Ten genes and two topologies: an exploration of higher relationships in skipper butterflies (Hesperiidae)

Ranjit Kumar Sahoo [Corresp., 1] , **Andrew D. Warren** [2] , **Niklas Wahlberg** [3, 4] , **Andrew V. Z. Brower** [5] , **Vladimir A Lukhtanov** [6] , **Ullasa Kodandaramaiah** [1]

[1] School of Biology, Indian Institute of Science Education and Research Thiruvananthapuram, Thiruvananthapuram, Kerala, India

[2] McGuire Center for Lepidoptera and Biodiversity, Florida Museum of Natural History, University of Florida, PO Box 112710, 3215 Hull Rd., UF Cultural Plaza, Gainesville, Florida 32611-2710, United States

[3] Department of Biology, Lund University, Lund, Sweden

[4] Department of Biology, University of Turku, Turku, Finland

[5] Evolution and Ecology Group, Department of Biology, Middle Tennessee State University, Murfreesboro, Tennessee, United States

[6] Department of Insect Systematics, Zoological Institute of Russian Academy of Sciences, Universitetskaya nab, St. Petersburg, Russia

Corresponding Author: Ranjit Kumar Sahoo
Email address: sahoork@iisertvm.ac.in

Despite multiple attempts to infer the higher level phylogenetic relationships of skipper butterflies (Family Hesperiidae), uncertainties in the deep clade relationships persist. The most recent phylogenetic analysis included fewer than 30% of known genera and data from three gene markers. We here reconstruct the higher-level relationships with a rich sampling of ten nuclear and mitochondrial markers (7726 base pairs) from 270 genera and find two distinct but equally plausible topologies among subfamilies at the base of the tree. In one set of analyses, the nuclear markers suggest two contrasting topologies, one of which is supported by the mitochondrial dataset. However, another set of analyses suggests mito-nuclear conflict as the reason for topological incongruence. Neither topology is strongly supported, and we conclude that there is insufficient phylogenetic evidence in the molecular dataset to resolve these relationships. Nevertheless, taking morphological characters into consideration, we suggest that one of the topologies is more likely.

1 **Title**: Ten genes and two topologies: an exploration of higher relationships in
2 skipper butterflies (Hesperiidae).

5 **Authors**: Ranjit Kumar Sahoo[1], Andrew D. Warren[2], Niklas Wahlberg[3, 4], Andrew V. Z.
6 Brower[5], Vladimir A. Lukhtanov[6], Ullasa Kodandaramaiah[1]

9 [1]School of Biology, Indian Institute of Science Education and Research Thiruvananthapuram,
10 Kerala 695016, India.

12 [2]McGuire Center for Lepidoptera and Biodiversity, Florida Museum of Natural History,
13 University of Florida, PO Box 112710, 3215 Hull Rd., UF Cultural Plaza, Gainesville, FL
14 32611-2710 USA

16 [3]Department of Biology, University of Turku, 20014 Turku, Finland

17 [4]Department of Biology, Sölvegatan 37, Lund University, 223 62 Lund, Sweden

19 [5]Evolution and Ecology Group, Department of Biology, Middle Tennessee State University,
20 Murfreesboro, TN 37037 USA

22 [6]Department of Insect Systematics, Zoological Institute of Russian Academy of Sciences,
23 Universitetskaya nab. 1, 199034 St. Petersburg, Russia

25 Corresponding Author:

26   Ranjit Kumar Sahoo[1]

27   School of Biology, Indian Institute of Science Education and Research Thiruvananthapuram,

28   Kerala 695016, India.

29   sahoork@iisertvm.ac.in

30

31

32

33

34

35

36

37

38   **Abstract**:

39   Despite multiple attempts to infer the higher level phylogenetic relationships of skipper
40   butterflies (Family Hesperiidae), uncertainties in the deep clade relationships persist. The most
41   recent phylogenetic analysis included fewer than 30% of known genera and data from three gene
42   markers. We here reconstruct the higher-level relationships with a rich sampling of ten nuclear
43   and mitochondrial markers (7726 base pairs) from 270 genera and find two distinct but equally
44   plausible topologies among subfamilies at the base of the tree. In one set of analyses, the nuclear
45   markers suggest two contrasting topologies, one of which is supported by the mitochondrial
46   dataset. However, another set of analyses suggests mito-nuclear conflict as the reason for
47   topological incongruence. Neither topology is strongly supported, and we conclude that there is
48   insufficient phylogenetic evidence in the molecular dataset to resolve these relationships.
49   Nevertheless, taking morphological characters into consideration, we suggest that one of the
50   topologies is more likely.

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

66

67

68

69

70

71

72

73

74

75

76

77

78

79

80

81

82

83

84

85

86

87

88 **Introduction:**

89

90 A robust phylogeny is the key to understanding historical macroevolutionary processes that have
91 shaped extant diversity. For instance, a phylogenetic hypothesis is needed to address questions
92 regarding patterns of morphological evolution, coevolution, and historical biogeography, as well
93 as for a higher-level classification system. Among invertebrates, butterflies have been the most
94 popular study systems in evolutionary biology (Boggs et al. 2003). Relationships among and
95 within butterfly families have been largely studied by phylogenetic analyses of DNA sequence
96 data (Campbell, Brower & Pierce, 2000; Caterino et al., 2001; Braby, Vila & Pierce, 2006;
97 Nazari, Zakharov & Sperling, 2007; Wahlberg et al., 2009; Simonsen et al., 2011; Heikkila et al.,
98 2012; Wahlberg et al., 2014; Espeland et al., 2015). Yet, the higher level relationships among
99 skipper butterflies, with more than 4000 species in about 567 genera (Warren, Ogawa & Brower,
100 2008) and representing a fifth of the world's butterfly fauna (Hernández-Roldán, Bofill &
101 Dapporto, 2014), are still unsatisfactorily resolved.

102

103 Until recently, the higher-level classification of the family that has been generally followed was
104 that proposed by Evans (1949) based on morphological characters. However, a major problem
105 with skipper systematics is the remarkable uniformity of morphological structure among skipper
106 taxa, which makes phenotype-based grouping extremely challenging (Voss, 1952; Warren,
107 Ogawa & Brower, 2008). Following multiple attempts over the last several decades (Voss, 1952;
108 Ackery, 1984; Scott, 1985; Scott and Wright, 1990; Chou, 1994, 1999; Mielke, 2005), a recent
109 study employing molecular data suggested a classification that included five subfamilies (Warren,
110 Ogawa & Brower, 2008). This classification relied on analyses of one mitochondrial and two

111   nuclear markers, a dataset of 2085 bp (base pairs). A subsequent analysis that added
112   morphological data (49 characters) to the same molecular data led to a revised classification that
113   included seven subfamilies (Warren, Ogawa & Brower, 2009).

114

115   Warren and colleagues (Warren, Ogawa & Brower, 2008; Warren, Ogawa & Brower, 2009) used
116   Maximum Parsimony analyses with nodal support estimated through Bremer Support values
117   (Bremer, 1994). Despite some strongly supported monophyletic taxa being recovered, many
118   putative higher clades were unresolved. Specifically, uncertainty remained about relationships
119   among the major clades within the subfamily Pyrginae. Furthermore, support for relationships
120   among the monophyletic subfamilies Heteropterinae, Trapezitinae and Hesperiinae was weak to
121   moderate. The status of Euschemoninae as sister to rest of the family, except Coeliadinae,
122   received very low nodal support (Bremer support = 1), although this placement is corroborated
123   by the early developmental characters of Euschemoninae, which are similar to those of
124   Coeliadinae and Eudaminae (Warren, Ogawa & Brower, 2009).

125

126   Yuan et al. (2015) investigated relationships among a small subset of hesperiid taxa - 23 genera
127   from China - using 1458 bp of mitochondrial sequence data. Their Maximum Likelihood tree
128   also indicated uncertainty in the position of Eudaminae and Pyrginae. Another study based on
129   complete mitochondrial genomes of six skipper butterflies representing five subfamilies (sensu
130   Warren, Ogawa & Brower, 2009) failed to support the monophyly of Pyrginae (Kim et al., 2014).

131

132   In summary, existing studies on the higher-level relationships within this speciose butterfly
133   family have indicated significant conflicts (Warren, Ogawa & Brower, 2008; Warren, Ogawa &
134   Brower, 2009; Kim et al., 2014; Yuan et al., 2015), and we currently lack a robust higher level
135   phylogenetic hypothesis for evolutionary studies or a sub-family level classification. The reasons
136   for conflicting topologies across studies and poor nodal support could be a) incongruence among
137   gene trees due to incomplete lineage sorting (Pollard et al., 2006; Whitfield & Lockhart, 2007), b)
138   ancestral introgression (Eckert & Carstens, 2008), c) differences in characteristics of the datasets
139   used (Nabholz et al., 2011), d) poor taxon sampling, e) insufficient data to resolve deeper nodes
140   (Wolf et al., 2002; Rokas et al., 2003) or, f) a near-hard polytomy due to a rapid-radiation
141   (Kodandaramaiah et al., 2010).

142

143   In order to bring further understanding to the higher-level phylogeny of skipper butterflies, we
144   assembled sequences of ten gene regions from 270 genera and analyzed a 7726 bp dataset using
145   both parsimony and model-based tree reconstruction methods. We also compiled the complete

146   mitochondrial genome of 15 skipper species across five subfamilies from GenBank to compare
147   the tree from the mitochondrial genome with that of single mitochondrial and combined-nuclear
148   genes. Consistent with the existing conflict across studies (Warren, Ogawa & Brower, 2008;
149   Warren, Ogawa & Brower, 2009; Kim et al., 2014; Yuan et al., 2015), our analyses showed
150   conflicting topologies at the deeper nodes of the phylogeny. To understand the reasons for the
151   uncertainty in the phylogenetic estimation, we followed an integrative approach with systematic
152   data encoding and tree comparison.

153

154   **Methods:**

155

156   *Taxon and gene sampling:*

157   Our analyses were based on 311 ingroup specimens representing 270 hesperiid genera and 12
158   outgroup taxa (5 Papilionidae, 2 Hedylidae and 5 Pieridae). This dataset builds on the previous
159   study by Warren, Ogawa & Brower (2009) that included sequences of three protein-coding genes
160   (mtDNA COI, EF1a and wingless). We sequenced an additional part of COI and seven more
161   genes (ArgKin, CAD, GAPDH, IDH, MDH, RpS2 and RpS5) using protocols and primers from
162   Wahlberg & Wheat (2008). A new primer-pair was designed (Table S1) to amplify the gene IDH
163   for certain taxa. We have also included 96 additional specimens representing 71 genera to the
164   present analyses. Our taxon sampling accounts for 60-70% of the genera of Coeliadinae (6
165   genera), Eudaminae (38 genera), Heteropterinae (7 genera), and Trapezitinae (11 genera); 40-50
166   % of Pyrginae (73 genera) and Hesperiinae (134 genera), and 100 % of Euschemoninae (1
167   genus). The sequences for outgroups were acquired from GenBank. We included the
168   morphological, behavioral and ecological data matrix used in Warren, Ogawa & Brower (2009)
169   in certain analyses. The sequences generated during this study were deposited in GenBank. The
170   molecular data matrix in our study comprised 7726 characters, more than three times that of the
171   previous dataset (Warren, Ogawa & Brower, 2008; Warren, Ogawa & Brower, 2009).

172

173   *Dataset encoding*:

174   Along with the analysis of the concatenated dataset (nt_123), we generated context specific
175   datasets from the concatenated gene matrix for various analyses designed to identify potential
176   sources of conflict and/or poor nodal support. In one analysis, accounting for the impact of
177   compositional heterogeneity, we assigned ambiguity to all the sites that potentially experience
178   synonymous change (degen_1) (Regier et al., 2010; Zwick et al., 2012). We also checked the
179   extent of substitution saturation in each gene matrix using DAMBE v6.4.20 (Xia, 2013), which
180   showed saturation in 3rd codon positions (Figure S1). To account for substitution saturation and

181   degeneracy, we removed the 3rd codon positions and the 1st codon positions coding for Arginine
182   or Lysine (noLRall1+nt2) (Regier et al., 2008). In further analyses, we removed the 3rd coding
183   positions from the concatenated dataset (nt_12) or only from the mitochondrial genes
184   (nuclear_123+CO_12).

185

186   We also analyzed all the nuclear genes together (nuclear_123) and reconstructed multi-gene and
187   single gene trees for comparison. In subsequent analyses, we also combined the morphological,
188   ecological and behavioral characters from Warren, Ogawa & Brower (2009) with certain
189   molecular datasets – nt_123 and nuclear_123. In addition, we analysed an assembly of all protein
190   coding sequences (13 genes) from the mitochondrial genomes of 15 skipper butterflies acquired
191   from GenBank.

192

193   *Phylogenetic analyses*:

194   We performed Maximum Parsimony analyses in TNT v.1.1 (Goloboff, Farris & Nixon, 2008)
195   using 'New Technology' searches (Goloboff, 1999; Nixon, 1999) (consisting of tree fusion,
196   sectorial search, ratchet and tree drift) with 1000 random addition replicates; nodal supports were
197   derived from 1000 bootstrap replicates (Felsenstein, 1985). For Maximum Likelihood (ML) and
198   Bayesian Inference (BI) analyses, we used RAxML v8 (Stamatakis, 2014) and MrBayes v3.2
199   (Ronquist et al., 2012) respectively, on the XSEDE web server through the CIPRES Science
200   gateway (Miller, Pfeiffer & Schwartz, 2010). For ML analyses, we used the GTR model of
201   substitution with gamma model of rate heterogeneity (GTR+G) and different partition schemes,
202   either gene-based or based on rates of evolution calculated by the program TIGER (Cummins &
203   McInerney, 2011). In gene-based partitions each gene was considered as a separate partition,
204   while in TIGER-partitions, the characters were binned together based on their rate of evolution
205   regardless of gene origin (used as a partitioning strategy in Rota & Wahlberg, 2012). TIGER-
206   partitions for the dataset were derived from the program TIGER (Tree Independent Generation
207   of Evolutionary Rates) that calculates relative rates of evolution of each site in an alignment
208   (Cummins & McInerney, 2011). The data were then divided into seven partitions based on the
209   relative rates using an algorithm developed by Tobias Malm (J. Rota, T. Malm & N Wahlberg, in
210   prep.) such that the first partition consisted of the invariant and very slowly evolving sites and
211   the last partition consisted of sites evolving very quickly. To check whether model selection had
212   any impact on the tree reconstruction, we also performed the above ML analyses using the best
213   fitted model from PartitionFinder v1.1.1 (Lanfear et al., 2012) (detail in Figure S2). For each ML
214   analyses, the node supports were computed from 1000 bootstrap replicates. Single gene matrices
215   were analyzed with GTR+G model and the node supports were derived from 500 bootstrap
216   replicates. For single gene analyses, we dropped the taxa for which the corresponding gene
217   sequence was not available. For the mitogenome analysis, we performed ML tree searches with

218   codon-based partitions and estimated the nodal support from 100 bootstraps.

219

220   The BI analysis of the concatenated dataset with TIGER-partitions was performed with a mixed
221   model of substitution which samples all possible models in the GTR family in proportion to their
222   posterior distributions (Huelsenbeck et al., 2004) as implemented in MrBayes 3.2 (Ronquist et al.,
223   2012). We assigned the gamma model of rate heterogeneity to all the partitions; the first partition
224   was additionally assigned a proportion of invariable sites. The program MrBayes was set to
225   estimate the base frequencies and shape parameters from the data. Two independent runs with
226   two chains per run were performed for ~30 million generations, sampling trees every 10000
227   generations. The convergence of independent runs was analyzed from the values of potential
228   scale reduction factors (PSRF) (PSRF close to 1 determines convergence) (Gelman & Rubin,
229   1992); we also checked the plots of log-likelihoods and other parameters on Tracer v1.6
230   (Rambaut et al., 2014).

231

232   *Tree comparison*:

233   To investigate the differences among the trees from multiple analyses, we compared the trees
234   (except single gene trees) for their topological incongruence with and without the likelihood
235   scores. While the non-likelihood incongruence test reflects differences in branching patterns
236   among the topologies, the likelihood-based comparison calculates the difference between
237   competing hypotheses with distinct topologies for a given dataset (Planet, 2006). For the non-
238   likelihood based tree comparison at the deep level divergences, we used a Lento plot to depict
239   the conflict among different trees in a two-dimensional graph (Lento et al., 1995). We employed
240   the approximate unbiased (AU) test (Shimodaira, 2002) for likelihood-based tree comparisons.
241   To check for incongruence among gene trees, we used partitioned bremer support (PBS) analysis
242   that examines the contribution of each gene partition to the topological support of the consensus
243   tree (Baker & DeSalle, 1997).

244

245   **Results**:

246

247   *Multilocus tree estimation*:

248   Trees from the concatenated dataset, irrespective of parsimony or ML analysis or the partition
249   scheme used, showed identical relationships among the early branches representing the major
250   clades (sensu Warren, Ogawa & Brower, 2009) (Figure 1A) – (i) Coeliadinae was sister to the
251   rest of the family, (ii) Pyrginae was paraphyletic, (iii) Euschemoninae was sister to Eudaminae,

252   and (iv) Heteropterinae, Trapezitinae and Hesperiinae were all monophyletic. The topology
253   remained unchanged even after the addition of morphological characters to the concatenated
254   dataset.

255

256   However, the ML trees from the combined-nuclear dataset (nuclear_123) showed a contrasting
257   topology (Figure 1B) to that of the concatenated dataset, irrespective of the partitioning scheme:
258   Pyrginae was monophyletic and Euschemoninae was sister to rest of Hesperiidae except
259   Coeliadinae. The topology remained unchanged with the addition of morphological characters to
260   the dataset (nuclear_123+morph) as well as analyses of the subsets of the combined-nuclear
261   matrix (7gene_123, 8gene_123).

262

263   These dataset specific variation in tree topologies were consistent across different evolutionary
264   model and were also found in partitioning scheme from PartitionFinder (detail in Figure S2).

265

266   ML analyses of non-degenerated datasets (nt_12, degen_1, noLRall1+nt2) resulted in unresolved
267   tree topologies indicating insufficient phylogenetic signal in the 1st and 2nd codon positions of the
268   dataset. (Figure 1C).

269

270   BI analysis of the concatenated dataset showed a topology similar to that of Figure 1B with a few
271   changes (detail in Figure S3). Randomly sampled 100 trees from the MCMC generations, after
272   discarding burnin, showed the presence of only one topology (as in Figure 1B); however, a very
273   low proportion (6%) of an alternate topology (as in Figure 1A) was found in one of the runs.

274

275

276   *Tree Comparison*:

277   To test whether multiple tree topologies across ML analyses are equally likely given the datasets,
278   we performed an approximate unbiased (AU) test (Shimodaira, 2002) for both the concatenated
279   and combined-nuclear datasets independently. The AU test, without any partitioning scheme,
280   rejected (p<0.0001) the tree topologies from the non-degenerate datasets (nt_12, degen,
281   noLRall1+nt_2). Hence, we dropped the trees from non-degenerate data sets from further
282   analyses of tree topological similarity at higher taxonomic levels. However, two distinct
283   topologies (Figure 1A and 1B) were accepted as significant trees (P>0.05) for the combined-
284   nuclear dataset, whereas only one topology (Figure 1A) was significant (P>0.05) for the

285   concatenated dataset.

286

287   A visual comparison of the two distinct tree topologies (Figure 1A and 1B) showed that all the
288   subfamily clades (sensu Warren, Ogawa & Brower, 2009) except Pyrginae were monophyletic
289   (BS >98). Pyrginae was recovered either as monophyletic (BS=7-60) or paraphyletic. Similarly,
290   Euschemoninae and Eudaminae were either sisters (BS= 58-76) or non-sisters. However, due to
291   the presence of low BS values (<76) in 6 out of 13 deep nodes across the analyses, we were
292   uncertain whether there existed significant conflict among different tree topologies.

293

294   *Mito-genomic analysis:*

295   The ML tree from all 13 protein coding genes from complete mitochondrial genomes of 15
296   skipper butterflies had a tree-wide average BS of 80 (see Figure S4). Along with many well-
297   supported nodes (BS>97), low support values (BS 30 – 60) were also obtained for 5 out of 14
298   internal nodes. The lowest BS (=32) was obtained for the node containing Eudaminae, the
299   lineages of Pyrginae and the common ancestor of Heteropterinae and Hesperiinae. The node that
300   placed Eudaminae as sister to one of the Pyrginae clades had BS=42.

301

302   We found that the tree topology from the mitogenomic analysis corroborated the deep splits
303   found in the COI gene tree; Eudaminae nested within Pyrginae rendering the latter paraphyletic,
304   which was also the case for the concatenated dataset.

305

306   *Multiple plausible tree topologies*:

307   The presence of multiple tree topologies was not limited to the dataset-specific analyses.
308   Regardless of the partitioning scheme or the dataset (concatenated and combined-nuclear dataset)
309   used, almost equal numbers of contrasting tree topologies were present among the ML-bootstrap
310   trees across multiple analyses (see Figure S5); hence, it is likely that the topology of the best ML
311   tree is one among two equally likely topologies.

312

313   To investigate this further, we extracted 105 model-optimized trees from each of the ML
314   analyses of the concatenated (nt_123) and combined-nuclear (nuclear_123) data, separately from
315   datasets with both partitioning schemes (gene-partitions and TIGER-partitions). The clades in the
316   resulting trees were collapsed to the subfamily level except Pyrginae, which was collapsed to the
317   level of tribes (sensu Warren, Ogawa & Brower, 2009), and were plotted on a Lento plot (Figure

318   2) to check for support and conflict for each split. We observed conflicting splits among the trees
319   from the combined-nuclear dataset with TIGER-partitions, which indicated the presence of
320   multiple topologies; however, trees from the concatenated dataset had no conflicting splits,
321   indicating a single topology. But the case was different when gene-partitions were used –
322   multiple topologies resulted from concatenated dataset and only one of these topologies was
323   recovered from the combined-nuclear dataset. The AU test showed equal tree likelihoods (p >
324   0.05) for multiple topologies for both concatenated and combined-nuclear datasets given the
325   respective partition schemes as explained above.

326

327   Thus, we found two contrasting tree topologies, respectively supporting (i) monophyly of
328   Pyrginae and non-sister status of Eudaminae and Euschemoninae, and (ii) paraphyly of Pyrginae
329   and sister status of Eudaminae and Euschemoninae. In addition, we noticed that the position of
330   Eudaminae with respect to Pyrginae varied across these contrasting topologies. While
331   Eudaminae was sister to all Hesperiidae except Coeliadinae and Euschemoninae in the former,
332   the latter topology indicated that Eudaminae was sister to the clade containing Heteropterinae,
333   Trapezitinae and Hesperiinae.

334

335   We observed that the contrasting topologies had short branch lengths at the fluctuating clades:
336   the interrelationship among Eudaminae, the major clades of Pyrginae and the common ancestor
337   of Heteropterinae, Trapezitinae and Hesperiinae. The association of very low BS (< 60) with
338   these short branches suggests possible topological conflict among gene trees. To investigate this
339   hypothesis, we examined the relationships among the conflicting clades in the individual gene
340   trees. Because the single gene trees had very low BS values for most of the clades, we clustered
341   gene markers based on their topological congruency for the deeper clades and reanalyzed,
342   expecting an improvement of nodal support and consistency in signal. For example, we
343   combined all those gene markers from which Pyrginae was recovered as monophyletic and
344   expected an improved BS for the Pyrginae clade in the gene-cluster analysis. As expected, this
345   gene-cluster recovered Pyrginae as monophyletic with higher node support (BS=80). However,
346   none of the gene-clusters recovered the non-sister status of Eudaminae and Euschemoninae even
347   though their sister status was poorly supported in all the gene-cluster analyses (Figure S6). This
348   pattern of incongruence is not an artifact of missing data in our dataset, because sequential
349   removal of taxa with 80-40% of missing data from the analyses changed neither the topology nor
350   the support values at deeper nodes (Table S2); however, such sequential removal gradually
351   reduced the proportion of alternate tree topologies in the tree set from best ML tree search
352   (Figure S7). Hence, we could not accept the hypothesis that the two contrasting tree topologies
353   are due to incongruence in gene histories; this is also evident from PBS analysis where no pattern
354   of incongruence among gene trees were observed (Figure S8).

355

Discussion:

357

358 With a dataset of 7726 bp from 270 hesperiid genera, we present the most comprehensive
359 phylogeny of this important group of butterflies. Our analyses suggested that there are two
360 contrasting topologies for the higher-level skipper phylogeny. First, we reconstructed the
361 phylogenetic trees using the concatenated and combined-nuclear datasets; the resulting trees
362 were well-supported for higher-level relationships except at certain deep nodes. Tree
363 comparisons revealed that there are multiple tree topologies for the relationships among major
364 skipper lineages. We explicitly investigated gene-specific signals for the relationships among
365 major clades, clustered them based on their topological congruence and reanalyzed to check for
366 consistency.

367

Conflicting topologies?

369 Our analyses indicate the occurrence of two equally likely deep tree topologies (Figure 1A and
370 1B). Interestingly, the proximate reasons for the occurrence of these contrasting topologies
371 appear to vary depending on the partitioning scheme used for the analysis. However, neither
372 topology was strongly supported in any analysis and the results from our explorations of
373 incongruence among gene histories were not conclusive. Gene-cluster analyses improved the
374 nodal support for the monophyly of Pyrginae but were unable to recover the sister status of
375 Eudaminae and Euschemoninae with good support. Similarly, from the PBS analysis, no pattern
376 of incongruence in gene histories was observed. We conclude that there is insufficient
377 information in the molecular dataset to resolve these relationships despite the extensive
378 taxonomic sampling and large number of molecular characters.

379

380 The presence of conflicting topologies has also been reported from many other studies across
381 plants (Soltis et al., 2002; Burleigh and Mathews 2004; Ruhfel et al., 2014) and animals (Rokas
382 et al., 2003; Song et al., 2012). The possible reasons for topological incongruency are
383 phylogenetic noise or conflict among gene trees (Smith et al., 2015). In case of the former, a
384 concatenation approach is expected to give a better result (Rokas et al., 2003; Smith et al., 2015).
385 In the latter case, where the conflict is presumed to be a result of gene flow across taxa or
386 incomplete lineage sorting, coalescence based methods have been used for tree reconstruction
387 (Jarvis et al., 2014; Xi et al., 2014; Smith et al., 2015). However, all there was very low node
388 supports across gene trees, indicating that strong conflict across genes does not explain the
389 patterns found here. We predict that a phylogenomic approach would provide a better outlook to

390 this conflicting scenario or resolve the phylogeny, as such an approach has proved instrumental
391 in other studies (Dunn et al., 2008; Kocot et al., 2011; Smith et al., 2011; Johnson et al., 2013;
392 Jarvis et al., 2014; Richart et al., 2016).

393

394 *Systematic implications:*

395 Our study confirmed that all the subfamilies, possibly except Pyrginae, are monophyletic and
396 received high BS support across multiple analyses. We are uncertain about the monophyly of
397 Pyrginae, as our study reveals homoplastic character distributions that could potentially be
398 explained by the occurrence of ancestral introgression among its early lineages. Hence, when a
399 certain combination of genes was used for phylogenetic construction, Pyrginae was recovered
400 monophyletic (e.g., Figure 1B), as in the previous study (Warren, Ogawa & Brower, 2009). This
401 result appears to be supported by morphology. Similarly, we remain uncertain about the true
402 relationships among Eudaminae, Pyrginae and the clade containing Heteropterinae, Trapezitinae
403 and Hesperiinae, due to short branches that may be explained by their rapid divergence from
404 each other and possibly an introgression between Eudaminae and Euschemoninae. However, the
405 arrangement of Euschemoninae as sister to all Hesperiidae (except Coeliadinae), and Eudaminae
406 as sister to Pyrginae, is generally supported by morphology. We suggest that the relationships
407 shown in Figure 1B (also in Figure 3B) should be used as the preferred phylogenetic hypothesis
408 until a better-resolved phylogeny is available.

409

410 In addition, we observed unexpected placement of a few taxa within Pyrginae. For instance,
411 *Eracon*, which was previously classified under Pyrgini (Warren, Ogawa & Brower, 2009), was
412 found herein to group with Achlyodini. Likewise, *Clito* grouped within either Pyrgini or
413 Erynnini based on dataset specific analyses. Moreover, we note that both *Clito* and *Eracon*
414 sequences in our dataset have >70% missing sites. Hence, it is likely that the presence of
415 insufficient informative sites within these taxa might influence their true positions in the
416 phylogeny (Wiens 2003; Wiens and Morrill, 2011). Therefore, for systematic implications, we
417 pruned *Clito*, *Eracon* and three additional taxa with a large percentage of missing data from the
418 dataset and reanalysed with gene partitions (Figure 3). We observed no change in tree topology
419 or node support values as a result of pruning these taxa. Hence, although they may appear on the
420 tree in unorthodox positions, it is unlikely that presence of these taxa has any impact on our
421 interpretation of higher-level relationships in the dataset as a whole.

422

423 We observed that the genus *Cabirus*, previously included within Eudaminae, grouped within
424 Achlyodini (subfamily Pyrginae). Further study of the morphology of *Cabirus* is needed to
425 corroborate this placement, although its position outside of Eudaminae seems to be correct.

426 Three tribes within Hesperiinae – Aeromachini, Taractrocerini and Baorini – are monophyletic
427 with high BS values. However, we are uncertain about the phylogenetic status of other proposed
428 tribes within Hesperiinae due to prevalence of low BS values along the short internal branches.
429 This indicates the possible occurrences of rapid ancestral radiation within Hesperiinae and needs
430 further investigation.

431

432 **Conclusions:**

433

434 With a broad coverage of all known subfamilies, we present the higher level relationships among
435 skipper butterflies. Our analyses suggest possible conflicting topologies with respect to (i)
436 monophyly or paraphyly of Pyrginae and (ii) sister or non-sister status of Eudaminae and
437 Euschemoninae. However none of the topologies resulting from our alternative analyses is
438 strongly supported, and incongruences in signal among genes cannot satisfactorily resolve these
439 differences. We surmise that there is insufficient phylogenetic information in the current dataset
440 to resolve these relationships. It is unlikely that adding data from a few more genes will improve
441 the results, but data from entire genomes may result in a better-resolved phylogeny. However,
442 taking morphological characters into consideration, we suggest one of the topologies as most
443 likely (Figs. 1B & 3B), and that this topology will aid in future studies on this group.

444

445

446 **References:**

447 Ackery PR. 1984. *The Biology of Butterflies*. (Eds. Vane-Wright, R. I. & Ackery, P. R.) 9–21
448   Academic Press.

449 Baker RH, DeSalle R. 1997. Multiple sources of character information and the phylogeny of
450   Hawaiian drosophilids. *Systematic Biology* 46(4):654-673. DOI: 10.1093/sysbio/46.4.654.

451 Boggs CL, Watt WB, Ehrlich PR (Eds.) 2003. *Butterflies: ecology and evolution taking flight.*
452   Chicago: University of Chicago Press. 739 pp.

453 Braby MF, Vila R, Pierce NE. 2006. Molecular phylogeny and systematics of the Pieridae
454   (Lepidoptera: Papilionoidea): higher classification and biogeography. *Zoological Journal*
455   *of the Linnean Society.* 147(2):239-275. DOI: 10.1111/j.1096-3642.2006.00218.x.

456 Bremer K. 1994. Branch support and tree Stability. *Cladistics* 10(3):295–304. DOI:
457   10.1111/j.1096-0031.1994.tb00179.x.

458 Burleigh JG, Mathews S. 2004. Phylogenetic signal in nucleotide data from seed plants:
459   implications for resolving the seed plat tree of life. *Am. J. Bot*. 91(10): 1599-1613. DOI:
460   10.3732/ajb.91.10.1599.

461 Campbell DL, Brower AVZ, Pierce NE. 2000. Molecular evolution of the *wingless* gene and its
462   implications for the phylogenetic placement of the butterfly family Riodinidae
463   (Lepidoptera: Papilionoidea). *Molecular Biology and Evolution* 17(5):684-696.

464 Caterino MS, Reed RD, Kuo MM, Sperling FA. 2001. A partitioned likelihood analysis of
465   swallowtail butterfly phylogeny (Lepidoptera: Papilionidae). *Systematic Biology*
466   50(1):106–127. DOI: 10.1080/10635150119988.

467 Chou, I. 1994. *Monographia Rhopalocerorum Sinensium*. Henan Scientific and Technological
468   Publishing House.

469 Chou, I. 1999. *Classification and identification of Chinese butterflies*. Henan Scientific and
470   Technological Publishing House.

471 Cummins CA, McInerney JO. 2011. A method for inferring the rate of evolution of homologous
472   characters that can potentially improve phylogenetic inference, resolve deep divergence
473   and correct systematic biases. *Systematic Biology* 60(6):833-844. DOI:
474   10.1093/sysbio/syr064.

475 Dunn CW, Hejnol A, Matus DQ, Pang K, Browne WE, Smith SA, Seaver E, Rouse GW, Obst M,
476   Edgecombe GD, Sorensen MV, Haddock SHD, Schmidt-Rhaesa A, Okusu A, Kristensen
477   RM, Wheeler WC, Martindale MQ, Giribet G. 2008. Broad phylogenomic sampling
478   improves resolution of the animal tree of life. *Nature* 452(7188):745–9.
479   doi:10.1038/nature06614.

480 Eckert AJ, Carstens BC. 2008. Does gene flow destroy phylogenetic signal? The performance of
481   three methods for estimating species phylogenies in the presence of gene flow. *Molecular*
482   *Phylogenetics and Evolution* 49(3): 832-842. DOI: 10.1016/j.ympev.2008.09.008.

483 Espeland M, Hall JP, DeVries PJ, Lees DC, Cornwall M, Hsu YF, Wu LW, Campbell DL,
484   Talavera G, Vila R, Salzman S, Ruehr S, Lohman DJ, Pierce N. 2015. Ancient
485   Neotropical origin and recent recolonisation: Phylogeny, biogeography and
486   diversification of the Riodinidae (Lepidoptera: Papilionoidea). *Molecular Phylogenetics*
487   *and Evolution* 93:296–306. DOI: 10.1016/j.ympev.2015.08.006.

488 Evans WH. 1949. A catalogue of the Hesperiidae from Europe, Asia, and Australia in the British
489   Museum (Natural History). British Museum, London.

490 Felsenstein J. 1985. Phylogenies and the comparative method. *The American Naturalist*
491   125(1):1-15.

492   Gelman A, Rubin DB. 1992. Inference from iterative simulation using multiple
493   sequences. Statistical Science 7(4):457–472.

494   Goloboff PA. 1999. Analyzing large data sets in reasonable times: solutions for composite
495   optima. *Cladistics* 15(4):415–428. DOI: 10.1111/j.1096-0031.1999.tb00278.x.

496   Goloboff PA, Farris JS, Nixon KC. 2008. TNT, a free program for phylogenetic analysis.
497   *Cladistics* 24(5):774-786. DOI: 10.1111/j.1096-0031.2008.00217.x.

498   Heikkila M, Kaila L, Mutanen M, Pena C, Wahlberg N. 2012. Cretaceous origin and repeated
499   tertiary diversification of the redefined butterflies. *Proceedings of the Royal Society B:*
500   *Biological Sciences* 279(1731):1093–1099. DOI: 10.1098/rspb.2011.1430.

501   Hernández-Roldán JL, Bofill R, Dapporto L, Munguira ML, Vila R. 2014. Morphological and
502   chemical analysis of male scent organs in the butterfly genus Pyrgus (Lepidoptera:
503   Hesperiidae). *Organisms Diversity & Evolution* 14(3):269-278. DOI: 10.1007/s13127-
504   014-0170-x.

505   Huelsenbeck JP, Larget B, Alfaro ME. 2004. Bayesian phylogenetic model selection using
506   reversible jump markov chain monte carlo. Mol. Biol. Evol. 21(6):1123-1133.

507   Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, ... & Suh A. 2014. Whole-genome
508   analyses resolve early branches in the tree of life of modern birds. *Science*
509   346(6215):1320-31. DOI: 10.1126/science.1253451.

510   Johnson Johnson BR, Borowiec ML, Chiu JC, Lee EK, Atallah J, Ward PS. 2013.
511   Phylogenomics resolves evolutionary relationships among ants, bees, and wasps. *Curr*
512   *Biol*. 23:2058–062. doi:10.1016/j.cub.2013.08.050.

513   Kim MJ, Wang AR, Park JS, Kim I. 2014. Complete mitochondrial genomes of five skippers
514   (Lepidoptera: Hesperiidae) and phylogenetic reconstruction of Lepidoptera. *Gene*
515   549(1):97-112. DOI: 10.1016/j.gene.2014.07.052.

516   Kocot KM, Cannon JT, Todt C, Citarella MR, Kohn AB, Meyer A, Santos SR, Schander C,
517   Moroz LL, Lieb B, Halanych KM. 2011. Phylogenomics reveals deep molluscan
518   relationships. *Nature* 477:452–6. doi:10.1038/nature10382.

519   Kodandaramaiah U, Lees DC, Müller CJ, Torres E, Karanth KP, Wahlberg N. 2010.
520   Phylogenetics and biogeography of a spectacular Old World radiation of butterflies: the
521   subtribe Mycalesina (Lepidoptera: Nymphalidae: Satyrini). *BMC Evolutionary Biology*
522   10(1):172. DOI: 10.1186/1471-2148-10-172.

523   Lanfear R, Calcott B, Ho SYW, Guindon S. 2012. PartitionFinder: Combined Selection
524   of Partitioninggsdhjh Schemes and Substitution Models for Phylogenetic
525   Analyses. *Molecular Biology and Evolution* 29:1695–1701.

526 Lento GM, Hickson RE, Chambers GK, Penny D. 1995. Use of spectral analysis to test
527   hypotheses on the origin of pinnipeds. *Molecular biology and evolution* 12(1):28–52.
528   DOI: 10.1093/oxfordjournals.molbev.a040189.

529 Mielke OHH. 2005. Catalogue of the American Hesperioidea: Hesperiidae (Lepidoptera).
530   Sociedade Brasileira de Zoologia, Parana, Brazil. 6 vols.

531 Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference
532   of large phylogenetic trees. In: *Proceedings of the Gateway Computing Environments*
533   *Workshop* (GCE), 14 Nov. 2010, New Orleans, LA: 1 - 8.

534 Nabholz B, Kunstner A, Wang R, Jarvis ED, Ellegren H. 2011. Dynamic evolution of base
535   composition: causes and consequences in avian phylogenomics. *Molecular Biology and*
536   *Evolution* 28(8):2197–2210. DOI: 10.1093/molbev/msr047.

537 Nazari V, Zakharov EV, Sperling FA. 2007. Phylogeny, historical biogeography and taxonomic
538   ranking of Parnassiinae (Lepidoptera, Papilionidae) based on morphology and seven
539   genes. *Molecular Phylogenetics and Evolution* 42(1):131–156. DOI:
540   10.1016/j.ympev.2006.06.022.

541 Nixon KC. 1999. The parsimony ratchet, a new method for rapid parsimony analysis. *Cladistics*
542   15(4):407-414. DOI: 10.1111/j.1096-0031.1999.tb00277.x.

543 Planet PJ. 2006. Tree disagreement: measuring and testing incongruence in phylogenies. *Journal*
544   *of Biomedical Informatics* 39(1):86–102. DOI: 10.1016/j.jbi.2005.08.008.

545 Pollard DA, Iyer VN, Moses AM, Eisen MB. 2006. Widespread discordance of gene trees with
546   species tree in Drosophila: evidence for incomplete lineage sorting. *PLoS Genet*
547   2(10):e173. DOI: 10.1371/journal.pgen.0020173.

548 Rambaut A, Suchard MA, Xie D, Drummond AJ. 2014. Tracer v1.6. Available at
549   *http://beast.bio.ed.ac.um/Tracer*.

550 Regier JC, Shultz JW, Ganley AR, Hussey A, Shi D, Ball B, Zwick A, Stajich JE, Cummings
551   MP, Martin JW, Cunningham CW. 2008. Resolving arthropod phylogeny: exploring
552   phylogenetic signal within 41 kb of protein-coding nuclear gene sequence. *Systematic*
553   *Biology* 57(6):920–938. DOI: 10.1080/10635150802570791.

554 Regier JC, Shultz JW, Zwick A, Hussey A, Ball B, Wetzer R, Martin JW, Cunningham CW.
555   2010. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-
556   coding sequences. *Nature* 463(7284):1079–1083. DOI: 10.1038/nature08742.

557 Richart CH, Hayashi CY, Hedin M. 2016. Phylogenomic analyses resolve an ancient trichotomy
558   at the base of Ischyropsalidoidea (Arachnida, Opiliones) despite high levels of gene tree

559   conflict and unequal minority resolution frequencies. Mol. Phyl. Evol. 95:171-182. DOI:
560   10.1016/j.ympev.2015.11.010.

561   Rokas A, King N, Finnerty J, Carroll SB. 2003. Conflicting phylogenetic signals at the base of
562   the metazoan tree. *Evolution and Development* 5(4):346-359. DOI: 10.1046/j.1525142-
563   X.2003.03042.x.

564   Rokas A, Williams BL, King N, Carroll SB. 2003. Genome-scale approaches to resolving
565   incongruence in molecular phylogenies. *Nature* 425(6960):798-804. DOI:
566   10.1038/nature02053.

567   Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L,
568   Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic
569   inference and model choice across a large model space. *Systematic biology* 61(3):539-42.
570   DOI: 10.1093/sysbio/sys029.

571   Rota, J. & Wahlberg, N. 2012. Exploration of data partitioning in an eight-gene dataset:
572   phylogeny of metalmark moths (Lepidoptera, Choreutidae). Zoologica Scripta 41:
573   536-546. doi:10.1111/j.1463-6409.2012.00551.x.

574   Ruhfel BR, Gitzendanner MA, Soltis PS, Soltis DE, Burleigh JG. 2014. From algae to
575   angiosperms - inferring the phylogeny of green plants (*Viridiplantae*) from 360 plastid
576   genomes. *BMC Evol. Biol*. 14:23. DOI: 10.1186/1471-2148-14-23.

577   Scott JA. 1985. The phylogeny of butterflies (Papilionoidea and Hesperioidea). *J. Res.*
578   *Lepidoptera* 23(4):241–281.

579   Scott JA & Wright DM. 1990. *Butterflies of Europe* (eds Kudrna, O.). Aula-Verlag.

580   Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. *Systematic*
581   *Biology* 51(3):492-508. DOI: 10.1080/10635150290069913.

582   Simonsen TJ, Zakharov EV, Djernaes M, Cotton AM, Vane-Wright RI, Sperling FA. 2011.
583   Phylogenetics and divergence times of Papilioninae (Lepidoptera) with special reference
584   to the enigmatic genera Teinopalpus and Meandrusa. *Cladistics* 27(2):113–137. DOI:
585   10.1111/j.1096-0031.2010.00326.x.

586   Smith SA, Wilson NG, Goetz FE, Feehery C, Andrade SCS, Rouse GW, Giribet G, Dunn CW.
587   2011. Resolving the evolutionary relationships of molluscs with phylogenomic tools.
588   *Nature* 480:364–7. doi:10.1038/nature10526.

589   Smith SA, Moore MJ, Brown JW and Yang Y. 2015. Analysis of phylogenomic datasets reveals
590   conflict, concordance, and gene duplications with examples from animals and plants.
591   *BMC Evol. Biol.* 15:150. DOI: 10.1186/s12862-015-0423-0.

592   Soltis DE, Soltis PS, Zanis MJ. 2002. Phylogeny of seed plants based on evidence from eight
593   genes. *Am. J. Bot.* 89(10): 1670-1681. DOI: 10.3732/ajb.89.10.1670.

594   Song S, Liu L, Edwards SV, Wu S. 2012. Resolving conflict in eutherian mammal phylogeny
595   using phylogenomics and the multispecies coalescent model. *Proc. Nat. Acad. Sci.*
596   109(37):14942-14947.

597   Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
598   large phylogenies. *Bioinformatics* 30:1312–1313. DOI: 10.1093/bioinformatics/btu033.

599   Voss EG. 1952. On the classification of the Hesperiidae. *Annals of the Entomological Society of*
600   *America* 45(2):246–258. DOI: 10.1093/aesa/45.2.246.

601   Wahlberg N, Leneveu J, Kodandaramaiah U, Pena C, Nylin S, Freitas AVL, Brower AVZ. 2009.
602   Nymphalid butterflies diversify following near demise at the Cretaceous/Tertiary
603   boundary. *Proceedings of the Royal Society B: Biological Sciences* 276:4295-4320. DOI:
604   10.1098/rspb.2009.1303.

605   Wahlberg N, Rota J, Braby MF, Pierce NE, Wheat CW. 2014. Revised systematics and higher
606   classification of pierid butterflies (Lepidoptera: Pieridae) based on molecular data.
607   *Zoologica Scripta* 43(6):641–650. DOI: 10.1111/zsc.12075.

608   Wahlberg N, Wheat CW. 2008. Genomic outposts serve the phylogenomic pioneers: designing
609   novel nuclear markers for genomic DNA extractions of Lepidoptera. *Systematic Biology*
610   57(2):231-242. DOI: 10.1080/10635150802033006.

611   Warren AD, Ogawa JR, Brower AVZ. 2008. Phylogenetic relationships of subfamilies and
612   circumscription of tribes in the family Hesperiidae (Lepidoptera: Hesperioidea).
613   *Cladistics* 24(5):642-676. DOI: 10.1111/j.1096-0031.2008.00218.x.

614   Warren AD, Ogawa JR, Brower AVZ. 2009. Revised classification of the family Hesperiidae
615   (Lepidoptera: Hesperioidea) based on combined molecular and morphological data.
616   *Systematic Entomology* 34(3):467–523. DOI: 10.1111/j.1365-3113.2008.00463.x.

617   Whitfield JB, Lockhart PJ. 2007. Deciphering ancient rapid radiations. *Trends in Ecology &*
618   *Evolution* 22(5):258–265. DOI: 10.1016/j.tree.2007.01.012.

619   Wiens, JJ. 2003. Missing data, incomplete taxa, and phylogenetic accuracy. *Systematic*
620   *Biology* 52(4):528-538. DOI: 10.1080/10635150390218330.

621   Wiens, JJ and Morrill MC. 2011. Missing data in phylogenetic analysis: reconciling results from
622   simulations and empirical data. *Systematic Biology* 60(5):719-731.
623   DOI: 10.1093/sysbio/syr025.

624 Wolf YI, Rogozin IB, Grishin NV, Koonin EV. 2002. Genome trees and the tree of life. *Trends*
625   *in Genetics* 18(9):472-479. DOI: 10.1016/S0168-9525(02)02744-0.

626 Xi Z, Liu L, Rest JS, Davis CC. 2014. Coalescent versus concatenation methods and the
627   placement of amborella as sister to water lilies. *Systematic Biology* 63(6): 919-932. DOI:
628   10.1093/sysbio/syu055.

629 Xia X. 2013. DAMBE5: A comprehensive software package for data analysis in molecular
630   biology and evolution. *Mol Biol Evol* 30(7): 1720-1728. DOI: 10.1093/molbev/mst064.

631 Yuan X, Gao K, Yuan F, Wang P, Zhang Y. 2015. Phylogenetic relationships of subfamilies in
632   the family Hesperiidae (Lepidoptera: Hesperioidea) from China. *Scientific Reports*
633   5:11140. DOI: 10.1038/srep11140.

634 Zwick A, Reiger JC, Zwickl DJ. 2012. Resolving discrepancy between nucleotides and amino
635   acids in deep-level arthropod phylogenomics: differentiating serine codons in 21-amino
636   acid models. *PLoS ONE* 7(11):e47450. DOI: 10.1371/journal.pone.0047450.

637

638

639

Figure 1

Comparison of best ML trees across analyses.

The best ML trees from the concatenated dataset (A), combined-nuclear dataset (B) and degenerated dataset (degen_1) (C). The values at nodes in (A) and (B) represent support from 1000 bootstrap trees analyzed with TIGER-partitions / gene-partitions. In (C), the values at the nodes are the output from the analyses using TIGER-partitions only and the nodes without any value have bootstrap support < 20. The other degenerated datasets (nt_12, noLRall1+nt2) have similar or more fluctuating rearrangements for the nodes with BS<20.

Note: Auto Gamma Correction was used for the image. This only affects the reviewing manuscript. See original source image if needed for review.

Figure 2

Comparison of the supports/conflicts on the Lento plots.

The Lento plots were drawn from 105 ML trees recovered during best ML tree search across different datasets using two different partitioning schemes. The X-axis represents each non-trivial clade, with filled circles indicating the clade composition; the Y-axis shows relative support (values above zero) or conflict (values below zero). Splits are colour coded to aid comparison across analyses. The splits which are not coloured are found only in that particular analysis and not recovered from others.
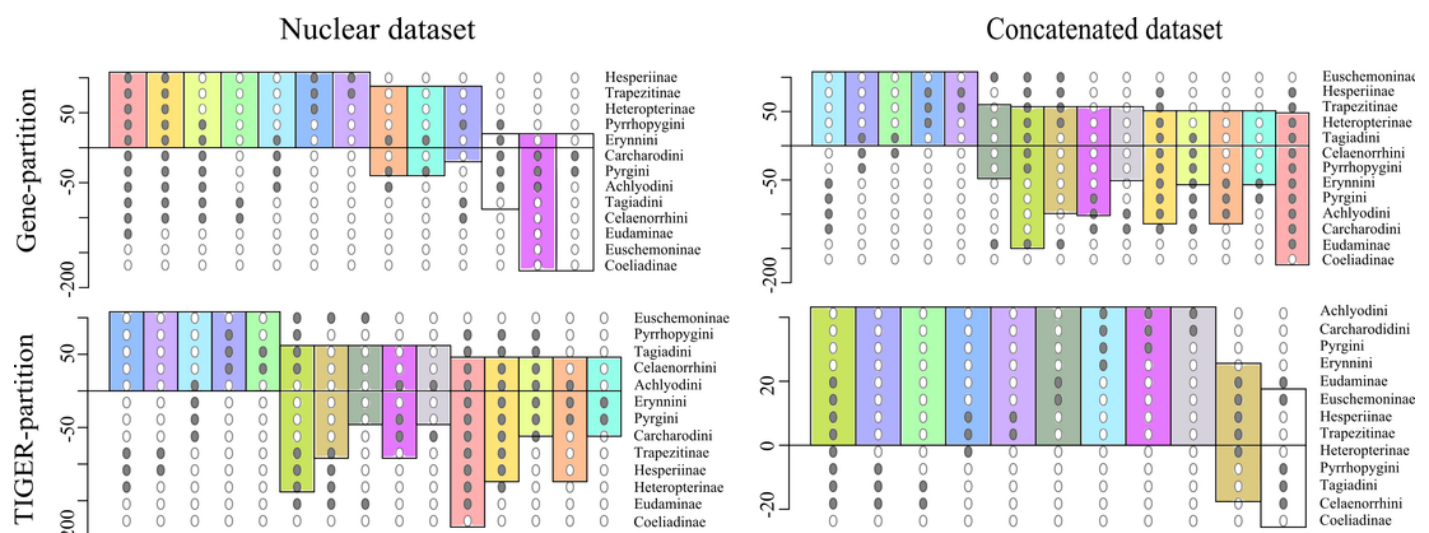
Figure 3

The ML trees from a reduced dataset.

The ML trees from the analyses of (A) the concatenated dataset with gene partitions, and (B) the combined nuclear dataset with gene partitions. *Clito*, *Eracon* and three additional taxa were removed prior to the analyses (see text for detail). The size of the circle at the node corresponds to the bootstrap support that was derived from 1000 pseudo-replicates. All taxa are colour coded based on their subfamily status, except the taxa within subfamily Pyrginae which are coloured based on their tribe. Silhouette © PhyloPic.