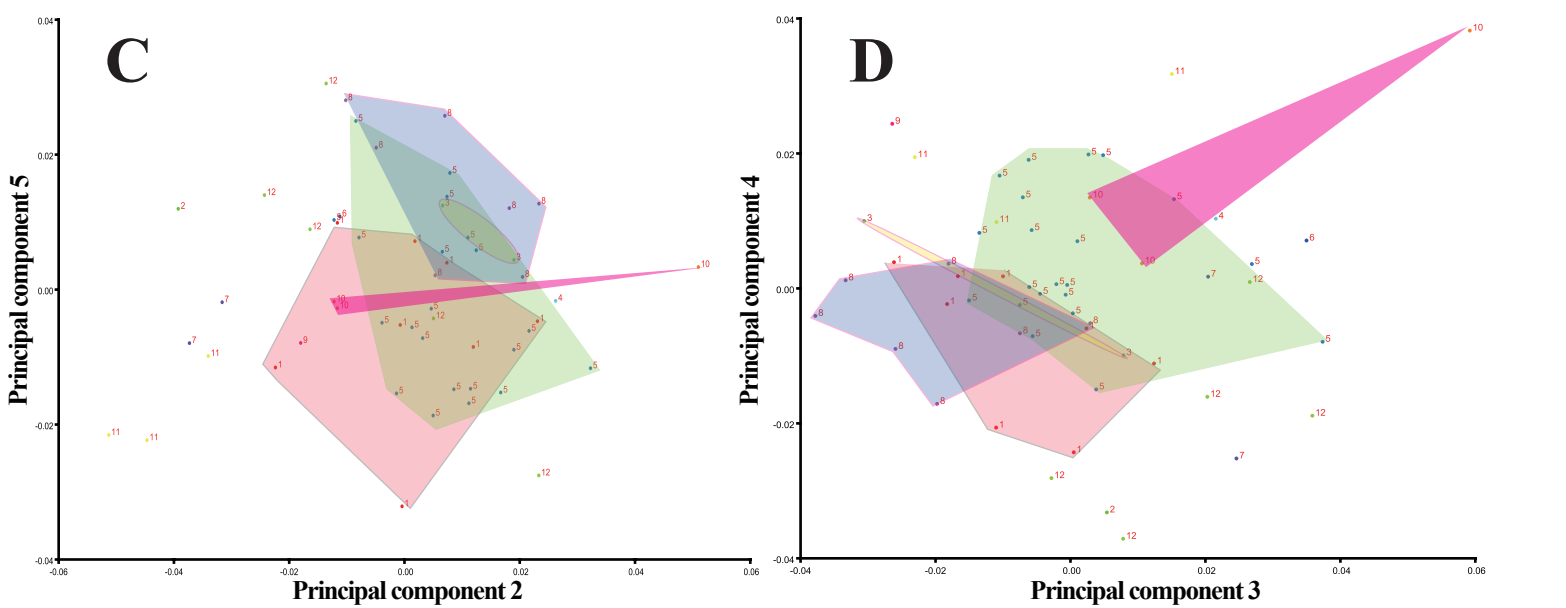
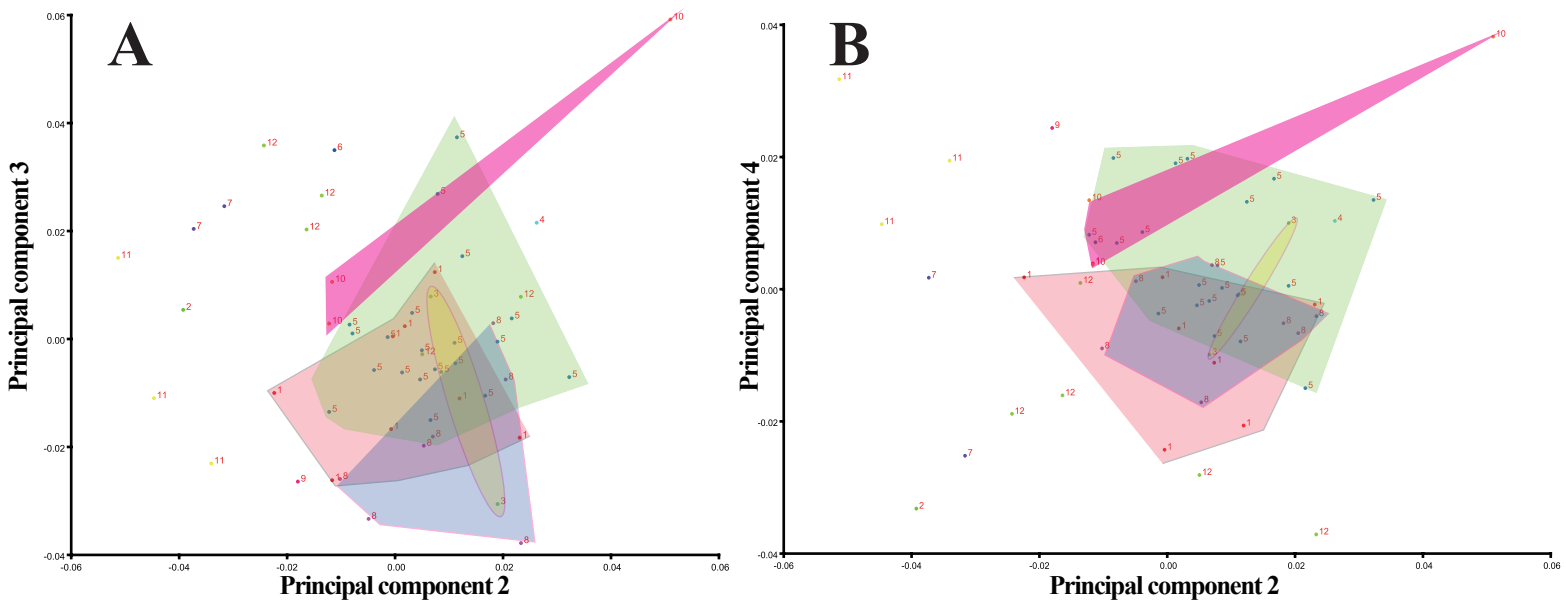
D EF-1a



E Wlg



FigureS9



LEGEND

