bioRxiv preprint doi: https://doi.org/10.1101/2021.05.28.446102; this version posted May 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

# Revision of the genus *Dichaetophora* Duda (Diptera: Drosophilidae), part I: DNA bar-coding and molecular phylogenetic reconstruction

Takehiro K. Katoh[1,2], Ji-Min Chen[2], Jin-Hua Yang[2], Guang Zhang[2], Lu Wang[2], Awit Suwito[3], Masanori J. Toda[4,*], Ya-Ping Zhang[2,5,*], Jian-Jun Gao[1,2,*]

[1]Yunnan Key Laboratory of Plant Reproductive Adaptation and Evolutionary Ecology, Yunnan University, Kunming, Yunnan 650091, China

[2]State Key Laboratory for Conservation and Utilization of Bioresources in Yunnan, Yunnan University, Kunming, Yunnan 650091, China

[3]Zoological Division (Museum Zoologicum Bogoriense), Research Center for Biology-LIPI, Jl. Raya Jakarta-Bogor Km. 46, Cibinong 16911, West Java, Indonesia

[4]Hokkaido University Museum, Hokkaido University, Nishi 8, Kita 10, Kita-ku, Sapporo 060-0810, Japan

[5]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Kunming, Yunnan 650223, China

*Corresponding authors: gaojj@ynu.edu.cn, zhangyp@mail.kiz.ac.cn, hutian@lemon.plala.or.jp

**Abstract.** The genus *Dichaetophora* Duda is of 69 formally described, Old World species assigned into five species groups, i.e., *agbo*, *tenuicauda*, *acutissima*, *sinensis* and *trilobita*. Most of these species were delimitated morphologically, with the within-genus relationship established largely via cladistic analyses of morphological characters. In the present study, we first conducted species-delimitation with aids of morphological data as well DNA barcodes (nucleotide sequences of the mitochondrial *COI*, i.e., cytochrome *c* oxidase subunit I, gene), for a huge sample of *Dichaetophora* and allied taxa (genus *Mulgravea* and subgenus *Dudaica* of *Drosophila*) collected from a wide geographical range. Then, multiple-locus phylogenetic reconstruction was conducted based on elaborate taxon sampling from the known and newly recognized species in the above taxa, with the maximum likelihood (ML) and Bayesian inference (BI) methods. As a result, 189 species (186 of *Dichaetophora*, 2 of *Mulgravea*, and 1 for *Dudaica*) were newly recognized. In our ML and BI trees, several well-supported species clusters equivalent to the species groups *agbo* (excluding of *neocirricauda*), *tenuicauda*, *sinensis* (inclusive of *neocirricauda*) and *trilobita* of *Dichaetophora*, were recovered, with the sister-relationship between the third and fourth proved. Other well-supported clusters include 1) a clade comprising of *Di. acutissima* group and *Dudaica*, with the former proved to be paraphyletic to the latter; 2) genus *Mulgravea*; 3) a clade comprising exclusively of newly recognized *Dichaetophora* species, and was placed as sister to *Mulgravea*. Three of the remaining five representatives of *Dichaetophora* species form a solid cluster, leaving the positions of the last two unresolved. The present study greatly renewed out knowledge about the species diversity in a pan-*Dichaetophora* clade, providing us with an unprecedented historical framework for further taxonomy revision of this clade, and valuable baseline knowledge for future reconstruction of the history of its adaptive diversification in the particular microhabitats.

**Key words:** *Dichaetophora*, DNA bar-coding, Drosophilidae, molecular phylogenetic reconstruction, taxonomy, species delimitation

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.28.446102; this version posted May 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## Introduction

Duda (1940) established *Dichaetophora* as a subgenus in the genus *Drosophila* Fallén (Drosophilidae: Drosophilinae) with *Dr. aberrans* Lamb 1914 from Seychelles as the type species. After that, a number of species were added to this subgenus from Africa (3 spp.; Burla 1954; Graber, 1957) and East Asia (6 spp.; Lee, 1964, Okada, 1965, 1966, 1968; Kang *et al.* 1967). These East Asian species, however, were later transferred to the genus *Nesiodrosophila* Wheeler & Takada (Okada, 1976, 1977, 1984a), which was established with *Ne. lindae* Wheeler & Takada as the type species by Wheeler & Takada (1964). Then, a large number of *Nesiodrosophila* species were found from the Old World: 14 spp. from the Oriental region (Lin & Ting, 1971; Okada, 1984a, 1988; Gupta & De, 1996), 3 spp. from the Palearctic region (Nishiharu, 1981; Toda, 1989), 15 spp. from the Australasian region (Bock, 1982; Okada, 1984a; Toda *et al*., 1987), and 1 sp. from the Afrotropical region (Okada, 1984a). In addition, another related taxonomic group, i.e., the *Drosophila tenuicauda* species group, was recognized within the subgenus *Lordiphosa* Basden of *Drosophila* (Toda, 1983; Okada, 1984b; Hu *et al*., 1999; Katoh *et al*., 2000): Hu & Toda (2001) inferred the sister relationship between the *tenuicauda* group and *Nesiodrosophila* from a phylogenetic analysis based on 68 morphological characters. Grimaldi (1990) elevated *Dichaetophora* and *Lordiphosa*, along with the subgenera *Hirtodrosophila* Duda and *Scaptodrosophila* Duda of *Drosophila*, to the generic rank, based on the results of a family-wide cladistic analysis on 217 adult morphological characters.

Taking into account these possible relationships proposed in the previous studies, Hu & Toda (2002) conducted a morphological cladistic analysis focusing on *Dichaetophora*, *Nesiodrosophila* and the *tenuicauda* group. Based on the result that these three taxa form a monophyletic taxon, all the species pertaining to them were merged into a revised *Dichaetophora*, within which three species groups were newly proposed: *agbo*, *tenuicauda* and *acutissima*. The first comprises of the four Afrotropical *Dichaetophora* species and all the species assigned to *Nesiodrosophila* by then), the latter two as a result of splitting of the previous *Lo. tenuicauda* group. Since then, two additional species groups were added to *Dichaetophora*: the *sinensis* group comprised of four Chinese species (Hu & Toda, 2005) and the *trilobita* group of six Oriental species (Yang *et al*., 2017).

The current genus *Dichaetophora* includes a total of 67 formally described species: 43 spp. of the *agbo* group, 10 spp. of the *tenuicauda* group, 4 spp. of the *acutissima* group, 4 spp. of the *sinensis* group, and 6 spp. of the *trilobita* group; 5 spp. distributed in the Palearctic region (East Asia), 6 spp. in the Palearctic (East Asia) and Oriental regions, 35 spp. in the Oriental region, 15 spp. in the Australasian region, 5 spp. in the Afrotropical region, and 1 sp. in the Palearctic (East Asia), Oriental and Australasian regions (DrosWLD-Species: https://bioinfo.museum.hokudai.ac.jp/db/index.php; TaxoDros: http://www.taxodros.uzh.ch/). Our intensive surveys of drosophilid faunas in the Oriental region during the past two decades uncovered very high species diversity of drosophilid flies from previously less explored microhabitats. For *Dichaetophora*, an unexpectedly large number of putatively new species (*ca*. 150) were recognized among specimens collected by net sweeping mostly from herbaceous stands or forest floor, occasionally from tree-trunks, flowers, fallen fruits and fungi, or by light traps.

Grimaldi (1990) included two species, *aberrans* and *rotundicornis* (Okada), of the current *Dichaetophora* in his morphological cladistic analysis under an extensive taxon-sampling of most genera and subgenera of the family Drosophilidae. However, the two species were placed in different lineages distant from each other in the resulting tree. When

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.28.446102; this version posted May 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Hu & Toda (2002) redefined the genus *Dichaetophora*, they suggested its relationships with the genera *Jeannelopsis* Séguy, *Sphaerogastrella* Duda, *Mulgravea* Bock and *Liodrosophila* Duda because an important diagnostic character "the oviscapt with apical ovisensillum robust and the largest, distinguishable from the others" for *Dichaetophora* is shared as a synapomorphy (ap. 213) of Grimaldi (1990) by these genera. This character is shared by the subgenus *Dudaica* Strand of *Drosophila* as well (Katoh *et al*., 2018). On the other hand, Hu & Toda (2005) suggested the sister relationship between *Hirtodrosophila* and the monophyletic *Dichaetophora* comprised of the *agbo*, *tenuicauda*, *acutissima* and *sinensis* groups. The close relationship between *Dichaetophora* (the *tenuicauda* and *acutissima* groups) and *Hirtodrosophila* was suggested in molecular phylogenetic analyses by Katoh *et al*. (2000) and Russo *et al*. (2013) as well. Yassin (2013) constructed a family-wide Bayesian phylogenetic tree, based on a multi-locus (seven nuclear and one mitochondrial genes) dataset of DNA sequences from 190 species of 33 drosophilid genera, and inferred that *Dichaetophora* (the *agbo*, *tenuicauda* and *acutissima* groups) formed a cluster with the genera *Hirtodrosophila*, *Mycodrosophila*, *Zygothrica* Wiedemann, *Dettopsomyia* Lamb and *Jeannelopsis*. However, the statistical supports (Bootstrap, Bremer and/or Posterior Probability values) for these relationships were all low. Thus, the phylogenetic position of *Dichaetophora* and the relationship within this genus are still to be investigated, especially by means of molecular phylogenetic methods.

In the present study, we first conduct a species delimitation in *Dichaetophora* and its relatives based on morphological and DNA barcode data, employing a huge amount of specimens of known and putatively new species of these taxa collected by our surveys for ten-odd years in the Oriental and East Asian regions. We then conduct a multi-locus molecular phylogenetic analysis by sampling as many species as possible from the refined members within *Dichaetophora* and possibly allied genera/subgenera such as those mentioned above, with "peripheral" out-group species representing some major lineages within the subfamily Drosophilinae. The phylogeny to be reconstructed in this study will provide a framework for revising the taxonomy of *Dichaetophora* and allied taxa in a subsequent study of this serial work and baseline knowledge for future evolutionary studies on this speciose clade adapted to particular microhabitats.

## Materials and methods

### *Fly samples and species delimitation*

Before phylogenetic reconstruction, species delimitation was conducted based on morphological and DNA bar-code data with a large, pooled sample of drosophilid flies preliminarily recognized as of genera *Dichaetophora*, *Mulgravea* and subgenus *Dudaica* of genus *Drosophila*. All these were collected from China and adjacent countries (Japan, Malaysia, Indonesia, Vietnam, and Myanmar) or even Australia (Table S1), mostly by net sweeping above herbaceous-layer along forest edges or in the forests, sometimes by aspirating from flowers or by using light traps set in the tree canopy. Specimens were preserved either in alcohol (70% or 100%, for morphological or molecular study, respectively), or Kahle's solution (to maintain the pigmentation of specimens for long time in the laboratory; for specimens collected from Myanmar and Australia only) before this study.

The specimens were first identified as morpho-species of *Dichaetophora* (assigned into species groups as far as possible, referring to Hu & Toda 2005's and Yang et al. 2017's diagnoses), *Mulgravea* or *Drosophila* (*Dudaica*). For this, the genitalia and/or some other

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.28.446102; this version posted May 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

body parts (e.g., head, leg, and mouthparts, etc.) of representative specimens were separated from the main body and examined under microscopes. Some of the specimens thus identified were selected and subjected to DNA barcode sequencing, considering the total number, gender and geographical origin of the available specimens of each morpho-species. We used the same methods as in Yang et al. (2017) for fly tissue sampling, DNA extraction (using TIANamp® Genome DNA Kit) and PCR [using TIANGEN® *Taq* DNA polymerase and Folmer et al.'s (1994) primer pair LCO1490/HCO2198]. The PCR products were subjected to sequencing in the TSINGKE Biological Technology (http://www.tsingke.net) with an ABI 3700 sequencer, with trace files edited subsequently in the SeqMan module of the DNAStar package version 7.1.0 (DNAStar Inc., Madison, WI).

Newly determined barcodes were aligned with 78 *COI* sequences downloaded from GenBank (Table S1) in MEGA 7 (Kumar et al. 2016) with the ClustalW method. A neighbor-joining (NJ) tree was then built with the resulting sequence alignment in MEGA 7 (options: model = *p*-distance, variance estimation method = bootstrap method of 1,000 replicates, gaps treatment = pair-wise deletion). A primary species delimitation was then conducted using the ABGD (automatic barcode gap discovery) algorithm (Puillandre et al. 2012) for MOTU (molecular operational taxonomic unit) recognition through the web interface (http://wwwabi.snv.jussieu.fr/public/abgd/abgdweb.html), with the "simple distance" (i.e., *p*-distance) option under default settings (Pmin = 0.001, Pmax = 0.1, Steps = 10, X (a proxy for minimum gap width) = 1.5, Nb bins (for distance distribution) = 20). The species delimitation was finalized then by synthesizing the morphological and DNA sequence evidences.

*Taxon sampling for phylogenetic reconstruction*

We reconstructed the phylogenetic relationship of *Dichaetophora*, and reexamine its relationship to numbers of allied taxa, including *Mulgravea* and *Dudaica*, treating *Dichaetophora*, *Mulgravea*, and *Dudaica* as tentative in-group taxa considering the peculiar morphological like among these three taxa (all referring to as "pan-*Dichaetophora*" taxa hereinafter), and the remaining ones involved as out-groups. For the in-group taxa, either known or newly recognized species were sampled considering respective representativeness of respective species clusters in the tree of barcodes, and of respective species groups prejudged in light of morphological features (Table 1). For out-groups, a total of 37 species were sampled from 13 genera in the subfamily Drosophilinae [e.g., *Scaptodrosophila*, *Chymomyza*, *Impatiophila*, *Drosophila* (varied subgenera), etc. (Table 1, Table S2)].

*Genetic markers and DNA sequencing*

We adopted the D7 region of the *28S rRNA* locus (Friedrich & Tautz 1997) as a practicable marker, and tried to set up some additional, single-copy, orthologous, protein-coding gene (PCG) markers, referring to the alignments of our unpublished transcriptome sequences of ten drosophilid species (three of *Dichaetophora* and the remaining of allied genera). The PCG markers were evaluated considering sequence variation within the coding part of the target regions of the alignments.

We then designed PCR/sequencing primer pairs for each of these markers, chose optimal reaction condition for each pair, and evaluated their performance with DNA samples of eight *Dichaetophora* species (*acutissima*, *pseudocyanea*, *tenuicauda*, *facilis*, *ogasawarensis*, *lindae*, *neocirricauda*, and *trilobita*) covering all the five species groups in this genus. Then, the target regions of the selected nuclear markers as well the *28S rRNA* marker were amplified and sequenced for the in- and out-group species with the designed primer pairs (with the same

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.28.446102; this version posted May 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

procedures described in the DNA bar-coding section, in the TSINGKE Biological Technology), and the resulting trace files edited with SeqMan or MEGA7. The protein-coding part(s) of the newly collected sequences of each locus were aligned with some homologous sequences available from GenBank or the Ehime-fly (*COI* sequence only) using the ClustalW method, and then concatenated for all loci, also in MEGA 7.

*Data partitioning and multiple-locus phylogenetic reconstruction*

We used the program PartitionFinder 2 (Lanfear et al. 2016) to search optimal partitioning scheme for the alignment of the concatenated sequences, and select the best fit nucleotide substation model for each suggested partition. The search was conducted with the greedy algorithm (Lanfere et al. 2012) under the BIC (Bayesian Information Criterion; Schwarz 1978), with data blocks defined in light of gene locus and each codon position for the PCG sequences. The model was set as "models=all" for the ML, but "model=mrbayes" for the BI analyses. The ML tree was constructed using RAxML 1.0.1 (Kozlov et al. 2019), with node confidences evaluated through 1,000 non-parametric bootstrap replicates. The BI analysis was performed in MrBayes 3.2.7 (Ronquist et al. 2012) through two independent runs each includes four MCMC chains, with chains sampled every 100 generations. The convergence of runs was evaluated in Tracer V1.6 (Rambaut et al. 2014) after discarding the initial 25% samples as burn-in.

## Results

*Species delimitation*

We finally determined *COI* sequences for 1,013 specimens of pan-*Dichaetophora* taxa (1,004 of *Dichaetophora*, 5 of *Mulgravea*, and 4 of *Dudaica*, Table S1). The NJ barcode tree built with the alignment of these barcodes and 57 GenBank *COI* sequences (23 of *Di. trilobita* group, 34 of *Dudaica*, Table S1) is shown in Fig. 1 (partially compressed) and Fig. S1 (intact), with bootstrap supported < 50 not shown. The clustering pattern of the sequences is largely in line with the morphological link among the corresponding specimens, though numbers of cryptic species are newly prompted. In our ABGD analysis with all but one (#00676, only 212 nucleotide sites in length) of the *COI* sequences, a total of 177 to 311 MOTUs were recognized within the prior maximum intraspecific divergence (*P*) of 0.001000–0.021544 (Table 1). $P = 0.004642$ was then selected as an optimal considering the morphological concordance (i.e., known or newly recognized morpho-species were well resolved in this P value) and the bar-coding gap of 0.04−0.05 indicating by the distribution of pairwise *p*-distances (Fig. 2). With $P = 0.004642$, the numbers of MOTUs identified in the initial and recursive partitions were 222 and 226, respectively. The former (i.e. 222) was taken as optimal number of hypothetical species based on stability (Table 1). The hypothetical species suggested by ABGD was largely in agreement with the pattern of sequence clustering in the NJ tree (Fig. 1). We thus adopted all the species delimitation in ABGD analysis, except for that: two morphologically well supported species, *Di.* sp.DLS3a and *Di.* sp.DLS3b, were accepted, though ABGD analysis failed to distinguish between them. As a result, a total of 223 pan-*Dichaetophora* species (34 known and 168 new) were recognized (Fig. 1, Fig. S2, Table 2, and Table S1). The final result of the species delimitation is summarized group by group in the following.

1) *Di. agbo* group (Fig. 1A−F,J): A total of 123 species were newly recognized in addition to the seven known, representative species, including *Di. neocirricauda*, which was

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.28.446102; this version posted May 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

placed as most closely related to members of the *Di. sinensis* group, i.e., *Di.* sp.1 aff. *hainanensis* (BP = 99), and then to *Di.* sp.2 aff. *hainanensis* (BP = 81) (Fig. 1F). Though *Di.* sp.PE6b was weakly suggested as being paraphyletic with respect to *Di.* sp.PE6c representing by a single sequence of female specimen (#06262; Fig. 1B, Fig. S1). The two forms can be readily distinguished from each other in light of feature in female terminalia: the apical, robust ovisensillum and the subapical, prominent lateral ovisensillum acute apically in *Di.* sp.PE6b, but blunt in *Di.* sp.PE6c (not shown).

2) *Di. tenuicauda* group (Fig. 1G,H): Thirty two species were newly recognized. The two forms, *Di.* sp.DLS3a and *Di.* sp.DLS3b, are accepted considering the distinct differentiation between them in male (but not female) genitalia, for example, aedeagus and aedeagal apodeme (Fig. 1H). Our NJ tree and ABGD analysis failed to distinguish between these two forms. In contrast, distinct divergence was observed between them in light of the nuclear gene markers used in our phylogenetic reconstruction (see below for detail). The "paraphyletic" status of *Di.* sp.Y1 (Fig. S1, Fig. 1H) is attributable to the short length (212 nucleotides) of the sequence #00676". The *p*-distance between the two sets of the sequences (Part 1 *vs*. Part 2, *p*-distance = 0.00077 ± 0.00033) is rather inconspicuous.

3) *Di. acutissima* group (Fig. 1E,F): Fourteen species were newly recognized. All members of this group were placed into two well supported clusters, one (*Di. acutissima* + *Di. harpophallata* + seven newly recognized species, BP = 93), the other is of *Di. cyanea*, *Di. pseudocyanea* and four newly recognized species (BP = 97).

4) *Di. sinensis* group (Fig. 1J): Four species were newly recognized in addition to *Di. sinensis*. The *Di.* sp.PE12 and *Di. sinensis* were suggested as sister to each other (BP = 99). Among the remaining three species, *Di.* sp.1 aff. *hainanensis* form a pair with *Di. neocirricauda* of the *Di. agbo* group (BP = 99), both further form a cluster with *Di.* sp.2 aff. *hainanensis* (BP = 81). The position of *Di.* sp. aff. *abnormis* showing highly concordant morphological concordance (except trumpet-like dilation of aedeagus) to the known members in the *sinensis* group, is poorly resolved.

5) There are 12 newly recognized *Dichaetophora* species not yet assigned into any existing species group in light of morphological or DNA barcode data. These are "*Di.* sp.2a *acutissima* group" (Fig. 1I), "*Di.* sp.2b *acutissima* group" (Fig. 1J), *Di.* sp.WTS2 (Fig. 1I), *Di.* sp.17 (Fig. 1I), *Di.* sp.XM1a (Fig. 1I), *Di.* sp.XM1b (Fig. 1I), *Di.* sp.18 (Fig. 1I), *Di.* sp.JC1 (Fig. 1F), "*Di.* sp.HM1 *tenuicauda* group" (Fig. 1F), "*Di.* sp.HM2 *tenuicauda* group" (Fig. 1E), "*Di.* sp. 8 Malaysia" (Fig. 1E) and "*Di.* sp.6" (Fig. 1J). The first six of these (referring to as "*Dichaetophora* Part 6" hereinafter) forms share some morphological features, and so do the latter five (referring to as "*Dichaetophora* Part 7" hereinafter).

6) The two new species (*Mu.* sp.a aff. *detriculata* and *Mu.* sp.b aff. *detriculata*) recognized morphologically as members of *Mulgravea* form a pair (BP = 80) (Fig. 1F), both were placed as close to a known species of the same genus, *Mu. detriculata* (BP < 50). One new species [*Dr.* (*Du.*). sp. aff. *qiongzhouensis*] morphologically recognized as of *Dudaica* was placed as sister to the known species *Dr.* (*Du.*) *dissimilis* (BP = 99), both forms a cluster with three other known species of this subgenus (BP = 64; Fig. 1E).

*Phylogenetic reconstruction*

We finally chose 11 nuclear, PCG makers of desirable performance (*AdSS*, *ATPsynB*, *bur*, *ced-6*, *eIF3-S8*, *Pdi*, *Pgi*, *RpL3*, *RpS17*, *sina* and *VhaSFD*; Table S2). DNA sequences were determined with the primer pairs of all the 12 nuclear genes, for 105 species [82 of *Dichaetophora*, three of *Dr.* (*Dudaica*), four of *Mulgravea*, and the remaining 16 of the

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.28.446102; this version posted May 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

out-group taxa]. The alignment of the data sets of the newly determined and the "extracted" sequences of 21 other species spans 4,571 nucleotide sites, among which 2,005 are variable, 1,710 are parsimony informative. The optimal partitioning strategy for the concatenated data set is shown in Table 4, together with the selected nucleotide substitution model for each partition. The resulting ML tree is rooted at the midpoint of the branch connecting the "*Scaptodrosophila* + *Colocasiomyia* + *Chymomyza* + *Impatiophila*" cluster and the collection of the remaining taxa (Fig. 3).

The ML and Bayesian trees lend relatively strong support to the monophyly of the "pan-*Dichaetophora*" assemblage (BP = 67, PP = 1.00). There are several more or less strongly supported clusters of the out-group species: that of Hawaiian *Drosophila* + *Scaptomyza* (BP = 100, PP = 1.00), that of *Li. aerea* + *Hy. guttata* + *Za. indianus* (BP = 41, PP = 1.00), that of *Dr.* (*Dr.*) *albomicans* + *Dr.* (*Do.*) *busckii* (BP = 61, PP = 1.00), the subgenus *Siphlodora* (BP = 100, PP = 1.00) and the subgenus *Sophophora* (BP = 100, PP = 1.00). The *Zygothrica* genus group (*Hirtodrosophila* + *Zygothrica* + *Mycodrosophila*) is moderately supported (BP = 59, PP = 0.74), and put as closest to the pan-*Dichaetophora* assemblage among the out-group taxa (BP = 28, PP = 0.63).

A total of seven major species-clusters are recovered within the pan-*Dichaetophora* assemblage (Fig. 3). The first comprises of all the representative species of the *Di. agbo* group, except for *Di. neocirricauda*. Two sub-clusters are found within this cluster, one comprising of *Di. ogasawarensis*, *Di. lindae*, *Di. delicata* and 13 newly recognized one (BP = 63, PP = 0.99), the other comprising of *Di. sakagamii* and 23 newly recognized one (BP = 79, PP = 1.00). The relationship within either of the two sub-clusters is well-resolved.

The second species-cluster is of the *Di. acutissima* group (four known + five newly recognized species) plus the subgenus *Dudaica* (BP = 96; PP = 1.00). Within this cluster, the *Di. acutissima* group is suggested as paraphyletic with respect to *Dudaica*: five representatives of this group (e.g., *Di. cyanea* and *Di. pseudocyanea*) form a solid sub-cluster together with the representatives of *Dudaica* (BP = 100, PP = 1.00). This sub-cluster is strongly suggested as mutually monophyletic with the collection of the remaining representatives of the *Di. acutissima* group.

The third cluster (the *Di. tenuicauda* group) is strongly suggested as monophyletic (BP = 100, PP = 1.00) and placed as sister to the *Di. acutissima* group + *Dudaica* cluster. The relationship among the 14 representatives of the *Di. tenuicauda* group is well resolved: *Di. facilis* and two newly recognized members form a sub-cluster (BP = 100, PP = 1.00) basal to the sub-cluster of all the remaining representatives (BP = 85, PP = 1.00). The relationship within this later sub-cluster is almost fully resolved. In addition, *Di*. sp.DLS3a and *Di*. sp.DLS3b are separated by species-level divergence (*p*-distance = 0.013040 ± 0.002124) calculated with the concatenated nuclear DNA sequences though they are indistinguishable from each other in light of the divergence in *COI* sequences, indicating probable mitochondrial introgression between these two forms.

The fourth species cluster (i.e., *Di. sinensis* group + *Di. neocirricauda*; BP = 95, PP = 1.00). The sister relationship between this cluster and the fifth cluster, i.e., that of the solid *Di. trilobita* group, is strongly supported (BP = 100, PP = 1.00). The sixth species cluster (BP = 53, PP = 1.00) consists of two lineages: one is the genus *Mulgravea* representing by five species, including the type species *Mu. minima* and the newly recognized *Mu.* sp.a aff. *detriculata* (BP = 92, PP = 1.00), the other is of the "*Dichaetophora* Part 6" lineage (BP = 98, PP = 1.00).

The monophyly of the "*Dichaetophora* Part 7" is not supported, though three out of its

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.28.446102; this version posted May 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

five members ("*Di.* sp.HM1 *tenuicauda* group", "*Di.* sp.JC1" and "*Di.* sp.HM2 *tenuicauda* group") form a solid lineage (BP = 100, PP = 1.00), the positions of the remaining two are not settled.

## Discussion

### *The boundary of, and phylogeny within* Dichaetophora

The compilation between the known and newly recognized *Dichaetophora* species puts the total number in this genus (with *Mulgravea* and *Dudaica* species not included) about triple the previously known (255 *vs.* 69). The recognition of the morphological links between the genus *Mulgravea* or the subgenus *Dudaica* of genus *Drosophila* to *Dichaetophora* enable us to better design the taxon-sampling in phylogenetic reconstruction, and more accurately define the boundary of this genus. As a result, the "*Dichaetophora* Part 6 + *Mulgravea*" clade was recovered and the nature of this clade as a pan-*Dichaetophora* lineage confirmed. In addition, though the cluster of the five species of the "*Dichaetophora* Part 7" is not supported, the nature that these species are of the pan-*Dichaetophora* lineage is well supported.

Our phylogenetic reconstruction lent strong support to an expanded *Di. tenuicauda* group. This is in accordance with the results of Hu & Toda's (2002, 2005) morphological cladistic analysis. The placement of *Di. neocirricauda*, presumably also *Di. cirricauda* into the *Di. sinensis* group in the cluster 4 is rather comprehensible considering that, these two species are morphologically well coincident with the Hu & Toda's (2005) diagnosis for the *Di. sinensis* group [e.g., aedeagus apically with trumpet-like dilation as figured for *Di. cirricauda* in Okada's (1988) Fig. 9H]. The sister relationship between a revised *Di. sinensis* group (inclusive of *Di. neocirricauda* and *Di. cirricauda*) and the *Di. trilobita* group is thus logical taking into account the morphologically compatibility between the two groups (e.g., very large ocellar triangle, ≥40 medial sensilla on cibarium, and ventral surface of prementum forming discrete bump; Yang et al. 2017). However, further study with more extensive taxon sampling, especially from the *Di. trilobita* species group is needed to test if the two groups are mutually monophyletic.

The monophyly of the "*Di. acutissima* group + *Dr.* (*Dudaica*)" clade is as expected considering that, members in *Dudaica* (eight in total, all from the Oriental region; Duda 1926, Grimaldi 1990, Katoh et al. 2018) are morphologically highly analogous to those in *Dichaetophora* (ocellar setae outside the triangle made by ocelli, anterolateral corners of cibarium slightly protruded, number of pseudotracheae on labellum less than six). Moreover, essential morphological evidences from leg, wing and male/female genitalia in supporting the clade are found (apical seta on tibia of foreleg distinctly stout, dm-cu vein of wing clouded, surstylus without spines on out mesal surface, and aedeagus without basal processes, apical ovisensillum on oviscapt robust and the largest, distinguishable from the other). Our result about the sister relationship between the *Di. tenuicauda* group and the (*Di. acutissima* group + *Dudaica*) clade is readily acceptable taking into account the morphological similarity among the three taxa.

In our phylogenetic reconstruction, four *Mulgravea* species (including the type species of this genus, i.e., *Mu. minima*) were employed. Species in this genus are morphologically largely analogous in light of Hu & Toda's (2002) diagnosis of the latter one, except for the property that "ocellar setae outside triangle made by ocelli". Some *Mulgravea* species exhibit special morphological similarity to those of the "*Dichaetophora* Part 6" cluster. However, the definition of the "*Mulgravea* + *Dichaetophora* Part 6" clade is somewhat reluctant and further

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.28.446102; this version posted May 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

phylogenetic study combining morphological and DNA sequence data may needed to determine the phylogenetic positions of these two taxa in Drosophilidae.

*The species diversity, ecology and geography of the pan-*Dichaetophora *clade*

Our species delimitation recognized as many as 189 new members in addition to the 91 known members of the pan-*Dichaetophora* clade, raising this assemblage of 280 species as the seventh-ranked speciose generic taxa in the family Drosophilidae, or the fourth-ranked one within the tribe Drosophilini Okada (*sensu* Yassin, 2013), following the genera *Drosophila* (1,256 spp.), Hawaiian *Drosophila* (i.e., *Idiomyia* Grimshaw; 412 spp.) and *Scaptomyza* Hardy (298 spp.): *Drosophila* is evidently polyphyletic (see O'Grady and Desalle, 2018), including some independent lineages such as the subgenera *Drosophila* (432 spp.), *Siphlodora* (337 spp.) and *Sophophora* (344 spp.), and *Idiomyia* and *Scaptomyza* are well known drosophilids having experienced explosive evolutionary radiation in the Hawaiian archipelago (e.g., Zimmerman, 1958; Hardy, 1965; Carson *et al*., 1970; Kambysellis *et al*., 1995; O'Grady and Desalle, 2018).

Larvae of drosophilid flies depend on microbes (especially yeasts) causing fermenting of fruits, saps, and fungi, etc., or decay of leaves, thus are regarded as important in saprophytic food chains (Throckmorton 1975). Flies of pan-*Dichaetophora* species had been frequently collected from herb-layer habitats in the forest by net sweeping (e.g., Nishiharu 1981, Toda et al. 1987, Hu et al. 2019, Hu & Toda, 2005, Yang et al. 2017; Katoh *et al.* 2018). Some species in the *Di. agbo* group proved to use decaying leaves or stems of herbs (Nishiharu 1981). Actually a great majority of the samples employed in the present study were collected by net sweeping above herbaceous-layer, usually together with flies of the *Lordiphosa* proper, *Liodrosophila*, *Dettopsomyia*, etc. In addition, adult flies of a few *Dichaetophora* species (e.g. *Di. pseudocyanea*) were obtained via laboratory culturing of decaying herbs (JJ Gao, unpublished data). Therefore, it is very likely that the extraordinary species diversity achieved by the pan-*Dichaetophora* clade is attributed to its successful exploitation of the within- or along-forest, herbaceous-layer microhabitats.

Our phylogenetic reconstruction recovered seven major species clusters within the pan-*Dichaetophora* assemblage, one of the major drosophilid elements associated with herb habitats in the Old World (Fig. 3, Fig. S2). Though the relationships among these clusters are not well resolved, each of them is strongly supported, with the within-cluster relationship well resolved. The dominance of the Oriental elements in each of these clusters indicates that ancient adaptive radiation may occurred in the ancestor of the pan-*Dichaetophora* assemblage in the oriental region, following by subsequent dispersals from Oriental to Palearctic, Australasian and Afrotropical regions. However, more effort is needed to acquire accurate information about the species diversity of the pan-*Dichaetophora* lineage in the Afrotropical and Australasian regions, so as to test the above evolutionary hypothesis. Despite the above, our phylogenetic reconstruction will provide a framework for revising the taxonomy of *Dichaetophora* and allied taxa in a subsequent study of this serial work and baseline knowledge for future evolutionary studies on this speciose clade adapted to particular microhabitats.

## Acknowledgements

We thank Administrations of the Xishuangbanna, Gaoligongshan, Jianfengling and Taibaishan National Nature Reserves, the Xishuangbanna Tropical Botanical Garden, CAS, the Fushan

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.28.446102; this version posted May 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Botanical Garden, and the Forestry Bureau of Xizang Autonomous Region for administrative supports during field survey in each Garden/Reserve. We also thank Dr. Shun-Chern Tsaur, Dr. Chau-Ti Ting, Dr. Fu-Guo Robert Liu, Mr. Zhi-Wei Chang, Dr. Hong-Wei Chen, Dr. Xi-Peng Chen, Mr. Zhao Fu, Ms. Nan-Nan Li for their helps in field works in China; Dr. Shin-ichi Tanabe, Dr. Maklarin B. Lakim and Dr. Maryati Bte Mohamed for help in field works in Sabah, Malaysia under research permissions (UPE:40/200/19 SJ. 732 and UPE: 40/200/19 SJ.1194 and 1195) of Economic Planning Unit of Malaysian Government; Prof. E. S. Ratna of Bogor Agricultural University for help in field works in West Kalimantan, Indonesia, Prof. Masayoshi Watada for his help in field works in Ehime, Japan, Dr. Artyom Kopp and Dr. Toru Katoh for discussion about phylogenetic reconstruction. This work was supported by NSFC (Nos 31160429, 32060112 and 31760617) and the fund of the Ministry of Science and Technology of China (Nos 2011FY120200 and 2012FY110800).

## References

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

Burla H. 1954. Zur Kenntnis der Drosophiliden der Elfenbeinküste (Französisch West-Afrika). *Rev. Suisse Zool.* 61(Suppl.):1–218.

Carson, H.L., Hardy, D.E., Spieth, H.T. & Stone, W.S. (1970) The evolutionary biology of the Hawaiian Drosophilidae. *Essays in Evolution and Genetics in Honor of Theodosius Dobzhansky* (ed. by M.K. Hecht and W.C. Steere), pp. 437-543. Appleton Century Crofts, New York.

DrosWLD-Species. Available from: https://bioinfo.museum.hokudai.ac.jp/db/index.php

Duda O. 1926. Fauna sumatrensis: Drosophilidae (Dipt.). *Suppl. Ent.* 14:42–116.

Duda O. 1940. Revision der afrikanischen Drosophiliden (Diptera). II. *Annis. hist-nat. Mus. natn. hung*. 33:19–53.

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. 1994. DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I for diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* 3:294–299.

Friedrich M, Tautz D. 1997. Evolution and phylogeny of the Diptera: a molecular phylogenetic analysis using 28S rDNA sequences. *Syst. Biol.* 46:674–698; doi:10.1093/sysbio/46.4.674

Graber H. 1957. Ergebnisse der schweizerischer Virunga-Expedition (Leiter: Prof. Dr. Amold Heim). Afrikanische Drosophiliden als Blütenbesucher. *Zool. Jahrb. Abt. F. Syust.* 85:305–316.

Grimaldi DA. 1990. A phylogenetic, revised classification of genera in the Drosophilidae (Diptera). *Bull. Am. Mus. Nat. Hist.* 197:1–139.

Gupta JP, De A. 1996. Records of Drosophilidae with description of two new species from Bhutan (Insecta: Diptera). *Entomon* 21:177–186.

Hardy, D.E. (1965) Diptera: Cyclorrhapha II, Series Schizophora. Section Acalypteratae I. Family Drosophilidae. In: *Insects of Hawaii*, Vol. 12 (ed. by E.C. Zimmerman), vii + 814 pp. The University Press of Hawaii, Honolulu.

Hu YG, Toda MJ, Watabe HA. 1999. A revision of the *Lordiphosa tenuicauda* species group, with descriptions of eight new species from China (Diptera: Drosophilidae) *Entomol. Sci.*

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.28.446102; this version posted May 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2(1):105–199.

Hu YG, Toda MJ. 2001. Polyphyly of *Lordiphosa* and its relationships in Drosophilinae (Diptera: Drosophilidae). *Syst. Entomol.* 26:15–31; doi:10.1046/j.1365-3113.2001.00135.x

Hu YG, Toda MJ. 2002. Cladistic analysis of the genus *Dichaetophora* Duda (Diptera: Drosophilidae) and a revised classification. *Insect Syst. Evol.* 33:91–102; doi:10.1163/187631202X00064

Hu YG, Toda MJ. 2005. A new species group in the genus *Dichaetophora* Duda (Diptera: Drosophilidae) based on a phylogenetic analysis, with descriptions of four new species from China. *Zool. Sci.* 22:1265–1276; doi:10.2108/zsj.22.1265

Kambysellis, M.P., Ho, K.F., Craddock, E.M., Piano, F., Parisi, M. & Cohen, J. (1995) Pattern of ecological shifts in the diversification of Hawaiian *Drosophila* inferred from a molecular phylogeny. *Current Biology*, **5**, 1129–1139.

Kang YS, Lee TJ, Bahng KW. 1967. On *Drosophila quelpartiensis* sp. nov. with supplementary notes on the female of *Drosophila pseudonokogiri* Kang, Lee and Bahng. *Korean J. Zool.* 10:71–72.

Katoh T, Tamura K, Aotsuka T. 2000. Phylogenetic position of the subgenus *Lordiphosa* of the genus *Drosophila* (Diptera: Drosophilidae) inferred from alcohol dehydrogenase (*Adh*) gene sequences. *J. Mol. Evol.* 51:122–130; doi:10.1007/s002390010072

Katoh TK, Zhang G, Toda MJ, Suwito A, Gao JJ. 2018. A revision of the subgenus *Dudaica* Strand of the genus *Drosophila* Fallén, with descriptions of six new species (Diptera, Drosophilidae) *ZooKeys* 781:19–50; doi:10.3897/zookeys.781.27354

Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*. 35:4453–4455; doi:10.1093/bioinformatics/btz305

Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33:1870–1874; doi:10.1093/molbev/msw054

Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. 2016. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* 34:772–773; doi:10.1093/molbev/msw260

Lee, T.J. (1964) Taxonomy, and geographical distribution of Drosophilidae (Diptera) in Korea. *The Chungang University Theses Collection*, **9**, 425–459.

Lin FJ, Ting JI. 1971. Several additions to the fauna of Taiwan Drosophilidae (Diptera). *Bull. Insf. Zool. Academia Sinica.* 10:17–35.

Nishiharu S. 1981. Three new species of the genus *Nesiodrosophila* from Japan with a note on *N. surukella* Okada. *Kontyû* 49(1):19–26.

O'Grady, P.M. & DeSalle, R. (2018) Phylogeny of the genus *Drosophila*. *Genetics*, **209**, 1−25. https://doi.org/10.1534/genetics.117.300583

Okada T, Chung YJ. 1960. Three species of Drosophilidae from south Korea. *Akitu* 9: 25–30.

Okada T. 1965. Drosophilidae (Diptera) of southeast Asia collected by the Thai-Japanese expedition 1961–62. *Nature and Life in Southeast Asia* 3:439–466.

Okada T. 1966. Diptera from Nepal. Cryptochaetidae. Diastatidae & Drosophilidae. *Bull. Br.*

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.28.446102; this version posted May 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

*Mus. Nat. Hist. (Ent.) Suppl.* 6:1–129.

Okada T. 1968. Addition to the fauna of the family Drosophilidae of Japan and adjacent countries (Diptera). II. Genera *Paramycodrosophila*, *Mycodrosophila*, *Liodrosophila*, and *Drosophila*, including a few subgenus *Psilodorha*. *Kontyû* 36:324–340.

Okada, T. (1976) New distribution records of the drosophilids in Oriental Region. *Makunagi (Acta Dipterologica)*, **8**, 1–9. (In Japanese)

Okada, T. (1977) Family Drosophilidae. *A Catalog of the Diptera of the Oriental Region, Vol. III Suborder Cyclorrhapha (excluding Division Aschiza)* (ed. by Hardy, D.E. and Delfinado, M.), pp. 342–387. The University Press of Hawaii, Honolulu.

Okada T. 1984a. Descriptions of new species of the genus *Nesiodrosophila* Wheeler et Takada (Diptera, Drosophilidae), with taximetrical analyses. *Kontyû* 52:21–36.

Okada T. 1984b. New or little known species of *Drosophila* (*Lordiphosa*) with taximetrical analyses (Diptera, Drosophilidae), *Kontyû* 52:565–575.

Okada T. 1988. Family Drosophilidae (Diptera) from the Lund University Ceylon Expedition in 1962 and Borneo collections in 1978–1979. *Ent. Scand. Suppl.* 30:111–151.

Puillander N, Lambert A, Brouillet S, Achaz G: ABGD, Automatic barcode gap discovery for primary species delimitation. *Mol. Ecol.* 21:1864–1877. doi:10.1111/j.1365-294X.2011.05239.x

Rambaut A, Suchard MA, Xie D, Drummond AJ. 2014. Tracer v1.6. Available from: http://tree.bio.ed.ac.uk/software/tracer

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61:539–542; doi:10.1093/sysbio/sys029

Russo CA, Mello B, Frazão A, Voloch CM. 2013. Phylogenetic analysis and a time tree for a large drosophilid data set (Diptera: Drosophilidae). *Zool. J. Linn. Soc.* 169:765–775; doi:10.1111/zoj.12062

Schwarz G. 1978. Estimating the dimension of a model. *Ann. Stat.* 6:461–464; doi:10.1214/aos/1176344136

Throckmorton, L.H. (1975) The phylogeny, ecology and geography of *Drosophila*. *Handbook of Genetics*, Vol. 3 (ed. by R. C. King), pp. 421–469. Plenum Press, New York.

Toda MJ. 1983. Two species of the subgenus *Lordiphosa* Basden of the genus *Drosophila* (Diptera, Drosophilidae) from Japan. *Kontyû* 51(3):468–473.

Toda, M.J. (1989) Three new and one newly recorded species of Drosophilidae (Diptera) from *Japan. Japanese Journal of Entomology*, **57**, 375–382.

Toda MJ, Riihimaa A, Fuyama Y. 1987. Additional notes on drosophilid flies (Diptera, Drosophilidae) in the Bonin Islands, with descriptions of two new species. *Kontyû* 55:140–258.

Wheeler MR, Takada H. 1964. Diptera: Drosophilidae. *Ins. Micronesia*. *Bernice P. Bishop Museum* 14:164–242.

Yang JH, Toda MJ, Suwito A, Hashim R, Gao JJ. 2017. A new species group in the genus *Dichaetophora*, with descriptions of six new species from the Oriental region (Diptera, Drosophilidae). *ZooKeys* 665:121–146; doi:10.3897/zookeys.665.11609

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.28.446102; this version posted May 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Yassin A. 2013. Phylogenetic classification of the Drosophilidae Rondani (Diptera): the role of morphology in the postgenomic era. *Syst. Entomol.* 38:349–364; doi:10.1111/j.1365-3113.2012.00665.x

Zimmerman, E.C. (1958) Three hundred species of *Drosophila* in Hawaii? A challenge to geneticists and evolutionists. *Evolution*, **12**, 557−558.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.28.446102; this version posted May 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

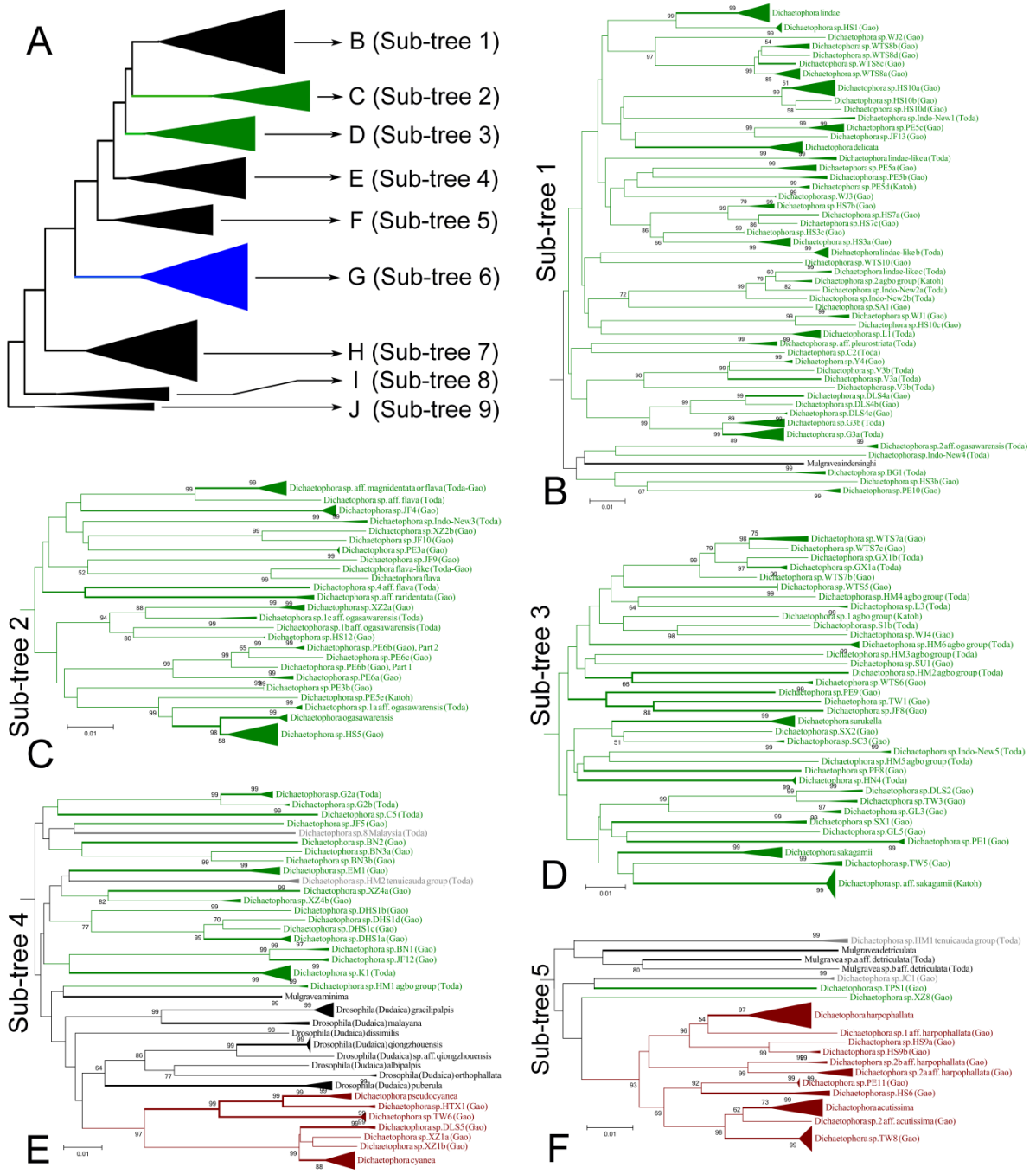


**Figure 1.** Compressed NJ tree of barcodes (see Figure S1 for the intact one). The skeleton of the tree is shown in panel A, with sub-trees shown individually in B–J. In the sub-tree 9 (panel H), morphological differentiation between *Di.* sp.DLS3a and *Di.* sp.DLS3b, two morpho-species indistinguishable from each other in light of DNA barcodes, was exemplified with lateral silhouettes of aedeagus (aed) and aedeagal apodeme (aed a). Species to be used in subsequent phylogenetic reconstruction are highlighted with thick branch(s), and different branch colors (and colors of names) are applied to distinguish among species assigned into different groups [based on morphological feature, relation to known species in the NJ and ML trees (if available) for newly recognized species]: green, *agbo* group; blue, *tenuicauda* group; brown, *acutissima* group; red, *sinensis* group; purple, *trilobita* group; gray, species not assigned into any species groups.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.28.446102; this version posted May 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

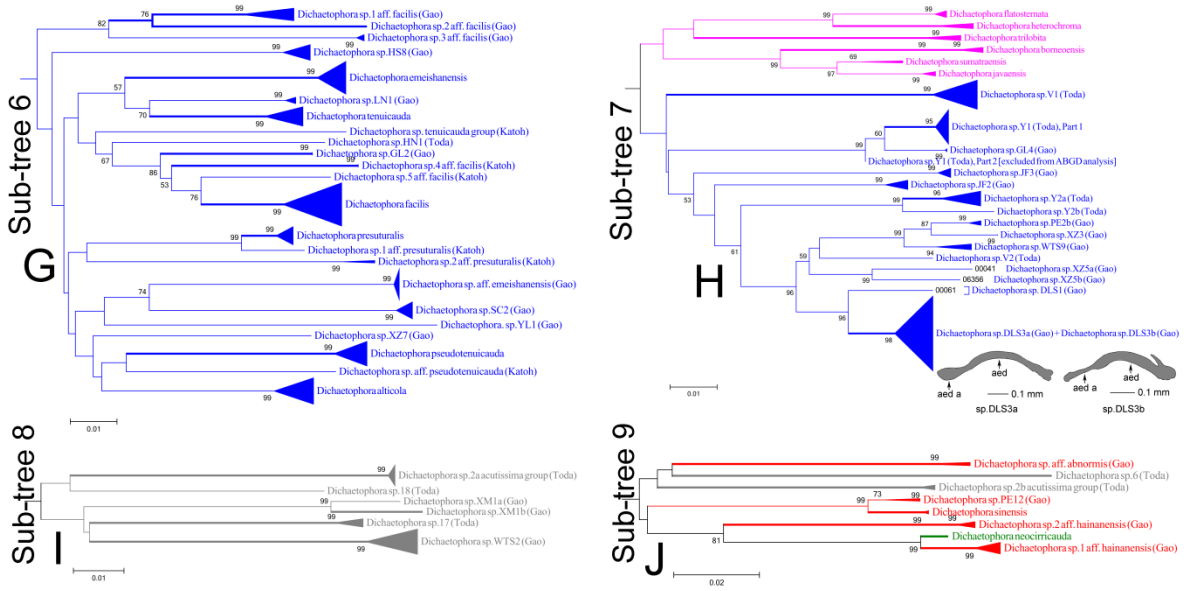


**Figure 1 (continued).**

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.28.446102; this version posted May 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

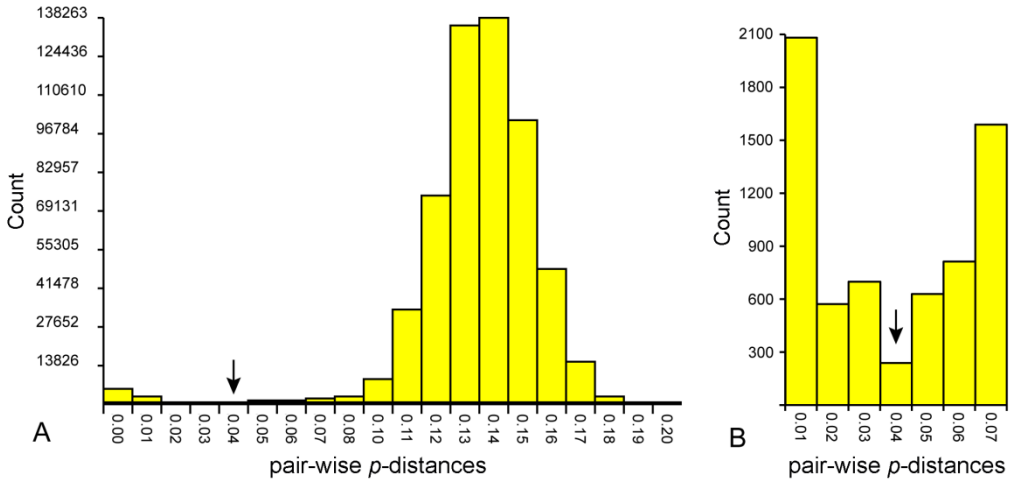


**Figure 2.** Histogram of the distribution of pairwise *p*-distances of pan-*Dichaetophora* (i.e., *Dichaetophora* plus quasi-*Dichaetophora*) species generated in the ABGD analysis. A, histogram covering all intervals of the *p*-distances; B, histogram covering the intervals between *p*-distance = 0.01 to 0.07. Arrows indicate the bar-coding gap.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.28.446102; this version posted May 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

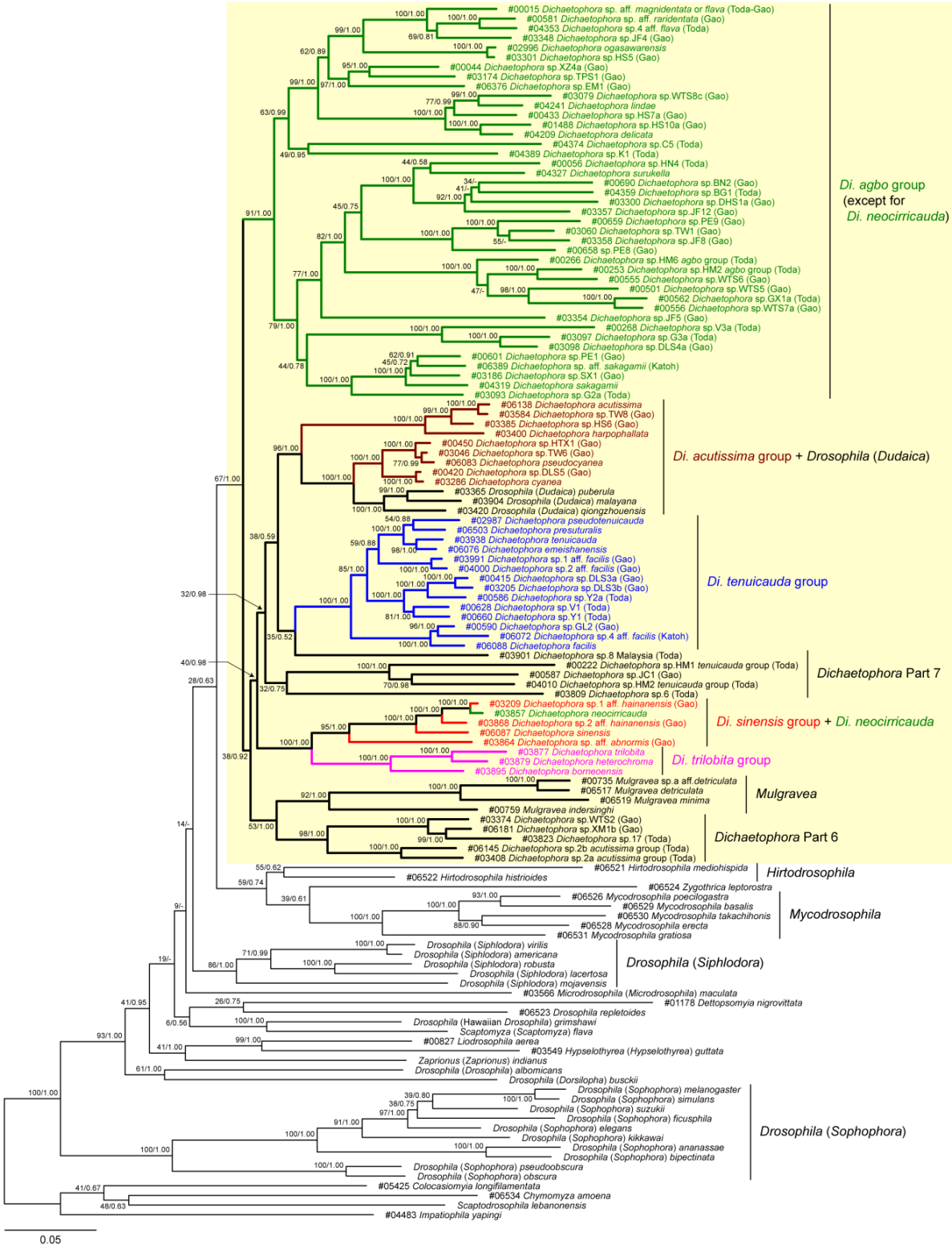


**Figure 3.** Maximum likelihood tree of the genus *Dichaetophora*. Numbers beside nodes are corresponding ML bootstrap percentage/Bayesian posterior probabilities ("-" indicates unavailable information). Different branch color (and color of names) are applied to distinguish among species of different species group within the "pan-*Dichaetophora*" clade (colored part of the tree): green, *agbo* group; blue, *tenuicauda* group; brown, *acutissima* group; red, *sinensis* group; purple, *trilobita* group; black, newly recognized, but ungrouped species. Voucher numbers are given for species involved in DNA sequencing. For each newly recognized species, a tentative name is given, with author name provided in parentheses.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.28.446102; this version posted May 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

**Table 1.** The result of species delimitation by prior intraspecific divergences (*P*) in the ABGD analysis.

| Prior intraspecific divergence (*P*) | Number of MOTUs | |
|---|---|---|
| | Initial partition | Recursive partition |
| 0.001000 | 309 | 311 |
| 0.001668 | 309 | 310 |
| 0.002783 | 309 | 310 |
| 0.004642 | 222 | 226 |
| 0.007743 | 222 | 224 |
| 0.012915 | 173 | 179 |
| 0.021544 | 173 | 177 |

**Table 2.** A summary of numbers of species involved in DNA bar-coding in the present study.

| Genus (subgenus) and species group | Number of known species [a] | Species involved in DNA bar-coding/phylogenetic reconstruction | | | Total number of species [c] |
|---|---|---|---|---|---|
| | | Known species | Newly recognized species [b] | Total | |
| *Dichaetophora*, *agbo* group | 45 | 7/6 | 123/36 | 130/42 | 168 |
| *Dichaetophora*, *tenuicauda* group | 10 | 6/5 | 33/9 | 39/14 | 43 |
| *Dichaetophora*, *acutissima* group | 4 | 4/4 | 14/5 | 18/9 | 18 |
| *Dichaetophora*, *sinensis* group | 4 | 1/1 | 4/3 | 5/4 | 8 |
| *Dichaetophora*, *trilobita* group | 6 | 6/3 | 0/0 | 6/3 | 6 |
| "*Dichaetophora* Part 6" | 0 | 0/0 | 7/5 | 7/5 | 7 |
| "*Dichaetophora* Part 7" | 0 | 0/0 | 5/5 | 5/5 | 5 |
| **Subtotal** | **69** | **24/19** | **186/63** | **210/82** | **255** |
| *Mulgravea* | 14 | 3/3 | 2/1 | 5/4 | 16 |
| *Drosophila* (*Dudaica*) | 8 | 7/3 | 1/0 | 8/3 | 9 |
| **Subtotal** | **22** | **10/6** | **3/1** | **13/7** | **25** |
| **Total** | **91** | **34/25** | **189/64** | **223/89** | **280** |

[a] Referring to the number of known species in the corresponding genus or subgenus.
[b] Referring to the newly recognized species in the present study.
[c] Referring to the sum of all the known and newly recognized species of the corresponding taxon.

**Table 3.** PCR/sequencing primer pairs designed for nuclear markers in the present study.

| Gene locus | Primer names and sequence (3'−5') | Annealing temperature (℃) |
|---|---|---|
| *Adenylosuccinate synthetase* (*AdSS*) | AdSS-f: TGGGYACCACCAAAAAGGG<br>AdSS-r: GGATACGTGCCAAARTCAATG | 50 |
| *ATP synthase, subunit B* (*ATPsynB*) | ATPsynB-f: GAGGARTGGTTCCAGTTYTT<br>ATPsynB-r: GCAATRTTYTCCTTCTTGGC | 50 |
| *Burgundy* (*bur*) | bur-f: GGCATYGATTTRATAGTGC<br>bur-r: GTTTCCGCRTTCKTGCTGAC | 48 |
| *Cell death protein 6* (*ced-6*) | ced-6-f: GARACGGGCACACAGGAGAA<br>ced-6-r: CCTGTAGGCCAAATCAAARG | 53 |
| *Eukaryotic translation initiation factor 3, subunit C* (*eIF3-S8*) | eIF3-S8-f: GYCAAATGCCATTCCAYATG<br>eIF3-S8-r: AAGTTGCCCTGCTTCATGTC | 52 |
| *Protein disulfide isomerase* (*Pdi*) | Pdi-f: GATTGGGACAARCARCCCGTC<br>Pdi-r: TTACAACTCRTCCTTCTTRGGC | 50 |
| *Phosphoglucose isomerase* (*Pgi*) | Pgi-f: GAAGGAGTTTACCAAYAAGG<br>Pgi-r: CCWACCCAATCCCARAAACC | 46 |
| *Ribosomal protein L3* (*RpL3*) | RpL3-f: AAAAGAAGGCGCACATCATG<br>RpL3-r: GATCTTCTTGTTGATCTCGG | 50 |
| *Ribosomal protein S17* (*RpS17*) | RpS17-f: TCGCGTCAGAACCAAGACWG<br>RpS17-r: CCTCYTCCTGCAGCTTAATGG | 53 |
| *Seven in absentia* (*sina*) | sina-f: AGTGCTGGGAACACATCMTC<br>sina-r: AACGCCCTCGTGTATGGAAC | 55 |
| *Vacuolar $H^+$-ATPase SFD subunit* (*VhaSFD*) | VhaSFD-f: TAYATGCARTCGCAAATGAT<br>VhaSFD-r: TRAGGAACTGCAAGTAGAAG | 44 |

**Table 4.** Optimal partitioning strategies and substitution models selected for Bayesian and maximum likelihood phylogenetic reconstruction with concatenated DNA sequences of 12 nuclear gene markers.

| Method | Partition # | Constituent block(s)[a] | No. of sites | Selected model[b] |
|---|---|---|---|---|
| Bayesian inference | 1 | *AdSS*_$CP_3$, *ATPsynB*_$CP_3$, *Pdi*_$CP_3$, *VhaSFD*_$CP_3$ | 436 | GTR+I+G |
| | 2 | *AdSS*_$CP_1$, *bur*_$CP_1$, *Pgi*_$CP_1$, *VhaSFD*_$CP_1$ | 416 | SYM+I+G |
| | 3 | *AdSS*_$CP_2$, *RpL3*_$CP_1$ | 191 | K80+I+G |
| | 4 | *ATPsynB*_$CP_1$, *ced-6*_$CP_1$, *Pdi*_$CP_1$ | 344 | GTR+I+G |
| | 5 | *ATPsynB*_$CP_2$, *ced-6*_$CP_2$ | 219 | SYM+I+G |
| | 6 | *bur*_$CP_3$, *ced-6*_$CP_3$ | 200 | SYM+I+G |
| | 7 | *bur*_$CP_2$, *eIF3-S8*_$CP_2$, *RpL3*_$CP_2$, *RpS17*_$CP_2$ | 436 | F81+I+G |
| | 8 | *eIF3-S8*_$CP_1$, *RpS17*_$CP_1$, *sina*_$CP_1$ | 491 | SYM+I+G |
| | 9 | *eIF3-S8*_$CP_3$, *Pgi*_$CP_3$ | 327 | GTR+I+G |
| | 10 | *Pdi*_$CP_2$, *Pgi*_$CP_2$, *VhaSFD*_$CP_2$ | 368 | GTR+I+G |
| | 11 | *RpL3*_$CP_3$, *RpS17*_$CP_3$, *sina*_$CP_3$ | 388 | GTR+I+G |
| | 12 | *sina*_$CP_2$ | 230 | JC69+I |
| | 13 | *28S rRNA* | 525 | GTR+I+G |
| Maximum likelihood | 1 | *AdSS*_$CP_3$, *ATPsynB*_$CP_3$, *Pdi*_$CP_3$, *VhaSFD*_$CP_3$ | 436 | GTR+I+G |
| | 2 | *AdSS*_$CP_1$, *bur*_$CP_1$, *Pgi*_$CP_1$ | 301 | SYM+I+G |
| | 3 | *AdSS*_$CP_2$, *Pdi*_$CP_1$, *RpL3*_$CP_1$ | 316 | TrNef+I+G |
| | 4 | *ATPsynB*_$CP_1$, *VhaSFD*_$CP_1$ | 216 | TIM1+I+G |
| | 5 | *ATPsynB*_$CP_2$, *ced-6*_$CP_2$ | 219 | SYM+I+G |
| | 6 | *bur*_$CP_3$, *ced-6*_$CP_3$, | 200 | SYM+I+G |
| | 7 | *bur*_$CP_2$, *eIF3-S8*_$CP_1$, *RpL3*_$CP_2$, *RpS17*_$CP_1$, *sina*_$CP_1$ | 667 | TrNef+I+G |
| | 8 | *ced-6*_$CP_1$ | 118 | TrN+I+G |
| | 9 | *eIF3-S8*_$CP_2$, *Pdi*_$CP_2$, *Pgi*_$CP_2$, *VhaSFD*_$CP_2$ | 576 | TVM+I+G |
| | 10 | *eIF3-S8*_$CP_3$, *Pgi*_$CP_3$ | 327 | GTR+I+G |
| | 11 | *RpL3*_$CP_3$, *RpS17*_$CP_3$, *sina*_$CP_3$ | 388 | GTR+I+G |
| | 12 | *RpS17*_$CP_2$, *sina*_$CP_2$ | 291 | JC69+I |
| | 13 | *28S rRNA* | 525 | TVM+I+G |

[a] Shown in a form of "gene loci_codon position", with codon position specified with subscript.
[b] Model and indices: F81, Felsenstein 1981 model; sGTR, General time reversible; K80, Kimura 2-parameter model; SYM, Symmetrical model; JC69, Jukes-Cantor 1969 model; TrNef, Tamura-Nei equal base frequency model; TIM1, Transition model; TVM, Transition model; TrN, Tamura-Nei model. I, proportion of invariable sites; G, Gamma distributed.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.28.446102; this version posted May 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## Supplementary information

**Table S1.** Known and newly collected DNA barcodes employed in species delimitation in the present study.

**Table S2.** Taxon sampling and information about DNA sequences (either cited or newly determined) employed in the present phylogenetic reconstruction. A hyphen ("-") is used to indicate a case of missing data.

**Figure S1.** Neighbor-joining tree built with 1,070 *COI* sequences. Numbers beside nodes are bootstrap percentages (BP, not show if < 50). Voucher number (or GenBank accession number) is shown for each sequence. Species to be used in the subsequent, molecular phylogenetic reconstruction are distinguished by thick branch(es). Species of different groups are distinguished with the same coloring scheme of branch and species name as in Figure 1.

**Figure S2.** Geographic distribution of all the pan-*Dichaetophora* species, i.e., those of the genus *Dichaetophora*, genus *Mulgravea* and the subgenus *Dudaica* of genus *Drosophila*. Known records of species are extracted from literature, and for known species, both known and new records (presented if available, and indicated with an asterisk) are shown.