

1

2

3

DrosoPhyla: resources for drosophilid phylogeny and systematics

4

5

6 Cédric Finet^{1*}, Victoria A. Kassner¹, Antonio B. Carvalho², Henry Chung³, Jonathan
7 P. Day⁴, Stephanie Day⁵, Emily K. Delaney⁶, Francine C. De Ré⁷, Héloïse D.
8 Dufour¹, Eduardo Dupim², Hiroyuki F. Izumitani⁸, Thaísa B. Gautério⁹, Jessa Justen¹,
9 Toru Katoh⁸, Artyom Kopp⁶, Shigeyuki Koshikawa^{10,11}, Ben Longdon¹², Elgion L.
10 Loreto⁷, Maria D. S. Nunes^{13,14}, Komal K. B. Raja^{15,16}, Mark Rebeiz⁵, Michael G.
11 Ritchie¹⁷, Gayane Saakyan⁶, Tanya Sneddon¹⁷, Machiko Teramoto^{10,18}, Venera
12 Tyukmaeva¹⁷, Thyago Vanderlinde², Emily E. Wey¹⁹, Thomas Werner¹⁵, Thomas M.
13 Williams¹⁹, Lizandra J. Robe^{7,9}, Masanori J. Toda²⁰, Ferdinand Marlétaz²¹

14

15

Author affiliations

16

17 ¹ Howard Hughes Medical Institute and Laboratory of Molecular Biology, University
18 of Wisconsin, Madison, USA

19 ² Departamento de Genética, Instituto de Biologia, Universidade Federal do Rio de
20 Janeiro, Rio de Janeiro, Brazil

21 ³ Department of Entomology, Michigan State University, East Lansing, USA

22 ⁴ Department of Genetics, University of Cambridge, Cambridge, United Kingdom

23 ⁵ Department of Biological Sciences, University of Pittsburgh, Pittsburgh,
24 Pennsylvania, USA

25 ⁶ Department of Evolution and Ecology, University of California-Davis, Davis, USA

26 ⁷ Programa de Pós-Graduação em Biodiversidade Animal, Universidade Federal de
27 Santa Maria, Rio Grande do Sul, Brazil

28 ⁸ Department of Biological Sciences, Faculty of Science, Hokkaido University,
29 Sapporo, Japan

30 ⁹ Programa de Pós-Graduação em Biologia de Ambientes Aquáticos Continentais,
31 Universidade Federal do Rio Grande, Rio Grande do Sul, Brazil

32 ¹⁰ The Hakubi Center for Advanced Research and Graduate School of Science, Kyoto
33 University, Kyoto, Japan

34 ¹¹ Present address: Faculty of Environmental Earth Science, Hokkaido University,
35 Sapporo, Japan

36 ¹² Present address: Centre for Ecology & Conservation, College of Life and
37 Environmental Sciences, University of Exeter, United Kingdom

38 ¹³ Department of Biological and Medical Sciences, Oxford Brookes University,
39 Oxford, United Kingdom

40 ¹⁴ Centre for Functional Genomics, Oxford Brookes University, Oxford, United
41 Kingdom

42 ¹⁵ Department of Biological Sciences, Michigan Technological University, Houghton,
43 Michigan, USA

44 ¹⁶ Present address: Department of Pathology and Immunology, Baylor College of
45 Medicine, Houston, Texas, USA

46 ¹⁷ School of Biology, University of St Andrews, St Andrews, Scotland, United
47 Kingdom

48 ¹⁸ Present address: National Institute for Basic Biology, Okazaki, Japan

49 ¹⁹ Department of Biology, University of Dayton, Dayton, Ohio, USA

50 ²⁰ Hokkaido University Museum, Hokkaido University, Sapporo, Japan

51 ²¹ Centre for Life's Origins and Evolution, Department of Genetics, Evolution and
52 Environment, University College London, London, United Kingdom

53

54

55 *Corresponding author:

56 Cédric Finet : cedric.finet@ens-lyon.org

57

58

59 **Abstract**

60 The vinegar fly *Drosophila melanogaster* is a pivotal model for invertebrate
61 development, genetics, physiology, neuroscience, and disease. The whole family
62 Drosophilidae, which contains over 4,400 species, offers a plethora of cases for
63 comparative and evolutionary studies. Despite a long history of phylogenetic
64 inference, many relationships remain unresolved among the genera, subgenera and
65 species groups in the Drosophilidae. To clarify these relationships, we first developed
66 a set of new genomic markers and assembled a multilocus dataset of 17 genes from
67 704 species of Drosophilidae. We then inferred a species tree with highly supported
68 groups for this family. Additionally, we were able to determine the phylogenetic
69 position of some previously unplaced species. These results establish a new
70 framework for investigating the evolution of traits in fruit flies, as well as valuable
71 resources for systematics.

72

73 **Key words**

74 Drosophilidae; Phylogenomics; Systematics

75

76 **Significance statement**

77 Comparative studies require a robust phylogenetic framework for investigating trait
78 diversity. The family Drosophilidae comprises more than 4,400 species including the
79 model organism *Drosophila melanogaster*. Work on numerous *Drosophila* species is
80 providing ways to understand evolutionary mechanisms. Yet, the relationships among
81 major lineages in the Drosophilidae remain unresolved. To clarify these relationships,
82 we first developed a set of new genomic markers and assembled a multilocus dataset
83 of 17 genes from 704 species of Drosophilidae. We then inferred species and
84 composite group trees with high support for this family. Our study timely establishes
85 a phylogenetic framework for comparative studies and provides an easily extendable
86 dataset for further advances in Drosophilidae systematics.

87

88 **Introduction**

89 The vinegar fly *Drosophila melanogaster* is a well-established and versatile model
90 system in biology (Hales et al. 2015). The story began at the start of the 20th century

91 when the entomologist Charles Woodworth bred *D. melanogaster* in captivity, paving
92 the way to William Castle's seminal work at Harvard in 1901 (Sturtevant A. H. 1959).
93 But it is undoubtedly with Thomas Hunt Morgan and his colleagues that *D.*
94 *melanogaster* became a model organism in genetics (Morgan 1910). Nowadays, *D.*
95 *melanogaster* research encompasses diverse fields, such as biomedicine (Ugur et al.
96 2016), developmental biology (Hales et al. 2015), growth control (Wartlick et al.
97 2011), gut microbiota (Trinder et al. 2017), innate immunity (Buchon et al. 2014),
98 behaviour (Cobb 2007), and neuroscience (Bellen et al. 2010).

99

100 By the mid-20th century, evolutionary biologists have widened *Drosophila* research
101 by introducing many new species of Drosophilidae in comparative studies. For
102 example, the mechanisms responsible for morphological differences of larval denticle
103 trichomes (Sucena et al. 2003; McGregor et al. 2007), adult pigmentation (Jeong et al.
104 2008; Yassin, Delaney, et al. 2016), sex combs (Tanaka et al. 2009), and genital shape
105 (Glassford et al. 2015; Peluffo et al. 2015) have been thoroughly investigated across
106 Drosophilidae. Comparative studies brought new insights into the evolution of
107 ecological traits, such as host specialization (Lang et al. 2012; Yassin et al. 2016),
108 niche diversification (Chung et al. 2014), species distribution (Kellermann et al.
109 2009), pathogen virulence (Longdon et al. 2015), and behavior (Dai et al. 2008;
110 Karageorgi et al. 2017).

111

112 More than 150 genomes of *Drosophila* species are now sequenced (Adams et al.
113 2000; Clark et al. 2007; Wiegmann and Richards 2018; Kim et al. 2021), allowing the
114 comparative investigation of gene families (Sackton et al. 2007; Almeida et al. 2014;
115 Finet et al. 2019) as well as global comparison of genome organization (Bosco et al.
116 2007; Bhutkar et al. 2008). For all these studies, a clear understanding of the historical
117 relationships between species is necessary to interpret the results in an evolutionary
118 context. A robust phylogeny is then crucial to confidently infer ancestral states,
119 identify synapomorphic traits, and reconstruct the history of events during the
120 evolution and diversification of Drosophilidae.

121

122 Fossil-based divergence time estimation suggest that the family Drosophilidae
123 originated at least 30-50 Ma (Throckmorton 1975; Grimaldi 1987; Wiegmann et al.
124 2011). To date, the family comprises more than 4,400 species (DrosWLD-Species

125 2021) classified into two subfamilies, the Drosophilinae Rondani and the Steganinae
126 Hendel. Each of these subfamilies contains several genera, which are traditionally
127 subdivided into subgenera, and are further composed of species groups. Nevertheless,
128 the monophyletic status of each of these taxonomic units is frequently controversial or
129 unassessed. Part of this controversy is related to the frequent detection of paraphyletic
130 taxa within Drosophilidae (Throckmorton 1975; Katoh et al. 2000; Robe et al. 2005;
131 Robe et al. 2010b; Da Lage et al. 2007; Van Der Linde et al. 2010; Russo et al. 2013;
132 Yassin 2013; Katoh et al. 2017; Gautério et al. 2020), although the absence of a
133 consistent phylogenetic framework for the entire family makes it difficult to assess
134 alternative scenarios.

135

136 Despite the emergence of the *Drosophila* genus as a model system to investigate the
137 molecular genetics of functional evolution, relationships within the family
138 Drosophilidae remain poorly supported. The first modern phylogenetic trees of this
139 family relied on morphological characters (Throckmorton 1962; Throckmorton 1975;
140 Throckmorton 1982), followed by a considerable number of molecular phylogenies
141 that mainly focused on individual species groups (reviewed in (Markow and O’Grady
142 2006; O’Grady and DeSalle 2018)). For the last decade, only a few large-scale studies
143 have attempted to resolve the relationships within Drosophilidae as a whole. For
144 example, supermatrix approaches brought new insights, such as the identification of
145 the earliest branches in the subfamily Drosophilinae (Van Der Linde et al. 2010;
146 Yassin et al. 2010), the paraphyly of the subgenus *Drosophila* (*Sophophora*) (Gao et
147 al. 2011), the placement of Hawaiian clades (O’Grady et al. 2011; Lapoint et al. 2013;
148 Katoh et al. 2017), and the placement of Neotropical Drosophilidae (Robe et al.
149 2010c). Most of the aforementioned studies have suffered from limited taxon or gene
150 sampling. Recent studies improved the taxon sampling and the number of loci
151 analysed (Morales-Hojas and Vieira 2012; Russo et al. 2013; Izumitani et al. 2016).
152 To date, the most taxonomically-broad study is a revision of the Drosophilidae that
153 includes 30 genera in Steganinae and 43 in Drosophilinae, but only considering a
154 limited number of genomic markers (Yassin 2013).

155

156 To clarify the phylogenetic relationships in the Drosophilidae, we built a
157 comprehensive dataset of 704 species that include representatives from most of the
158 major genera, subgenera, and species groups in this family. We developed new

159 genomic markers and compiled available ones from previously published
160 phylogenetic studies. We then inferred well-supported trees at the group- and species-
161 level for this family. Additionally, we were able to determine the phylogenetic
162 position of several species of uncertain affinities. Our results establish a new
163 framework for investigating the systematics and diversification of fruit flies and
164 provide a valuable genomic resource for the *Drosophila* community.

165

166 **Results and Discussion**

167 **A multigene phylogeny of 704 drosophilid species**

168 We assembled a multilocus dataset of 17 genes (14,961 unambiguously aligned
169 nucleotide positions) from 704 species of Drosophilidae. Our phylogeny recovers
170 many of the clades or monophyletic groups previously described in the Drosophilidae
171 (Figure 1). While the branching of the species groups is generally well-supported, we
172 observe that some of the deepest branches of the phylogenetic tree remain poorly
173 supported or unresolved, especially in Bayesian analyses (Figures S1 and S2). This
174 observation prompted us to apply a composite taxon strategy that has been used to
175 resolve challenging phylogenetic relationships (Finet et al. 2010; Campbell and
176 Lapointe 2011; Sigurdson and Green 2011; Charbonnier et al. 2015; Mengual et al.
177 2017; Fan et al. 2020). This approach limits branch lengths in selecting slow-evolving
178 sequences, and decreases the percentage of missing data, improving phylogenetic
179 reconstruction for sparse data matrices (Campbell and Lapointe 2009). We defined 63
180 composite groups as the monophyletic groups identified in the 704-taxon analysis
181 (Figure 1, Table S1), and added these to the sequences of 20 other ungrouped taxa to
182 perform additional phylogenetic evaluations. The overall bootstrap values and
183 posterior probabilities were higher for the composite tree (Figures 2A, S3 and S4). In
184 addition, we applied the summary method ASTRAL to our composite dataset to infer
185 a species tree from a collection of input trees. However, the resulting tree is less
186 resolved than the one obtained by concatenation (Figure S5).

187

188 Incongruence among phylogenetic markers can be related to incomplete lineage
189 sorting, introgression, hybridization or other processes and can be detrimental to
190 accurate species tree reconstruction (Jeffroy et al. 2006; Kapli et al. 2020). In order to
191 estimate the presence of incongruent signal in our dataset, we first investigated the

192 qualitative effect of single marker removal on the topology of the composite tree
193 (Figure S6). We found the overall topology is very robust to marker sampling, with
194 only a few minor changes for each dataset. For instance, the *melanogaster* subgroup
195 sometimes clusters with the *eugracilis* subgroup instead of branching off prior to the
196 *eugracilis* subgroup (Figures 2 and S6). The position of the genus *Dettopsomyia* and
197 that of the *angor* and *histrion* groups is also very sensitive to single marker removal,
198 which could explain the low support values obtained (Figures 2 and S6). To a lesser
199 extent, the position of *D. fluvialis* can vary as well depending on the removed marker
200 (Figures 2 and S6). We also quantitatively investigated the incongruence present in
201 our dataset by calculating genealogical concordance. The gene concordance factor is
202 defined as the percentage of individual gene trees containing that node for every node
203 of the reference tree. Similarly, the fraction of nodes supported by each marker can be
204 determined. The markers we developed in this study show concordance rates ranging
205 from 46.2 to 90.9% (Figure 3, Table 2). With an average concordance rate of 65%,
206 these new markers appear as credible phylogenetic markers, without significantly
207 improving the previous markers (average concordance rate of 64.8%).

208

209 Multiple substitutions at the same position is another classical bias in phylogenetic
210 reconstruction, capable of obscuring the genuine phylogenetic signal (Jeffroy et al.
211 2006). We quantified the mutational saturation for each phylogenetic marker. On
212 average, the newly developed markers are moderately saturated (Figures 3 and S7,
213 Table 2). These markers are indeed less saturated than the *Amyrel*, *COI*, and *COII*
214 genes that have been commonly applied for phylogenetic inference in Drosophilidae
215 (Baker and Desalle 1997; O'Grady et al. 1998; Remsen and O'Grady 2002; Bonacum
216 et al. 2005; Da Lage et al. 2007; Robe et al. 2010a; Gao et al. 2011; O'Grady et al.
217 2011; Russo et al. 2013; Yassin 2013).

218

219 In the following sections of the paper, we will highlight and discuss some of the most
220 interesting results we obtained. Our analyses either confirm or challenge previous
221 phylogenies and shed light on several unassessed questions, contributing to an
222 emerging picture of phylogenetic relationships in Drosophilidae.

223

224 **The Steganinae subfamily**

225 To avoid long branch attraction due to some divergent steganine sequences, we
226 compiled a more specific and comprehensive dataset from 164 taxa of Steganinae
227 (versus 80 taxa in the 704-taxon analysis). Whereas morphology-based studies
228 suggest the monophyly of Steganinae (Okada 1989; Grimaldi 1990), molecular
229 phylogenetic have led to contradictory results (Remsen and O’Grady 2002; Otranto et
230 al. 2008; Van Der Linde et al. 2010; Russo et al. 2013; Yassin 2013). Our study
231 identifies the Steganinae as monophyletic for both datasets (Figures 1 and S8) and
232 supports a recent phylogenomic study of Steganinae (Dias et al. 2020). The topology
233 within the Steganinae substantially differs from the division of the subfamily into two
234 monophyletic tribes: Steganini and Gitonini (Yassin 2013). Our study does not
235 recover the monophyly of the genera *Leucophenga* and *Parastegana*, only due to the
236 placement of the two species *Leucophenga maculata* and *Parastegana femorata*.
237 Future studies are needed to disentangle possible contamination and true phylogenetic
238 position. We also found the branching of some *Colocasiomyia* species within the
239 Steganinae (Figure S8). This finding, which challenges previous published
240 cladograms of *Colocasiomyia* (Grimaldi 1991; Sultana et al. 2006) and our 704-taxon
241 analysis (Figure 1), is likely an artifact of reconstruction.

242

243 **The *Sophophora* subgenus and closely related taxa**

244 We found that the *obscura-melanogaster* clade is the sister group of the lineages
245 formed by the Neotropical *saltans* and *willistoni* groups, and the *Lordiphosa* genus
246 (bootstrap percentage [BP] = 73) (Figures 2A and S3). Thus, our study recovers the
247 relationship between the groups of the *Sophophora* subgenus (Gao et al. 2011; Russo
248 et al. 2013; Yassin 2013) and supports the paraphyletic status of *Sophophora*
249 regarding *Lordiphosa* (Katoh et al. 2000). However, we noted substantial changes
250 within the topology presented for the *melanogaster* species group. The original
251 description of *Drosophila oshimai* noted a likeness to *Drosophila unipectinata*, thus
252 classifying *D. oshimai* into the *suzukii* species subgroup (Choo and Nakamura 1973).
253 The phylogenetic tree we obtained does not support this classification (Figure 2A). It
254 rather defines *D. oshimai* as the representative of a new subgroup (Bayesian posterior
255 probability [PP] = 1, BP = 96) that diverged immediately after the split of the
256 *montium* group. The position of *D. oshimai* therefore challenges the monophyly of the
257 *suzukii* subgroup. Interestingly, the paraphyly of the *suzukii* subgroup has also been
258 suggested in previous studies (Lewis et al. 2005; Russo et al. 2013). Another

259 interesting case is the positioning of the *denticulata* subgroup that has never been
260 tested before. Our analysis convincingly places its representative species *Drosophila*
261 *denticulata* as the fourth subgroup to branch off within the *melanogaster* group (PP =
262 1, BP = 82). Last, the topology within the *montium* group drastically differs from the
263 most recent published phylogeny (Conner et al. 2021). Despite substantial sampling in
264 the subgenus *Sophophora*, our study would benefit from the addition of
265 representatives of the *dentissima*, *dispar*, *fima*, *populi*, *setifemur* groups, as well as the
266 genus *Zapriothrica*, to draw a more complete picture of the relationships within
267 *Sophophora*.

268 The genus *Collessia* comprises five described species that can be found in Australia,
269 Japan, and Sri Lanka, but its phylogenetic status was so far quite ambiguous (Okada
270 1967; Bock 1982; Okada 1988). In addition, Grimaldi (1990) proposed that
271 *Tambourella ornata* should belong to the genus *Collessia*. These two genera are
272 similar in the wing venation and pigmentation pattern (Okada 1984).

273 Our phylogenetic analysis identifies *Collessia* as sister group to the species
274 *Hirtodrosophila duncani* (PP = 1, BP = 100). Interestingly, this branching is also
275 supported by morphological similarities shared between the genera *Collessia* and
276 *Hirtodrosophila*. The species *C. kirishimana* and *C. hiharai* were indeed initially
277 described as *Hirtodrosophila* species (Okada 1967) but later assigned to the genus
278 *Collessia* (Okada 1984), based on the similarity in wing coloration with *C. superba*.
279 However, the affiliation of *C. kirishimana* to *Collessia* would require further
280 investigations. The species *H. duncani* is morphologically disparate for
281 *Hirtodrosophila* and might be removed from this genus in the future (Grimaldi 2018).
282 The clade *Collessia-H. duncani* is sister to the *Sophophora-Lordiphosa* lineage in the
283 ML inference (BP = 100) but to the Neotropical *Sophophora-Lordiphosa* clade in the
284 Bayesian inference (PP = 0.92).

285

286 **The early lineage of *Microdrosophila* and *Dorsilopha***

287 Within the tribe Drosophilini, all the remaining taxa (composite taxa + ungrouped
288 species) other than those of the *Sophophora-Lordiphosa* and *Collessia-H. duncani*
289 lineage form a large clade (PP = 1, BP = 100). Within this clade, the genus
290 *Microdrosophila*, the subgenus *Dorsilopha*, and *Drosophila ponera* group into a
291 lineage (PP = 0.97, BP = 82) that appears as an early offshoot in our composite tree
292 (Figure 2), reminiscent of the placement of *Dorsilopha* found in Yassin (2013). It is

293 nevertheless noteworthy that the placement of the *Dorsilopha* + *Microdrosophila*
294 clade differs in our supermatrix tree (Figure 1) and resembles the placement of
295 *Microdrosophila* in Yassin (2013). In spite of scarce genomic data, we added the
296 genus *Styloptera* which has been previously found close to the genus *Dorsilopha*
297 (Yassin 2013). The position of *Styloptera* varies according to the analysis (Figure S9
298 and online supplementary tree files) without grouping with *Dorsilopha*. Generating
299 genomic data for the genus *Styloptera* will be necessary to unambiguously place this
300 genus. *Drosophila ponera* is an enigmatic species collected in La Réunion (David and
301 Tsacas 1975), whose phylogenetic position has never or rarely been investigated. In
302 spite of morphological similarities with the *quinaria* group, the authors suggested to
303 keep *D. ponera* as ungrouped with respect to a divergent number of respiratory egg
304 filaments (David and Tsacas 1975). To our knowledge, our study is the first attempt
305 to phylogenetically position this species. We found that *D. ponera* groups with the
306 *Dorsilopha* subgenus (PP = 0.99, BP = 75) within this early-diverging lineage.

307

308 **The Hawaiian drosophilid clade and the *Siphlodora* subgenus**

309 The endemic Hawaiian Drosophilidae contain approximately 1,000 species that split
310 into the genera *Idiomyia* (or Hawaiian *Drosophila* according to Grimaldi (1990)) and
311 the genus *Scaptomyza* (O'Grady et al. 2009). Generally considered as sister to the
312 *Siphlodora* subgenus (Robe et al. 2010b; Russo et al. 2013; Yassin 2013), these
313 lineages represent a remarkable framework to investigate evolutionary radiation and
314 subsequent diversification of morphology (Stark and O'Grady 2010), pigmentation
315 (Edwards et al. 2007), ecology (Magnacca et al. 2008), and behavior (Kaneshiro
316 1999). Although the relationships within the *Siphlodora* clade are generally in
317 agreement with previous studies (Tatarenkov et al. 2001; Robe et al. 2010b; Russo et
318 al. 2013; Yassin 2013), its sister clade does not seem to be restricted to the Hawaiian
319 Drosophilidae. In fact, according to our phylogenies, it also includes at least four
320 other species of the genus *Drosophila* (Figures 2A, S3, and online supplementary tree
321 files). We propose that this broader clade, rather than the Hawaiian clade *sensu*
322 *stricto*, should be seen as a major lineage of Drosophilidae.

323 This broader clade is strongly supported (PP = 1, BP = 100) and divided into two
324 subclades, one comprises the genera *Idiomyia* and *Scaptomyza* (PP = 0.99, BP = 97)
325 and the other includes *D. annulipes*, *D. adamsi*, *D. maculinotata* and *D. nigrosparsa*
326 (PP = 0.99, BP = 75). The latter subclade, also suggested by Katoh et al. (2007) and

327 Russo et al. (2013), is interesting with respect to the origin of Hawaiian drosophilids.
328 Of the four component species, *D. annulipes* was originally described as a member of
329 the subgenus *Spinulophila*, which was synonymized with *Drosophila* and currently
330 corresponds to the *immigrans* group, although Wakahama et al. (1983) and Zhang and
331 Toda (1992) cast doubt on its systematic position. The fact that *D. annulipes* does not
332 belong to the *immigrans* species group implies that the subgenus *Drosophila* is
333 paraphyletic rather than polyphyletic. As for *D. adamsi*, Da Lage et al. (2007)
334 suggested it may be close to the *Idiomyia-Scaptomyza* clade, which is supported by
335 our analyses. On the other hand, Prigent et al. (2013) based on morphological
336 characters and Prigent et al. (2017) based on DNA barcoding have proposed that *D.*
337 *adamsi* defines a new species group along with *D. acanthomera* and an undescribed
338 species. *Drosophila adamsi* resembles *D. annulipes* in the body color pattern (Fig.
339 2F,E,H), suggesting their close relationship: Adams (1905) described, “mesonotum
340 with five longitudinal, brown vittae, the central one broader than the others and
341 divided longitudinally by a hair-like line, ...; scutellum yellow, with two sublateral,
342 brownish lines, ...; pleurae with three longitudinal brownish lines”, for *Drosophila*
343 *quadrimaculata* Adams, 1905, which is a homonym of *Drosophila quadrimaculata*
344 Walker, 1856 and has been replaced with the new specific epithet “*adamsi*” by
345 Wheeler (1959). Another species, *D. nigrosparsa*, belongs to the *nigrosparsa* species
346 group, along with *D. secunda*, *D. subarctica* and *D. vireni* (Bächli et al. 2004).
347 Moreover, Máca (1992) pointed out the close relatedness of *D. maculinotata* to the
348 *nigrosparsa* group. It is noteworthy that the *nigrosparsa* species group is thought to
349 be basal to *Siphlodora* in regard to the morphology of male genitalia (Yassin 2013).

350

351 **The *Drosophila* subgenus and closely related taxa**

352 Although general relationships within the *Drosophila* subgenus closely resemble
353 those recovered by previous studies (Hatadani et al. 2009; Robe et al. 2010b; Robe et
354 al. 2010c; Izumitani et al. 2016), there are some outstanding results related to other
355 genera or poorly studied *Drosophila* species.

356 *Samoaia* is a small genus of seven described species endemic to the Samoan
357 Archipelago (Malloch 1934; Wheeler and Kambysellis 1966), particularly studied for
358 their body and wing pigmentation (Dufour et al. 2020). In our analysis, the genus
359 *Samoaia* is found to group with the *quadri-lineata* species subgroup of the *immigrans*
360 group. This result is similar to conclusions formulated by some previous studies

361 (Tatarenkov et al. 2001; Robe et al. 2010b; Yassin et al. 2010; Yassin 2013), but
362 differs from other published phylogenies in which *Samoaia* is sister to most other
363 lineages in the subgenus *Drosophila* (Russo et al. 2013). It is noteworthy that our
364 sampling is the most substantial with four species of *Samoaia*.

365 The two African species *Drosophila pruinosa* and *Drosophila pachneissa*, which
366 were assigned to the *loiciana* species complex because of shared characters such as a
367 glaucous-silvery frons and rod-shaped surstylus (Tsacas 2002), are placed together
368 with the *immigrans* group (PP = 1, BP = 94). In previous large-scale analyses, *D.*
369 *pruinosa* was suggested to group with *Drosophila sternopleuralis* into the sister clade
370 of the *immigrans* group (Da Lage et al. 2007; Russo et al. 2013).

371 Among other controversial issues, the phylogenetic position of *Drosophila aracea*
372 was previously found to markedly change according to the phylogenetic
373 reconstruction methods (Da Lage et al. 2007). This anthophilic species lives in
374 Central America (Heed and Wheeler 1957). Its name comes from the behavior of
375 females that lay eggs on the spadix of plants in the family Araceae (Heed and
376 Wheeler 1957; Tsacas and Chassagnard 1992). Our analysis places *D. aracea* as the
377 sister taxon of the *bizonata-testacea* clade with high confidence (PP = 1, BP = 85).
378 No occurrence of flower-breeding behavior has been reported in the *bizonata-testacea*
379 clade, reinforcing the idea that *D. aracea* might have recently evolved from a
380 generalist ancestor (Tsacas and Chassagnard 1992).

381

382 **The *Zygothrica* genus group**

383 The fungus-associated genera *Hirtodrosophila*, *Mycodrosophila*, *Paraliodrosophila*,
384 *Paramycodrosophila*, and *Zygothrica* contain 449 identified species (DrosWLD-
385 Species 2021) and have been associated with the *Zygothrica* genus group (Grimaldi
386 1990). Although the *Zygothrica* genus group was recurrently recovered as
387 paraphyletic (Da Lage et al. 2007; Van Der Linde et al. 2010; Russo et al. 2013;
388 Yassin 2013), two recent studies suggest, on the contrary, its monophyly (Gautério et
389 al. 2020; Zhang et al. 2021). Our study does not support the monophyly of the
390 *Zygothrica* genus group in virtue of the polyphyletic status of *Hirtodrosophila* and
391 *Zygothrica*: some representatives (e.g., *H. duncani*) cluster with *Collessia*, while
392 others (e.g., *Hirtodrosophila* IV and *Zygothrica* II) appear closely related to the
393 genera *Dichaetophora* and *Mulgravea*. Furthermore, the placement of the *Zygothrica*
394 genus group recovered in our study also differs from some previous estimates. In fact,

395 the broadly defined *Zygothrica* genus group, which includes *Dichaetophora* and
396 *Mulgravea* (PP = 0.95, BP = 64), appears as sister to the clade composed of the
397 subgenus *Drosophila* and the *Hypselothyrea/Liodrosophila* + *Sphaerogastrella* +
398 *Zaprionus* clade (PP = 1, BP = 56) (Figures 2A and S3). This placement is similar to
399 the ones obtained in different studies (Van Der Linde et al. 2010; Russo et al. 2013),
400 but contrasts with the close relationship of the *Zygothrica* genus group to the
401 subgenus *Siphlodora* + *Idiomyia/Scaptomyza* proposed in two recent studies
402 (Gautério et al. 2020; Zhang et al. 2021). Given the moderate bootstrap value, the
403 exact status of the *Zygothrica* genus group remains as an open question.

404 Furthermore, within the superclade of the broadly defined *Zygothrica* genus group
405 (Figures 1 and 2A), the genus *Hirtodrosophila* is paraphyletic and split into four
406 independent lineages, reinforcing previous suggestions based on multilocus
407 approaches (Van Der Linde et al. 2010; Gautério et al. 2020; Zhang et al. 2021). This
408 also occurred with the genus *Zygothrica*, which split into two independent clades
409 (Figure 2A). The *leptorostra* subgroup (*Zygothrica* II) clusters with the subgroup
410 *Hirtodrosophila* IV (PP = 1, BP = 100), whereas the *Zygothrica* I subgroup clusters
411 with the species *Hirtodrosophila levigata* (PP = 0.99, BP = 98).

412

413 **DrosoPhyla: a powerful tool for systematics**

414 Besides bringing an updated and improved phylogenetic framework to Drosophilidae,
415 our approach also addresses several questions that were previously unassessed or
416 controversial at the genus, subgenus, group, or species level. We are therefore
417 confident that it may become a powerful tool for future drosophilid systematics.
418 According to diversity surveys (O’Grady and DeSalle 2018), ~25% of drosophilid
419 species remain to be discovered, potentially a thousand species to place in the tree of
420 Drosophilidae. While whole-genome sequencing is becoming widespread, newly
421 discovered species often come down to a few specimens pinned or stored in ethanol –
422 non-optimal conditions for subsequent genome sequencing and whole-genome studies
423 (Korlević et al. 2021). An alternative promising approach to PCR is exome capture
424 using baits to hybridize to genomic regions of interest, which has been used with
425 other insects (Branstetter et al. 2017). Nevertheless, based on a few short genomic
426 markers, our approach is compatible with taxonomic work, and gives good resolution.

427

428 **Acknowledgements**

429 We thank Jean-Luc Da Lage, Beatriz Goni, John Jaenike, Louis Bernard Klaczko,
430 Adriana Ludwig, Suzana Vaz, and Carlos Vilela for providing fly specimens. We
431 thank Virginie Orgogozo and Noah Whiteman for giving early access to the genome
432 of *D. pachea* and *S. flava*, respectively. We thank Masafumi Inoue, Stéphane Prigent,
433 Yasuo Hoshino, and the Japan Drosophila Database for providing photos. We thank
434 Amir Yassin for fruitful discussions and comments on the manuscript. We thank the
435 Sean Carroll laboratory for discussions and financial support. This paper is dedicated
436 to the memory of the French biologist Jean David and his great legacy to the biology
437 of *Drosophila*.

438

439 **Material and Methods**

440 **Taxon sampling**

441 The species used in this study were sampled from different locations throughout the
442 world (Table S1). The specimens were field-collected by the authors, purchased from
443 the National Drosophila Species Stock Center (<http://blogs.cornell.edu/drosophila/>)
444 and the Kyoto Stock Center (<https://kyotofly.kit.jp/cgi-bin/stocks/index.cgi>), or
445 obtained from colleagues. Individual flies were preserved in 100% ethanol and
446 identified based on morphological characters.

447

448 **Data collection**

449 Ten genomic markers were amplified by PCR using degenerate primers developed for
450 the present study (Table 1). Genomic DNA was extracted from a single adult fly as
451 follows: the fly was placed in a 0.5-mL tube and mashed in 50 μ L of squishing buffer
452 (Tris-HCl pH=8.2 10 mM, EDTA 1 mM, NaCl 25 mM, proteinase K 200 μ g/mL) for
453 20-30 seconds, the mix was incubated at 37°C for 30 minutes, then the proteinase K
454 was inactivated by heating at 95°C for 1-2 minutes. A volume of 1 μ L was used as
455 template for PCR amplification. Nucleotide sequences were also retrieved from the
456 NCBI database for the five nuclear markers *28S ribosomal RNA (28S)*, *alcohol*
457 *dehydrogenase (Adh)*, *glycerol-3-phosphate dehydrogenase (Gpdh)*, *superoxide*
458 *dismutase (Sod)*, *xanthine dehydrogenase (Xdh)*, and the two mitochondrial markers
459 *cytochrome oxidase subunit 1 (COI)* and *cytochrome oxidase subunit 2 (COII)*. The
460 sequences reported in this paper have been deposited in GenBank under specific

461 accession numbers: *Amyrel* (MW392482-MW392524), *Ddc* (MW403139-
462 MW403307), *Dll* (MW403308-MW403483), *eb* (MW415022-MW415267), *en*
463 (MW418945-MW419079), *eve* (MW425034-MW425273), *hh* (MW385549-
464 MW385782), *Notum* (MW429853-MW430003), *ptc* (MW442160-MW442361), *wg*
465 (MW392301-MW392481).

466

467 **Phylogenetic reconstruction**

468 Alignments for each individual gene were generated using MAFFT 7.45 (Kato and
469 Standley 2013) assuming a gap opening penalty of 1.53 and other default parameters
470 (no offset and extra round of refinement). Unreliably aligned positions were excluded
471 using trimAl with parameters -gt 0.5 and -st 0.001 (Capella-Gutiérrez et al. 2009).
472 The possible contamination status was verified by inferring independent trees for each
473 gene using RAxML 8.2.4 under the GTR+ Γ_4 model (Stamatakis 2014). Thus, any
474 sequence leading to the suspicious placement of a taxonomically well-assigned
475 species, in terms of both topology and bootstrap value, was removed from the dataset.
476 Moreover, almost identical sequences leading to very short tree branches were
477 carefully examined and excluded if involving non-closely related taxa. In-house
478 Python scripts were used to concatenate the aligned and filtered sequences, and the
479 resulting dataset was used for phylogenetic reconstruction. Maximum-likelihood
480 (ML) searches were performed using IQ-TREE 2.0.6 (Minh, Schmidt, et al. 2020)
481 under the GTR model, with the FreeRate model of rate heterogeneity across sites with
482 four categories, and ML estimation of base frequencies from the data (GTR+R+FO).
483 The edge-linked proportional partition model was used with one partition for each
484 gene.

485

486 **Composite taxa**

487 This strategy started from clustering the species by unambiguous monophyletic
488 genera, groups, or subgroups identified in the 704-taxon analysis. After this, the least
489 diverging sequence or species recovered for each taxonomic unit for each marker was
490 selected to ultimately yield a unique composite taxon by concatenation. The
491 composite matrix was also used for conducting ML and Bayesian phylogenetic
492 inference using IQ-TREE under a partitioned GTR+R+FO model (parameters: -m

493 GTR+FO+R -B 1000 -bnni -p) , and PhyloBayes under a GTR+Γ model (parameters:
494 -ncat 1 -gtr) (Lartillot et al. 2009), respectively.

495

496 **Saturation and concordance analysis**

497 For each marker gene, the saturation was computed by performing a simple linear
498 regression of the percent identity for each pair of taxa (observed distance) onto the
499 ML patristic distance (inferred distance) (Philippe et al. 1994) estimated using the
500 ETE 3 library (Huerta-Cepas et al. 2016). We also calculated per gene and per site
501 concordance factors using IQ-TREE under the GTR+R+FO model as recently
502 described (Minh, Hahn, et al. 2020). We also applied ASTRAL to estimate species
503 tree from individual species tree, using default parameters and the same input single
504 gene trees (Zhang et al. 2018).

505

506 **Data availability statement**

507 The data underlying this article are available on Zenodo (10.5281/zenodo.5091961).

508

509 **Author contributions**

510 C.F. and H.D.D. initiated the project. M.J.T. provided most of the specimens. C.F.
511 and F.M. established the methodological approaches. The generation of new
512 sequences is primarily attributable to C.F., V.A.K., H.D.D., then to most authors of
513 the paper. C.F. gathered and formatted the data. F.M. conducted all analyses. C.F.,
514 M.J.T., L.J.R. and F.M. wrote the first version of the manuscript, and all authors
515 contributed edits and further elaborations.

516

517 **Competing interests**

518 The authors have no competing interests.

519

520 **References**

- 521 Adams CF. 1905. Diptera Africana, I. Kansas Univ. Sci. Bull. 3:149–188.
522 Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer
523 SE, Li PW, Hoskins RA, Galle RF, et al. 2000. The genome sequence of
524 *Drosophila melanogaster*. Science 287:2185–2195.
525 Almeida FC, Sánchez-Gracia A, Campos JL, Rozas J. 2014. Family size evolution in

526 drosophila chemosensory gene families: A comparative analysis with a critical
527 appraisal of methods. *Genome Biol. Evol.* 6:1669–1682.

528 Baker RH, Desalle R. 1997. Multiple sources of character information and the
529 phylogeny of Hawaiian Drosophilids. *Syst. Biol.* 46:654–673.

530 Bellen HJ, Tong C, Tsuda H. 2010. 100 years of Drosophila research and its impact
531 on vertebrate neuroscience: a history lesson for the future. *Nat. Rev. Neurosci.*
532 11:514–522.

533 Bhutkar A, Schaeffer SW, Russo SM, Xu M, Smith TF, Gelbart WM. 2008.
534 Chromosomal rearrangement inferred from comparisons of 12 drosophila
535 genomes. *Genetics* 179:1657–1680.

536 Bock I. 1982. Drosophilidae of Australia V. Remaining genera and synopsis (Insecta:
537 Diptera). *Aust. J. Zool.* 89:1–164.

538 Bonacum J, O’Grady PM, Kambysellis M, DeSalle R. 2005. Phylogeny and age of
539 diversification of the planitibia species group of the Hawaiian Drosophila. *Mol.*
540 *Phylogenet. Evol.* 37:73–82.

541 Bosco G, Campbell P, Leiva-Neto JT, Markow TA. 2007. Analysis of Drosophila
542 species genome size and satellite DNA content reveals significant differences
543 among strains as well as between species. *Genetics* 177:1277–1290.

544 Branstetter MG, Danforth BN, Pitts JP, Faircloth BC, Ward PS, Buffington ML,
545 Gates MW, Kula RR, Brady SG. 2017. Phylogenomic Insights into the Evolution
546 of Stinging Wasps and the Origins of Ants and Bees. *Curr. Biol.* 27:1019–1025.

547 Buchon N, Silverman N, Cherry S. 2014. Immunity in *Drosophila melanogaster* —
548 from microbial recognition to whole-organism physiology. *Nat. Rev. Immunol.*
549 14:796–810.

550 Campbell V, Lapointe FJ. 2009. The use and validity of composite taxa in
551 phylogenetic analysis. *Syst. Biol.* 58:560–572.

552 Campbell V, Lapointe FJ. 2011. Retrieving a mitogenomic mammal tree using
553 composite taxa. *Mol. Phylogenet. Evol.* 58:149–156.

554 Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: A tool for
555 automated alignment trimming in large-scale phylogenetic analyses.
556 *Bioinformatics* 25:1972–1973.

557 Charbonnier S, Audo D, Barriol V, Garassino A, Schweigert G, Simpson M. 2015.
558 Phylogeny of fossil and extant glypheid and litogastrid lobsters (Crustacea,
559 Decapoda) as revealed by morphological characters. *Cladistics* 31:231–249.

560 Choo J, Nakamura K. 1973. On a new species of *Drosophila* (Sophophora) from
561 Japan (Diptera). *Kontyû* 41:305–306.

562 Chung H, Loehlin DW, Dufour HD, Vaccarro K, Millar JG, Carroll SB. 2014. A
563 single gene affects both ecological divergence and mate choice in *Drosophila*.
564 *Science* 343:1148–1151.

565 Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman TC,
566 Kellis M, Gelbart W, Iyer VN, et al. 2007. Evolution of genes and genomes on
567 the *Drosophila* phylogeny. *Nature* 450:203–218.

568 Cobb M. 2007. A gene mutation which changed animal behaviour: Margaret Bastock
569 and the yellow fly. *Anim. Behav.* 74:163–169.

570 Conner WR, Delaney EK, Bronski MJ, Ginsberg PS, Wheeler TB, Richardson KM,
571 Peckenpaugh B, Kim KJ, Watada M, Hoffmann AA, et al. 2021. A phylogeny
572 for the *Drosophila montium* species group: A model clade for comparative
573 analyses. *Mol. Phylogenet. Evol.* 158:107061.

574 Dai H, Chen Y, Chen S, Mao Q, Kennedy D, Landback P, Eyre-Walker A, Du W,
575 Long M. 2008. The evolution of courtship behaviors through the origination of a
576 new gene in *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* 105:7478–7483.

577 David J, Tsacas L. 1975. Les *Drosophilidae* (Diptera) de l’Ile de la Réunion et de l’Ile
578 Maurice. I. Deux nouvelles espèces du genre *Drosophila*. *Bull. Mens. la Société*
579 *Linnéenne Lyon* 5:134–143.

580 Dias GR, Dupim EG, Vanderlinde T, Mello B, Carvalho AB. 2020. A phylogenomic
581 study of Steganinae fruit flies (Diptera: Drosophilidae): strong gene tree
582 heterogeneity and evidence for monophyly. *BMC Evol. Biol.* 20.

583 DrosWLD-Species. 2021. DrosWLD-Species.
584 <https://bioinfo.museum.hokudai.ac.jp/db/index.php>

585 Dufour HD, Koshikawa S, Finet C. 2020. Temporal flexibility of gene regulatory
586 network underlies a novel wing pattern in flies. *Proc. Natl. Acad. Sci.*
587 117:11589–11596.

588 Edwards KA, Doescher LT, Kaneshiro KY, Yamamoto D. 2007. A database of wing
589 diversity in the Hawaiian *Drosophila*. *PLoS One* 2:3487.

590 Fan L, Wu D, Goremykin V, Xiao J, Xu Y, Garg S, Zhang C, Martin WF, Zhu R.
591 2020. Phylogenetic analyses with systematic taxon sampling show that
592 mitochondria branch within Alphaproteobacteria. *Nat. Ecol. Evol.* 4:1213–1219.

593 Finet C, Slavik K, Pu J, Carroll SB, Chung H. 2019. Birth-and-Death Evolution of the

594 Fatty Acyl-CoA Reductase (FAR) Gene Family and Diversification of Cuticular
595 Hydrocarbon Synthesis in *Drosophila*. *Genome Biol. Evol.* 11:1541–1551.

596 Finet C, Timme RE, Delwiche CF, Marlétaz F. 2010. Multigene phylogeny of the
597 green lineage reveals the origin and diversification of land plants. *Curr. Biol.*
598 20:2217–2222.

599 Gao JJ, Hu YG, Toda MJ, Katoh T, Tamura K. 2011. Phylogenetic relationships
600 between Sophophora and Lordiphosa, with proposition of a hypothesis on the
601 vicariant divergences of tropical lineages between the Old and New Worlds in
602 the family Drosophilidae. *Mol. Phylogenet. Evol.* 60:98–107.

603 Gautério TB, Machado S, Loreto EL da S, Gottschalk MS, Robe LJ. 2020.
604 Phylogenetic relationships between fungus-associated Neotropical species of the
605 genera *Hirtodrosophila*, *Mycodrosophila* and *Zygothrica* (Diptera,
606 Drosophilidae), with insights into the evolution of breeding sites usage. *Mol.*
607 *Phylogenet. Evol.* 145.

608 Glassford WJ, Johnson WC, Dall NR, Smith SJ, Liu Y, Boll W, Noll M, Rebeiz M.
609 2015. Co-option of an Ancestral Hox-Regulated Network Underlies a Recently
610 Evolved Morphological Novelty. *Dev. Cell* 34:520–531.

611 Grimaldi D. 1987. Amber Fossil Drosophilidae (Diptera), with Particular Reference to
612 the Hispaniolan taxa. *Am. Museum Novit.* 2880:1–23.

613 Grimaldi D. 1991. Systematics of the genus *Colocasiomyia* de Meijere (Diptera:
614 Drosophilidae): cladistics, a new generic synonym, new records, and a new
615 species from Nepal. *Insect Syst. Evol.* 22:417–426.

616 Grimaldi DA. 1990. A Phylogenetic, Revised Classification of Genera in the
617 Drosophilidae (Diptera). *Bull. Am. Museum Nat. Hist.* 197.

618 Grimaldi DA. 2018. *Hirtodrosophila* of North America (Diptera: Drosophilidae). *Bull.*
619 *Am. Museum Nat. Hist.* 421.

620 Hales KG, Korey CA, Larracuente AM, Roberts DM. 2015. Genetics on the fly: A
621 primer on the *drosophila* model system. *Genetics* 201:815–842.

622 Hatadani LM, McInerney JO, Medeiros HF de, Junqueira ACM, Azeredo-Espin AM
623 de, Klaczko LB. 2009. Molecular phylogeny of the *Drosophila tripunctata* and
624 closely related species groups (Diptera: Drosophilidae). *Mol. Phylogenet. Evol.*
625 51:595–600.

626 Heed WB, Wheeler MR. 1957. Thirteen new species in the genus *Drosophila* from the
627 Neotropical region. *Univ. Texas Publ.* 5721:17–38.

628 Huerta-Cepas J, Serra F, Bork P. 2016. ETE 3: Reconstruction, Analysis, and
629 Visualization of Phylogenomic Data. *Mol. Biol. Evol.* 33:1635–1638.

630 Izumitani HF, Kusaka Y, Koshikawa S, Toda MJ, Katoh T. 2016. Phylogeography of
631 the subgenus *Drosophila* (Diptera: Drosophilidae): Evolutionary history of faunal
632 divergence between the old and the new worlds. *PLoS One* 11:e0160051.

633 Jeffroy O, Brinkmann H, Delsuc F, Philippe H. 2006. Phylogenomics: the beginning
634 of incongruence? *Trends Genet.* 22:225–231.

635 Jeong S, Rebeiz M, Andolfatto P, Werner T, True J, Carroll SB. 2008. The Evolution
636 of Gene Regulation Underlies a Morphological Difference between Two
637 *Drosophila* Sister Species. *Cell* 132:783–793.

638 Kaneshiro KY. 1999. Sexual selection and speciation in hawaiian *Drosophila*
639 (*Drosophilidae*): A model system for research in *Tephritidae*. In: *Fruit Flies*
640 (*Tephritidae*): Phylogeny and Evolution of Behavior.

641 Kapli P, Yang Z, Telford MJ. 2020. Phylogenetic tree building in the genomic age.
642 *Nat. Rev. Genet.* 21:428–444.

643 Karageorgi M, Bräcker LB, Lebreton S, Minervino C, Cavey M, Siju KP, Grunwald
644 Kadow IC, Gompel N, Prud'homme B. 2017. Evolution of Multiple Sensory
645 Systems Drives Novel Egg-Laying Behavior in the Fruit Pest *Drosophila suzukii*.
646 *Curr. Biol.* 27:847–853.

647 Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version
648 7: Improvements in performance and usability. *Mol. Biol. Evol.* 30:772–780.

649 Katoh T, Izumitani HF, Yamashita S, Watada M. 2017. Multiple origins of hawaiian
650 *Drosophilids*: Phylogeography of *Scaptomyza hardy* (Diptera: Drosophilidae).
651 *Entomol. Sci.* 20:33–44.

652 Katoh T, Nakaya D, Tamura K, Aotsuka T. 2007. Phylogeny of the *Drosophila*
653 *immigrans* species group (Diptera: Drosophilidae) based on *Adh* and *Gpdh*
654 sequences. *Zoolog. Sci.* 24:913–921.

655 Katoh T, Tamura K, Aotsuka T. 2000. Phylogenetic position of the subgenus
656 *lordiphosa* of the genus *Drosophila* (Diptera: Drosophilidae) inferred from
657 alcohol dehydrogenase (*Adh*) gene sequences. *J. Mol. Evol.* 51:122–130.

658 Katoh TK, Zhang G, Toda MJ, Suwito A, Gao JJ. 2018. A revision of the subgenus
659 *dudaica* strand of the genus *Drosophila fallén*, with descriptions of six new
660 species (Diptera, Drosophilidae). *Zookeys* 2018:19–50.

661 Kellermann V, Van Heerwaarden B, Sgrò CM, Hoffmann AA. 2009. Fundamental

662 evolutionary limits in ecological traits drive drosophila species distributions.
663 Science 325:1244–1246.

664 Kim BY, Wang JR, Miller DE, Barmina O, Delaney E, Thompson A, Comeault AA,
665 Peede D, D’Agostino ERR, Pelaez J, et al. 2021. Highly contiguous assemblies
666 of 101 drosophilid genomes. eLife 10:e66405.

667 Korlević P, McAlister E, Mayho M, Makunin A, Flicek P, Lawniczac MKN. A
668 minimally morphologically destructive approach for DNA retrieval and whole
669 genome shotgun sequencing of pinned historic Dipteran vector species. BioRxiv.

670 Da Lage JL, Kergoat GJ, Maczkowiak F, Silvain JF, Cariou ML, Lachaise D. 2007. A
671 phylogeny of Drosophilidae using the Amyrel gene: Questioning the Drosophila
672 melanogaster species group boundaries. J. Zool. Syst. Evol. Res. 45:47–63.

673 Lang M, Murat S, Clark AG, Gouppil G, Blais C, Matzkin LM, Guittard É,
674 Yoshiyama-Yanagawa T, Kataoka H, Niwa R, et al. 2012. Mutations in the
675 neverland gene turned Drosophila pachea into an obligate specialist species.
676 Science 337:1658–1661.

677 Lapoint RT, O’Grady PM, Whiteman NK. 2013. Diversification and dispersal of the
678 Hawaiian Drosophilidae: The evolution of Scaptomyza. Mol. Phylogenet. Evol.

679 Lartillot N, Lepage T, Blanquart S. 2009. PhyloBayes 3: A Bayesian software
680 package for phylogenetic reconstruction and molecular dating. Bioinformatics
681 25:2286–2288.

682 Lewis RL, Beckenbach AT, Mooers A. 2005. The phylogeny of the subgroups within
683 the melanogaster species group: Likelihood tests on COI and COII sequences
684 and a Bayesian estimate of phylogeny. Mol. Phylogenet. Evol. 37.

685 Van Der Linde K, Houle D, Spicer GS, Stepan SJ. 2010. A supermatrix-based
686 molecular phylogeny of the family Drosophilidae. Genet. Res. (Camb). 92:25–
687 38.

688 Longdon B, Hadfield JD, Day JP, Smith SCL, McGonigle JE, Cogni R, Cao C,
689 Jiggins FM. 2015. The Causes and Consequences of Changes in Virulence
690 following Pathogen Host Shifts. PLoS Pathog. 11:e1004728.

691 Magnacca KN, Foote D, O’Grady PM. 2008. A review of the endemic Hawaiian
692 Drosophilidae and their host plants. Zootaxa 1728:1–58.

693 Malloch JR. 1934. Part VI. Diptera. In: Insects of Samoa. p. 267–312.

694 Markow T a., O’Grady P. 2006. Drosophila: A Guide to Species Identification and
695 Use. Elsevier.

696 McGregor AP, Orgogozo V, Delon I, Zanet J, Srinivasan DG, Payre F, Stern DL.
697 2007. Morphological evolution through multiple cis-regulatory mutations at a
698 single gene. *Nature* 448:587-590.

699 Mengual X, Kerr P, Norrbom AL, Barr NB, Lewis ML, Stapelfeldt AM, Scheffer SJ,
700 Woods P, Islam MS, Korytkowski CA, et al. 2017. Phylogenetic relationships of
701 the tribe Toxotrypanini (Diptera: Tephritidae) based on molecular characters.
702 *Mol. Phylogenet. Evol.* 113:84–112.

703 Minh BQ, Hahn MW, Lanfear R. 2020. New methods to calculate concordance
704 factors for phylogenomic datasets. *Mol. Biol. Evol.* 37:2727–2733.

705 Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler
706 A, Lanfear R, Teeling E. 2020. IQ-TREE 2: New Models and Efficient Methods
707 for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* 37:1530–1534.

708 Morales-Hojas R, Vieira J. 2012. Phylogenetic Patterns of Geographical and
709 Ecological Diversification in the Subgenus *Drosophila*. *PLoS One* 7:e49552.

710 Morgan TH. 1910. Sex Limited Inheritance in *Drosophila*. *Science* 32:120–122.

711 O’Grady PM, Clark JB, Kidwell MG. 1998. Phylogeny of the *Drosophila saltans*
712 species group based on combined analysis of nuclear and mitochondrial DNA
713 sequences. *Mol. Biol. Evol.* 15:656–664.

714 O’Grady PM, DeSalle R. 2018. Phylogeny of the genus *Drosophila*. *Genetics* 209:1–
715 25.

716 O’Grady PM, Lapoint RT, Bonacum J, Lasola J, Owen E, Wu Y, DeSalle R. 2011.
717 Phylogenetic and ecological relationships of the Hawaiian *Drosophila* inferred
718 by mitochondrial DNA analysis. *Mol. Phylogenet. Evol.*

719 O’Grady PM, Magnacca K, Lapoint RT. 2009. *Drosophila*. In: Gillespie R, Clague D,
720 editors. *Encyclopedia of Islands*. University of California press, Berkeley, CA. p.
721 232–235.

722 Okada T. 1967. A revision of the subgenus *Hirtodrosophila* of the Old World, with
723 descriptions of some new species and subspecies (Diptera, Drosophilidae,
724 *Drosophila*). *Mushi* 41:1–36.

725 Okada T. 1984. The Genus *Collessia* of Japan (Diptera: Drosophilidae). *Proc.*
726 *Japanese Soc. Syst. Zool.* 29:57–58.

727 Okada T. 1988. Family Drosophilidae (Diptera) from the Lund University Ceylon
728 Expedition in 1962 and Borneo collections in 1978-1979. *Entomol. Scand.*
729 30:109–149.

- 730 Okada T. 1989. A Proposal of Establishing Tribes for the Family Drosophilidae with
731 Key to Tribes and Genera (Diptera) : Taxonomy and Systematics. Zool. Sci.
732 6:391–399.
- 733 Otranto D, Stevens JR, Testini G, Cantacessi C, MácA J. 2008. Molecular
734 characterization and phylogenesis of Steganinae (Diptera, Drosophilidae)
735 inferred by the mitochondrial cytochrome c oxidase subunit 1. Med. Vet.
736 Entomol. 22:37–47.
- 737 Peluffo AE, Nuez I, Debat V, Savisaar R, Stern DL, Orgogozo V. 2015. A major
738 locus controls a genital shape difference involved in reproductive isolation
739 between *Drosophila yakuba* and *Drosophila santomea*. G3 Genes, Genomes,
740 Genet. 5:2893–2901.
- 741 Philippe H, Sörhannus U, Baroin A, Perasso R, Gasse F, Adoutte A. 1994.
742 Comparison of molecular and paleontological data in diatoms suggests a major
743 gap in the fossil record. J. Evol. Biol. 7:247–265.
- 744 Prigent SR, Le Gall P, Mbunda SW, Veuille M. 2013. Seasonal and altitudinal
745 structure of drosophilid communities on Mt Oku (Cameroon volcanic line).
746 Comptes Rendus - Geosci. 345:316–326.
- 747 Prigent SR, Suwalski A, Veuille M. 2017. Connecting systematic and ecological
748 studies using DNA barcoding in a population survey of Drosophilidae (Diptera)
749 from Mt Oku (Cameroon). Eur. J. Taxon. 2017.
- 750 Remsen J, O’Grady P. 2002. Phylogeny of Drosophilinae (Diptera: Drosophilidae),
751 with comments on combined analysis and character support. Mol. Phylogenet.
752 Evol. 24.
- 753 Robe L. J., Cordeiro J, Loreto ELS, Valente VLS. 2010a. Taxonomic boundaries,
754 phylogenetic relationships and biogeography of the *Drosophila willistoni*
755 subgroup (Diptera: Drosophilidae). Genetica 138.
- 756 Robe Lizandra J., Loreto ELS, Valente VLS. 2010b. Radiation of the "Drosophila"
757 subgenus (Drosophilidae, Diptera) in the Neotropics. J. Zool. Syst. Evol. Res.
758 48:310–321.
- 759 Robe LJ, Valente VLS, Budnik M, Loreto ÉLS. 2005. Molecular phylogeny of the
760 subgenus *Drosophila* (Diptera, Drosophilidae) with an emphasis on Neotropical
761 species and groups: A nuclear versus mitochondrial gene approach. Mol.
762 Phylogenet. Evol. 36:623–640.
- 763 Robe Lizandra J., Valente VLS, Loreto ELS. 2010c. Phylogenetic relationships and

764 macro-evolutionary patterns within the *Drosophila tripunctata* “radiation”
765 (Diptera: Drosophilidae). *Genetica* 138:725–735.

766 Russo CAM, Mello B, Frazão A, Voloch CM. 2013. Phylogenetic analysis and a time
767 tree for a large drosophilid data set (Diptera: Drosophilidae). *Zool. J. Linn. Soc.*
768 169:765–775.

769 Sackton TB, Lazzaro BP, Schlenke TA, Evans JD, Hultmark D, Clark AG. 2007.
770 Dynamic evolution of the innate immune system in *Drosophila*. *Nat. Genet.*
771 39:1461–1468.

772 Sigurdson T, Green DM. 2011. The origin of modern amphibians: A re-evaluation.
773 *Zool. J. Linn. Soc.* 162:457–469.

774 Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-
775 analysis of large phylogenies. *Bioinformatics* 30:1312–1313.

776 Stark JB, O’Grady PM. 2010. Morphological variation in the forelegs of the hawaiian
777 drosophilidae. I. The AMC clade. *J. Morphol.* 271:86–103.

778 Sturtevant A. H. 1959. Thomas Hunt Morgan. In: A biographical memoir of national
779 academy of sciences. Vol. 33. p. 283–325.

780 Sucena E, Delon I, Jones I, Payre F, Stern DL. 2003. Regulatory evolution of
781 *shavenbaby/ovo* underlies multiple cases of morphological parallelism. *Nature*
782 424:935–938.

783 Sultana F, Hu YG, Toda MJ, Takenaka K, Yafuso M. 2006. Phylogeny and
784 classification of *Colocasiomyia* (Diptera, Drosophilidae), and its evolution of
785 pollination mutualism with aroid plants. *Syst. Entomol.* 31:684–702.

786 Tanaka K, Barmina O, Kopp A. 2009. Distinct developmental mechanisms underlie
787 the evolutionary diversification of *Drosophila* sex combs. *Proc. Natl. Acad. Sci.*
788 U. S. A. 106:4764–4769.

789 Tatarenkov A, Zurovcová M, Ayala FJ. 2001. *Ddc* and *amd* sequences resolve
790 phylogenetic relationships of *Drosophila*. *Mol. Phylogenet. Evol.* 20:321–325.

791 Throckmorton L. 1962. The problem of phylogeny in the genus *Drosophila*. *Univ.*
792 *Texas Publ.* 2:207–343.

793 Throckmorton L. 1975. The phylogeny, ecology and geography of *Drosophila*. In:
794 King R, editor. *Handbook of genetics*. New York. p. 421–469.

795 Throckmorton L. 1982. Pathways of evolution in the genus *Drosophila* and the
796 founding of the repleta group. In: Barker J, Starmer W, editors. *Ecological*
797 *Genetics and Evolution: the Cactus-Yeast-Drosophila Model System*. Academic

798 Press, New York. p. 33–47.

799 Trinder M, Daisley BA, Dube JS, Reid G. 2017. *Drosophila melanogaster* as a high-
800 throughput model for host-microbiota interactions. *Front. Microbiol.* 8:751.

801 Tsacas L. 2002. Le nouveau complexe africain *Drosophila loiciana* et l'espèce
802 apparentée *D. matileana* n. sp. (Diptera: Drosophilidae). *Ann. la Société*
803 *Entomol. Fr.* 38:57–70.

804 Tsacas L, Chassagnard M-T. 1992. Les relations Araceae-Drosophilidae. *Drosophila*
805 *aracea* une espèce anthophile associée à l'aracée *Xanthosoma robustum* au
806 Mexique (Diptera: Drosophilidae). *Ann. la Société Entomol. Fr.* 28:421–439.

807 Ugur B, Chen K, Bellen HJ. 2016. *Drosophila* tools and assays for the study of human
808 diseases. *Dis. Model. Mech.* 9:235–244

809 Wakahama K-I, Shinohara T, Hatsumi M, Uchida S, Kitagawa O. 1983. Metaphase
810 chromosome configuration of the immgrans species group of *Drosophila*.
811 *Japanese J. Genet.* 57:315–326.

812 Wartlick O, Mumcu P, Jülicher F, Gonzalez-Gaitan M. 2011. Understanding
813 morphogenetic growth control - lessons from flies. *Nat. Rev. Mol. Cell Biol.*
814 12:594–604

815 Wheeler MR, Kambysellis MP. 1966. Notes on the Drosophilidae (Diptera) of Samoa.
816 *Univ. Texas Publ.* 6615.

817 Wiegmann BM, Richards S. 2018. Genomes of Diptera. *Curr. Opin. Insect Sci.*
818 25:116–124.

819 Wiegmann BM, Trautwein MD, Winkler IS, Barr NB, Kim J-W, Lambkin C, Bertone
820 M a, Cassel BK, Bayless KM, Heimberg AM, et al. 2011. Episodic radiations in
821 the fly tree of life. *Proc. Natl. Acad. Sci. U. S. A.* 108:5690–5695.

822 Yassin A. 2013. Phylogenetic classification of the Drosophilidae Rondani (Diptera):
823 The role of morphology in the postgenomic era. *Syst. Entomol.* 38:349–364.

824 Yassin A, Debat V, Bastide H, Gidaszewski N, David JR, Pool JE. 2016. Recurrent
825 specialization on a toxic fruit in an island *Drosophila* population. *Proc. Natl.*
826 *Acad. Sci. U. S. A.* 113:4771–4776.

827 Yassin A, Delaney EK, Reddiex AJ, Seher TD, Bastide H, Appleton NC, Lack JB,
828 David JR, Chenoweth SF, Pool JE, et al. 2016. The *pdm3* Locus Is a Hotspot for
829 Recurrent Evolution of Female-Limited Color Dimorphism in *Drosophila*. *Curr.*
830 *Biol.* 26:2412–2422.

831 Yassin A, Da Lage J-L, David JR, Kondo M, Madi-Ravazzi L, Prigent SR, Toda MJ.

832 2010. Polyphyly of the *Zaprionus* genus group (Diptera: Drosophilidae). Mol.
833 Phylogenet. Evol. 55:335–339.

834 Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: Polynomial time
835 species tree reconstruction from partially resolved gene trees. BMC
836 Bioinformatics 19.

837 Zhang W, Toda MJ. 1992. A new species-subgroup of the *Drosophila immigrans*
838 species group (Diptera, Drosophilidae) with description of two new species from
839 China and revision of taxonomic terminology. Japanese J. Entomol. 60:839–850.

840 Zhang Y, Izumitani HF, Katoh TK, Finet C, Toda MJ, Watabe H, Katoh Toru. 2021.
841 Phylogeny and evolution of mycophagy in the *Zygothrica* genus group (Diptera:
842 Drosophilidae). Mol. Phylogenet. Evol. 163: 107257.

843

844 **Figure legends**

845 **Figure 1.** Phylogram of the 704-taxon analyses. IQ-TREE maximum-likelihood
846 analysis was conducted under the GTR+R+FO model. Support values obtained after
847 100 bootstrap replicates are shown for selected supra-group branches, and infra-group
848 branches within the *melanogaster* group (all the support values are shown online).
849 Black dots indicate support values of $PP > 0.9$ and $BP > 90$; grey dots $0.9 \geq PP > 0.75$
850 and $90 \geq BP > 75$; black squares only $BP > 90$; grey squares only $90 \geq BP > 75$.
851 Scale bar indicates the number of changes per site. Groups and subgroups are
852 numbered or abbreviated as follows: (1) *montium*, (2) *takahashii* sgr, (3) *suzukii* sgr,
853 (4) *eugracilis* sgr, (5) *melanogaster* sgr, (6) *ficuspila* sgr, (7) *elegans* sgr, (8)
854 *rhopaloo* sgr, (9) *ananassae*, (10) *Collessia*, (11) *mesophragmatica*, (12) *dreyfusi*,
855 (13), *coffeata*, (14) *canalina*, (15) *nannoptera*, (16) *annulimana*, (17) *flavopilosa*,
856 (18) *flexa*, (19) *angor*, (20) *Dorsilopha*, (21) *ornatifrons*, (22) *histrion*, (23)
857 *macroptera*, (24) *testacea*, (25) *bizonata*, (26) *funnebris*, (27) *Samoia*, (28)
858 *quadrilineata* sgr, (29) *Liodrosophila*, (30) *Hypselothyrea*, (31) *Sphaerogastrella*,
859 (32) *Zygothrica* I, (33) *Paramycodrosophila*, (34) *Hirtodrosophila* III, (35)
860 *Hirtodrosophila* II, (36) *Hirtodrosophila* I, (37) *Dettopsomyia*, (38) *Mulgravea*, (39)
861 *Hirtodrosophila* IV, (40) *Zygothrica* II, *Chy*: *Chymomyza*; *Colo*: *Colocasiomyia*;
862 *Dichae*: *Dichaetophora*; *immigr*: *immigrans*; *Lord*: *Lordiphosa*; *Mic*:
863 *Microdrosophila*; *Myc*: *Mycodrosophila*; *pol*: *polychaeta*; *salt*: *saltans*; *Scap*:
864 *Scaptodrosophila*; *trip*: *tripunctata*; *will*: *willistoni*.

865

866 **Figure 2.** (A) Phylogram of the 83-taxon analyses. The overall matrix represents
867 14,961 nucleotides and 83 taxa, including 63 composite ones. Support values obtained
868 after 100 bootstrap replicates and Bayesian posterior probabilities are shown for
869 selected branches and mapped onto the ML topology (all the support values are
870 shown in Figure S1). The dotted line indicates that the placement of *Dettopsomyia*
871 varies between ML and Bayesian trees. Scale bar indicates the number of changes per
872 site. (B-H) Photos of species of particular interest in this paper. (B) *Drosophila*
873 *oshimai* female (top) and male (bottom) (Japan, courtesy of Japan Drosophila
874 Database), (C-D) *Collessia kirishimana* (Japan, courtesy of Masafumi Inoue), (E-F)
875 *Drosophila annulipes* (Japan, courtesy of Yasuo Hoshino), (G) *Drosophila pruinosa*
876 (São Tomé, courtesy of Stéphane Prigent), (H) *Drosophila adamsi* (Cameroun,
877 courtesy of Stéphane Prigent).

878

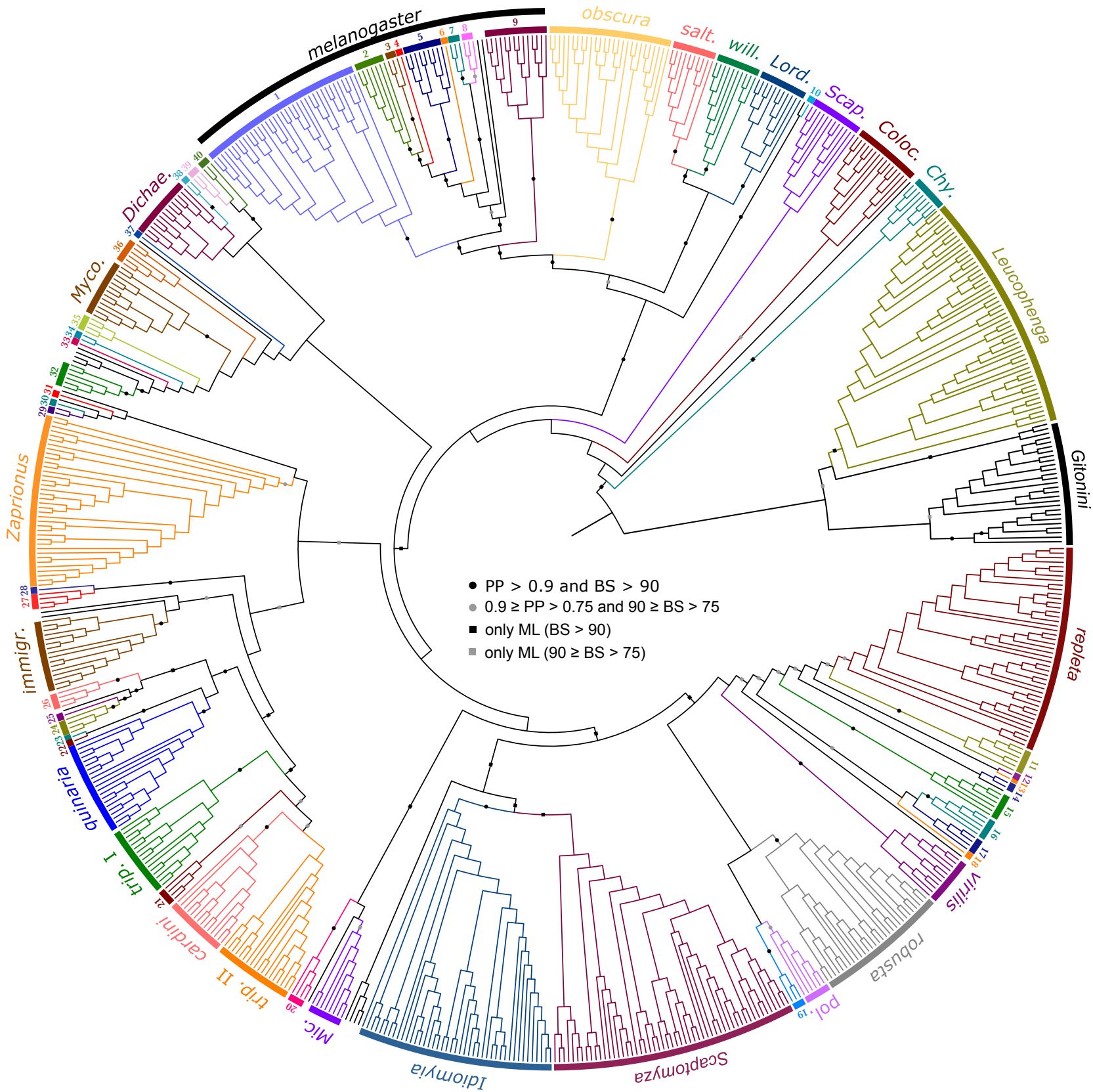
879 **Figure 3.** Concordance *versus* mutational saturation of the phylogenetic markers. The
880 y-axis indicates the percentage of concordant nodes, and the x-axis indicates the
881 saturation level. In comparison with published markers (black dots), the markers
882 developed in this study (orange dots) generally show moderate saturation levels and
883 satisfying concordance.

884

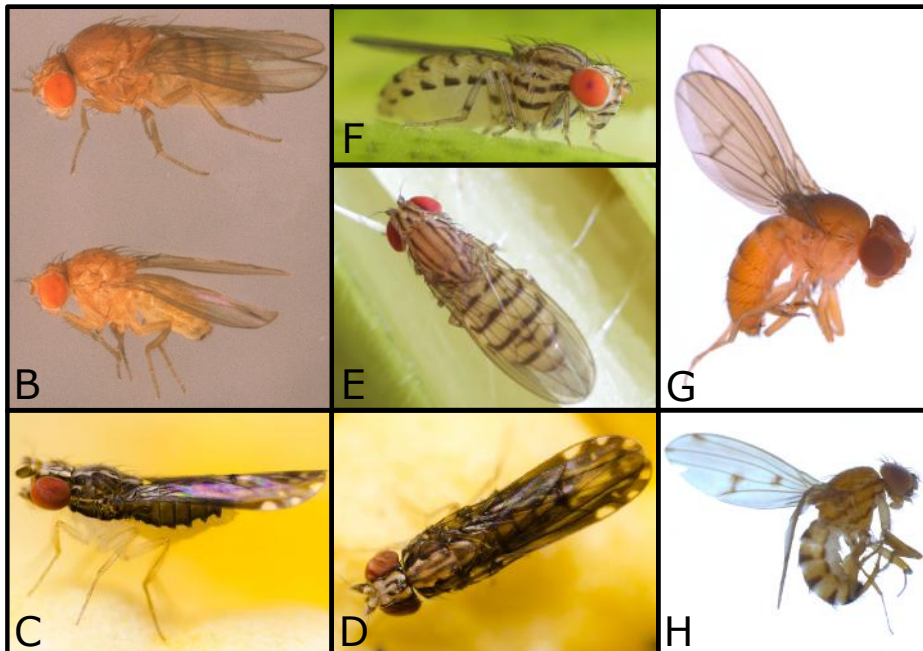
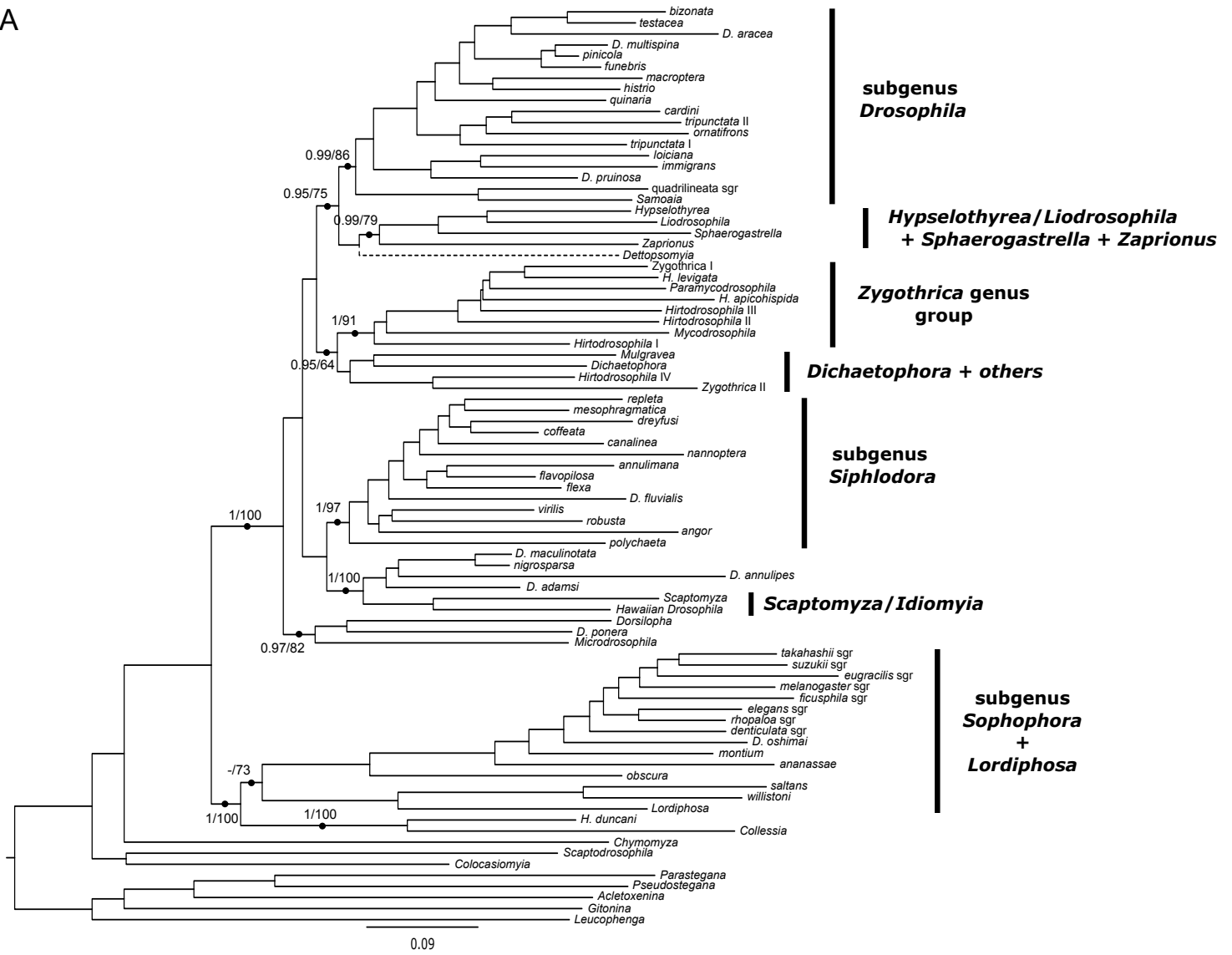
885 **Table legends**

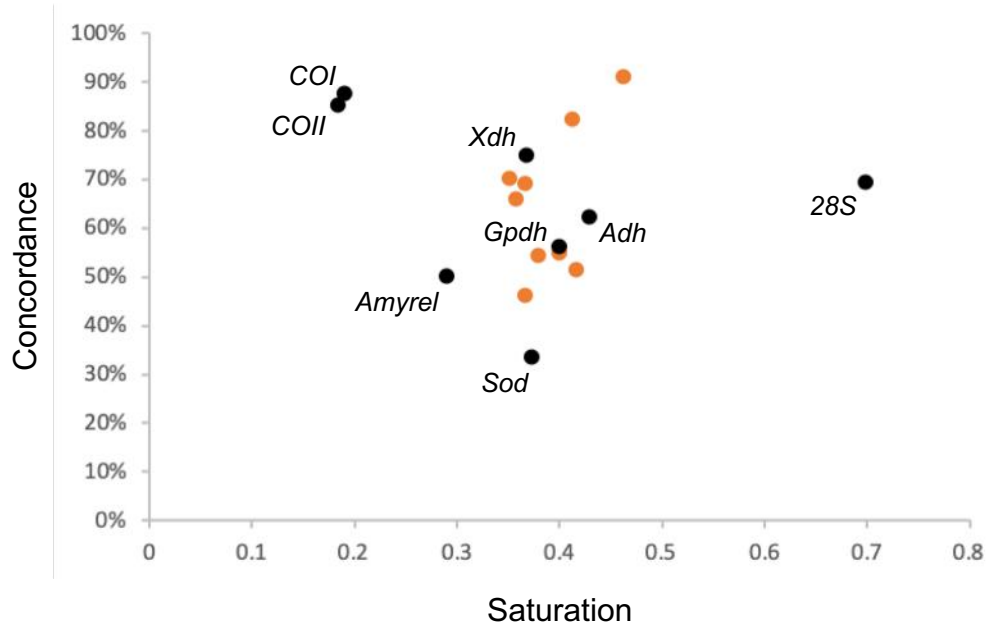
886 **Table 1.** List of PCR primers used in this study.

887 **Table 2.** Dataset statistics.



A





Genomic Locus	Primer	Primer Sequence (5'-3')	Annealing	size	References
<i>Amyrel</i>	zone2bis	GTAAATNGGNNCCACGCGAAG	53°C	1,000 bp	Da Lage et al. (2007)
	relrev+	GTTCCCCAGCTCTGCAGCC			
	reludir	TGGATGCNGCCAAGCACATGGC		1,000 bp	
	relavbis	GCATTTGTACCGTTTGTGTCGTTATCG			
<i>Distal-less</i>	dll-F	TGATACCAATACTGSGGCACATA	56°C	600 bp	this study
	dll-R	ATGATGAARGCMGCTCAGGG			
<i>Dopa decarboxylase</i>	ddc-F	TTCCASGAGTACTCCATGTCCTCG	58°C	1,200 bp	this study
	ddc-R	GGCAGGATGTKATGAAGGACATTGAG			
<i>ebony</i>	eb-F	CCCATSACCTCKGTGGAGCCGTA	59°C	900 bp	this study
	eb-R	CTGCATCGCATCTTYGAGGAGCA			
<i>engrailed</i>	en-F	AATCAGCGCCCAGTCCACCAG	65°C	1,500 bp	this study
	en-R	GCCACATCTCGTTCTTGCCGC			
<i>even-skipped</i>	eve-F	TGCCTVTCCAGTCCRGAYAACTC	55°C	1,000 bp	this study
	eve-R	TACGCCTCAGTCTTG TAGGG			
<i>hedgehog</i>	hh-F	ACCTTG TABARGGCATTGGCATAACCA	56°C	600 bp	this study
	hh-R	ATCGGWGATCGDGTGCTRAGCATG			
<i>Notum</i>	not-F	TGGA ACTAYATHCAYGADATGGGCGG	56°C	800 bp	this study
	not-R	GAGCAGYTCVAGRAADCGCATCTC			
<i>patched</i>	ptc-F1	ACCCAGCTGCGCATSAGRAAGG	54°C	600 bp	this study
	ptc-F2	ACCCAGCTGCGCATSAGRAACG			
	ptc-R	GCTGACGGCSGSTATGCGG			
<i>wingless</i>	wg-F	AGCACGTYCARGCRGAGATGCG	58°C	400 bp	this study
	wg-R	ACTGTTKGGCGAYGGCATRTTGGG			

Name	# sequences	# sites	Informative sites (%)	Inferred distance	Observed distance	saturation	# concording nodes	# missing nodes	Concordance (%)
<i>28S</i>	49/83	848	18.4	0.200	0.189	0.700	25/80	44	69.4
<i>Adh</i>	53/83	724	54.4	0.886	0.331	0.430	28/80	35	62.2
<i>Amyrel</i>	48/83	1475	53.5	2.458	0.545	0.290	18/80	44	50.0
<i>COI</i>	51/83	1438	33.8	1.119	0.666	0.191	35/80	40	87.5
<i>COII</i>	57/83	688	37.8	1.004	0.169	0.185	40/80	33	85.1
<i>Gpdh</i>	26/83	859	35.0	0.784	0.286	0.400	9/80	64	56.3
<i>Sod</i>	22/83	574	49.3	1.072	0.333	0.373	4/80	68	33.3
<i>Xdh</i>	19/83	2088	42.4	0.919	0.314	0.368	9/80	68	75.0
<i>Ddc</i>	52/83	1162	42.3	1.003	0.262	0.358	27/80	39	65.9
<i>Dll</i>	56/83	377	30.8	0.629	0.229	0.463	40/80	36	90.9
<i>eb</i>	67/83	891	46.7	1.247	0.318	0.380	32/80	21	54.2
<i>en</i>	51/83	1119	51.1	1.009	0.307	0.371	18/80	41	46.2
<i>eve</i>	66/83	806	48.6	1.083	0.303	0.367	40/80	22	69.0
<i>hh</i>	63/83	486	62.6	1.203	0.352	0.400	29/80	27	54.7
<i>Notum</i>	51/83	672	62.6	1.005	0.352	0.417	18/80	45	51.4
<i>ptc</i>	60/83	430	55.8	1.076	0.323	0.413	42/80	29	82.4
<i>wg</i>	57/83	324	51.5	1.223	0.321	0.352	33/80	33	70.2

Supplementary Figure and Table Legends

Figure S1. Phylogram of the 204-taxon analysis. IQ-TREE maximum-likelihood analyses were conducted using the GTR+R+FO model. Support values obtained after 100 bootstrap replicates are shown for all branches. Scale bar indicates the number of changes per site.

Figure S2. Phylogram of the 204-taxon analysis. PhyloBayes Bayesian analyses were conducted using the GTR+G model. Bayesian posterior probabilities are shown for all branches. Scale bar indicates the number of changes per site.

Figure S3. Phylogram of the 83-taxon analyses. (Left) IQ-TREE maximum-likelihood analyses were conducted using the GTR+R+FO model. Support values obtained after 100 bootstrap replicates are shown for all branches. Scale bar indicates the number of changes per site. (Right) PhyloBayes Bayesian analyses were conducted using the GTR+G model. Bayesian posterior probabilities are shown for all branches. Scale bar indicates the number of changes per site.

Figure S4. Comparison of support values between the non-composite and composite maximum-likelihood trees. All support values were obtained after 100 bootstrap replicates. The first value refers to the composite approach (83 taxa), and the second value in parentheses refers to the non-composite approach (704 taxa).

Figure S5. Phylogram of the 83-taxon ASTRAL analysis. Branch support values measure the support for a quadripartition (the four cluster around a branch) and not the bipartition, as is commonly done. Scale bar indicates the number of changes per site.

Figure S6. The impact of marker sampling on the tree topology. The composite tree was built on 17 different datasets that correspond to the whole dataset minus one marker sequentially removed. The changes in relation to the ML composite tree depicted in Figure 2 are shown in red. Scale bar indicates the number of changes per site.

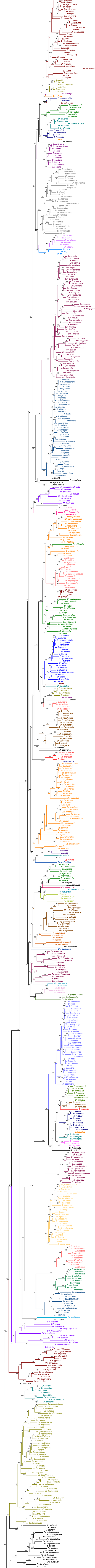
Figure S7. Mutational saturation of the 17 phylogenetic markers. The x-axis indicates the distance inferred from the ML composite tree, whereas the y-axis indicates the observed distance between two taxa. The slope of the red line is an indicator of the saturation level, low values meaning high saturation. The black line corresponds to the absence of multiple substitutions.

Figure S8. Phylogram of the Steganinae subfamily. This ML tree was built on a dataset that includes 164 steganine taxa. IQ-TREE maximum-likelihood analysis was conducted under the GTR+R+FO model. Support values obtained after 100 bootstrap replicates are shown for selected branches (all the support values are available online). Scale bar indicates the number of changes per site.

Figure S9. Addition of missing taxa with scarce genomic data to the composite tree. We added the published sequences of the genera *Jeannelopsis*, *Lissocephala*, *Neotanygastrella*, *Phorticella*, *Styloptera* (Yassin 2013), the subgenus *Dudaica* (Kato et al. 2018), and several *Hirtodrosophila* and *Zygothrica* species (Gautério et al. 2020) to our 83-taxon composite dataset to draw a more comprehensive picture of the Drosophilinae, especially the tribe *Colocasiomyini*.

Table S1. Taxon sampling and presence/absence of markers per taxon. Markers generated in this study are indicated in black, markers retrieved from GenBank are indicated in grey, missing data are indicated in white.

Figure S1



0.10

Figure S2

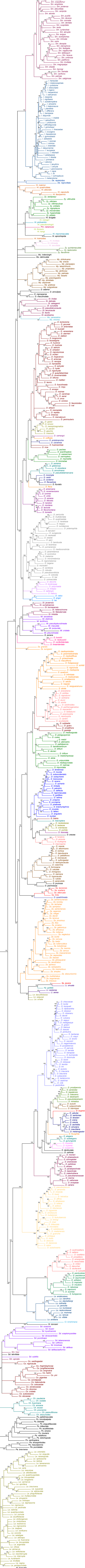
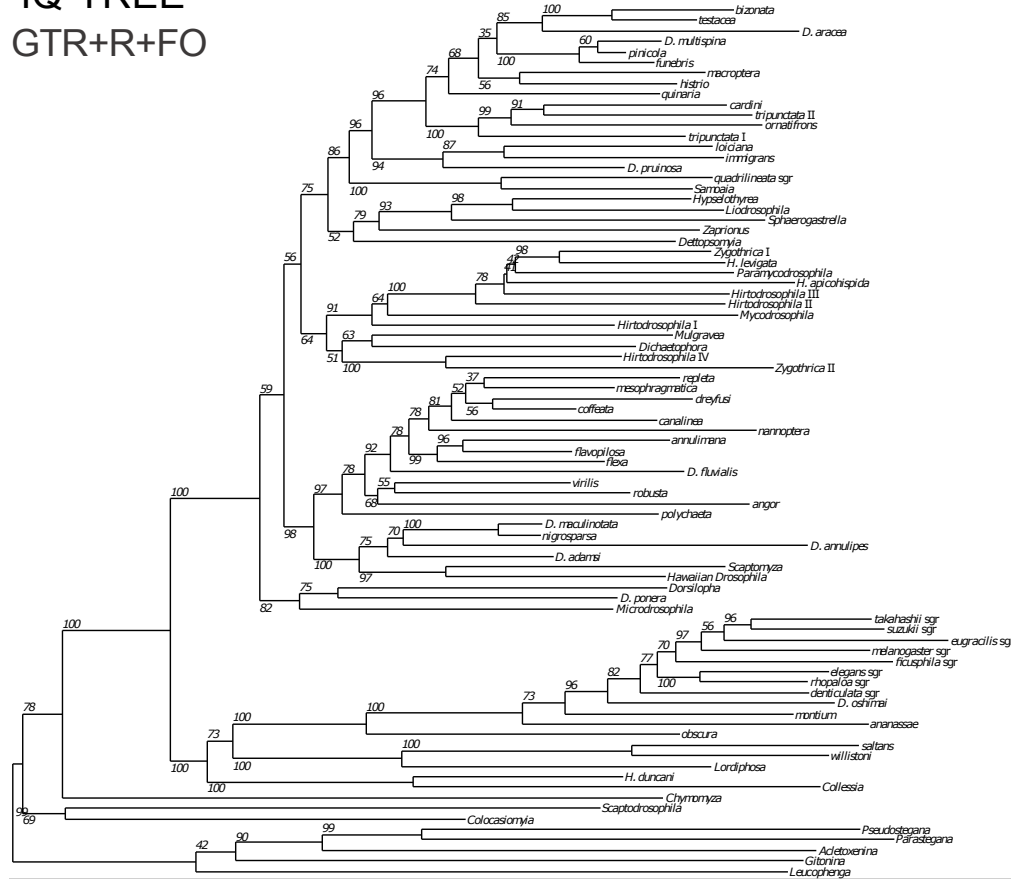


Figure S3

IQ-TREE
GTR+R+FO



PhyloBayes
GTR+G

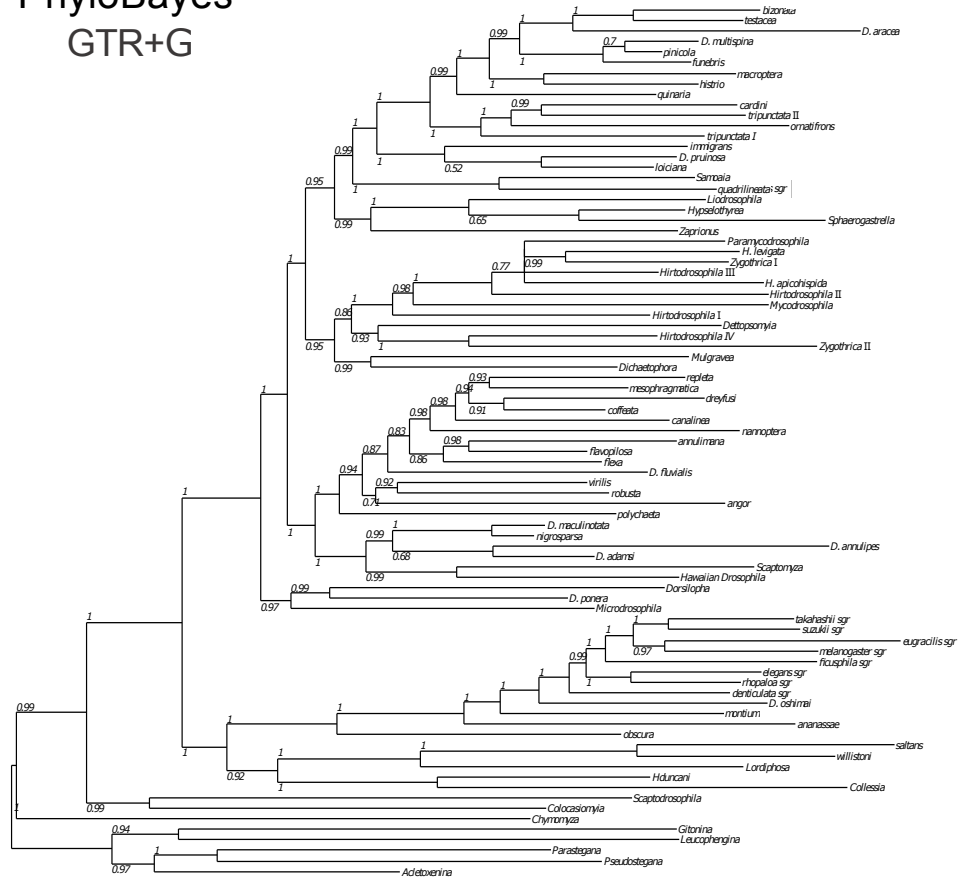


Figure S4

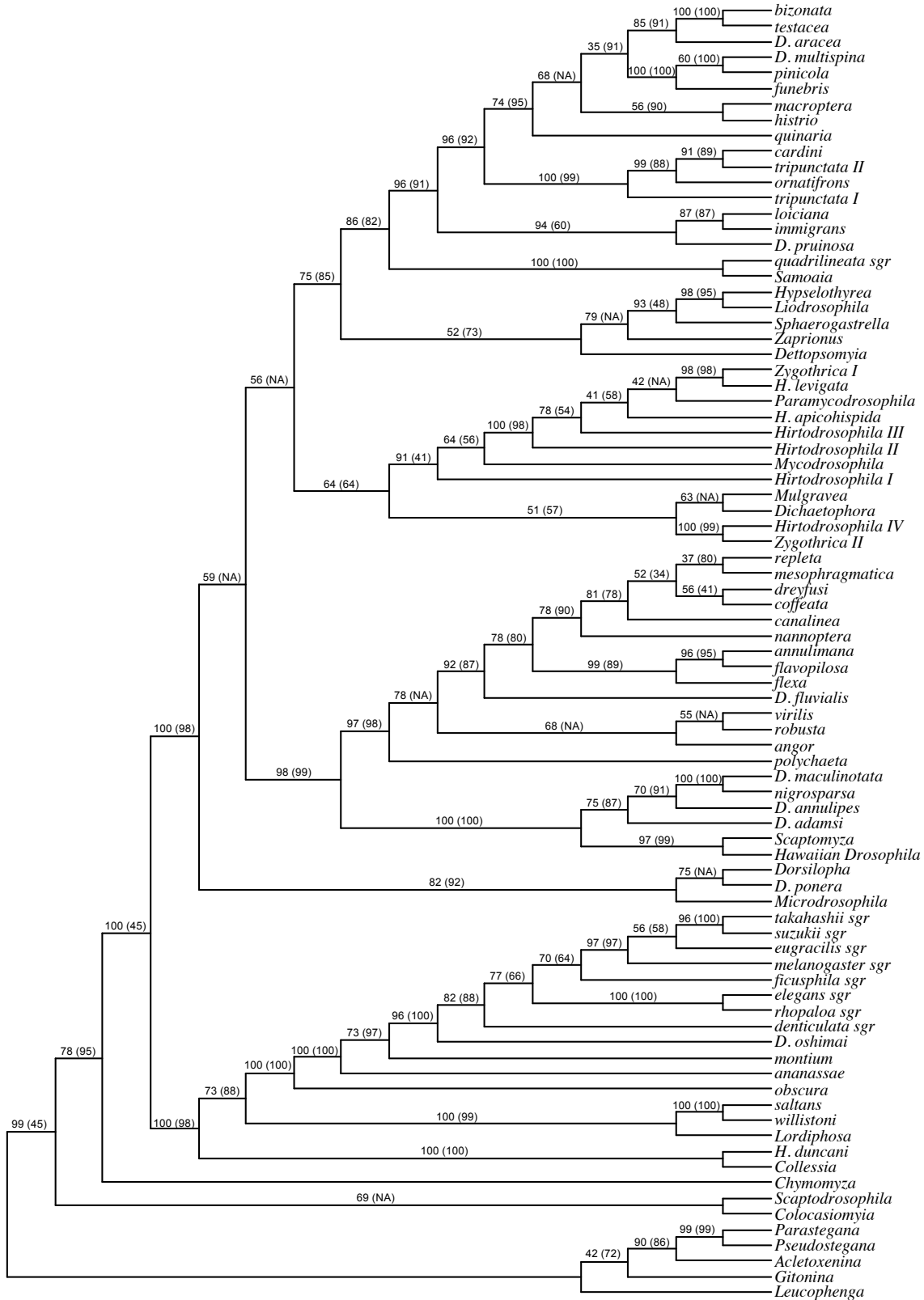


Figure S5

ASTRAL

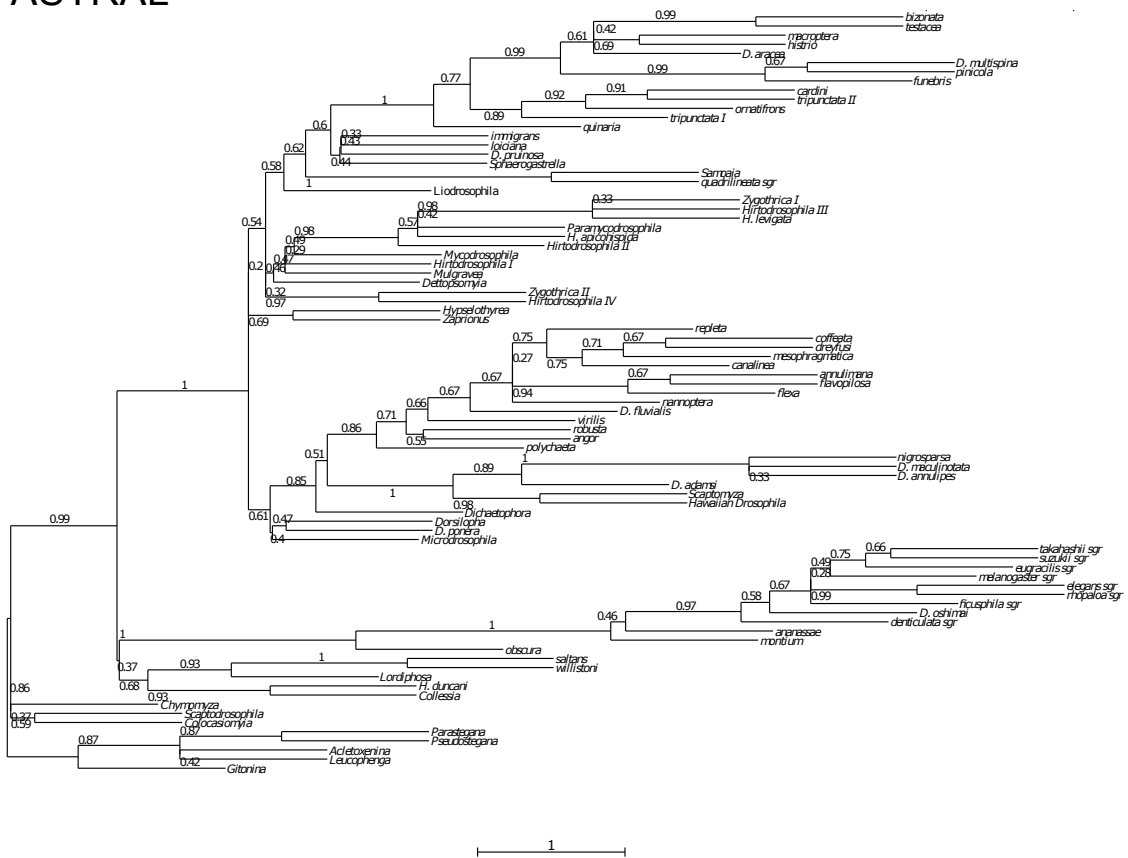


Figure S6

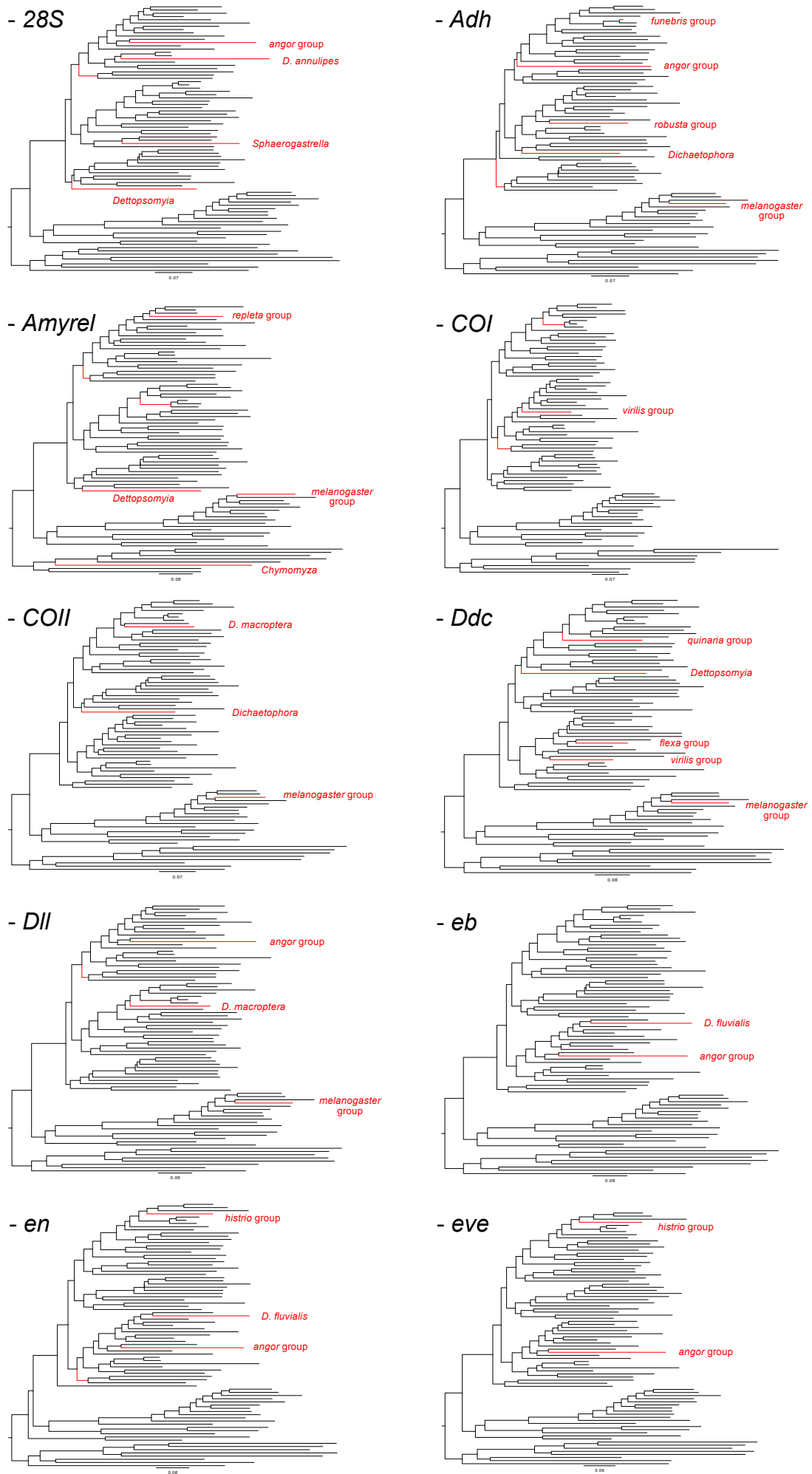


Figure S7

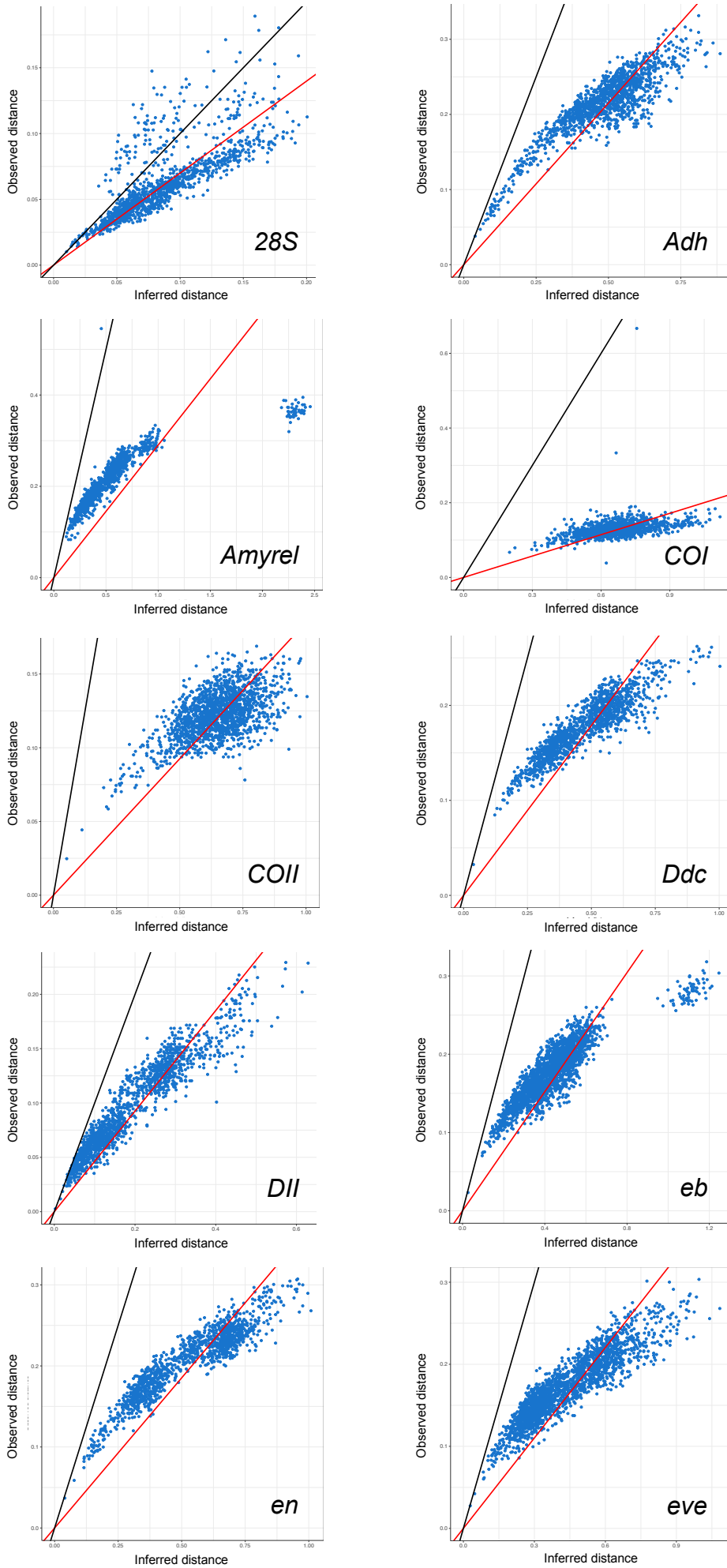


Figure S9

