Zoologica Scripta

Phylogenetic relationships in the subfamily Psychodinae (Diptera, Psychodidae)

Anahí Espíndola, Sven Buerki, Anouchka Jacquier, Jan Ježek & Nadir Alvarez

Submitted: 21 December 2011
Accepted: 9 March 2012

doi:10.1111/j.1463-6409.2012.00544.x

Espíndola, A., Buerki, S., Jacquier, A., Ježek, J. & Alvarez, N. (2012). Phylogenetic relationships in the subfamily Psychodinae (Diptera, Psychodidae). —*Zoologica Scripta*, *00*, 000–000.

Thanks to recent advances in molecular systematics, our knowledge of phylogenetic relationships within the order Diptera has dramatically improved. However, relationships at lower taxonomic levels remain poorly investigated in several neglected groups, such as the highly diversified moth-fly subfamily Psychodinae (Lower Diptera), which occurs in numerous terrestrial ecosystems. In this study, we aimed to understand the phylogenetic relationships among 52 Palearctic taxa from all currently known Palearctic tribes and subtribes of this subfamily, based on mitochondrial DNA. Our results demonstrate that in light of the classical systematics of Psychodinae, none of the tribes *sensu* Ježek or *sensu* Vaillant is monophyletic, whereas at least five of the 12 sampled genera were not monophyletic. The results presented in this study provide a valuable backbone for future work aiming at identifying morphological synapomorphies to propose a new tribal classification.

Corresponding author: *Anahí Espíndola, Laboratory of Evolutionary Entomology, Institute of Biology, University of Neuchâtel. Emile-Argand 11, 2000 Neuchâtel, Switzerland. E-mail: anahi.espindola@gmail.com*

Present address for Anahí Espíndola, Department of Ecology and Evolution, Biophore Building, University of Lausanne, 1015 Lausanne, Switzerland

Sven Buerki, Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3DS, UK. E-mail: s.buerki@kew.org

Anouchka Jacquier, Laboratory of Evolutionary Entomology, Institute of Biology, University of Neuchâtel. Emile-Argand 11, 2000 Neuchâtel, Switzerland. E-mail: anouchka.jacquier@gmail.com

Jan Ježek, Department of Entomology, National Museum of Prague (Natural History), Praha 4, Czech Republic. E-mail: jan.jezek@quick.cz

Nadir Alvarez, Department of Ecology and Evolution, Biophore Building, University of Lausanne, 1015 Lausanne, Switzerland. E-mail: nadir.alvarez@unil.ch

Introduction

The insect order Diptera (true flies) includes over 150 000 species and is considered one of the largest and most diverse groups of organisms (Yeates *et al.* 2007). Despite the extreme abundance and species richness of Diptera, the systematics of flies remains poorly understood. Recently, the popularization of molecular methods has triggered the reconstruction of phylogenetic relationships for this group. These methods have allowed testing previous classifications and have enabled using phylogenies as guidelines to establish new taxonomies when necessary (e.g. Friedrich & Tautz 1997; Bertone *et al.* 2008), providing, for instance, the first complete, resolved and dated phylogenetic inference of the order at the family level (Wiegmann *et al.* 2011).

Because Diptera comprise ca. 180 families, of which many are poorly known in terms of number and distribution of species, the inference of the evolutionary history of many lineages has been difficult. Psychodidae (Lower Diptera) is one of the worldwide families currently lacking such phylogenetic studies. It comprises ca. 3000 species (Ježek & Barták 2000) and is divided into six subfamilies (Bruchomyiinae, Trichomyiinae, Horaiellinae, Phlebotominae, Psychodinae and Sycoracinae). Current molecular chronograms estimate the origin of the family during the Cretaceous (ca. 95 Mya, Bertone & Wiegmann 2009), whereas fossil evidence [Blagoderov *et al.* (2007); not taken into account for the dating of Diptera by Wiegmann *et al.* (2011)] suggests an earlier origin during the Late Triassic-Early Jurassic (ca. 201 Mya). Phylogenetic relationships

within the family are only known for the subfamily Phlebotominae, a well-studied group because of its ability to transmit human diseases (Depaquit *et al.* 1998; Beati *et al.* 2004). In this study, to contribute to the knowledge of the evolution of the family, we investigated the phylogeny of the cosmopolitan and species-rich subfamily Psychodinae (Vaillant 1971) that is frequently found in decomposing trophic networks (Jung 1956).

The Psychodinae show broad morphological and ecological diversity, allowing them to disperse worldwide and to colonize and survive in remote places such as oceanic islands and subantarctic regions (Withers 1988). In the last 40 years, Vaillant (1971) and Ježek (1984a) have provided a species-level taxonomic framework for this subfamily, especially for Palearctic taxa. For instance, Vaillant (1971) attempted to describe and normalize characters for classification of larvae and proposed identification keys to relate them to their respective imagoes, whereas Ježek (1984a) solved nomenclatural problems and described species from poorly investigated geographic regions. Supra-specific entities have, however, remained subject to controversy, and most of them contradicted the first classification of Enderlein (1936) (see Table 1 for a summary of the classification of Psychodinae). For instance, Vaillant (1971) recommended dividing the Palearctic Psychodinae into four tribes: Telmatoscopini, Psychodini, Pericomini and Brunettiini. Unfortunately, most of Vaillant's work lacked nomenclatural (several types were wrongly cited) and cladistic rigour (relationships among species and genera were not clarified), and Ježek (1983, 1990a,b) thus proposed an alternative classification based on a cladistic approach (Table 1).

Although the circumscription of suprageneric levels of classification within the subfamily remains a challenging task, the strong taxonomic background available for Palearctic species (as well as the cladistic hypotheses provided by Ježek 1983, 1990a,b) allows testing the current Psychodinae systematics with a molecular approach. In this study, based on a sampling covering all Palearctic tribes of Psychodinae, we used molecular sequence data of two mitochondrial regions to infer phylogenetic relationships within the subfamily. The aims of this survey were (i) to evaluate the validity of Palaearctic tribes *sensu* Ježek and *sensu* Vaillant; (ii) when possible, to test the monophyly of genera; (iii) to propose guidelines for future systematics within Psychodinae.

Material and methods

Sampling

We sampled 52 species [88 specimens, distributed in 26 genera, representing all tribes in the European Psychodinae (Ježek 2009)] (Table 1; see Table S1 for information on amplification success of samples). These were obtained from the personal collection of Dr. J. Ježek. Samples were collected during several field trips, from years 1995 to 2009 in Eastern Europe, and were preserved in 70% ethanol. Outgroup species were added to the dataset based on recent phylogenetic studies on Diptera (see e.g. Bertone *et al.* 2008; Wiegmann *et al.* 2011). The outgroup sequences were obtained from GenBank (Table S1) and consisted of eight samples from seven species belonging to the infraorder Culicomorpha (Culicidae: *Anopheles funestus*, *A. gambiae*, *A. quadrimaculatus*, *A. darlingi*, *Aedes albopictus* and *A. aegypti*; Ceratopogonidae: *Culicoides arakawae*) and one species belonging to the Psychodidae subfamily Phlebotominae (*Phlebotomus riouxi*).

DNA extraction, amplification and sequencing

Insect DNA was extracted using the QIAgen DNEasy Animal tissue extraction kit (Qiagen, Hombrechtikon, Switzerland). Two mitochondrial (hereafter mtDNA) fragments were amplified with two primer pairs described in Simon *et al.* (1994): CB-J-11338 –CAC ATT CAA CCA GAA TGA TAT TT– and N1-N-12051 –GAT TTT GCT GAA GGT GAA TCA GA– for the first fragment and N1-J-12248 –AAG CTA ATC TAA CTT CAT AAG– and LR-N-12866 –ACA TGA TCT GAG TTC AAA CCG G– for the second fragment. While the first fragment amplified portions of genes Cytochrome B (cytB) and NADH1, the second partially covered genes NADH1 and 16S (see below and Fig. S1, for more details). Amplifications were performed in a 20 μL master mix consisting of 0.5× buffer, 100 mM dNTP, 0.5 μM of each primer, 0.12 mM $MgCl_2$, 1 unit of GoTaq *Taq*-polymerase (Promega, Dübendorf, Switzerland) and 4 μL DNA. Reactions were run in a Biometra Thermocycler (Biometra, Goettingen, Germany) with an initial denaturation of 1:30 min at 95 °C, followed by 35 cycles of 1 min of denaturation at 95 °C, 1 min of annealing at temperatures comprised between 42 °C and 52 °C, depending on samples and primers, and 30 s of elongation at 72 °C. Cycles were followed by 5 min of final elongation at 72 °C. Purification of PCR products and Sanger sequencing were done by Macrogen Inc. (Seoul, South Korea) and Fasteris SA (Geneva, Switzerland) with the same primers as used for the PCR amplification.

Sequence alignment and phylogenetic reconstruction

The program CHROMASPRO 1.41 (Technelysium Pty Ltd., Brisbane, Australia) was used to assemble complementary strands and verify software base-calling. The two amplified mtDNA fragments were initially aligned applying the Clustal-Wallis algorithm (implemented in BioEdit 7.0.4.1; Hall 1999) and further visually adjusted. Fragments from

Table 1 Taxonomic classification of all sampled taxa. Names are given following the taxonomy proposed by Ježek (1984a,b) and FAUNA EUROPAEA (Fauna Europaea Web Service, 2004). Supra-generic classification is given following Ježek (1984a,b) and Vaillant (1971). Phylogenetic clustering based on this work is given in the last column. A star (*) indicates the best-match subtribal or tribal classification following Ježek (1984a,b)

		Vaillant (1971)	Ježek (1984a,b)		
Species name (Ježek)	Species name (FAUNA EUROPAEA)	Tribe	Tribe	Sub-tribe	This study
Threticus incurvus	*Threticus incurvus* Krek 1972	Telmatoscopini	Paramormiini	Trichopsychodina*	I
Threticus lucifugus	*Threticus lucifugus* Walker 1856	Telmatoscopini	Paramormiini	Trichopsychodina*	I
Threticus silvaticus	*Threticus silvaticus* Ježek 1986	Telmatoscopini	Paramormiini	Trichopsychodina*	I
Psychodocha cinerea	*Psychoda cinerea* Banks 1894	Psychodini	Psychodini*		II
Psychodocha gemina	*Psychoda gemina* Eaton 1904	Psychodini	Psychodini*		II
Atrichobrunettia graeca	*Atrichobrunettia graeca* Ježek & Goutner 1993	Brunettiini	Mormiini	Brunettiina*	II
Psychodula minuta	*Psychoda minuta* Banks 1895	Psychodini	Psychodini*		II
Copropsychoda brevicornis	*Psychoda brevicornis* Tonnoir 1940	Psychodini	Psychodini*		II
Logima albipennis	*Psychoda albipennis* Zetterstedt 1852	Psychodini	Psychodini*		II
Logima satchelli	*Psychoda satchelli* Quate 1955	Psychodini	Psychodini*		II
Ypsydocha setigera	*Psychoda setigera* Tonnoir 1923	Psychodini	Psychodini*		II
Logima erminea	*Psychoda erminea* Eaton 1898	Psychodini	Psychodini*		II
Psycha grisescens	*Psychoda grisescens* Tonnoir 1923	Psychodini	Psychodini*		II
Psychoda crassipenis	*Psychoda crassipenis* Tonnoir 1940	Psychodini	Psychodini*		II
Psychoda phalaenoides	*Psychoda phalaenoides* Linnaeus 1758	Psychodini	Psychodini*		II
Psychoda uniformata	*Psychoda uniformata* Haseman 1907	Psychodini	Psychodini*		II
Psychomora mycophila	*Psychoda mycophila* Vaillant 1990	Psychodini	Psychodini*		II
Tinearia alternata	*Tinearia alternata* Say 1824	Psychodini	Psychodini*		II
Tinearia lativentris	*Tinearia lativentris* Berden 1952	Psychodini	Psychodini*		II
Trichopsychoda hirtella	*Trichopsychoda hirtella* Tonnoir 1919	Telmatoscopini	Paramormiini	Trichopsychodina*	III
Sciria advena	*Telmatoscopus advenus* Eaton 1893	Telmatoscopini	Paramormiini	Paramormiina*	IV
Paramormia polyascoidea	*Paramormia polyscoidea* Krek 1970	Telmatoscopini	Paramormiini	Paramormiina*	IV
Paramormia ustulata	*Paramormia ustulata* Walker 1856	Telmatoscopini	Paramormiini	Paramormiina*	IV
Clogmia albipunctata	*Clogmia albipunctatus* Williston 1893	Telmatoscopini	Paramormiini	Paramormiina*	IV
Clytocerus longicorniculatus	N/A	Pericomini	Pericomini*		IV
Clytoceru ocellaris	*Clytocerus) ocellaris* Meigen 1818	Pericomini	Pericomini*		IV
Clytocers rivosus	*Clytocerus rivosus* Tonnoir 1919	Pericomini	Pericomini*		IV
Oomoria andrenipes	*Mormia andrenipes* Strobl 1910	Pericomini	Mormiini	Mormiina*	IV
Jungiela procera	*Jungiella procera* Krek 1971	Telmatoscopini	Paramormiini	Paramormiina*	IV
Jungiella valachia	*Jungiella valachia* Vaillant 1963	Telmatoscopini	Paramormiini	Paramormiina*	IV
Parajungiella consors	*Jungiella consors* Eaton 1893	Telmatoscopini	Paramormiini	Paramormiina*	IV
Panimerus denticulatus	*Panimerus denticulatus* Krek 1972	Telmatoscopini	Paramormiini*		IV
Telmatoscopus britteni	*Telmatoscopus britteni* Tonnoir 1940	Telmatoscopini	Paramormiini	Paramormiina*	IV
Berdeniella illiesi	*Berdeniella illiesi* Wagner 1973	Pericomini	Pericomini*		IV
Berdeniella stavniensis	*Berdeniella stavniensis* Krek 1969	Pericomini	Pericomini*		IV
Berdeniella unispinosa	*Berdeniella unispinosa* Tonnoir 1919	Pericomini	Pericomini*		IV
Pericoma blandula	*Pericoma blandula* Eaton 1893	Pericomini	Pericomini*		IV
Pericoma exquisita	*Pericoma exquisita* Eaton 1893	Pericomini	Pericomini*		IV
Pericoma rivularis	*Pericoma rivularis* Berden 1955	Pericomini	Pericomini*		IV
Pneumia compta	*Satchelliella compta* Eaton 1893	Pericomini	Pericomini*		IV
Pneumia gracilis	*Satchelliella gracilis* Eaton 1893	Pericomini	Pericomini*		IV
Pneumia nubila	*Satchelliella mutua* Eaton 1893	Pericomini	Pericomini*		IV
Pneumia pilularia	*Satchelliella pilularia* Tonnoir 1940	Pericomini	Pericomini*		IV
Pneumia plumicornis	*Satchelliella plumicornis* Tonnoir 1922	Pericomini	Pericomini*		IV
Pneumia trivialis	*Satchelliella trivialis* Eaton 1893	Pericomini	Pericomini*		IV
Saraiella rotunda	*Saraiella rotunda* Krek 1970	Pericomini	Pericomini*		IV
Ulomyia cognata	*Ulomyia cognata* Eaton 1893	Pericomini	Pericomini*		IV
Ulomyia fuliginosa	*Ulomyia fuliginosa* Meigen 1818	Pericomini	Pericomini*		IV
Ulomyia undulata	*Ulomyia undulata* Tonnoir 1919	Pericomini	Pericomini*		IV
Pneumia stammeri	*Satchelliella stammeri* Jung 1954	Pericomini	Pericomini*		IV

the two primer pairs covered the following mitochondrial regions: (i) partial CytB gene; (ii) tRNA-Ser (iii) complete NADH1 gene (with the exception of ~275 bp in the middle of the gene completed with unknown (N) bases in the alignment; i.e. the two primer pairs did not overlap); (iv) tRNA-Leu; (v) partial 16s rRNA (see Fig. S1). Whereas CytB and NADH1 were considered as two separate partitions, RNA regions (both transfer and ribosomal RNAs) were merged into a third data partition (hereafter trRNA) to avoid considering partitions shorter than 250 bp. When present, gaps were coded using the simple gap-coding algorithm of Simmons & Ochoterena (2000) using FastGap 1.2 (Borchsenius 2009). Gap characters were added as a fourth data partition (see below). Finally, the number of constant, variable and parsimony informative characters was calculated on the complete and partial (i.e. containing all samples presenting <50% of missing data; see below) datasets with PAUP* 4b10 (Swofford 2003) for all regions and for each codon position in the coding regions.

The suitability of the DNA regions used in this study was evaluated by quantifying the level of saturation in the phylogenetic signal of each partition. This was performed by plotting the uncorrected pairwise distances between taxa against Maximum Likelihood (ML) distances derived in PAUP* 4b10 (Swofford 2003) using the best-fit model (see below). The coefficient of determination (R^2) of the correlation between the two distances was also calculated using the R *stats* package (R Development Team, 2009). Such an approach was applied to each DNA partition separately and to the concatenated dataset. High R^2 values would be interpreted as an indication of lack of saturation and thus appropriateness of the use of a given partition at a given phylogenetic level (Van de Peer *et al.* 2002). Differences in the scales of the two axes indicate the suitability of the use of a model of evolution different from the uncorrected pairwise distance.

Phylogenetic inferences based on DNA sequences can be biased by two major types of issues: (i) incongruences in the evolution of molecular regions and (ii) a missing data effect (Wiens & Morrill 2011, and references therein). To confirm that such biases were not present in our study, we performed several analytical approaches. The datasets compared for answering point (i) were those consisting of single-gene and concatenated matrices. The datasets analysed for investigating point (ii) were those including all specimens (hereafter, the complete dataset) and those harbouring only those specimens presenting <50% missing data in all genes (hereafter, the partial dataset).

To analyse the incongruence between the molecular partitions, two methods were applied. First, an incongruence length difference (ILD) test (Farris *et al.* 1994) was performed using PAUP* 4b10 (Swofford 2003) with 1000 replicates. In this analysis, if the calculated probability value is higher than 0.05, it is statistically correct to assume that the information provided by the different partitions is congruent (Swofford 2003). Second, phylogenetic inferences were conducted using three phylogenetic approaches (see below) for the concatenated and single-gene datasets. Here, we evaluated the general inferred topologies, considering the node supports obtained for each of them (see below). Once the topologies estimated, we proceeded to their comparison by (i) their visual analysis, with incongruences recognized if inconsistencies were supported by bootstrap values >75% or Bayesian Posterior Probabilities >0.8; and (ii) the evaluation of normalized partition metric (NPM) distances between all the obtained trees (calculated following Buerki *et al.* 2011).

A similar approach to the one used for evaluating the presence of partition incongruence was applied to test the effect of missing data on our phylogenetic inferences. The phylogenies inferred in the complete and partial datasets using three different phylogenetic approaches (see below) were afterwards compared.

Model-based (i.e. Bayesian inference and ML) and parsimony [i.e. Maximum Parsimony (MP)] inferences were performed for all datasets that let us also identify the effect that different phylogenetic algorithms have in the obtained inferences. While the same run parameters were applied for the analysis of all datasets (concatenated, single-gene, complete and partial), the concatenated datasets were analysed using a partitioned approach in the case of the Bayesian analysis (see below).

A partitioned Bayesian analysis was performed using a Markov-Chain Monte-Carlo (MCMC) approach, as implemented in MrBayes v3.1.2 (Ronquist & Huelsenbeck 2003) and following recommendations by Nylander *et al.* (2004). Best-fit models for each region were selected using MrModeltest v.3.0 (Nylander 2004) based on the Akaike information criterion (Akaike 1973). For the three DNA partitions, the best-fit model was the general time reversible (GTR) model with an alpha parameter for the shape of the gamma distribution to account for rate heterogeneity among sites (Yang 1993) and a rate for invariable sites. Even if the models of evolution were similar for the three partitions, rate parameters were allowed to vary independently in the three DNA partitions (all parameters were unlinked except the topology following recommendations made in the MrBayes manual). As datasets were compatible and followed the same model of evolution, all gaps were merged into a single partition and the *restriction* option was applied to this particular partition. Two Metropolis-coupled MCMCs with incremental heating temperature of 0.5 were run for 20 million generations,

with the parameters and the resulting trees being sampled every 1000 generations. The analysis was repeated twice, starting with random trees. Convergence was accepted when standard deviations attained values below 0.01 and when the Potential Scale Reduction Factor index (Gelman & Rubin 1992) approached 1.0. We considered the MCMC sampling sufficient when the Effective Sampling Size (ESS) was higher than 200 –checked on Tracer v1.4 (Rambaut & Drummond 2004). After a burn-in period corresponding to 5000 trees, a 50% majority-rule consensus tree (and its associated Bayesian posterior probabilities; BPP) was reconstructed (based on the remaining 15 000 trees) using MrBayes v3.1.2 (Ronquist & Huelsenbeck 2003). The analysis was run on the Bioportal cluster from the University of Oslo (Kumar *et al.* 2009).

The ML analysis was performed using RAxML 7.2.6 (Stamatakis 2006) with 10 000 rapid bootstrap analyses followed by the search of the best-scoring ML tree in one single run. The three mtDNA regions were considered as one single partition (by applying the GTRCAT model) and the gap partition was not considered. The analysis was performed using the facilities offered by the CIPRES portal (San Diego, CA, USA).

Finally, the combined dataset was analysed using a parsimony ratchet slightly similar to that proposed by Nixon (1999) and implemented in TNT v1.1 (Goloboff *et al.* 2008). Based on recommendations of Nixon (1999), 10 independent searches (with 200 iterations and 15% of the characters perturbed) were conducted. The most parsimonious trees were afterwards combined to create majority-rule and strict consensus trees. Branch supports for both the strict and majority-rule topologies were estimated calculating bootstraps, based on 1000 replicates, and using TNT v.1.1.

The definition of the phylogenetic clades was done by considering the following criteria: (i) the monophyly of the clade is retrieved by at least two among the three algorithms; (ii) the basal node of the clade is supported with a BPP >0.65 and bootstrap values >55%.

Congruence between Bayesian, ML and MP inferences was evaluated by comparing topologies and node supports for the complete and partial datasets, and for each partition. Moreover, distances between the topologies were calculated using the NPM distances, as done by Buerki *et al.* (2011).

Results

Among the 96 samples (88 in the ingroup and eight outgroups), 84 properly amplified CytB, 94 amplified NDH1 and 90 amplified trRNA. The combined aligned dataset was 1548 bp: 273 bp for CytB, 978 bp for NADH1 and 297 bp for trRNA (alignments available at TreeBase: http://www.treebase.org/treebase-web/search/study/summary.html?id=11642). The gap partition comprised 50 characters (coded as presence/absence). The proportion of variable sites ranged from 37 to 64% depending on the region and dataset considered, with more than two-thirds being parsimony-informative (Table 2). The third codon-position was the most variable in the two protein-coding regions (i.e. at most 54% of all parsimony-informative sites were found in the third codon-position). The phylogenetic signal did not appear to be saturated (Fig. S2), indicating that the genetic information retrieved was appropriate for the analysis at the evolutionary scale investigated here.

The evaluation of the incongruence between topologies obtained from single-gene, concatenated, complete and

Table 2 Various statistics related to the molecular datasets. Constant (C), variable (V), parsimony informative (PI) and total number of sites, in base pairs, (a) for NADH1, CytB and trRNA and for each codon position in the two coding regions (b) NADH1 and CytB. Percentage of the total base-pairs is indicated in parentheses; gaps and total sequence length are shown in the last columns

(a)

Region	C	V	PI	Total	Gaps
NADH1	528	450	330		
	53.99%	46.01%	46.01%		
$NADH1_{partial}$	614	364	306		
	62.78%	37.22%	31.29%	978	9
CytB	98	175	159		
	35.90%	64.10%	58.24%		
$CytB_{partial}$	116	157	137		
	42.49%	57.51%	50.18%	273	0
trRNA	156	141	98		
	52.53%	47.47%	33%		
$trRNA_{partial}$	171	126	92		
	57.58%	42.42%	30.98%	297	41
Total				1548	50

(b)

NADH1

Position	C	V	PI	Total
1	216 (51.06)	110 (19.81)	72 (17.65)	326
2	146 (34.51)	180 (32.43)	132 (32.35)	326
3	61 (14.43)	265 (47.76)	204 (50)	326
Total	423	555	408	978

CytB

Position	C	V	PI	Total
1	46 (40)	45 (28.48)	43 (30.07)	91
2	63 (54.78)	28 (17.72)	22 (15.38)	91
3	6 (5.22)	85 (53.80)	78 (54.55)	91
Total	115	158	143	273

partial datasets indicated that neither gene partitioning nor missing data lead to any significant bias in the resulting phylogenetic inferences. First, the ILD test provided *P* values higher than 0.05 ($P = 0.08$), indicating that no statistically significant incongruence between partitions was observed. Second, distances between the topologies obtained for all datasets were relatively low (mean similarities ranged at 75%), and were largely explained by polytomies (Tables S2 and S3). Importantly, phylogenetic reconstructions based on single-gene matrices were less resolved when compared with concatenated datasets, providing more support to a total-evidence approach (Table S3).

In what concerned the three phylogenetic approaches applied to the complete dataset, they also yielded largely congruent topologies (Fig. 1 and Table S2), even though MP topologies presented low bootstrap supports, and were more polytomized (Table 3).

The inferences done on the complete dataset identified four main clades (Fig. 1 and Table 3), hereafter referred to as I to IV (Fig. 1). While clade I and III were supported by all phylogenetic methods, clades II and IV were supported by ML and Bayesian methods and presented low bootstrap values and polytomies in the MP approaches. We did not consider finer grained clade definitions to be sufficiently well sampled in our study to be generally informative. Clade I included all *Threticus* specimens, which belong to subtribe Trichopsychodina (Table 1). Clade II was formed by all Psychodini samples

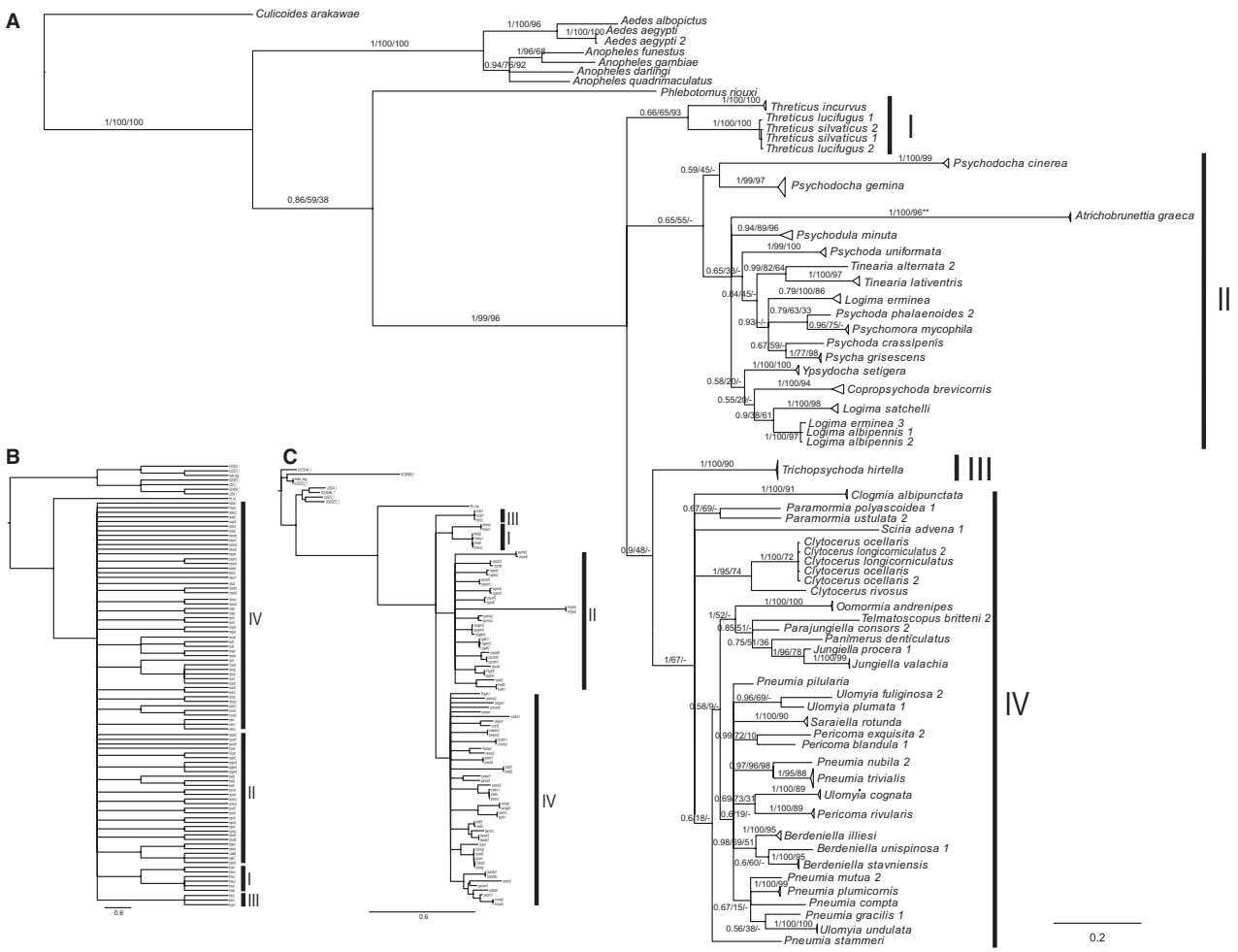

Fig. 1 Phylogenetic inference of subfamily Psychodinae. Vertical bars on the right illustrate the informal classification considered in the present study. —A. 50% majority-rule consensus Bayesian topology. Node supports are shown on branches and correspond to Bayesian Posterior Probability, Maximum Likelihood (ML) bootstraps and Maximum Parsimony (MP) bootstraps values respectively. Monophyletic species appear collapsed and are represented by triangles. —B. 50% majority-rule MP topology with unsupported nodes collapsed. —C. ML topology with unsupported nodes collapsed.

Table 3 Clade definition, best-match taxonomic classification at the tribal or subtribal levels, following Ježek (1984a,b), support values and mean normalized partition metric distance of the given topology to the remaining trees. Inferences are classified following the nature of the dataset (i.e. complete or partial). Support values are given for the 50% majority-rule consensus Bayesian tree (BPP), the Maximum Likelihood (ML) and the bootstrapped MP strict and 50% majority-rule (majrule) consensus trees

		Clade I	Clade II	Clade III	Clade IV	
		Trichopsychodina	Psychodini + Brunettiina	Trichopsychodina	Paramormiina + Pericomini + Mormiina	Mean Distance
Complete dataset	BPP	0.66	0.65	1	1	0.22
	ML	65	55	100	67	0.24
	strict MP	93	Polytomy	90	Polytomy	0.38
	majrule MP	93	Polytomy	90	Polytomy	0.22
Partial dataset	BPP	1	1	1	0.98	0.28
	ML	100	100	100	77	0.24
	strict MP	100	99	100	Polytomy	0.22
	majrule MP	100	99	100	Polytomy	0.22

BPP, Bayesian posterior probabilities; MP, Maximum Parsimony.

and by the only Brunettina (tribe Mormiini) specimen included in our study (Table 1). Clade III comprised all specimens assigned to *Trichopsychoda* and represented the remaining Trichopsychodina samples of this study (Table 1). Finally, clade IV contained all Pericomini and Paramormiini samples as well as the unique representative of subtribe Mormiina (tribe Mormiini, Table 1).

At the generic level, seven of the 12 genera for which more than one species were sampled were monophyletic under at least one phylogenetic criterion (Table 4). The remaining five genera (i.e. *Logima, Pneumia*, *Pericoma*, *Psychoda, Ulomyia*; Table 4) were para- or polyphyletic under all criteria. The polyphyletic genera were retrieved in two subclades of clade II for *Logima*, and IV for *Pneumia*. The paraphyletic genera *Pericoma* and *Ulomyia* were restricted to clade IV, while *Psychoda* was comprised in several subclades of clade II.

At the species level, two species in clade II contained especially long branches: *Psychodocha cinerea* and *Atrichobrunettia graeca*, which could suggest an accelerated rate of evolution, or a long-branch-attraction bias due to the lack of sister lineages included in this study.

Discussion

Dataset robustness

The main clade subdivision unravelled in our analyses (Fig. 1) and later discussed is robust and independent of the data partitioning and the amount of missing data present in our study (Table 3 and Table S2). The choice of the genetic regions used for our inferences is appropriate to the evolutionary scale of this study, since partitions harbour high amounts of variable positions (Table 2) and are not saturated (Fig. S2). From this perspective, our results appear to represent a solid basis for discussing the Palearctic Psychodinae tribes.

All phylogenetic algorithms (Bayesian, ML and MP) applied to all datasets (concatenated and single-gene; complete and partial) reached highly congruent results (Table 3 and Table S2), pointing out the presence of a strong phylogenetic signal within the data. However, some NPM distances between trees were fairly high. This was particularly the case in the set of all MP topologies, which were systematically more polytomized, and were closer to each other when compared with all model-based methods (Table S3). In fact, while mean distances between MP topologies were of 0.15, mean distances between the latter and model-based methods were higher (0.25), suggesting that the polytomized nature of the MP inferences were strongly and negatively affecting the final mean distance values. It thus appears that topological differences (and thus distances) were mainly explained by the phylogenetic method applied. Although MP topologies were polytomized, their clustering was in agreement with those yielded with the other two phylogenetic methods.

Partial contradiction of previous classifications

Results presented in this study suggest the monophyly of subfamily Psychodinae, indicating the polyphyly of all tribes (with the exception of Psychodini, paraphyletic with subtribe Brunettiina; see below) and challenge the monophyly of several genera (Table 4).

On the basis of current sampling, we can suggest the monophyly for some taxa, while the results of polyphyly of other taxa should be considered as conclusive. From this perspective, five genera are polyphyletic (*Logima* and *Pneumia*, *Pericoma*, *Psychoda* and *Ulomyia*) (Table 4). Further investigations are required before revising their generic status. However, the splitting of *Psychoda s.l.* proposed by Ježek (1984a,b, 1990a,b); Table 1) appears to be phylogenetically supported, although it should be refined in future

Table 4 (a) Summary of genera *sensu* Ježek (1984a,b) for which monophyly could be retrieved with Bayesian inference (BI), Maximum Parsimony (MP) and Maximum Likelihood (ML) approaches. Number of species sampled per genus is also indicated. N/A: monophyly not tested because only one species sampled. (b) Counts of monophyletic and paraphyletic genera

(a)

		Monophyly		
Genus	Number of species	Bayesian	MP	ML
Atrichobrunettia	1	N/A	N/A	N/A
Berdeniella	3	Yes	Yes	Yes
Clogmia	1	N/A	N/A	N/A
Clytocerus	3	Yes	Yes	Yes
Copropsychoda	1	N/A	N/A	N/A
Jungiella	2	Yes	Yes	Yes
Logima	3	No	No	No
Oomormia	1	N/A	N/A	N/A
Panimerus	1	N/A	N/A	N/A
Parajungiella	1	N/A	N/A	N/A
Paramormia	2	Yes	No	Yes
Pericoma	3	No	No	No
Pneumia	8	No	No	No
Psycha	1	N/A	N/A	N/A
Psychoda	3	No	No	No
Psychodocha	2	Yes	No	No
Psychodula	1	N/A	N/A	N/A
Psychomora	1	N/A	N/A	N/A
Saraiella	1	N/A	N/A	N/A
Sciria	1	N/A	N/A	N/A
Telmatoscopus	1	N/A	N/A	N/A
Threticus	3	Yes	Yes	Yes
Tinearia	2	Yes	Yes	Yes
Trichopsychoda	1	N/A	N/A	N/A
Ulomyia	4	No	No	No
Ypsydocha	1	N/A	N/A	N/A

(b)

Monophyly	Count
Yes for BI, MP or ML (or all)	7 (5)
No for neither BI, MP nor ML	5
N/A	14
Total	26

taxonomic revisions. In that respect, the phylogenetic inferences will greatly help guiding the recognition of morphological synapormophies for each entity via, for instance, the mapping of characters on the topologies.

The phylogenetic tree depicted here does not fully agree with any current taxonomic treatment. However, one of the four tribes defined by Ježek (Psychodini, clade IId; Fig. 1) matched our results when merged with the subtribe Brunettiina. Similar conclusions can be drawn when considering the two tribes Brunettiini and Psychodini (Table 1) proposed by Vaillant (1971). In contrast, tribes Pericomini and Paramormiini are paraphyletic, which is also true for subtribes Paramormiina (clade IV) and Trichopyschodina (clades I and III) (Fig. 1 and Table 3). However, even if Trichopsychodina is a paraphyletic entity, representatives of this tribe comprise the monophyletic genera *Threticus* (clade I) and *Trichopsychoda* (clade III), which might constitute valuable information to circumscribe future taxa. These new findings suggest that part of the classification of Ježek (based on cladistic analyses) reflects the evolutionary history of Psychodinae and should be considered when proposing a new classification.

Although our results are based on few molecular regions, the evidence provided by the data indicates that in the future, taxonomists should adopt a broader concept of tribe Psychodini by including subtribe Brunettiina *sensu* Ježek (or tribe Brunettiini *sensu* Vaillant). This new concept will better reflect the evolutionary relationships than the current taxonomic treatments. We also advise considering genera *Threticus* and *Trichopsychoda* as references for defining strongly supported entities on which to base the characterization of future tribes or subtribes. In contrast, new morphological characters should be explored to define supra specific entities in tribe Pericomini and subtribe Paramormiina (Fig. 1 and Table 3). With the exception of the poorly sampled subtribes, Brunettina and Mormiina, none of the published taxonomic entities corresponds to any of the phylogenetic groups discovered here (Table 3).

The present study should be considered as a first step towards a complete understanding of the relationships within Psychodinae. In the future, and to resolve poorly supported phylogenetic relationships and recognize morphological synapomorphies (at the basis of a new formal classification), studies should (i) increase their taxon sampling by increasing the number of species at the different taxonomic levels and biogeographic areas, (ii) sequence more markers, both mitochondrial (Cameron *et al.* 2007) and nuclear (Wiegmann *et al.* 2011), (iii) compile an extensive morphological matrix. The sampling of a higher number of regions would allow for solving basal unclear relationships and, if combined with fossil data, could inform on the dating and rate of radiation of the group.

Acknowledgements

The authors thank the technical help of Laure Benoit. They also thank Sarah M. Hird for correcting the English language, and three anonymous reviewers and the journal Editor Per Ericson for helping to improve a previous version of this manuscript. This study was funded by the Swiss National Science Foundation (project No. 3100A0-116778). Financial support to SB was provided by Marie-Curie Intra-European Fellowship (CRADLE; no 253866).

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B.N. Petrov & F. Csaki (Eds.) *Second International Symposium on Information Theory* (pp. 267–281). Budapest: Akademiai Kiado.

Beati, L., Caceres, A. G., Lee, J. A. & Munstermann, L. E. (2004). Systematic relationships among *Lutzomyia* sand flies (Diptera: Psychodidae) of Peru and Colombia based on the analysis of 12S and 28S ribosomal DNA sequences. *International Journal of Parasitology*, *34*, 225–234.

Bertone, M. A. & Wiegmann, B. M. (2009). True flies (Diptera). In S. B. Hedges & S. Kumar (Eds) *The Timetree of Life* (pp. 270–277). UK: Oxford University Press.

Bertone, M. A., Courtney, G. W. & Wiegmann, B. M. (2008). Phylogenetics and temporal diversification of the earliest true flies (Insecta: Diptera) based on multiple nuclear genes. *Systematic Entomology*, *33*, 668–687.

Blagoderov, V., Grimaldi, D. A. & Fraser, N. C. (2007). How time flies for flies: diverse Diptera from the Triassic of Virginia and early radiation of the order. *American Museum Novitates*, *3572*, 1–39.

Borchsenius, F.2009. Fastgap 1.2. Denmark: Department of Biological Sciences, University of Aarhus. Available via http://192.38.46.42/aubot/fb/FastGap_home.htm.

Buerki, S., Forest, F., Salamin, N. & Alvarez, N. (2011). Comparative performance of supertree algorithms in large data sets using the soapberry family (Sapindaceae) as a case study. *Systematic Biology*, *60*, 32–44.

Cameron, S. L., Lambkin, C. L., Barker, S. C. & Whiting, M. (2007). A mitochondrial genome phylogeny of the Diptera: whole genome sequence data successfully resolves relationships over broad timescales with high precision. *Systematic Entomology*, *32*, 40–59.

Depaquit, J., Perrotey, S., Lecointre, G., Tillier, A., Tillier, S., Ferte, H., Kaltenbach, M. & Léger, N. (1998). Molecular systematics of phlebotominae: a pilot study. Paraphyly of the genus *Phlebotomus*. *Comptes Rendus de l'Académie des Sciences III*, *321*, 849–855.

Farris, J. S., Källersjö, M., Kluge, A. G. & Bult, C. (1994). Testing significance of incongruence. *Cladistics*, *31*, 5–319.

Fauna Europaea Web Service. (2004). Fauna Europaea version 1.1. Available via http://www.faunaeur.org/.

Friedrich, M. & Tautz, D. (1997). Evolution and phylogeny of the Diptera: a molecular phylogenetic analysis using 28S rDNA sequences. *Systematic Biology*, *46*, 674–698.

Gelman, A. & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7, 457–472.

Goloboff, P., Farris, J. & Nixon, K. (2008). TNT, a free program for phylogenetic analysis. *Cladistics*, *24*, 774–786.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, *41*, 95–98.

Ježek, J. (1983). Intergeneric relationships of selected tribes of the subfamily Psychodinae (Diptera, Psychodidae). *Acta Entomologica Musei Nationalis Pragae*, *41*, 255–259.

Ježek, J. (1984a). Nomenclatorical changes of some higher taxa of Palaearctic Psychodinae (Diptera, Psychodidae). *Acta Faunistica Entomologica Musei Nationalis Pragae*, *17*, 155–170.

Ježek, J. (1984b). Six new genera of the tribe Psychodini End. (Diptera, Psychodidae). *Acta Faunistica Entomologica Musei Nationalis Pragae*, *17*, 133–154.

Ježek, J. (1990a). Contribution to the taxonomy of some genera of paramormiine moth flies (Diptera, Psychodidae) with description of a new genus *Karakovounimerus*. *Acta Entomologica Musei Nationalis Pragae*, *43*, 129–157.

Ježek, J. (1990b). Redescriptions of nine common Palaearctic and Holarctic species of Psychodini End. (Diptera: Psychodidae). *Acta Entomologica Musei Nationalis Pragae*, *43*, 33–83.

Ježek, J. (2009). Psychodidae Newman 1834. Available via http://zoology.fns.uniba.sk/diptera.

Ježek, J. & Barták, M. (2000). *Psychodidae*. *Folia Facultatis Scientiarium Naturalium Universitatis Masarykianae Brunensis*, *104*, 93–100.

Jung, H. F. (1956). Beiträge zur Biologie, Morphologie und Systematik der europäischen Psychodiden (Diptera). *Deutsche Entomologische Zeitschrift*, *2*, 97–257.

Kumar, S., Skjaeveland, A., Orr, R. J., Enger, P., Ruden, T., Mevik, B. H., Burki, F., Botnen, A., Shalchian-Tabrizi, K. (2009). AIR: a batch-oriented web program package for construction of supermatrices ready for phylogenomic analyses. *BMC Bioinformatics*, *10*, 7.

Nixon, K. C. (1999). The parsimony ratchet, a new method for rapid parsimony analysis. *Cladistics*, *15*, 407–414.

Nylander, J. A. A. (2004). MrModeltest v2. Uppsala, Sweden: Program distributed by the author.

Nylander, J. A. A., Ronquist, F., Huelsenbeck, J. P. & Nieves Aldrey, J. L. (2004). Bayesian phylogenetic analysis of combined data. *Systematic Biology*, *53*, 47–67.

R Development Team (2009). *R: A language and environment for statistical computing, reference index version 2.2.1*. R Foundation for Statistical Computing, Vienna, Austria. Available via http://www.R-project.org.

Rambaut, A. & Drummond, A. J. (2004). *Tracer v1.4*. Edinburgh. Available via http://tree.bio.ed.ac.uk/software/tracer.

Ronquist, F. & Huelsenbeck, J. P. (2003). Mrbayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, *19*, 1572–1574.

Simmons, M. P. & Ochoterena, H. (2000). Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology*, *49*, 369–381.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H. & Flook, P. (1994). Evolution, weighting, and phylogenetic utility of mitochondrial gene-sequences and a compilation of conserved polymerase chain-reaction primers. *Annals of the Entomological Society of America*, *87*, 651–701.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, *22*, 2688–2690.

Swofford, D. L. (2003). *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4*. Sunderland, Massachusetts: Sinauer.

Vaillant, F. 1971. 9d. Psychodidae-Psychodinae. In E. Lindner (Ed.) *Die Fliegen der Palaearktischen Region* (p. 358). Stuttgart: E. Schweizerbart'sche Verlagsbuchhandlung.

Van de Peer, Y., Frickey, T., Taylor, J. S. & Meyer, A. (2002). Dealing with saturation at the amino acid level: a case study

based on anciently duplicated zebrafish genes. *Gene*, *295*, 205–211.

Wiegmann, B. M., Trautwein, M. D., Winkler, I. S., Barr, N. B., Kim, J. W., Lambkin, C., Bertone, M. A., Cassel, B. K., Bayless, K. M., Heimberg, A. M., Wheeler, B. M., Peterson, K. J., Pape, T., Sinclair, B. J., Skevington, J. H., Blagoderov, V., Caravas, J., Kutty, S. N., Schmidt-Ott, U., Kampmeier, G. E., Thompson, F. C., Grimaldi, D. A., Beckenbach, A. T., Courtney, G. W., Friedrich, M., Meier, R., Yeates, D. K. (2011). Episodic radiations in the fly tree of life. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 5690–5695.

Wiens, J. J. & Morrill, M. C. (2011). Missing data in phylogenetic analysis: reconciling results from simulations and empirical data. *Systematic Biology*, *60*, 719–731.

Withers, P. (1988). Moth flies. Diptera: Psychodidae. *Dipterist Digest*, *4*, 1–83.

Yang, Z. H. (1993). Maximum-likelihood-estimation of phylogeny from DNA-sequences when substitution rates differ over sites. *Molecular Biology and Evolution*, *10*, 1396–1401.

Yeates, D. K., Wiegmann, B. M., Courtney, G. W., Meier, R., Lambkin, C. & Pape, T. (2007). Phylogeny and systematics of Diptera: two decades of progress and prospects. *Zootaxa*, *1668*, 565–590.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Amplified regions and primer positions.

Fig. S2. Saturation plots for each partition (a: NADH1, b: CytB, c: trRNA) and for the total concatenated matrix (d).

Table S1. Regions amplified and GenBank Accession numbers for each sample.

Table S2. Support for each clade obtained by analysis of the different datasets, and mean NPM distance of the considered topology to all remaining trees.

Table S3. NPM distances calculated on the trees obtained for all the analysed datasets.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.