Systematic Entomology

Systematic Entomology (2018), **43**, 447–459 DOI: 10.1111/syen.12283

New data, same story: phylogenomics does not support Syrphoidea (Diptera: Syrphidae, Pipunculidae)

THOMAS PAULI[1,2], TREVOR O. BURT[2], KAREN MEUSEMANN[1,2,3], KEITH BAYLESS[4], ALEXANDER DONATH[2], LARS PODSIADLOWSKI[5], CHRISTOPH MAYER[2], ALEXEY KOZLOV[6], ALEXANDROS VASILIKOPOULOS[2], SHANLIN LIU[7,8,9], XIN ZHOU[10,11], DAVID YEATES[3], BERNHARD MISOF[2], RALPH S. PETERS[2] and XIMO MENGUAL[2]

[1]Institut für Biologie I (Zoologie), Albert-Ludwigs-Universität Freiburg, Freiburg, Germany, [2]Zoologisches Forschungsmuseum Alexander Koenig, Leibniz Institut für Biodiversität der Tiere, Bonn, Germany, [3]Australian National Insect Collection, CSIRO National Research Collections Australia, Canberra, Australia, [4]Department of Entomology, North Carolina State University, Raleigh, NC, U.S.A., [5]Institute of Evolutionary Biology and Ecology, Zoology and Evolutionary Biology, University of Bonn, Bonn, Germany, [6]Scientific Computing Group, Heidelberg Institute for Theoretical Studies, Heidelberg, Germany, [7]China National GeneBank-Shenzhen, BGI-Shenzhen, Shenzhen, China, [8]BGI-Shenzhen, Shenzhen, China, [9]Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark, [10]Beijing Advanced Innovation Center for Food Nutrition and Human Health, China Agricultural University, Beijing, China and [11]Department of Entomology, China Agricultural University, Beijing, China

Abstract. The Syrphoidea (families Pipunculidae and Syrphidae) has been suggested to be the sister group of the Schizophora, the largest species radiation of true flies. A major challenge in dipterology is inferring the phylogenetic relationship between Syrphoidea and Schizophora in order to understand the evolutionary history of flies. Using newly sequenced transcriptomic data of Syrphidae, Pipunculidae and closely related lineages, we were able to fully resolve phylogenetic relationships of Syrphoidea using a supermatrix approach with more than 1 million amino acid positions derived from 3145 genes, including 19 taxa across nine families. Platypezoidea were inferred as a sister group to Eumuscomorpha, which was recovered monophyletic. While Syrphidae were also found to be monophyletic, the superfamily Syrphoidea was not recovered as a monophyletic group, as Pipunculidae were inferred as sister group to Schizophora. Within Syrphidae, the subfamily Microdontinae was resolved as sister group to the remaining taxa, Syrphinae and Pipizinae were placed as sister groups, and the monophyly of Eristalinae was not recovered. Although our results are consistent with previously established hypotheses on Eumuscomorphan evolution, our approach is new to dipteran phylogeny, using larger-scale transcriptomic data for the first time for this insect group.

Introduction

Flies (Diptera) have been the focus of many studies due to their roles as biological control agents, pollinators, plant pests, organic matter decomposers or vectors of human and animal diseases, and their value in postmortem interval estimation in forensic science and model organisms for genetics (Pape, 2009). During the 20th century, extraordinary advances were made concerning the inference of phylogenetic relationships between Diptera (reviewed by Yeates & Wiegmann, 2012), and in the early part of the 21st century major goals have been accomplished by collaborative projects of dipterologists (see, e.g., the FLYTREE project, http://wwx.inhs.illinois.edu/research/flytree/). New molecular tools and morphological

Correspondence Thomas Pauli, Institute for Biologie I (Zoology), University of Freiburg, Hauptstraße 1, 79104 Freiburg (Breisgau), Germany. E-mail: tpauli.bio@gmail.com; Ximo Mengual, Zoological Research Museum Alexander Koenig, Adenauerallee 160, 53113 Bonn, Germany. E-mail: x.mengual@leibniz-zfmk.de

methods generated a more robust general evolutionary scenario for Diptera (Wiegmann *et al.*, 2011). For example, there is strong support and a large number of putative synapomorphies for Brachycera, Cyclorrhapha, Schizophora, and Calyptratae (e.g. Hennig, 1971, 1973; Griffiths, 1972; McAlpine, 1989; Woodley, 1989; Sinclair, 1992; Cumming *et al.*, 1995; Yeates & Wiegmann, 1999; Collins & Wiegmann, 2002; Wiegmann *et al.*, 2003, 2011; Yeates *et al.*, 2007 and references therein; Rotheray & Gilbert, 2008; Lambkin *et al.*, 2013).

However, we are still lacking crucial information to understand the evolution of important morphological characters and biological strategies, most notably a robust hypothesis regarding the phylogenetic relationships of the major lineages of a paraphyletic grade that is now commonly referred to as 'lower Cyclorrhapha', previously named 'Aschiza' (Griffiths, 1972; Hennig, 1973; Wada, 1991; Cumming *et al.*, 1995; Collins & Wiegmann, 2002; Moulton & Wiegmann, 2004; Wiegmann *et al.*, 2011; Lambkin *et al.*, 2013). The evolution of 'lower Cyclorrhapha' has been considered one of the most challenging riddles in dipterology regarding family-level relationships. Schizophora (flies with ptilinal fissure or ptilinum and a full circumversion of the male genitalia completed within the puparium) comprise more than half of the family-level diversity in Diptera (Yeates & Wiegmann, 1999), with some 80 recognized families (Lambkin *et al.*, 2013), and account for more than a third of extant fly diversity and 3% of all animal diversity (Wiegmann *et al.*, 2011). Moreover, they represent one of the episodic rapid radiations within Diptera characterized by very low extinction rates compared with immediately adjacent lineages (Wiegmann *et al.*, 2011). Thus, to infer the phylogenetic relationships between 'lower Cyclorrhapha' and Schizophora is a major step towards understanding the evolutionary history of flies.

'Lower Cyclorrhapha' families have traditionally been grouped into two superfamilies: Syrphoidea and Platypezoidea (including Opetiidae, Platypezidae, Lonchopteridae, Ironomyiidae, and Phoridae) (McAlpine, 1989; Cumming *et al.*, 1995; Grimaldi & Cumming, 1999; Collins & Wiegmann, 2002; Moulton & Wiegmann, 2004; Yeates & Wiegmann, 2005; Wiegmann *et al.*, 2011). Syrphoidea comprise only two families: Syrphidae (hoverflies or flower flies) and Pipunculidae (big-headed flies). The name was originally proposed by Coquillett (1901) and included other families, i.e. Platypezidae and Conopidae, but subsequent authors redefined it into the current concept providing several autapomorphies (Crampton, 1944; Griffiths, 1972; Hennig, 1976; McAlpine, 1989). Syrphoidea have been considered the sister clade of Schizophora (McAlpine, 1989; Cumming *et al.*, 1995; Zatwarnicki, 1996; Yeates & Wiegmann, 1999; Yeates *et al.*, 2007; Woodley *et al.*, 2009), together forming the clade Eumuscomorpha with evident synapomorphies (Wada, 1991; Cumming *et al.*, 1995; Zatwarnicki, 1996; Yeates & Wiegmann, 1999; Collins & Wiegmann, 2002; Rotheray & Gilbert, 2008). The sister group relationship between Syphidae and Pipunculidae has been contradicted by molecular analyses, resolving the Syrphidae and Pipunculidae into separate lineages, i.e. suggesting paraphyly of Syrphoidea (Collins & Wiegmann, 2002; Moulton & Wiegmann, 2004; Wiegmann *et al.*, 2011; Young *et al.*, 2016). In contrast, Skevington & Yeates (2000) studied the syrphoidean relationships with 12S and 16S rDNA data and results showed that Syrphoidea are monophyletic, although the number of included outgroups was limited. More recently, Tachi (2014) proposed a novel interpretation for the evolution of the metathorax and established a morphological synapomorphy for Pipunculidae + Schizophora.

Flower flies (Syrphidae) have traditionally been divided into three subfamilies based on adult morphological characters, i.e. Microdontinae, Eristalinae and Syrphinae (Vockeroth & Thompson, 1987; Thompson & Rotheray, 1998). Pipizinae, however, have recently been elevated to subfamily level, based on molecular, morphological and biological data (Mengual *et al.*, 2015). Recent phylogenetic studies on Syrphidae resolve Microdontinae as sister group of the remaining flower flies, with Pipizinae and Syrphinae forming one clade and inferring Eristalinae as paraphyletic (Skevington & Yeates, 2000; Ståhls *et al.*, 2003; Mengual *et al.*, 2015; Young *et al.*, 2016).

Based on large-scale transcriptomic data and using the workflow recently established for insect phylogenomics in the international research initiative 1KITE (Misof *et al.*, 2014; Peters *et al.*, 2017), we explore the phylogenetic relationships of the 'lower Cyclorrhapha' with Schizophora and of the subgroups of Syrphidae in order to resolve two major questions: (i) are Syrphoidea monophyletic; and (ii) are our results congruent with previous studies regarding the classification of syrphid subfamilies? The results of the present study will provide more robust and stronger evidence to understand the phylogenetic relationships of these groups and might help to shed light on the evolution of Diptera.

Material and methods

Taxonomic sampling

The taxon sampling was chosen in order to cover as much taxonomic diversity from the 'lower Cyclorrhapha' families as possible, as well as representatives of all syrphid subfamilies. A total of 19 taxa were studied in the present analysis, which are all the representatives of the studied families that were sequenced *de novo* within the 1KITE project. We included *Nephrocerus atrapilus* Skevington (Pipunculidae), and members of the syrphid subfamilies Syrphinae [i.e. *Baccha elongata* (Fabricius), *Melanostoma scalare* (Fabricius), *Episyrphus balteatus* (De Geer), *Leucozona lucorum* (Linnaeus)], Pipizinae [*Pipiza noctiluca* (Linnaeus)], Eristalinae [*Ferdinandea cuprea* (Scopoli), *Syritta pipiens* (Linnaeus), *Merodon equestris* (Fabricius), *Eristalis pertinax* (Scopoli)] and Microdontinae [*Archimicrodon brachycerus* (Knab & Malloch)]. We constrained *Heteropsilopus ingenuus* (Erichson) (Eremoneura: Empidoidea: Dolichopodidae) as the root in our phylograms, and used three Platypezoidea species as outgroup taxa based on the results of Wiegmann *et al.* (2011), i.e. *Lonchoptera bifurcata* (Fallén) (Lonchopteridae), *Platypeza anthrax* Loew (Platypezidae) and *Megaselia abdita* Schmitz (Phoridae). We also included a few schizophoran taxa, namely *Sapromyza sciomyzina* Schiner (Lauxaniidae), *Zacompsia fulva* Coquillett

(Uliidae), *Lenophila dentipes* (Guérin-Méneville) (Platystomatidae) and *Meroplius fasciculatus* (Brunetti) (Sepsidae). We did not include any representative of the Calyptratae as they have been proven to be a well-supported monophyletic group within Schizophora (Kutty *et al.*, 2010; Wiegmann *et al.*, 2011; Lambkin *et al.*, 2013). Information regarding specimen collection, preservation, number of specimens and sexes is listed in Table 1 and Table S1. Some of the analysed species are illustrated in Fig. 1. Following Meier (2017), the bibliographic references of identification methods for the studied taxa are also given in Table 1 and Table S1.

Specimens were ground alive in RNA*later*® stabilization solution (Invitrogen, Thermo Fisher Scientific, Carlsbad, CA, U.S.A.) after identification. Thus, no parts of specimens were kept as morphological vouchers. Nevertheless, the so-called COI barcodes (the 5′ region of the mitochondrial cytochrome *c* oxidase subunit I gen) may be retrieved from the transcriptomes and might be used as surrogate vouchers for identification purposes.

Transcriptome sequencing, assembly and contamination check

RNA extraction, NGS library preparation and sequencing of the prepared libraries on Illumina HiSeq sequencers were carried out following the protocols given by Peters *et al.* (2017). cDNA libraries were paired-end sequenced on Illumina HiSeq2000 sequencing platforms (Illumina Inc., San Diego, CA, U.S.A.) with read length of 150 bp. Per species, we collected about 2.5 Gbp of raw sequence data. All raw reads were quality-controlled, trimmed, assembled and screened for possible contaminant sequences (which were then removed) as described by Peters *et al.* (2017) (see Table 2 for the amount of removed contaminants from each species dataset). Both the raw reads and the assembled transcriptomes are archived at the National Center for Biotechnology Information (NCBI) under the Umbrella BioProject ID PRJNA183205 (‘The 1KITE project: evolution of insects’). For a full list of accession numbers, see Table 3.

Identification of single-copy genes in the sequenced transcriptomes

An orthologue set was generated from official gene sets of fully sequenced genomes based on OrthoDB version 7 (Waterhouse *et al.*, 2013). We requested single-copy genes with the hierarchical level at Mecopterida from the following species: *Bombyx mori* Linnaeus (Lepidoptera: Bombycidae), *Danaus plexippus* (Linnaeus) (Lepidoptera: Nymphalidae), *Aedes aegypti* (Linnaeus) (Diptera: Culicidae), *Drosophila melanogaster* Macquart (Diptera: Drosophilidae) and *Glossina morsitans* Westwood (Diptera: Glossinidae). This set of species will be called ‘reference species’ from here on. For all remaining species, the number of copies was set to unknown (‘?’). The orthologue set contained 3145 single-copy orthologue genes (OGs) present in all five reference species. Official gene sets on amino acid and nucleotide level were adjusted accounting for only the longest isoform per gene in the orthologue set when isoform information was known; otherwise the whole set of isoform sequences was used (see detailed information in Table 4). We then used ORTHOGRAPH (Petersen *et al.*, 2017) version 0.5.9 (February 2016, https://github.com/mptrsen/Orthograph/) to build a database of orthologues and subsequently infer orthologue sequences in our transcriptome data. We used default settings except for ‘substitute-u-with X’ (activated) and ‘extend-orf’ (see the ORTHOGRAPH manual for details). The number of orthologues found as well as statistics on the length of orthologue sequences are given in Table S2. When summarizing ORTHOGRAPH results, we replaced putative internal stop-codons and the amino acid Selenocysteine U with ‘X’ (amino acid level) or ‘NNN’ (nucleotide level) with a custom-made Perl script. We deposited the orthologue set on Mendeley (https://doi.org/10.17632/b5bdz64sp2.2).

Alignment of single-copy genes in the sequenced transcriptomes

The amino acid sequences of each of the 3145 OGs were aligned using the L-INS-i algorithm of MAFFT version 7.123 (Katoh & Standley, 2013). We then followed the procedure of Misof *et al.* (2014) in order to assess the quality of the amino acid multiple sequence alignments (MSAs). We performed alignment refinement of identified outlier sequences, and then removed final outliers from both amino acid MSAs and nucleotide sequences. We also removed the reference species from all amino acid MSAs and nucleotide files accordingly, and then all columns exclusively containing gaps or ‘X’ in each MSA. Next, we used a modified version of PAL2NAL version 14 (Suyama *et al.*, 2006; Misof *et al.*, 2014) in order to generate the corresponding nucleotide MSAs with the amino acid MSAs as blueprint.

Downstream analyses

We checked the amino acid MSAs of each OG for ambiguously aligned regions with the software ALISCORE version 1.2 (Misof & Misof, 2009; Kück *et al.*, 2010; Misof *et al.*, 2014). We used the –e option for gappy datasets and forced a comparison of all sequence pairs. Apart from these options, default parameters were used. Simultaneously, we annotated protein domains within the amino acid sequences with the aid of the Pfam-A database release 28.0 and the program PFAMSCAN.PL version 1.5, released 2013 (Finn *et al.*, 2014), which makes use of HMMSCAN, from the HMMER software package version 3.1b1 (Eddy, 1998). Pfam-A includes information on protein domains, families and clans. We received coordinates for protein domains, families, clans and unannotated regions (voids) at the amino acid and nucleotide levels for each OG. We then merged information from the domain annotation with information on ambiguously aligned regions, which we finally removed with custom-made Perl scripts (Misof *et al.*, 2014). We generated two corresponding supermatrices (amino acid and nucleotide levels)

Fig. 1. Legend on next page. [Colour figure can be viewed at wileyonlinelibrary.com].

Table 1. Sampled species included in this study.

Family	Genus	Species	Author	Identification method or reference
Syrphidae	*Archimicrodon*	*brachycerus*	(Knab & Malloch, 1912)	Direct comparison with material in the Australian National Insect Collection.
Syrphidae	*Baccha*	*elongata*	(Fabricius, 1775)	Van Veen (2004)
Syrphidae	*Episyrphus*	*balteatus*	(De Geer, 1776)	Van Veen (2004)
Syrphidae	*Eristalis*	*pertinax*	(Scopoli, 1763)	Van Veen (2004)
Syrphidae	*Ferdinandea*	*cuprea*	(Scopoli, 1763)	Van Veen (2004), Ricarte *et al.* (2010).
Dolichopodidae	*Heteropsilopus*	*ingenuus*	(Erichson, 1842)	Direct comparison with material in the Australian National Insect Collection.
Platystomatidae	*Lenophila*	*dentipes*	(Guérin-Méneville, 1843)	Direct comparison with material in the Australian National Insect Collection.
Syrphidae	*Leucozona*	*lucorum*	(Linnaeus, 1758)	Van Veen (2004)
Lonchopteridae	*Lonchoptera*	*bifurcata*	(Fallén, 1810)	Klymko & Marshall (2008)
Phoridae	*Megaselia*	*abdita*	Schmitz, 1959	Reared from laboratory culture
Syrphidae	*Melanostoma*	*scalare*	(Fabricius, 1794)	Van Veen (2004)
Syrphidae	*Merodon*	*equestris*	(Fabricius, 1794)	Van Veen (2004)
Sepsidae	*Meroplius*	*fasciculatus*	(Brunetti, 1909)	Reared from laboratory culture
Pipunculidae	*Nephrocerus*	*atrapilus*	Skevington, 2005	Skevington (2005)
Syrphidae	*Pipiza*	*noctiluca*	(Linnaeus, 1758)	Vujić *et al.* (2013)
Platypezidae	*Platypeza*	*anthrax*	Loew, 1870	Direct comparison with original descriptions, museum specimens and known distribution
Lauxaniidae	*Sapromyza*	*sciomyzina*	Schiner, 1868	Direct comparison with material in the Australian National Insect Collection.
Syrphidae	*Syritta*	*pipiens*	(Linnaeus, 1758)	Lyneborg & Barkemeyer (2005)
Uliidae	*Zacompsia*	*fulva*	Coquillett, 1901	Steyskal (1987)

with data blocks based on both protein domains/families/clans when available and based on gene boundaries for voids. Finally, we used MARE version 0.1.2-rc (Misof *et al.*, 2013) on the amino acid supermatrix to exclude all data blocks with information content (IC) of zero from both the amino acid and nucleotide supermatrix (see Misof *et al.*, 2014). A heat map of the IC per data block is shown in Fig. S1 before and after removal of data blocks with IC = 0. Finally, the datasets contain 3800 data blocks spanning an alignment length of 1 040 563 sites at the amino acid level and 3 121 689 sites at the nucleotide level. We refer to these datasets as 'original datasets' in the following text. Additionally, we generated an optimized dataset (also known as a 'decisive dataset'; Dell'Ampio *et al.*, 2014; Misof *et al.*, 2014) which includes only data blocks for which sequences of all eumuscomorphan taxa and at least one outgroup taxon were present to avoid a possible bias from nonrandomly distributed missing data (see Misof *et al.*, 2013; Dell'Ampio *et al.*, 2014). In order to generate the optimized dataset, 1900 data blocks were removed from the original dataset. Consequently, 1900 data blocks remained in the optimized data set, comprising a total of 678 763 amino acid columns.

We computed the species-pairwise amino acid site coverage for the original dataset and the optimized dataset with ALISTAT version 1.6 (available on https://github.com/thomaskf/AliStat). A heat map of the coverage of sequence pairs is shown for both datasets (Fig. S2).

Optimized partitioning scheme and selection of substitution models

We used the software PARTITIONFINDER version 2.1.1 (Lanfear *et al.*, 2016) to search for an optimal partitioning scheme and the best-fitting substitution models for both the original and the optimized amino acid dataset. We included the following amino acid substitution models in the search for the best-fitting model: BLOSUM62 (Henikoff & Henikoff, 1992), DCMUT (Kosiol & Goldman, 2005), LG (Le & Gascuel, 2008), JTT (Jones *et al.*, 1992), WAG (Whelan & Goldman, 2001) and LG4X (Le *et al.*,

Fig. 1. Images of some species analysed in the present study. (a) *Heteropsilopus ingenuus*, by Simon Grove (Tasmanian Museum & Art Gallery); (b) *Lonchoptera bifurcata*, by Steve Marshall (University of Guelph); (c) *Platypeza anthrax*, by Tom Murray (Groton, MA, USA); (d) *Nephrocerus atrapilus*, by Jeff. H. Skevington (Canadian National Collections); (e) *Zacompsia fulva*, by Graham Montgomery (Ithaca, NY, USA); (f) *Megaselia abdita*, by Karl Wotton (University of Exeter); (g) *Baccha elongata*, by Marion Friedrich (https://arthropodafotos.de); (h) *Pipiza noctiluca*, by Marion Friedrich (https://arthropodafotos.de); (i) *Eristalis pertinax*, by Marion Friedrich (https://arthropodafotos.de); (j) *Leucozona lucorum*, by Marion Friedrich (https://arthropodafotos.de); (k) *Episyrphus balteatus*, by Marion Friedrich (https://arthropodafotos.de); (l) *Ferdinandea cuprea*, by Marion Friedrich (https://arthropodafotos.de); (m) *Meroplius fasciculatus*, by Sepsidnet (sepsidnet-rmbr.nus.edu.sg, version 05/2013); (n) *Sapromyza sciomyzina*, by Shaun Winterton (California Department of Food and Agriculture); and (o) *Merodon equestris*, by Marion Friedrich (https://arthropodafotos.de). [Colour figure can be viewed at wileyonlinelibrary.com].

Table 2. Information on sequences removed during various contamination filtering steps.

		Number of sequences				Number of sequences			Transcriptome size (bp)			
Library ID	Species	After assembly	After local VecScreen[a]	After cross-contamination screen	Cross-contamination (%)	Filtered out by NCBI	NCBI filter (%)	Published	After assembly	Published	Removed (bp)	Removed (%)
INShkeTBBRAAPEI-57	*Archimicrodon brachycerus*	38 516	38 514	38 438	0	0	0.00	38 438	31 665 803	31 623 146	42 657	0.13
INSinlTBCRAAPEI-71	*Baccha elongata*	20 734	20 731	20 454	1	0	0.00	20 454	19 296 774	19 172 757	124 017	0.64
INSnfrTAWRAAPEI-11	*Episyrphus balteatus*	27 850	27 843	27 726	0	0	0.00	27 726	23 724 833	23 662 281	62 552	0.26
INSofmTCQRAAPEI-55	*Eristalis pertinax*	25 819	25 818	25 484	1	0	0.00	25 484	29 504 948	29 345 114	159 834	0.54
RINSinlTCKRAAPEI-84	*Ferdinandea cuprea*	16 960	16 957	16 253	4	0	0.00	16 253	16 145 965	15 595 385	550 580	3.41
INSswpTAIRAAPEI-19	*Heteropsilopus ingenuus*	42 069	42 059	41 994	0	0	0.00	41 994	28 959 014	28 910 029	48 985	0.17
INSobdTCQRAAPEI-33	*Lenophila dentipes*	22 376	22 377	22 220	1	0	0.00	22 220	24 630 453	24 560 553	69 900	0.28
INSinlTATRAAPEI-37	*Leucozona lucorum*	23 539	23 535	23 334	1	75	0.32	23 259	22 876 283	22 719 511	156 772	0.69
INSinlTAWRAAPEI-44	*Lonchoptera bifurcata*	31 098	31 084	30 186	3	125	0.41	30 061	16 980 873	16 617 294	363 579	2.14
RINSinlTBGRAAPEI-126	*Megaselia abdita*	35 081	35 078	33 812	4	294	0.87	33 518	18 583 506	18 123 224	460 282	2.48
RINSinlTBKRAAPEI-17	*Melanostoma scalare*	21 908	21 907	21 138	4	0	0.00	21 138	18 816 279	18 390 919	425 360	2.26
RINSinlTDARAAPEI-71	*Merodon equestris*	24 713	24 708	24 087	3	12	0.05	24 075	30 881 213	30 422 312	458 901	1.49
INSytvTAARAAPEI-9	*Meroplius fasciculatus*	32 242	32 242	32 156	0	12	0.04	32 144	31 668 358	31 624 136	44 222	0.14
INSobdTEHRAAPEI-44	*Nephrocerus atrapilus*	25 221	25 217	25 064	1	0	0.00	25 064	23 841 976	23 773 606	68 370	0.29
INSinlTAKRAAPEI-21	*Pipiza noctiluca*	29 415	29 417	28 617	3	32	0.11	28 585	25 926 727	25 356 664	570 063	2.20
RINSinlTBORAAPEI-22	*Platypeza anthrax*	12 016	12 016	11 959	0	0	0.00	11 959	17 850 390	17 824 408	25 982	0.15
INSerlTATRABPEI-16	*Sapromyza sciomyzina*	32 219	32 218	32 135	0	0	0.00	32 135	26 756 277	26 709 610	46 667	0.17
RINSinlTDHRAAPEI-93	*Syritta pipiens*	24 052	24 049	23 636	2	83	0.35	23 553	17 249 261	17 046 115	203 146	1.18
RINSinlTCRRAAPEI-37	*Zacompsia fulva*	27 480	27 471	27 036	2	7	0.03	27 029	25 322 071	25 141 355	180 716	0.71

[a]Note that the number of sequences can increase after local VecScreen because of sequences that are split into two parts by removing the contaminant.
NCBI, National Center for Biotechnology Information.

Table 3. National Center for Biotechnology Information (NCBI) accession numbers of the sequenced and assembled transcriptomes.

Library ID	Species	NCBI taxonomy ID	BioProject accession	BioSample accession	Experiment accession	Run accession	TSA project accession	TSA version
INShkeTBBRAAPEI-57	*Archimicrodon brachycerus*	1572511	267912	SAMN03223074	SRX798043	SRR1695319	GCFH00000000	GCFH01000000
INSinlTBCRAAPEI-71	*Baccha elongata*	226178	267916	SAMN03223078	SRX798047	SRR1695323	GCFG00000000	GCFG01000000
INSnfrTAWRAAPEI-11	*Episyrphus balteatus*	286459	267940	SAMN03223102	SRX798065	SRR1695341	GCFM00000000	GCFM01000000
INSofmTCQRAAPEI-55	*Eristalis pertinax*	1572519	267941	SAMN03223103	SRX798066	SRR1695342	GCGL00000000	GCGL01000000
RINSinlTCKRAAPEI-84	*Ferdinandea cuprea*	226145	267945	SAMN03223107	SRX798069	SRR1695345	GCHQ00000000	GCHQ01000000
INSswpTAIRAAPEI-19	*Heteropsilopus ingenuus*	188255	267950	SAMN03223112	SRX798074	SRR1695350	GCGO00000000	GCGO01000000
INSobdTCQRAAPEI-33	*Lenophila dentipes*	1572560	267953	SAMN03223115	SRX798077	SRR1695353	GCFT00000000	GCFT01000000
INSinlTATRAAPEI-37	*Leucozona lucorum*	414824	267954	SAMN03223116	SRX798078	SRR1695354	GCME00000000	GCME01000000
INSinlTAWRAAPEI-44	*Lonchoptera bifurcata*	385268	267957	SAMN03223119	SRX798081	SRR1695357	GCMR00000000	GCMR01000000
RINSinlTBGRAAPEI-126	*Megaselia abdita*	88686	267960	SAMN03223122	SRX798084	SRR1695360	GCMU00000000	GCMU01000000
RINSinlTBKRAAPEI-17	*Melanostoma scalare*	92598	267962	SAMN03223124	SRX798086	SRR1695362	GCGT00000000	GCGT01000000
RINSinlTDARAAPEI-71	*Merodon equestris*	511117	267963	SAMN03223125	SRX798087	SRR1695363	GCMV00000000	GCMV01000000
INSytvTAARAAPEI-9	*Meroplius fasciculatus*	1572525	267964	SAMN03223126	SRX798088	SRR1695364	GCNI00000000	GCNI01000000
INSobdTEHRAAPEI-44	*Nephrocerus atrapilus*	1572530	267971	SAMN03223133	SRX798095	SRR1695371	GCGE00000000	GCGE01000000
INSinlTAKRAAPEI-21	*Pipiza noctiluca*	1162231	267984	SAMN03223147	SRX798105	SRR1695381	GCMW00000000	GCMW01000000
RINSinlTBORAAPEI-22	*Platypeza anthrax*	1572536	267985	SAMN03223148	SRX798106	SRR1695382	GCGU00000000	GCGU01000000
INSerlTATRABPEI-16	*Sapromyza sciomyzina*	1572539	267992	SAMN03223155	SRX798113	SRR1695389	GCEW00000000	GCEW01000000
RINSinlTDHRAAPEI-93	*Syritta pipiens*	34682	268001	SAMN03223164	SRX798121	SRR1695397	GCND00000000	GCND01000000
RINSinlTCRRAAPEI-37	*Zacompsia fulva*	1572576	268007	SAMN03223170	SRX798127	SRR1695403	GCNH00000000	GCNH01000000

TSA, Transcriptome shotgun assembly.

Table 4. Official gene sets (OGS) used for the orthologue set and the reciprocal search within ORTHOGRAPH.

Order	Family	Species	OGS version	Filename	Adjusted number of genes included	Download location	Download date
Diptera	Culicidae	*Aedes aegypti*	AAEGY1.3	Corresp-Aedes-CDS.fa	16 957	Vectorbase	March 2015
Diptera	Culicidae	*Aedes aegypti*	AAEGY1.3	Corresp-Aedes-pep.fa	16 957	Vectorbase	March 2015
Diptera	Drosophilidae	*Drosophila melanogaster*	DMELA5.51	Corresp-dmel-CDS.fa	13 969	Flybase	March 2015
Diptera	Drosophilidae	*Drosophila melanogaster*	DMELA5.51	Corresp-dmel-pep.fa	13 969	ftp://cegg.unige.ch/OrthoDB7/	March 2015
Diptera	Glossinidae	*Glossina morsitans*	GMORS1.1	Corresp-Glossina-CDS.fa	12 406	Vectorbase	March 2015
Diptera	Glossinidae	*Glossina morsitans*	GMORS1.1	Corresp-Glossina-pep.fa	12 406	Vectorbase	March 2015
Lepidoptera	Bombycidae	*Bombyx mori*	BMORI2.0	Corresp-bmori-CDS.fa	14 623	SilkDB	March 2015
Lepidoptera	Bombycidae	*Bombyx mori*	BMORI2.0	Corresp-bmori-pep.fa	14 623	SilkDB	March 2015
Lepidoptera	Nymphalidae	*Danaus plexippus*	DPLEX 2.0	Corresp-Danaus-CDS.fa	15 129	Monarchbase	March 2015
Lepidoptera	Nymphalidae	*Danaus plexippus*	DPLEX 2.0	Corresp-Danaus-pep.fa	15 129	Monarchbase	March 2015

For amino acid level, available peptide files and for nucleotide level, respective CDS files (coding sequences) were downloaded and adjusted.

2012). This was done to avoid substitution models that are specific for HIV, chloroplasts and mitochondria. We included parameters of among-site variation (+G) and parameters of amino acid frequency estimation (+F). We used log-likelihood scores from RAXML version 8.2.7 (Stamatakis, 2014) with the corrected Akaike information criterion (Hurvich & Tsai, 1989) to let PARTITIONFINDER combine data blocks and to select the best-fitting model for each final partition. Maximum likelihood (ML) trees were used as starting trees for the search of the best log-likelihood scores. Branch lengths were set to 'linked'. We used the heuristic rcluster search algorithm (Lanfear *et al.*, 2014) and used the overall subset scaling parameters, base frequencies and alpha parameter of the gamma distribution as clustering parameters (-weights 1,1,0,1). The parameter rcluster-max was set to 10 595 and rcluster-percent was set to 100 for a highly exhaustive analysis. For the original nucleotide dataset, we modelled within each partition the first, second and third codon positions separately from each other to account for different rates of heterogeneity among codon positions. We exclusively applied the substitution model GTR + G, as other nucleotide models are not available in EXAML. We deposited the supermatrices of the original amino acid dataset, optimized amino acid dataset and original nucleotide dataset alongside the respective partitioning schemes on Mendeley (https://doi.org/10.17632/b5bdz64sp2.2).

Phylogenetic tree inference

We inferred phylogenetic trees with a partitioned analyses approach as implemented in the software EXAML version 3.0.17 (Kozlov *et al.*, 2015) on the original amino acid and corresponding nucleotide dataset, and on the optimized amino acid dataset. We performed 50 tree searches for each dataset, with 25 maximum parsimony starting trees and 25 random starting trees. Bootstrap support was derived from 50 slow, nonparametric bootstrap replicates. We checked for bootstrap convergence *a posteriori* according to the bootstopping criterion (Pattengale *et al.*, 2010) implemented in RAXML version 8.2.7 (weighted Robinson Fould distance building an extended

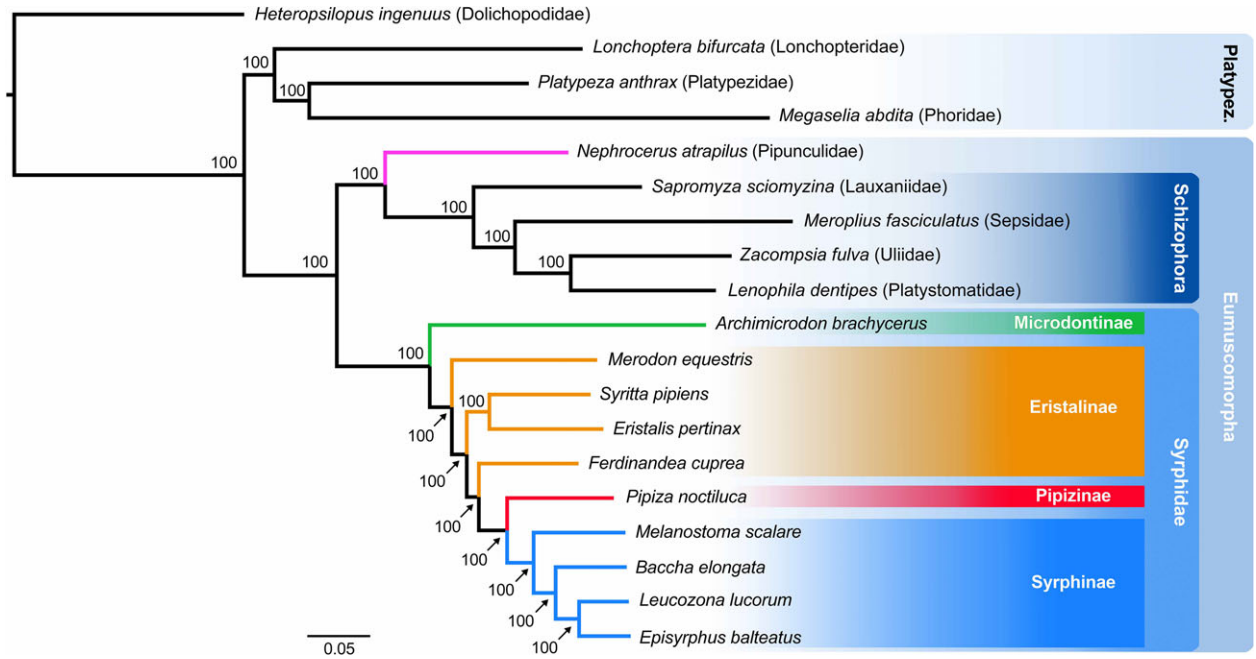

Fig. 2. Best scoring Maximum Likelihood tree based on the original amino-acid dataset (1,040,563 amino-acid sites with 3,184 partitions). Bootstrap support values are depicted at the nodes. [Colour figure can be viewed at wileyonlinelibrary.com].

majority-rule (MRE) consensus tree; autoMRE, threshold 0.03, with 1000 permutations; Stamatakis, 2014). Subsequently, bootstrap support was mapped onto the best ML tree. The tree was drawn with the software FIGTREE version 1.4.2 (Rambaut, 2009) and rooted with *H. ingenuus*. We graphically edited the tree with PHOTOSHOP CS5 version 12.1.

Results

Phylogenetic relationships

All inferred ML trees show a bootstrap support of 100% for all clades (Fig. 2). Results were fully congruent between analyses of the original amino acid dataset (spanning an alignment length of 1 040 563 sites with 3184 partitions; Fig. 2) and the optimized amino acid dataset (alignment length 678 763 sites, 1500 partitions; Fig. S3). The inferred phylogeny based on the original nucleotide sequence dataset (alignment length 3 121 689 sites; 9552 partitions, Fig. S4) differed from the amino acid datasets in the placement of *L. bifurcata* (Lonchopteridae), which was resolved as sister group of the remaining taxa, besides the root (*H. ingenuus*, Dolichopodidae). For additional information on intermediate steps, see 'Extended results' in the Supplementary Information File S1.

Members of Platypezoidea were resolved as a clade that is sister group to the monophyletic Eumuscomorpha in both topologies based on amino acid datasets (Fig. 2 and Fig. S3). Within Eumuscomorpha, schizophoran families were grouped together. Monophyly of Syrphoidea was not recovered with any of our datasets, as Pipunculidae were resolved as the sister group of Schizophora. Within flower flies, *A. brachycerus* was placed as the sister group of the remaining studied syrphids, and *Pipiza noctiluca* and Syrphinae were found to be sister groups. However, *M. scalare* and *B. elongata* (members of the tribe Bacchini, Syrphinae) were not recovered together. The monophyly of Eristalinae was not recovered in the present analysis, as members of the eristaline tribes Merodontini (*M. equestris*), Milesiini (*S. pipiens*), Eristalini (*E. pertinax*) and Rhingiini (*F. cuprea*) were not placed in a common clade.

Discussion

The present results are highly concordant with previous molecular studies (Collins & Wiegmann, 2002; Moulton & Wiegmann, 2004; Wiegmann *et al.*, 2011; Young *et al.*, 2016), and support the interpretation of the metathorax evolution suggested by Tachi (2014) based on the identity of metapleural elements.

The outgroup topology differs between phylograms inferred from amino acid (Fig. 2 and Fig. S3) and nucleotide datasets (Fig. S4). A possible explanation for this might be the heterogeneity of the third codon position, the overparametrization of the used nucleotide model (i.e. GTR), or the underparametrization of the selected amino acid models. We chose the more conservative analysis based on the amino acid dataset to display our results as the monophyly of Platypezoidea is consistent with other studies (Wiegmann *et al.*, 2011).

These results indicate that a reinterpretation of the morphological autapomorphies of Syrphoidea is necessary. Among those morphological autapomorphies suggested by previous authors were: acrostichal and dorsocentral setae differentiated (Grimaldi & Cumming, 1999), frons without macrosetae, sixth and seventh

abdominal segments asymmetrically developed on the left side with reduced terga, eighth sternum enlarged and asymmetrical (Griffiths, 1972), hypopygium strongly deflexed under right side of the abdomen and directed anteriorly (Griffiths, 1972; Cumming *et al.*, 1995; Zatwarnicki, 1996), puparium more or less globose and with a peculiar operculum and cleavage lines, wing with apices of veins R_{4+5} and M joined or nearly so (McAlpine, 1989), and larvae with antennae and maxillary organs at the apex of a pair of fleshy projections arising from the apicodorsal margin of the oral pocket (Rotheray & Gilbert, 2008).

The phylogenetic placement of Syrphidae as the sister group of Pipunculidae + Schizophora may have implications for our interpretation of the evolution of embryonic development in Diptera, as suggested by Wiegmann *et al.* (2011). In schizophoran flies such as house flies, blow flies, and vinegar flies (Muscidae, Calliphoridae, and Drosophilidae), the *bicoid* gene is expressed in an anterior-to-posterior gradient and establishes anteroposterior (AP) polarity of the embryo (Johnston & Nüsslein-Volhard, 1992; Sommer & Tautz, 1991; Schröder & Sander, 1993). Stauber *et al.* (1999, 2000) showed that *bicoid* is also present in nonschizophoran cyclorrhaphan flies, i.e. *M. abdita* (Phoridae), and Lemke *et al.* (2008) reported the occurrence of *bicoid* in two additional nonschizophoran cyclorrhaphan families, Lonchopteridae and Platypezidae. AP polarity of the *E. balteatus* (Syrphidae) embryo appears to be determined by two distinct factors at the anterior pole, but not *bicoid* (Lemke & Schmidt-Ott, 2009). *Episyrphus* has some ancestral traits of early embryonic development not present in other cyclorrhaphan flies, i.e. an anterodorsal serosa anlage (mid-dorsal in *Drosophila*) (Rafiqi *et al.*, 2008), a strong influence of *caudal* on the anteroposterior axis, the lack of *bicoid* to establish the AP polarity in the embryo (as in the beetle *Tribolium*), and *hunchback* expression in the serosa anlage, which has been reported for noncyclorrhaphan insects but is absent in *Drosophila*, *Musca* and *Megaselia* (Lemke & Schmidt-Ott, 2009). The current body of evidence is not sufficient to make precise statements regarding the evolution of embryonic development, and embryos from additional syrphid species are necessary to better understand the establishment of the AP polarity, as only the embryo of *E. balteatus* has been studied. But based on the current limited data, the most likely scenario is that syrphids may have undergone loss of several typical cyclorrhaphan developmental features or that there is a reversal of an ancestral mode of development in *Episyrphus* (Wiegmann *et al.*, 2011). In a similar rationale, Lemke *et al.* (2009) speculated that *Episyrphus* retained the ancestral cyclorrhaphan mechanism of AP axis specification.

Flower fly subfamilies were resolved in agreement with previous studies. The placement of Microdontinae, represented by *A. brachycerus*, as sister group of all remaining syrphids (Fig. 2) was originally suggested by Thompson (1969) using morphology and a Hennigian argumentation. This phylogenetic placement of microdontines was later recovered with only molecular data or with molecular and morphological data combined (Skevington & Yeates, 2000; Ståhls *et al.*, 2003; Mengual *et al.*, 2015; Young *et al.*, 2016). The ecological specializations of the microdontines, i.e. adults are not associated with flowers and larvae are predators or parasitoids of ant brood (Reemer, 2013; Pérez-Lachaud *et al.*, 2014), made Thompson (1969) and Speight (1987, 2016) consider them as a different family, not a subfamily. However, we agree with Reemer & Ståhls (2013) and Young *et al.* (2016) that this remains a subjective decision.

Monophyly of Eristalinae was not recovered in our analyses, which is in agreement with previous studies (Skevington & Yeates, 2000; Ståhls *et al.*, 2003; Hippa & Ståhls, 2005; Mengual *et al.*, 2015; Young *et al.*, 2016). The current concept of this subfamily includes nine tribes (i.e. Brachyopini, Callicerini, Cerioidini, Eristalini, Merodontini, Milesiini, Rhingiini, Sericomyiini and Volucellini) and several larval feeding modes, such as saprophagy, mycophagy, phytophagy and predation (Thompson, 1972; Thompson & Rotheray, 1998; Rotheray & Gilbert, 1999). The subfamily and its tribes are mostly characterized by adult morphological traits, and some larval characters are useful to distinguish tribes. Our results agree with those of Young *et al.* (2016), who also recovered members of Merodontini placed as a sister group of remaining syrphids exclusive of Microdontinae, but they disagree in the placement of the tribe Eristalini. Using anchored hybrid enrichment techniques and 343 molecular nuclear loci (the matrix comprises a total of 217 702 nt columns), Young *et al.* (2016) obtained a fully resolved tree with only one node having less than 70% bootstrap support. That clade comprised two branches, the first with a single member of Eristalini, *Helophilus fasciatus* Walker, and the second with members of tribes Milesiini, Callicerini, Brachyopini and Rhingiini plus Pipizinae + Syrphinae (Young *et al.*, 2016; Fig. 1). In our study, Eristalini (represented by *E. pertinax*) and Milesiini (represented by *S. pipiens*) form a maximally resolved clade, leaving *F. cuprea* (Rhingiini) as sister group to Pipizinae + Syrphinae (Fig. 2).

As expected, members of Pipizinae and Syrphinae were grouped into one clade, suggesting a common predatory ancestor. Both subfamilies have predaceous larvae that feed on soft-bodied arthropods, mostly aphids (Hemiptera: Aphididae), but syrphine larvae also feed on scale insects, psyllids, white flies, thrips and larvae of other insects (Rotheray, 1993; Rojo *et al.*, 2003), and recently a pipizine species was found associated with ants (Downes *et al.*, 2017), but there are still many larvae where the biology is unknown. Interestingly, a few secondarily phytophagous species (leaf miners, stem borers and pollen feeders) are found among Syrphini (Nishida *et al.*, 2002; Weng & Rotheray, 2008; Reemer & Rotheray, 2009: Zuijen & Nishida, 2011, Dumbardon-Martial, 2016). Monophyly of Syrphinae has been recovered by almost all previous phylogenetic studies based on molecular and/or morphological data (Rotheray & Gilbert, 1999; Skevington & Yeates, 2000; Ståhls *et al.*, 2003; Hippa & Ståhls, 2005; Mengual *et al.*, 2015; Young *et al.*, 2016), but the current classification of tribes within this subfamily is not supported (Rotheray & Gilbert, 1999; Ståhls *et al.*, 2003; Mengual, 2015; Mengual *et al.*, 2008, 2012, 2015; Miranda *et al.*, 2016; Young *et al.*, 2016). Our dataset was not designed to evaluate the monophyly or relationships among Syrphinae tribes and, consequently, we should not evaluate our results regarding them. But it is interesting to point out that the tribe Bacchini was recovered as nonmonophyletic, since *M. scalare* and *B, elongata*, both members of Bacchini, do not group in a clade. Our results agree

with those of Mengual *et al.* (2008) and Young *et al.* (2016) on the need for a revision of the tribal classification of Syrphinae that may shed light on the evolution of the larval feeding modes within this subfamily.

The present results are not novel, as the nonmonophyly of Syrphoidea has been suggested before, but it is the first time that transcriptomic data have been used to infer the phylogenetic relationship between Syrphidae and Pipunculidae. Moreover, our results on the relationships among syrphid subfamilies support the conclusions from previous studies. We have created a large dataset with *de novo* transcriptomes but the evidence is the same. The transcriptomic dataset presented here can help in the design of new bait kits for anchored hybrid enrichment of forthcoming projects on these dipteran families (for an approach as described by Mayer *et al.*, 2016). Future research on the taxa targeted in this study should have two clear objectives: (i) the phylogenetic relationships and monophyly of Platypezoidea families; and (ii) the study of synapomorphies for Pipunculidae + Schizophora. In the same line, future studies on the systematics of flower flies should target phylogenetic relationships of genera of Microdontinae, a new subfamilial division in that all the eristaline clades are recognized, and the definition of groupings within Syrphinae. Systematics work of this type can be extremely fruitful, providing fundamental insights into the embryological development of Cyclorrhapha, the evolution of larval feeding modes in Syrphidae, and the evolution of functional and morphological traits of these ecologically and economically important groups.

Author contributions

The study was conceived by XM, RSP and TP. Material for transcriptome sequencing was provided by DKY, KMB, RSP, XM, KM, and was processed by SL and XZ. BM, AD, LP and KM handled transcriptome data, performed the contamination check and cleaning and were responsible for deposition of all data at NCBI. The orthologue set was compiled by AV and KM. All further analyses were performed by TP, TOB, KM, CM, RSP and AK. All authors read, commented on and approved the final manuscript.

Supporting Information

Additional Supporting Information may be found in the online version of this article under the DOI reference: 10.1111/syen.12283

File S1. Extended Results.

Figure S1. MARE heatmaps of A) the original supermatrix before removal of data blocks with an IC = 0 and B) of the amino acid supermatrix after removal of data blocks with IC = 0. The heatmap shows species in rows and data block partitions in columns. Information content is coded in shades of blue with dark blue representing high information content, light blue representing low information content and red representing no information content. Missing data blocks are colored in white. A) Overall information content: 46.3%, matrix saturation 83%, 4,888 partitions. B) Overall information content: 59.5%, matrix saturation 86.5%, 3,800 partitions.

Figure S2. Heat map showing species-pairwise amino acid site coverage (AliStat results of species-pairwise comparison of all 19 species) for the supermatrix of A) the original dataset, B) the optimized dataset. Low shared site coverage in shades of red and high shared site coverage in shades of green. The completeness score for the entire alignment (Ca) is (A) Ca = 0.653416, (B) Ca = 0.742787.

Figure S3. Best scoring Maximum Likelihood tree based on the optimized amino-acid dataset (678,763 amino-acid sites with 1,500 partitions). Bootstrap support for all branches is 100.

Figure S4. Best scoring Maximum Likelihood tree based on the nucleotide dataset (3,121,689 nucleotide sites with 9,552 partitions). Bootstrap support for all branches is 100.

Table S1. Sampled species and material included in this study.

Table S2. Statistics of ortholog prediction with Orthograph.

Acknowledgements

We acknowledge Rudolf Meier, Chris Manchester, David J. Ferguson, Carola Greve, Alexander Blanke, Marissa S. Peiró, Andrew McKenzie, Klaus Sander, Brian K. Cassel and Kwong Shiyang for providing specimens, help with collecting and determination of species. We are grateful to Robert Waterhouse for help with compiling the orthologue set. We furthermore acknowledge the Gauss Centre for Supercomputing e. V. for funding computing time on the GCS Supercomputer SuperMUC at the Leibniz Supercomputing Centre (LRZ) and Ondrej Hlinka for help with analyses on the CSIRO HPC cluster. We thank all members of the 1KITE Antliophora group (http://1kite.org/subprojects.html) for granting us access of the transcriptome data used in this study. We thank the photographers who very kindly allowed us to use their images in Fig. 1, and two anonymous reviewers and Jeff Skevington for suggestions and comments that improved our work.

TP thanks the ZFMK for funding support. TOB has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement no. 642241. KM and DY thank the Schlinger Endowment to the CSIRO National Research Collections Australia (2013–2016), and KM thanks J. Korb, University of Freiburg, for support.

References

Collins, K.P. & Wiegmann, B.M. (2002) Phylogenetic relationships of the lower Cyclorrhapha (Diptera: Brachycera) based on 28SrDNA sequences. *Insect Systematics and Evolution*, **33**, 445–456.

Coquillett, D.W. (1901) A systematic arrangement of the families of the Diptera. *Proceedings of the United States National Museum*, **23**, 653–658.

Crampton, G.C. (1944) Suggestions for grouping the families of acalypterate cyclorrhaphous Diptera on the basis of the male terminalia. *Proceedings of the Entomological Society Washington*, **46**, 152–159.

Cumming, J.M., Sinclair, B.J. & Wood, D.M. (1995) Homology and phylogenetic implications of male genitalia in Diptera-Eremoneura. *Insect Systematics and Evolution*, **26**, 120–151.

Dell'Ampio, E., Meusemann, K., Szucsich, N.U. *et al.* (2014) Decisive data sets in phylogenomics: lessons from studies on the phylogenetic relationships of primarily wingless insects. *Molecular Biology and Evolution*, **31**, 239–249.

Downes, M.F., Skevington, J.H. & Thompson, F.C. (2017) A new ant inquiline flower fly (Diptera: Syrphidae: Pipizinae) from Australia. *Australian Journal of Entomology*, **44**, 29–38.

Dumbardon-Martial, E. (2016) Pollen feeding in the larva of *Toxomerus pulchellus* (Diptera, Syrphidae). *Bulletin de la Société Entomologique de France*, **121**, 413–420.

Eddy, S.R. (1998) Profile hidden Markov models. *Bioinformatics*, **14**, 755–763.

Finn, R.D., Bateman, A., Clements, J. *et al.* (2014) Pfam: the protein families database. *Nucleic Acids Research*, **42**, 222–230.

Griffiths, G.C.D. (1972) *The Phylogenetic Classification of Diptera Cyclorrhapha with Special Reference to the Structure of the Male Postabdomen*. Series Entomologica 8. Dr W. Junk, The Hague.

Grimaldi, D.A. & Cumming, J.M. (1999) Brachyceran Diptera in cretaceous ambers and Mesozoic diversification of the Eremoneura. *Bulletin of the American Museum of Natural History*, **239**, 1–124.

Henikoff, S. & Henikoff, J.G. (1992) Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences of the United States of America*, **89**, 10915–10919.

Hennig, W. (1971) Insektfossilien aus der unteren Kreide. III. Empidiformia ("Microphorinae") aus der unteren Kreide und aus dem Baltischen Bernstein; ein Vertreter der Cyclorrapha aus der unteren Kreide. *Stuttgarter Beiträge zur Naturkunde*, **232**, 1–28.

Hennig, W. (1973) Diptera (Zweiflügler). *Handbuch der Zoologie*, Band IV, 2 Hälfte: Insecta, 2/31, Lfg. 20 (ed. by J.-G. Helmcke, D. Starck and H. Wermuth), p. 200. De Gruyter, Berlin. pp.

Hennig, W. (1976) Das Hypopygium von Lonchoptera lutea Panzer und die phylogenetischen Verwandtschaftsbeziehungen der Cyclorrhapha (Diptera). *Stuttgarter Beiträge zur Naturkunde*, **283**, 1–63.

Hippa, H. & Ståhls, G. (2005) Morphological characters of adult Syrphidae: descriptions and phylogenetic utility. *Acta Zoologica Fennica*, **215**, 1–72.

Hurvich, C.M. & Tsai, C.-L. (1989) Regression and time series model selection in small samples. *Biometrika*, **76**, 297–307.

Johnston, D.S. & Nüsslein-Volhard, C. (1992) The origin of pattern and polarity in the *Drosophila* embryo. *Cell*, **68**, 201–220.

Jones, D., Taylor, W. & Thornton, J. (1992) The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences*, **8**, 275–282.

Katoh, K. & Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, **30**, 772–780.

Klymko, J. & Marshall, S.A. (2008) Review of the Nearctic Lonchopteridae (Diptera), including descriptions of three new species. *The Canadian Entomologist*, **140**, 649–673.

Kosiol, C. & Goldman, N. (2005) Different versions of the Dayhoff rate matrix. *Molecular Biology and Evolution*, **22**, 193–199.

Kozlov, A.M., Aberer, A.J. & Stamatakis, A. (2015) ExaML version 3: a tool for phylogenomic analyses on supercomputers. *Bioinformatics*, **31**, 2577–2579.

Kück, P., Meusemann, K., Dambach, J., Thormann, B., von Reumont, B.M., Wägele, J.W. & Misof, B. (2010) Parametric and non-parametric masking of randomness in sequence alignments can be improved and leads to better resolved trees. *Frontiers in Zoology*, **7**, 10.

Kutty, S.N., Pape, T., Wiegmann, B.M. & Meier, R. (2010) Molecular phylogeny of the Calyptratae (Diptera: Cyclorrhapha) with an emphasis on the superfamily Oestroidea and the position of Mystacinobiidae and McAlpine's fly. *Systematic Entomology*, **35**, 614–635.

Lambkin, C., Sinclair, B.J., Pape, T. *et al.* (2013) The phylogenetic relationships among infraorders and superfamilies of Diptera based on morphological evidence. *Systematic Entomology*, **38**, 164–179.

Lanfear, R., Calcott, B., Kainer, D., Mayer, C. & Stamatakis, A. (2014) Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Evolutionary Biology*, **14**, 82.

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T. & Calcott, B. (2016) PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, **34**, 772–773.

Le, S.Q., Dang, C.C. & Gascuel, O. (2012) Modeling protein evolution with several amino acid replacement matrices depending on site rates. *Molecular Biology and Evolution*, **29**, 2921–2936.

Le, S.Q. & Gascuel, O. (2008) An improved general amino acid replacement matrix. *Molecular Biology and Evolution*, **25**, 1307–1320.

Lemke, S. & Schmidt-Ott, U. (2009) Evidence for a composite anterior determinant in the hover fly *Episyrphus balteatus* (Syrphidae), a cyclorrhaphan fly with an anterodorsal serosa anlage. *Development*, **136**, 117–127.

Lemke, S., Stauber, M., Shaw, P.J., Rafiqi, A.M., Prell, A. & Schmidt-Ott, U. (2008) *bicoid* occurrence and *Bicoid*-dependent *hunchback* regulation in lower cyclorrhaphan flies. *Evolution & Development*, **10**, 413–420.

Lyneborg, L. & Barkemeyer, W. (2005) The Genus Syritta: A World Revision of the Genus Syritta Le Peletier & Serville, 1828 (Diptera, Syrphidae). Entomonograph, Vol. 15. Apollo Books, Stenstrup.

Mayer, C., Sann, M., Donath, A. *et al.* (2016) BaitFisher: a software package for multispecies target DNA enrichment probe design. *Molecular Biology and Evolution*, **33**, 1875–1886.

McAlpine, J.F. (1989) Phylogeny and classification of the Muscomorpha. *Manual of Nearctic Diptera*. Vol. 3, Agriculture Canada Monograph No. 32 (ed. by J.F. McAlpine), pp. 1397–1505. Research Branch, Agriculture Canada, Ottawa.

Meier, R. (2017) Citation of taxonomic publications: the why, when, what and what not. *Systematic Entomology*, **42**, 301–304.

Mengual, X. (2015) The systematic position and phylogenetic relationships of *Asiobaccha* Violovitsh (Diptera, Syrphidae). *Journal of Asia-Pacific Entomology*, **18**, 397–408.

Mengual, X., Ståhls, G. & Rojo, S. (2008) First phylogeny of predatory flower flies (Diptera, Syrphidae, Syrphinae) using mitochondrial COI and nuclear 28S rRNA genes: conflict and congruence with the concurrent tribal classification. *Cladistics*, **24**, 543–562.

Mengual, X., Ståhls, G. & Rojo, S. (2012) Is the mega-diverse genus *Ocyptamus* (Diptera, Syrphidae) monophyletic? Evidence from molecular characters including the secondary structure of 28S rRNA. *Molecular Phylogenetics and Evolution*, **62**, 191–205.

Mengual, X., Ståhls, G. & Rojo, S. (2015) Phylogenetic relationships and taxonomic ranking of pipizine flower flies (Diptera: Syrphidae) with implications for the evolution of aphidophagy. *Cladistics*, **31**, 491–408.

Miranda, G.F.G., Skevington, J.H., Marshall, S.A. & Kelso, S. (2016) The genus *Ocyptamus* Macquart (Diptera: Syrphidae): a molecular phylogenetic analysis. *Arthropod Systematics and Phylogeny*, **74**, 161–176.

Misof, B., Liu, S., Meusemann, K. *et al.* (2014) Phylogenomics resolves the timing and pattern of insect evolution. *Science*, **346**, 763–767.

Misof, B., Meyer, B., von Reumont, B.M., Kück, P., Misof, K. & Meusemann, K. (2013) Selecting informative subsets of sparse supermatrices increases the chance to find correct trees. *BMC Bioinformatics*, **14**, 348.

Misof, B. & Misof, K. (2009) A Monte Carlo approach successfully identifies randomness in multiple sequence alignments: a more objective means of data exclusion. *Systematic Biology*, **58**, 21–34.

Moulton, J.K. & Wiegmann, B.M. (2004) Evolution and phylogenetic utility of CAD (rudimentary) among Mesozoic-aged Eremoneuran Diptera (Insecta). *Molecular Phylogenetics and Evolution*, **31**, 363–378.

Nishida, K., Rotheray, G. & Thompson, F.C. (2002) First non-predaceous syrphine flower-fly (Diptera: Syrphidae): a new leaf-mining *Allograpta* from Costa Rica. *Studia Dipterologica*, **9**, 421–436.

Pape, T. (2009) Economic importance of Diptera. *Manual of Central American Diptera*, Vol. **1** (ed. by B.V. Brown, A. Borkent, J.M. Cumming, D.M. Wood, N.E. Woodley and M.A. Zumbado), pp. 65–77. NRC-CNRC Research Press, Ottawa.

Pattengale, N.D., Alipour, M., Bininda-Emonds, O.R., Moret, B.M. & Stamatakis, A. (2010) How many bootstrap replicates are necessary? *Journal of Computational Biology*, **17**, 337–354.

Pérez-Lachaud, G.J., Reemer, M. & Lachaud, J.P. (2014) An unusual, but not unexpected, evolutionary step taken by syrphid flies: the first record of true primary parasitoidism of ants by Microdontinae. *Biological Journal of the Linnean Society*, **111**, 462–472.

Peters, R.S., Krogmann, L., Mayer, C. *et al.* (2017) Evolutionary history of the hymenoptera. *Current Biology*, **27**, 2013–2018.

Petersen, M., Meusemann, K., Donath, A. *et al.* (2017) Orthograph: a versatile tool for mapping coding nucleotide sequences to clusters of orthologous genes. *BMC Bioinformatics*, **18**, 111.

Rafiqi, A.M., Lemke, S.J., Ferguson, S., Stauber, M. & Schmidt-Ott, U. (2008) Evolutionary origin of the amnioserosa in cyclorrhaphan flies correlates with spatial and temporal expression changes of zen. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 234–239.

Rambaut, A. (2009) *FigTree: Tree Figure Drawing Tool, Version 1.4.2*. [WWW document]. URL http://tree.bio.ed.ac.uk/software/figtree/ [accessed on 17 December 2016].

Reemer, M. (2013) Review and phylogenetic evaluation of associations between Microdontinae (Diptera: Syrphidae) and ants (Hymenoptera: Formicidae). *Psyche*, **2013**, 1–9.

Reemer, M. & Rotheray, G.E. (2009) Pollen feeding larvae in the presumed predatory syrphine genus *Toxomerus* Macquart (Diptera, Syrphidae). *Journal of Natural History*, **43**, 939–949.

Reemer, M. & Ståhls, G. (2013) Generic revision and species classification of the Microdontinae (Diptera, Syrphidae). *ZooKeys*, **288**, 1–213.

Ricarte, A., Nedeljković, Z., Quinto, J. & Marcos-García, M.A. (2010) The genus Ferdinandea Rondani, 1844 (Diptera, Syrphidae) in the Iberian Peninsula: first records and new breeding sites. *Journal of the Entomological Research Society*, **12**, 57–69.

Rojo, S., Gilbert, F., Marcos-García, M.A., Nieto, J.M. & Mier, M.P. (2003) *A World Review of Predatory Hoverflies (Diptera, Syrphidae: Syrphinae) and their Prey*. Centro Iberoamericano de la Biodiversidad (CIBIO), Alicante.

Rotheray, G.E. (1993) Colour guide to hoverfly larvae (Diptera, Syrphidae) in Britain and Europe. *Dipterists Digest*, **9**, 1–156.

Rotheray, G.E. & Gilbert, F. (1999) Phylogeny of Palaearctic Syrphidae (Diptera): evidence from larval stages. *Zoological Journal of the Linnean Society*, **127**, 1–112.

Rotheray, G.E. & Gilbert, F. (2008) Phylogenetic relationships and the larval head of the lower Cyclorrhapha. *Zoological Journal of the Linnean Society*, **153**, 287–323.

Schröder, R. & Sander, K. (1993) A comparison of transplantable *bicoid* activity and partial *bicoid* homeobox sequences in several *Drosophila* and blowfly species (Calliphoridae). *Roux's Archives of Developmental Biology*, **203**, 34–43.

Sinclair, B. (1992) A phylogenetic interpretation of the Brachycera (Diptera) based on the larval mandible and associated mouthpart structures. *Systematic Entomology*, **17**, 233–252.

Skevington, J.H. (2005) Revision of Nearctic Nephrocerus Zetterstedt (Diptera: Pipunculidae). *Zootaxa*, **977**, 1–36.

Skevington, J.H. & Yeates, D.K. (2000) Phylogeny of the Syrphoidea (Diptera) inferred from mtDNA sequences and morphology with particular reference to classification of the Pipunculidae (Diptera). *Molecular Phylogenetics and Evolution*, **16**, 212–224.

Sommer, R. & Tautz, D. (1991) Segmentation gene expression in the housefly *Musca domestica*. *Development*, **113**, 419–430.

Speight, M.C.D. (1987) External morphology of adult Syrphidae (Diptera). *Tijdschrift Voor Entomologie*, **130**, 141–175.

Speight, M.C.D. (2016) Species accounts of European Syrphidae 2016. *Syrph the Net, the Database of European Syrphidae (Diptera)*, Vol. **93**, p. 288 pp. Syrph the Net Publications, Dublin.

Ståhls, G., Hippa, H., Rotheray, G., Muona, J. & Gilbert, F. (2003) Phylogeny of Syrphidae (Diptera) inferred from combined analysis of molecular and morphological characters. *Systematic Entomology*, **28**, 433–450.

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

Stauber, M., Jäckle, H. & Schmidt-Ott, U. (1999) The anterior determinant *bicoid* of *Drosophila* is a derived *Hox* class 3 gene. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 3786–3789.

Stauber, M., Taubert, H. & Schmidt-Ott, U. (2000) Function of *bicoid* and *hunchback* homologs in the basal cyclorrhaphan fly *Megaselia* (Phoridae). *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 10844–10849.

Steyskal, G.C. (1987) Otitidae. *Manual of Nearctic Diptera* (ed. by J.F. McAlpine), Vol. 2, pp. 799–808. Agriculture Canada Monograph No. 32, Research Branch, Agriculture Canada, Ottawa.

Suyama, M., Torrents, D. & Bork, P. (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*, **34**, 609–612.

Tachi, T. (2014) Homology of the metapleuron of Cyclorrhapha, with discussion of the paraphyly of Syrphoidea (Diptera: Aschiza). *Insect Systematics and Evolution*, **45**, 395–414.

Thompson, F.C. (1969) A new genus of Microdontine flies (Diptera: Syrphidae) with notes on the placement of the subfamily. *Psyche*, **76**, 74–85.

Thompson, F.C. (1972) A contribution to a generic revision of the Neotropical Milesinae (Diptera: Syrphidae). *Arquivos de Zoologica*, **23**, 73–215.

Thompson, F.C. & Rotheray, G.E. (1998) Family Syrphidae. *Manual of Palaearctic Diptera*, Vol. **3** (ed. by L. Papp and B. Darvas), pp. 81–139. Science Herald, Budapest.

Van Veen, M. (2004) *Hoverflies of northwest Europe: Identification Keys to the Syrphidae*. KNNV Publishing, Utrecht.

Vockeroth, J.R. & Thompson, F.C. (1987) Syrphidae. *Manual of Nearctic Diptera*, Vol. **2** (ed. by J.F. McAlpine, B.V. Peterson, G.E. Shewell, H.J. Teskey, J.R. Vockeroth and D.M. Wood), pp. 713–743. Canadian Government Publishing Centre, Ottawa.

Vujić, A., Ståhls, G., Ačanski, J., Bartsch, H., Bygebjerg, R. & Stefanović, A. (2013) Systematics of Pipizini and taxonomy of

European Pipiza Fall en: molecular and morphological evidence (Diptera, Syrphidae). *Zoologica Scripta*, **42**, 288–305.

Wada, S. (1991) Morphologische Indizien für das unmittelbare Schwester-gruppenverh ältnis der Schizophora mit den Syrphoidea ('Aschiza') in der phylogenetischen Systematik der Cyclorrhapha (Diptera: Brachycera). *Journal of Natural History*, **25**, 1531–1570.

Waterhouse, R.M., Tegenfeldt, F., Zdobnov, E.M. & Kriventseva, E.V. (2013) OrthoDB: a hierarchical catalog of animal fungal and bacterial orthologs. *Nucleic Acids Research*, **41**, D358–D365.

Weng, J.L. & Rotheray, G.E. (2008) Another non-predaceous syrphine flower fly (Diptera: Syrphidae): pollen feeding in the larva of *Allograpta micrura*. *Studia Dipterologica*, **15**, 245–258.

Whelan, S. & Goldman, N. (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Molecular Biology and Evolution*, **18**, 691–699.

Wiegmann, B.M., Trautwein, M.D., Winkler, I.S. *et al.* (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.

Wiegmann, B.M., Yeates, D.K., Thorne, J.L. & Kishino, H. (2003) Time flies, a new molecular time-scale for Brachyceran fly evolution without a clock. *Systematic Biology*, **52**, 745–756.

Woodley, N.E. (1989) Phylogeny and classification of the "Orthorraphous" Brachycera. *Manual of Nearctic Diptera*, Vol. **3** (ed. by J. McAlpine and D. Wood), pp. 1371–1395. Agriculture Canada Research Branch, Ottawa.

Woodley, N.E., Borkent, A. & Wheeler, T.A. (2009) Phylogeny of the Diptera. *Manual of Central American Diptera*, Vol. **2** (ed. by B.V. Brown, A. Borkent, J.M. Cumming, D.M. Wood, N.E. Woodley and M.A. Zumbado), pp. 79–94. NRC Research Press, Ottawa.

Yeates, D.K. & Wiegmann, B.M. (1999) Congruence and controversy: toward a higher-level phylogeny of Diptera. *Annual Review of Entomology*, **44**, 397–428.

Yeates, D.K. & Wiegmann, B.M. (2005) Phylogeny and evolution of Diptera: recent insights and new perspectives. *The Evolutionary Biology of Flies* (ed. by D.K. Yeates and B.M. Wiegmann), pp. 14–44. Columbia University Press, New York, New York.

Yeates, D.K. & Wiegmann, B.M. (2012) The impact of the manual of Nearctic Diptera on phylogenetic Dipterology. In: Ottawa Dipterology (1937–1989): Festschrift commemorating the coordinators of the manual of Nearctic Diptera and their contributions to building the Canadian National Collection of Insects. Part III. *The Canadian Entomologist*, **144**, 197–205.

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.

Young, A.D., Lemmon, A.R., Skevington, J.H. *et al.* (2016) Anchored enrichment dataset for true flies (order Diptera) reveals insights into the phylogeny of flower flies (family Syrphidae). *BMC Evolutionary Biology*, **16**, 143.

Zatwarnicki, T. (1996) A new reconstruction of the origin of the eremoneuran hypopygium and its implications for classification (Insecta: Diptera). *Genus*, **7**, 103–175.

Zuijen, M.P. & Nishida, K. (2011) Description of life history and immature stages of phytophagous flower fly, *Allograpta zumbadoi* Thompson (Diptera: Syrphidae: Syrphinae). *Studia Dipterologica*, **17**, 37–51.

Accepted 10 January 2018
First published online 7 March 2018