WILEY-BLACKWELL

Cladistics (2019) 1–18

Cladistics

10.1111/cla.12375

Phylogenomic analysis of Calyptratae: resolving the phylogenetic relationships within a major radiation of Diptera

Sujatha Narayanan Kutty[a], Karen Meusemann[b,c,d], Keith M. Bayless[e,f], Marco A. T. Marinho[g], Adrian C. Pont[h], Xin Zhou[i,j], Bernhard Misof[c], Brian M. Wiegmann[f], David Yeates[d], Pierfilippo Cerretti[k], Rudolf Meier[a]* and Thomas Pape[l]*

[a]*Department of Biological Sciences, National University of Singapore, 14 Science Dr 4, Singapore 117543, Singapore;* [b]*Biology I, Evolutionary Biology & Ecology, University of Freiburg, Hauptstraße 1, Freiburg (Brsg.), Germany;* [c]*Zoologisches Forschungsmuseum Alexander Koenig (ZFMK)/Zentrum für Molekulare Biodiversitätsforschung (ZMB), Bonn, Germany;* [d]*Australian National Insect Collection, CSIRO National Research Collections Australia (NRCA), Acton, ACT, Canberra, Australia;* [e]*Department of Entomology, California Academy of Sciences, San Francisco, CA, USA;* [f]*Department of Entomology, North Carolina State University, Raleigh, NC 27695, USA;* [g]*Departamento de Ecologia, Zoologia e Genética, Instituto de Biologia, Universidade Federal de Pelotas, Pelotas, RS, Brazil;* [h]*Oxford University Museum of Natural History, Parks Road, Oxford, OX1 3PW, UK;* [i]*Beijing Advanced Innovation Center for Food Nutrition and Human Health, China Agricultural University, Beijing 100193, China;* [j]*Department of Entomology, China Agricultural University, Beijing 100193, China;* [k]*Dipartimento di Biologia e Biotecnologie 'Charles Darwin', Sapienza Università di Roma, Rome, Italy;* [l]*Natural History Museum of Denmark, University of Copenhagen, Universitetsparken 15, Copenhagen, DK–2100, Denmark*

Accepted 3 January 2019

Abstract

The Calyptratae, one of the most species-rich fly clades, only originated and diversified after the Cretaceous–Palaeogene extinction event and yet exhibit high species diversity and a diverse array of life history strategies including predation, phytophagy, saprophagy, haematophagy and parasitism. We present the first phylogenomic analysis of calyptrate relationships. The analysis is based on 40 species representing all calyptrate families and on nucleotide and amino acid data for 1456 single-copy protein-coding genes obtained from shotgun sequencing of transcriptomes. Topologies are overall well resolved, robust and largely congruent across trees obtained with different approaches (maximum parsimony, maximum likelihood, coalescent-based species tree, four-cluster likelihood mapping). Many nodes have 100% bootstrap and jackknife support, but the true support varies by more than one order of magnitude [Bremer support from 3 to 3427; random addition concatenation analysis (RADICAL) gene concatenation size from 10 to 1456]. Analyses of a Dayhoff-6 recoded amino acid dataset also support the robustness of many clades. The backbone topology Hippoboscoidea+(Fanniidae+(Muscidae+((Anthomyiidae–Scathophagidae)+Oestroidea))) is strongly supported and most families are monophyletic (exceptions: Anthomyiidae and Calliphoridae). The monotypic Ulurumyiidae is either alone or together with Mesembrinellidae as the sister group to the rest of Oestroidea. The Sarcophagidae are sister to Mystacinobiidae+Oestridae. Polleniinae emerge as sister group to Tachinidae and the monophyly of the clade Calliphorinae+Luciliinae is well supported, but the phylogenomic data cannot confidently place the remaining blowfly subfamilies (Helicoboscinae, Ameniinae, Chrysomyinae). Compared to hypotheses from the Sanger sequencing era, many clades within the muscoid grade are congruent but now have much higher support. Within much of Oestroidea, Sanger era and phylogenomic data struggle equally with regard to finding well-supported hypotheses.

© The Willi Hennig Society 2019.

*Corresponding author:
E-mail address: meier@nus.edu.sg; tpape@snm.ku.dk

© The Willi Hennig Society 2019

Introduction

The Cretaceous–Palaeogene (K–Pg) extinction event disrupted plant and animal communities across the planet and led to the collapse of many ecosystems (Macleod et al., 1997; Nichols and Johnson, 2008). This may have provided an opportunity for a massive "maggot-based" evolutionary radiation that was conceivably initiated by a cyclorrhaphan fly with a novel mechanism for puparial escape (Grimaldi and Engel, 2005). The resulting radiation yielded Schizophora, which today account for more than a third (> 55 000) of the described species diversity of Diptera (Pape et al., 2011). Within this large assemblage of schizophoran flies, Calyptratae diversified into a very species-rich clade that contains approximately 22 000 named extant species. Calyptrates include many well-known taxa such as the tsetse flies (Glossinidae), houseflies (Muscidae), fleshflies (Sarcophagidae), botflies (Oestridae) and blowflies (Calliphoridae). They span a remarkable range of feeding and breeding strategies that include saprophagy (including coprophagy), phytophagy, haematophagy, parasitism (including parasitoids), and predation on vertebrates as well as invertebrate host associations. Calyptrate flies have successfully adapted to breed on or in a wide range of substrates including non-angiosperm tracheophytes (horsetails, ferns, gymnosperms), angiosperms (dicots, monocots), decaying organic matter, mammal dung, vertebrate and invertebrate carrion, necrotic tissues and wounds of living vertebrates, vertebrate blood, arthropods (terrestrial isopods, myriapods, scorpions, spiders, insects), earthworms, and live and dead snails. Estimates of the origin and timeline for this radiation suffer from large confidence intervals because of the remarkable paucity of calyptrate fossils (Wiegmann et al., 2011; Ding et al., 2015; Cerretti et al., 2017). von Tschirnhaus and Hoffeins (2009) report that many non-calyptrate schizophoran families have been found in Baltic amber as inclusions of mid-Eocene age and speculate that calyptrate flies, which are generally larger than their acalyptrate relatives, were "powerful and quick enough to escape the sticky resin" (p. 174). However, this cannot be the sole explanation as copal is known to harbour a large number of relatively large calyptrates (O'Hara et al., 2013). Instead, it is conceivable that calyptrates only originated or diversified significantly later.

Molecular studies of calyptrate phylogenetic relationships have used both multi-gene (Vossbrinck and Friedman, 1989; Bernasconi et al., 2000b; Nirmala et al., 2001; Song et al., 2008; Kutty et al., 2010) and mitogenomic approaches (Ding et al., 2015; Junqueira et al., 2016; Zhang et al., 2016). The most comprehensive molecular phylogenetic analysis was carried out by Kutty et al. (2010), who presented a hypothesis based on 275 exemplar species representing all major recognized families and utilized four mitochondrial and four nuclear markers. Although many higher-level relationships were resolved with moderate to high support, the non-monophyly of certain families (Anthomyiidae, Rhinophoridae, Tachinidae) and relationships specifically between families of the Oestroidea (e.g. position of Oestridae and Mesembrinellidae) conflicted with morphology-based hypotheses. Some of the hypotheses also differed depending on whether the data were analysed with maximum parsimony (MP) or maximum likelihood (ML). More recently, Zhang et al. (2016) explored mitochondrial genomes for reconstructing calyptrate relationships but the data were again unable to robustly resolve many deeper splits within the oestroid radiation.

Yet, a robust phylogeny for the calyptrate radiation is crucial for reconstructing the evolution of the diverse natural history of this group. Here we attempt to obtain such a robust phylogenetic hypothesis using phylogenomics, given that increasing the number of genetic markers has successfully resolved many higher-level relationships in a diverse array of arthropod lineages (Dell'Ampio et al., 2014; Misof et al., 2014; Peters et al., 2017; Pauli et al., 2018; Kutty et al., 2018) . We use a phylogenomic approach because calyptrate molecular phylogenetics face a suite of challenges that are commonly encountered in efforts to resolve higher-level evolutionary relationships (e.g. Moulton and Wiegmann, 2007; Su et al., 2008; Winterton et al., 2010; Wiegmann et al., 2011; Tóthová et al., 2013; Zhao et al., 2013; Young et al., 2016): many long-recognized higher-level taxa (often assigned family rank) are well supported based on all data sources and tree optimality criteria, while the interrelationships between many of these clades are not recovered consistently or only with low support. The precise reasons for such problems are not fully understood and probably differ between taxa and datasets. However, in the case of calyptrates it is likely that lack of resolution is at least partially related to multiple episodes of rapid radiation and diversification, when increased speciation rates, coupled with low extinction rates, allowed many new lineages to survive and spread (Cerretti et al., 2017).

Here, we test to what extent transcriptomic data, in particular a set of nuclear single-copy protein-coding genes obtained from transcriptomes, can help to resolve the higher-level relationships within Calyptratae. The dataset comprises 40 exemplar species representing 14 recognized families and 1456 nuclear single-copy protein-coding genes, which is here analysed using a wide variety of methods to explore the inherent signal within the data. As such, we carry out MP and ML analyses of the nucleotide and amino acid data and explore the use of species tree techniques (ASTRAL: Mirarab and Warnow, 2015). We also

recode the amino acids using Dayhoff-6 groups (Dayhoff et al., 1978; Embley et al., 2003; Hrdy et al., 2004; Susko and Roger, 2007) in order to explore whether amino acid heterogeneity may affect the results of phylogenetic analyses. Lastly, we use random addition concatenation analysis (Narechania et al., 2012) to investigate the relative strength of support for the nodes on our relationship hypothesis. Such measures of relative support become increasingly important in phylogenomic analyses because conventional support values (e.g. bootstraps) are maximized for a large number of nodes with different levels of support.

Calyptratae are a compelling testing ground for exploring the power of phylogenomics because the monophyly of this group has been established beyond doubt based on morphological characters (Griffiths, 1972; Hennig, 1973; Hackman and Väisänen, 1985; McAlpine, 1989; Michelsen, 1991; Lambkin et al., 2013) and molecular data (Wiegmann et al., 2011). Yet, our understanding of relationships among the constituent (sub)families ranges from well-established to completely unclear (Pape, 1986, 1992, 2001, 2006; McAlpine, 1989; Verves, 1989; Michelsen, 1991; Rognes, 1997; Bernasconi et al., 2000a,b; Gleeson et al., 2000; Pape and Arnaud, 2001; Carvalho and Couri, 2002; Stireman, 2002; Stevens, 2003; Nihei and Carvalho, 2004, 2007; Savage and Wheeler, 2004; Savage et al., 2004; Dittmar et al., 2006; Kutty et al., 2007; Petersen et al., 2007; Schuehli et al., 2007; Marinho et al., 2012, 2017; Cerretti et al., 2014b; Winkler et al., 2015; Buenaventura and Pape, 2017; Buenaventura et al., 2017). As illustrated by the large number of studies, this is not due to a lack of effort.

Currently, the best understood relationships are within Hippoboscoidea and the muscoid grade. The monophyly of the superfamily Hippoboscoidea is well established and supported by both morphological (Griffiths, 1972; Hennig, 1973) and molecular data (Dittmar et al., 2006; Petersen et al., 2007). Glossinidae are considered the sister group to Hippoboscidae (louse flies and bat flies) (Petersen et al., 2007). Streblidae have been suggested to be paraphyletic (Dittmar et al., 2006, 2015), and for reasons of nomenclatural priority the bat-fly clade is now best treated as a subfamily (Nycteribiinae). Molecular data confirmed the non-monophyly of the muscoid grade (Kutty et al., 2008) and placed monophyletic Oestroidea within this grade. This position has since been repeatedly corroborated (Kutty et al., 2010; Lambkin et al., 2013; Ding et al., 2015). Within the muscoid grade, the monophyly of three families (Fanniidae, Muscidae and Scathophagidae) has been confirmed based on both morphological and molecular evidence (Domínguez and Roig-Juñent, 2008; Kutty et al., 2008; Michelsen and Pape, 2017). Anthomyiidae, however, emerge as paraphyletic in molecular studies (Kutty et al., 2008, 2010) although morphology supports monophyly (Michelsen, 1991, 1996). Based on molecular data (Kutty et al., 2008), Fanniidae are considered the sister group to the remaining non-hippoboscoid calyptrates, and Muscidae are sister to a clade comprising Oestroidea+(Anthomyiidae–Scathophagidae). Relationships between the recognized muscid subfamilies (Achanthipterinae, Atherigoninae, Azeliinae, Cyrtoneurininae, Coenosiinae, Muscinae, Mydaeinae, Phaoniinae, Muscidae) were investigated by Kutty et al. (2014) and Haseyama et al. (2015), with both studies supporting the monophyly of many subfamilies while rejecting the traditional tribal classification. The two subfamilies of Scathophagidae (Scathophaginae; Delininae) are considered monophyletic based on morphological characters (Gorodkov, 1986; Vockeroth, 1989) and DNA sequence data (Bernasconi et al., 2000a; Kutty et al., 2007), although Kutty et al. (2008) did not recover the monophyly of these subfamilies.

Family-level relationships within the monophyletic Oestroidea remain controversial. Depending on genetic markers and taxon sampling, conflicting hypotheses emerge. Within Oestroidea, the monophyly of Sarcophagidae and the three recognized subfamilies Miltogramminae, Paramacronychiinae and Sarcophaginae are well corroborated based on morphology (Pape, 1992, 1996) and have been confirmed by molecular data (Kutty et al., 2010). However, the relationships between subfamilies remain controversial (Piwczyński et al., 2017). Based on morphology, the monophyly of the family Oestridae, which contains exclusively mammal parasites, is well supported (Wood, 1986; Pape, 1992, 2001). Monophyly has also been recently confirmed by Zhang et al. (2016) based on mitogenomic data, although relationships within the family were not conclusively resolved. The monophyly of Rhinophoridae finds its most convincing morphological support from features of the first instar larva (Crosskey, 1977; Pape, 1986, 1998; Pape and Arnaud, 2001; Cerretti et al., 2014a). Molecular sequence data provide growing support for the monophyly of Rhinophoridae, but the taxon sample in all published analyses remains poor (Kutty et al., 2010; Cerretti et al., 2017).

There is strong evidence based on morphology and molecular sequence data that the calliphorids (blowflies) are not monophyletic (Rognes, 1997; Kutty et al., 2010; Nelson et al., 2012; Singh and Wells, 2013; Cerretti et al., 2017) and one of the major tasks in calyptrate systematics is re-defining the (sub)family-level units. The currently recognized subfamilies are Ameniinae, Aphyssurinae, Bengaliinae, Auchmeromyinae, Calliphorinae, Chrysomyinae, Luciliinae, Helicoboscinae, Melanomyinae, Phumosiinae, Polleniinae and Toxotarsinae (Hennig, 1973; Rognes, 1991, 1997; Norris, 1999). Numerous DNA-based analyses have aimed at resolving the calliphorid assemblage, but

well-supported relationship hypotheses are elusive [see Nelson et al. (2012, table 1) for a compilation of DNA-based studies of calliphorid systematics]. For example, the placement of a monophyletic Rhiniinae/Rhiniidae remains unsettled, with Singh and Wells (2013) recovering support for a sister group relationship between Rhiniinae and traditional calliphorid subfamilies using one mitochondrial (COI) and three nuclear genes (CPS, EF1-α, 28S), while Cerretti et al. (2017) found Rhiniidae to be the sister group to Bengaliinae based on one mitochondrial (16S) and two nuclear genes (28S, CAD). However, all analyses suggest that Rhiniinae should be given family rank (see also Kutty et al., 2010; Marinho et al., 2017). Similarly, Mesembrinellidae were until recently generally considered to be a subfamily of Calliphoridae (Hennig, 1973; Pape, 1992; Rognes, 1997), but both morphological and molecular evidence supports the view of Guimarães (1977) that this lineage warrants the rank of full family (Kutty et al., 2010; Singh and Wells, 2013; Cerretti et al., 2017; Marinho et al., 2017). Marinho et al. (2017) addressed the internal relationships within Mesembrinellidae based on molecular markers, but the most probable sister group could not be identified due to the limited outgroup sampling.

The monophyly of the family Tachinidae is strongly supported by morphological characters (Wood, 1987; Pape, 1992; Tschorsnig and Richter, 1998; Cerretti et al., 2014b), and all known species are arthropod (and largely insect) parasitoids (Stireman et al., 2006). Although Kutty et al. (2010) did not recover a monophyletic Tachinidae, other molecular studies have confirmed the monophyly and placed the calliphorid subfamily Polleniinae as its sister group (Winkler et al., 2015; Cerretti et al., 2017). In contrast, the monophyly of the subfamilies Tachininae, Exoristinae, Phasiinae and Dexiinae is still contentious and has been extensively discussed in the literature (Herting, 1957, 1984; O'Hara, 1985; Tschorsnig, 1985; Wood, 1987; Tschorsnig and Richter, 1998; Stireman, 2002; Tachi and Shima, 2010; Cerretti et al., 2014b; Winkler et al., 2015).

The calyptrates also include two monospecific, family-ranked taxa. The first is the apterous bat-associated New Zealand endemic *Mystacinobia zelandica* (Mystacinobiidae), which has been placed within Oestroidea based on both morphological and molecular evidence (Griffiths, 1982; McAlpine, 1989; Rognes, 1997; Gleeson et al., 2000; Kutty et al., 2010). Kutty et al. (2010) found molecular support for *Mystacinobia* as the sister group to "McAlpine's fly" (Ulurumyiidae), while *Mystacinobia* emerged as the sister group to the Anthomyiidae–Scathophagidae clade in the analysis of Cerretti et al. (2017). Similarly, the position of *Ulurumyia macalpinei* ("McAlpine's fly", Ulurumyiidae) is unclear. Based on morphological traits, this family was initially considered to be closely related to Anthomyiidae or Calliphoridae (Ferrar, 1979), but detailed morphological studies exclude it from any of the established calyptrate families and strongly support a position at or near the base of Oestroidea (Michelsen and Pape, 2017). *Mystacinobia* and *Ulurumyia* emerged as weakly supported sister taxa in the molecular analysis of Kutty et al. (2010), but they were widely separated by both molecular and morphological data in the analyses of Cerretti et al. (2017).

Materials and methods

Taxon sampling and transcriptome sequencing

Our taxon sample comprises 40 calyptrate species representing all 14 widely recognized families of Calyptratae (Table 1). RNA extractions for 18 species were carried out at the National University of Singapore (NUS) (see NUS in Table 1; Tables S1 and S2), using a modified version of the Trizol RNA extraction protocol (Invitrogen, ThermoFisher Scientific, Waltham, MA, USA). RNA libraries (insert size 500 bp) were prepared by AITbiotech Pte Ltd (Singapore) using the Illumina TruSeq Stranded mRNA Library Prep Kit. Eight libraries were multiplexed and paired-end sequencing of 100/150/250 bp was carried out on Illumina Hiseq 2000/2500 sequencing platform facilities at the Genome Institute of Singapore (GIS) or Singapore Centre for Environmental Life Sciences Engineering (SCELSE), Singapore. RNA extractions for seven of the species were carried out at North Carolina State University, Raleigh (see NCSU in Table 1; Tables S1 and S2). RNA was extracted using the Qiagen RNAEasy Kit (Valencia, CA, USA). Transcript libraries obtained at NCSU were developed using the NEBNext (New England Biosciences, Ipswich, MA, USA) Ultra RNA Library Prep Kit for Illumina, following the manufacturer's guidelines. RNA was bound to Agencourt AMPure XP Beads (Beckman Coulter, Inc., Brea, CA, USA) on a magnetic plate and the sample underwent a series of washes. A reverse transcription reaction was performed, followed by PCR enrichment, yielding a size-selected non-directional cDNA library to be sequenced as paired-end reads on an Illumina system (Illumina, San Diego, CA, USA). Inserts were 300 bp long if sequenced on an Illumina MiSeq and 125 bp long if sequenced on an Illumina HiSeq 2500. Data for ten species were obtained from the 1KITE (1000 Insect Transcriptome Evolution) project (see 1KITE in Table 1; Tables S1 and S2). Extraction of RNA, construction of paired-end libraries and sequencing are described in detail in Peters et al. (2017) and Pauli et al. (2018). Finally, we included transcriptome data for one exemplar calyptrate species

Table 1
List of species, number of assembled contigs, number of genes from orthology prediction and number of genes in the MARE reduced dataset

	Family	Subfamily	Species name	Author name	Data generated	No. of assembled contigs	EOGs in full dataset	EOGs in MARE reduced dataset
Hippoboscoidea	Hippoboscidae	Ornithomyinae	*Ortholfersia macleayi*	Leach, 1817	1KITE	59 827	2102	1407
	Nycteribiidae	Nycteribiinae	*Penicillidia dufourii*	Westwood, 1835	1KITE	20 120	1965	1378
Muscoid grade	Anthomyiidae		*Eustalomyia vittipes*	Zetterstedt, 1845	1KITE	14 303	1557	1188
	Anthomyiidae		*Delia antiqua*	Meigen, 1826	TSA	29 451	1908	1360
	Fanniidae		*Fannia canicularis*	Linnaeus, 1761	NCSU	43 555	1678	1233
	Muscidae	Coenosiinae	*Lispe sydneyensis*	Schiner, 1868	1KITE	22 893	1961	1403
	Muscidae	Muscinae	*Mesembrina meridiana*	Linnaeus, 1758	1KITE	56 164	2011	1425
	Muscidae	Coenosiinae	*Coenosia* sp.	Meigen, 1826	NCSU	40 133	1467	1132
	Muscidae	Coenosiinae	*Schoenomyza* sp.	Haliday, 1833	NCSU	47 561	2001	1324
	Muscidae	Azeliinae	*Muscina stabulans*	Fallén, 1817	NUS	26 832	1603	1236
	Muscidae	Muscinae	*Polietes lardarius*	Fabricius, 1781	NUS	30 812	1895	1392
	Muscidae	Mydaeinae	*Mydaea urbana*	Meigen, 1826	NUS	48 007	1923	1400
	Muscidae	Phaoniinae	*Hebecnema umbratica*	Meigen, 1826	NUS	57 726	1798	1317
	Muscidae	Phaoniinae	*Phaonia angelicae*	Scopoli, 1763	NUS	36 430	2006	1428
	Muscidae	Muscinae	*Musca domestica*	Linnaeus, 1758	SRA	50 172	1805	1320
	Muscidae	Muscinae	*Stomoxys calcitrans*	Linnaeus, 1758	SRA	88 708	2030	1365
	Scathophagidae	Scathophaginae	*Cordilura* sp.	Fallén, 1810	NCSU	102 240	1253	994
	Scathophagidae	Scathophaginae	*Scathophaga stercoraria*	Linnaeus, 1758	NCSU	45 066	475	374
Oestroidea	Calliphorid	Calliphorinae	*Calliphora vomitoria*	Linnaeus, 1758	1KITE	26 705	1969	1412
	Calliphorid	Ameniinae	*Amenia* sp.	Robineau-Desvoidy, 1830	NUS	26 114	1770	1363
	Calliphorid	Helicoboscinae	*Eurychaeta muscaria*	Meigen, 1826	NUS	54 002	1813	1364
	Calliphorid	Polleniinae	*Pollenia* sp.	Robineau-Desvoidy, 1830	NCSU	75 648	1595	1213
	Calliphorid	Bengaliinae	*Verticia nigra*	Malloch, 1927	NUS	51 157	1885	1385
	Calliphorid	Chrysomyinae	*Chrysomya megacephala*	Fabricius, 1794	SRA	31 516	1615	1232
	Calliphorid	Luciliinae	*Lucilia sericata*	Meigen, 1826	SRA	43 387	1195	961
	Mesembrinellidae		*Mesembrinella bellardiana*	Aldrich, 1922	NUS	16 734	1398	1098
	Mystacinobiidae		*Mystacinobia zelandica*	Holloway, 1976;	NUS	27 274	2060	1434
	Oestridae	Cuterebrinae	*Cuterebra austeni*	Sabrosky, 1986	1KITE	9936	1079	859
	Rhiniidae		*Stomorhina subapicalis*	Macquart, 1847	1KITE	29 906	2009	1411
	Rhinophoridae		*Bixinia winkleri*	Cerretti, Lo Giudice & Pape, 2014	NUS	23 022	1655	1261
	Rhinophoridae		*Stevenia* sp.		NUS	53 069	1824	1363
	Sarcophagidae	Miltogramminae	*Gymnoprosopa* sp.	Coquillett, 1895	NCSU	62 747	882	701
	Sarcophagidae	Sarcophaginae	*Sarcophaga* sp.		NUS	49 581	1533	1195
	Tachinidae	Dexiinae	*Euthera bicolor*	Coquillett, 1902	1KITE	18 316	1729	1265
	Tachinidae	Phasiinae	*Cylindromyia* sp.		NUS	53 366	1972	1413
	Tachinidae	Phasiinae	*Gymnosoma nitens*	Meigen, 1824	NUS	34 808	1996	1423
	Tachinidae	Tachininae	*Mintho rufiventris*	Fallén, 1817	NUS	74 997	2093	1426
	Tachinidae	Exoristinae	*Pseudogonia rufifrons*	Wiedemann, 1830	NUS	37 795	1829	1363
	Tachinidae	Tachininae	*Triarthria setipennis*	Fallén, 1810	1KITE	24 801	2140	1429
	Ulurumyiidae		*Ulurumyia macalpinei*	Michelsen and Pape, 2017	NUS	34 760	1766	1340

from the NCBI Sequence Read Archive (see SRA in Table 1 and Table S3) and for one from the Transcriptome Shotgun Assembly Sequence Database (see TSA in Table 1 and Table S3).

Assemblies and orthology prediction

De novo assemblies of paired-end transcriptomes generated at NUS were carried out in CLC Genomics Workbench 7.5.1 (https://www.qiagenbioinformatics.com/) [initially trimmed (limit: 0.001); *de novo* option parameters (word size = 50, bubble size = 50–150, identity fraction = 1, length fraction = 1)]. All assembled contigs were checked for contaminant sequences through a VecScreen contamination screen by the National Center for Biotechnology Information. A cross-contamination check across libraries was carried out whenever possible for assembled transcripts sequenced in the same lane via an all-versus-all search using blast [BLAST 2.2.28+ (Camacho et al., 2009); contigs that shared high sequence similarity (blastn parameters = length $\geq$ 200 bp, identity $\geq$ 98%, mismatches $\leq$ 4) were excluded from further analysis]. For NCSU species assemblies obtained with Trinity v.2.2 (Grabherr et al., 2011), read quality was assessed using FastQC (Andrews, 2010) to ensure that per-base sequence quality was above 15, and all adapter sequences were removed as shown in “Overrepresented sequences” and “Adapter content”. If these metrics failed, Trimmomatic (Bolger et al., 2014) was used for one to three iterations with minimum quality set to 15 and a 4 bp sliding window size. The minimum length of filtered reads was set to 25 as Trinity uses 25-mers during assembly (Grabherr et al., 2011). All raw reads obtained from samples from 1KITE were quality-controlled, trimmed, assembled using SOAPdenovo-Trans 31kmer (v.1.02) (Xie et al., 2014), and screened for possible contaminant sequences (which were then removed) as described by Peters et al. (2017) (see Table S4 for the amount of removed contaminants from the assembled transcriptomes of each species). Both the raw reads and the assembled transcriptomes from 1KITE are archived at the National Center for Biotechnology Information (NCBI) under the Umbrella BioProject ID PRJNA183205 (‘The 1KITE project: evolution of insects’).

Orthology prediction, post-processing and dataset preparation followed the methods outlined in Kutty et al. (2018). Orthology prediction was carried out using Orthograph v.0.5.8 (Petersen et al., 2017) based on an orthologous reference set with 3288 single-copy protein-encoding genes comprising clusters of orthologous sequences of the following five reference species for which official gene sets (OGS) are available: *Anopheles gambiae*, *Tribolium destructor*, *Mayetiola destructor*, *Drosophila melanogaster* and *Bombyx mori* (see Table S5). Clusters of orthologous sequences were based on OrthoDB5 (Waterhouse et al., 2011) with hierarchy split set to be between Hymenoptera and remaining Holometabola; details are described in Kutty et al. (2018). Candidate orthologues of all species were accepted when the reciprocal best BLAST hit criterion was fulfilled against any of the five reference species of the orthologue set; all other settings were left as default. Results were summarized using custom perl scripts, and internal stop codons as well as Selenocysteine, if any, were replaced with “X” to avoid problems in downstream analyses (treated as “NNN” at the nucleotide level; see Kutty et al., 2018). The amino acid sequences were aligned using MAFFT v.7.123b (Katoh and Standley, 2013) with the *L-INS-i* algorithm. Multiple sequence alignments (MSAs) were then scrutinized for outliers. Alignment refinement and final removal of outlier sequences was carried out as described by Peters et al. (2017) on the amino acid and nucleotide levels. Additionally, we removed all sequences from all reference species included in the orthologue set except for those from *D. melanogaster*. Corresponding nucleotide sequence alignments for all single-copy genes were generated with Pal2Nal (Suyama et al., 2006) in a slightly modified version (see Misof et al., 2014) using the refined amino acid MSAs as blueprints. Sections that were classified as randomly or ambiguously aligned by Aliscore v.2.0 (Misof and Misof, 2009; Kück et al., 2010) at the amino acid level (parameters: maximal number of pairwise comparisons, option -e for gappy alignments derived from RNAseq data and otherwise defaults) were removed with ALICUT v.2.3 (https://www.zfmk.de/en/research/research-centres-and-groups/utilities). The corresponding regions in the nucleotide alignments were removed using a custom perl script. Before final concatenation in FASconCAT-G (Kück and Longo, 2014), the leading and trailing gaps in each MSA were recoded as “X” for the amino acids and “N” for nucleotides. All the above steps utilized custom perl scripts used in Misof et al. (2014) and available from http://software.zfmk.de/The-1KITE-project_evolution-of-insects.zip.

To improve the information content (IC) of our dataset, we applied MARE v.0.1.2-rc (Misof et al., 2013) to the supermatrix at the amino acid level with a taxon weighting (t) = 3 (otherwise defaults) to retain all species in the dataset. All phylogenomic analyses were carried out on this MARE reduced and optimized dataset containing 1456 gene partitions. A corresponding optimized nucleotide dataset was generated with custom made perl scripts as MARE is applicable for amino acid datasets only. An AliStat analysis (v.1.7: https://github.com/thomaskf/AliStat) was carried out on the amino acid dataset to determine the missing data per site as used in Misof et al. (2014).

Phylogenomic analyses

The amino acid and nucleotide datasets were analysed using MP and ML. Parsimony analyses were carried out in TNT v.1.5 (Goloboff and Catalano, 2016) (new technology search, level 10, hits 20, gaps coded as missing data). Node support was assessed by jackknife (JK) resampling (1000 replicates at 36% deletion). To assess the influence of codon positions on the results, additional parsimony analyses were carried out for (1) a nucleotide dataset from which the 3rd codon positions were excluded and (2) a nucleotide dataset with only the 2nd codon positions included. Saturation tests of the nucleotide positions (3rd position) were carried out in DAMBE (Xia, 2017) (transitions and transversions vs. divergence, genetic distance = GTR distances, minimum number of shared sites = default; Fig. S1). A random addition concatenation analysis (RADICAL) (Narechania et al., 2012) was carried out on the amino acid dataset (10 randomizations of 10 step loci additions) until all 1456 genes were concatenated. For RADICAL, PAUP (Swofford, 2000) was used for the parsimony tree reconstructions. The resultant parsimony trees were then compared to the topology obtained via analysis of the reference amino acid dataset. The AUC value (area under the curve) was calculated by the software based on the frequency that a node in the amino acid most parsimonious tree (MPT) appeared in the set of parsimony trees for each concatenation size (minimum number of randomly concatenated loci required for a node to always appear in the resulting trees). Bremer support values (Bremer, 1994) were determined in TNT v.1.6 for the MPT from the amino acid dataset based on constraint tree searches (three times).

For the ML analyses, data partitioning and model testing (models: see Table S6) were carried out in PartitionFinder 2.1.1 (Lanfear et al., 2014, 2017; Stamatakis, 2014) using the rcluster algorithm (rcluster-percent = 10, rclustermax = 1000, branchlengths = linked, analysis = raxml, model selection = AICc). The amino acid dataset with merged data partitions was then analysed in RAxML-HPC v.8.2.9 XSEDE (Stamatakis, 2014) on the CIPRES computer cluster (Cyberinfrastructure for Phylogenetic Research; San Diego Supercomputing Center) (Miller et al., 2010). The search for the best scoring ML tree (20 multiple inferences) was carried out under the best partitioning scheme and best model+GAMMA. Multiple nonparametric bootstrap replicates were carried out and convergence was ensured *a posteriori* based on the default extended majority rule (MRE)-based criterion (Pattengale et al., 2010) at 252 replicates. A multicoalescent species tree analysis was carried out in ASTRAL v.5.5.6 (Mirarab and Warnow, 2015) using gene trees (one tree search per gene) estimated by ML searches conducted in RAxML v.8.2.8 (Stamatakis, 2014) and otherwise defaults and additionally assessed branch support values for the tree were inferred by ASTRAL (Sayyari and Mirarab, 2016).

An additional recoded datamatrix named "aminoacid_recoded" was generated in which each amino acid in the original alignment was recoded to a six-state Dayhoff group using the "pgrecodeseq" command in the PHYLOGEARS v.2.0 tool package (Tanabe, 2008). This dataset was analysed using parsimony in TNT v.1.5 (Goloboff and Catalano, 2016) with the same parameters used as in the original dataset. For the likelihood analysis, the aminoacid_recoded dataset was analysed under the multistate option using the multistate data model GAMMA with GTR (-m MULTIGAMMA -K GTR). The datasets were evaluated to test if they have evolved under globally stationary, reversible and homogeneous (SRH) conditions (Jermiin et al., 2004). To test for possible compositional heterogeneity of amino acid matrices, we used SymTest v.2.0.47 (Jermiin and Ott, 2017, https://github.com/ottmi/symtest), which uses matched-pairs tests of homogeneity (for details see Misof et al., 2014). Tests were applied on (1) the amino acid supermatrix, and (2) the "aminoacid_recoded" matrix. Heatmaps were generated based on the p-values obtained from Bowker's matched-pairs test of symmetry (Bowker, 1948) in order to determine which sequence pairs matched SRH conditions.

Lastly, four-cluster likelihood mapping (FcLM) (Strimmer and von Haeseler, 1997) was used to assess the phylogenetic support and to identify potentially confounding signal for an important tree node: the sister-group relationship between *Cuterebra austeni* and *Mystacinobia zelandica* as inferred in the best tree. Four taxonomic groups were thus defined (see Table S7 for specific information about hypothesis and which species were part of which group). For testing, an optimized dataset was generated, including only those partitions of the supermatrix that contained sequences of at least one species from the four taxa specified for testing the respective hypothesis. The PTHREADS implementation of RAxML v.8.2.10 (Stamatakis, 2014) (option: -f q -Y groupfile -q partitionfile) was used to infer the support for each quartet using the partitioning scheme and substitution models inferred for the supermatrix at the amino acid level. Results were mapped into two-dimensional simplex graphs with a custom-made Perl script. Confounding signal can arise for a variety of reasons: (1) heterogeneous composition of amino acid sequences along the tree (i.e. among-lineage heterogeneity), (2) non-stationary substitution processes, and (3) non-random distribution of missing data on our phylogenetic inferences. To assess the possible impact of confounding signal, a permutation approach was implemented as suggested by Misof et al. (2014). It aims at eliminating

phylogenetic but keeping confounding signal. These tests were performed using FcLM, which was conducted using the same partitioning scheme as before, but using the LG substitution matrix across all partitions for permutations I, II, and III. Permutation tests were done for the one hypothesis and with the identical taxonomic groups as described above, using RAxML v.8.2.10.

Results

The concatenated supermatrix yielded 2732 single-copy protein-coding genes (alignment length: 755 910 sites at the amino acid level) after orthology prediction, alignment, alignment refinement, outlier check and outlier removal, and identification and deletion of ambiguously aligned sections from the MSAs. The coverage of the data matrix in terms of gene occurrence was 0.63 (IC = 0.40) and the coverage of the matrix by site was 0.59 (Ca completeness score, AliStat). The final MARE-optimized dataset comprised 1456 genes (alignment of 444 058 amino acid sites) for 41 species including the outgroup, with the gene occurrence coverage improved to 0.86 (IC = 0.60); the coverage according to sites increased to 0.75 (Fig. S2).

The MPT with node support based on the amino acid dataset and rooted with *D. melanogaster* is presented in Fig. 1. The nodes also provide the results of the RADICAL analysis (concatenation size and AUC values). The MP and ML trees are very similar (compare Figs 1 and 2) and the phylogram (Fig. 2) illustrates that the branch lengths of many poorly supported nodes within Oestroidea are particularly short. The underlying support for those branches with full jackknife support varies by more than one order of magnitude when it is assessed via Bremer supports or RADICAL. Hippoboscidae are placed as the sister group to the remaining calyptrates, with the subfamilies Ornithomyinae and Nycteribiinae (here including Streblidae) being represented by one species respectively (*Ortholfersia macleayi*, *Penicillidia dufourii*). *Fannia canicularis* (Fanniidae) is placed at the base of the muscoid grade. Muscidae are monophyletic, as are the subfamilies Muscinae and Coenosiinae, while Mydaeinae are paraphyletic with regard to Phaoniinae. Muscidae are the sister group to (Anthomyiidae–Scathophagidae)+Oestroidea. A monophyletic Scathophagidae is nested within paraphyletic Anthomyiidae. The superfamily Oestroidea is monophyletic and nested within the muscoid grade as the sister group to the Anthomyiidae–Scathophagidae clade. Within Oestroidea, the families Sarcophagidae, Tachinidae and Rhinophoridae are well supported, while the monophyly of Oestridae, Rhiniidae, Mesembrinellidae, Mystacinobiidae and Ulurumyiidae could not be tested (represented by one species each). Oestridae are placed as the sister group to Mystacinobiidae, with this clade being recovered as the sister group to Sarcophagidae with high support in both MP and ML analyses. Within Tachinidae, a monophyletic Tachininae is recovered as the sister group to Exoristinae. *Euthera bicolor* (probably Dexiinae, see below) is the sister group to Phasiinae. As expected, calliphorids are paraphyletic. *Pollenia* (Polleniinae) is strongly supported as the sister group to the tachinids, and strong support is also obtained for a clade consisting of the macrolarviparous Helicoboscinae and Ameniinae together with Rhinophoridae. *Stomorhina subapicalis* (Rhiinidae) is nested in a calliphorid clade as the strongly supported sister group to *Verticia nigra* (Bengaliinae). The calliphorids *Calliphora vomitoria* (Calliphorinae) and *Lucilia sericata* (Luciliinae) are strongly supported sister taxa, but the ‘backbone’ of the calliphorid assemblage is otherwise poorly supported.

Conflicts or instability in relationships between the various analyses are primarily observed in Oestroidea (see Table 2; nodes A–I on Fig. 1). *Mesembrinella bellardiana* (Mesembrinellidae) and *Ulurumyia macalpinei* (Ulurumyiidae) are sister groups on the MP tree (MP_JK = 70), while *Ulurumyia macalpinei* is recovered as sister to the remaining Oestroidea (ML_BS = 88) on the ML tree. The aminoacid_recoded dataset recovered the latter relationship with parsimony but again with low support (MP_JK = 53). Also, while *Amenia* sp. (Ameniinae) is sister to the rhinophorids on the MP tree (MP_JK = 88), it is sister to *Eurychaeta muscaria* (Helicoboscinae) in the ML tree (ML_boot = 83). A moderately well-supported sister-group relationship between these two species is recovered based on the aminoacid_recoded dataset (MP_JK = 91, ML_boot = 88) (see Fig. S7 for relationships from the aminoacid_recoded dataset). *Chrysomya megacephala* (Chrysomyinae) is recovered as the sister group to the Bengaliinae+Rhiniidae clade in both analyses, but with low support (MP_JK = 57, ML_boot = 67). Also, the sister-group relationship between Calliphorinae+Luciliinae and Chrysomyinae+(Bengaliinae+Rhiinidae) as well as the sister-group relationship between this group and the clade encompassing Helicoboscinae, Ameniinae and Rhinophoridae is weakly supported in both analyses (MP_JK = 57, ML_boot = 71). Phasiinae (Tachinidae) are recovered as monophyletic but with very low support (MP_JK = 50; ML_boot = 52), but support values improved when the aminoacid_recoded dataset was analysed (MP_JK = 78; ML_boot = 85). Finally, the sister-group relationship between the muscids *Phaonia angelicae* (Phaoniinae) and *Mydaea urbana* (Mydaeinae) is not robust (MP_JK = 69, ML_BS = 55), although Mydaeinae monophyly is supported by the analysis of the aminoacid_recoded dataset (MP_JK = 99, ML_BS = 80).

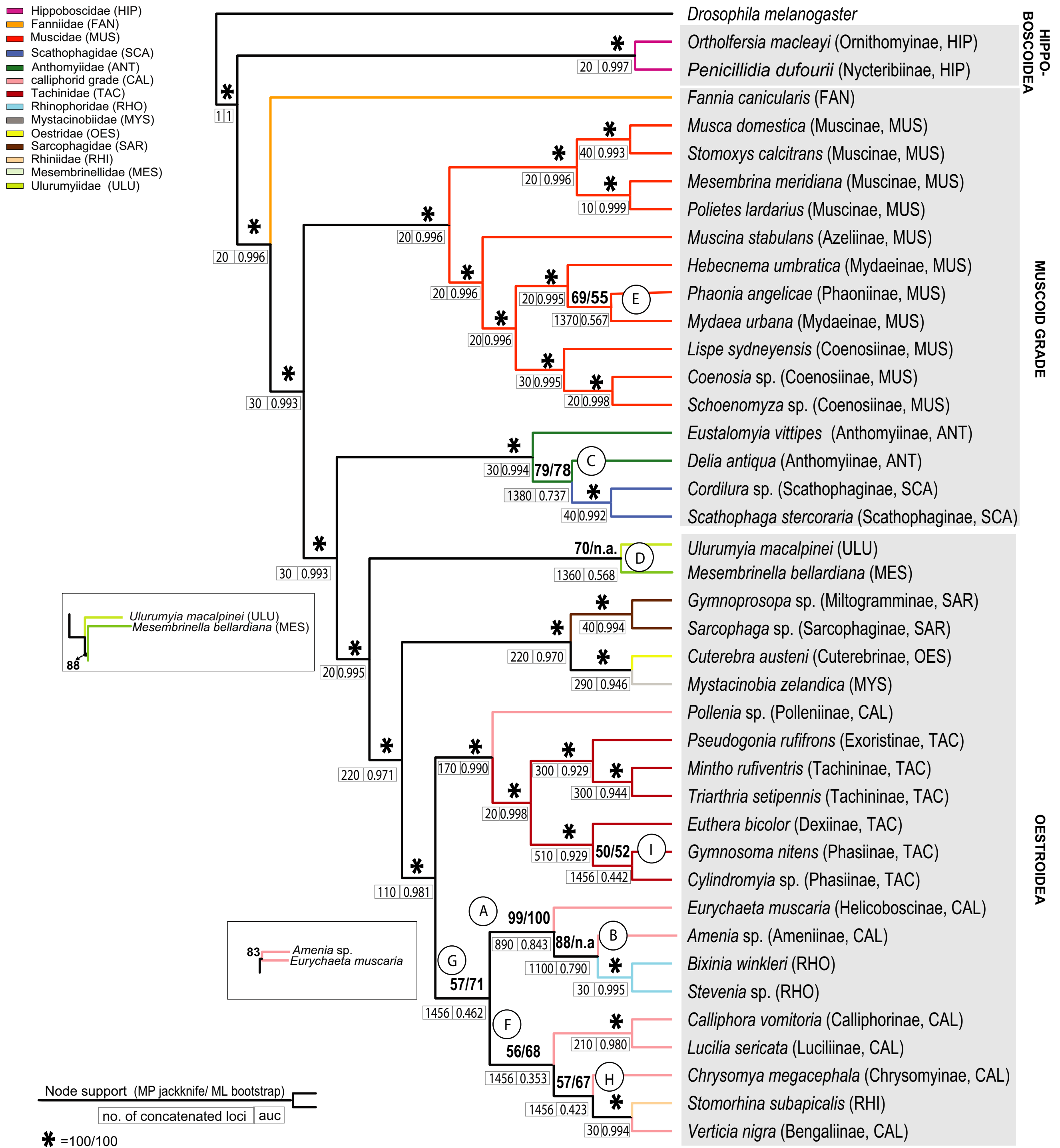

Fig. 1. Most parsimonious tree (MPT) inferred from the amino acid dataset with support values from both maximum parsimony (MP) and maximum likelihood (ML) analyses mapped above nodes (MP jackknife/ML bootstrap) and RADICAL support values presented below nodes. A–I indicate relationships that are unstable (refer to Table 2). The two relationships that are in conflict with the ML topology are illustrated in boxes.

Parsimony trees based on amino acids and nucleotides are mostly congruent (Fig. S3) when either 3rd codon positions or both 1st and 3rd codon positions are excluded (see Table 2). There are two conflicting nodes between the topologies based on amino acids and those based on the 2nd codon positions (within

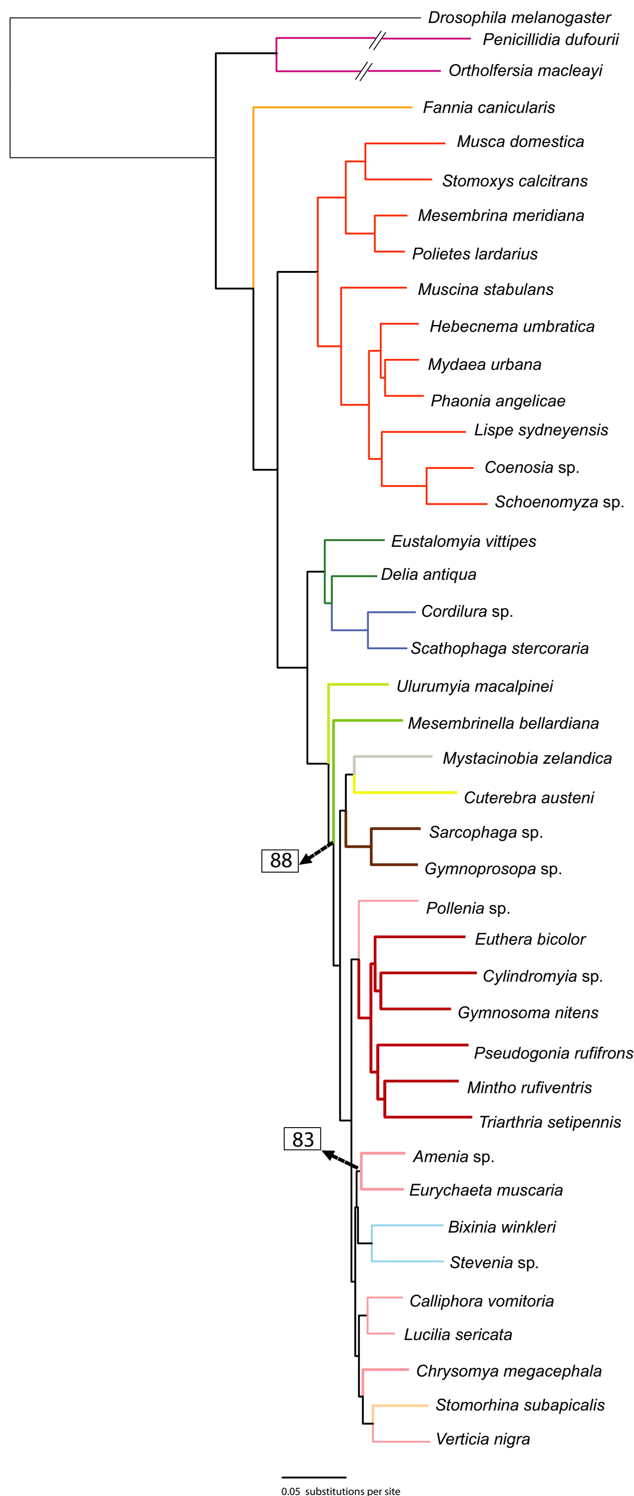

Fig. 2. Phylogram from ML analysis denoting a rapid radiation within the calyptrates. The particularly short branch lengths within the Oestroidea contribute to the difficulty in inferring relationships within this clade. Only nodes and support values for two relationships that are in conflict with the MP topology are highlighted (arrowed).

Tachinidae and the position of *Ulurumyia*). The coalescent-based species tree analysis (Fig. S3) is overall similar to the MP and ML trees and all nodes that are in conflict with concatenation-based topologies are weakly supported (Fig. S4). Heatmaps generated with SymTest for both the amino acid and the aminoacid_recoded datasets show reduced heterogeneity across lineages with the recoded matrix (Fig. S5). FcLM analyses and permutation tests on the sister-group relationship between Oestridae and *Mystacinobia zelandica* (see Fig. S6 and Table S7) confirmed strong signal for T3 topology (similar to Fig. 1) and minimal signal for the alternativee T1 topology. The permutations tests designed to identify confounding signal recovered < 30% support for the T3 topology and hence confirm that this novel relationship was not driven by confounding signal.

Discussion

We present the first phylogenomic hypothesis for Calyptratae based on transcriptome data. Most nodes obtained in the parsimony-, likelihood- and coalescent-based analyses are well supported and congruent. The trees based on amino acid data are overall also very similar to the trees based on nucleotide data. Not surprisingly, the highest level of congruence is observed for the tree based on 2nd codon positions (36 congruent nodes), although the tree based on all codon positions fares only marginally worse (31 congruent nodes).

High jackknife and bootstrap support and the need for additional support measures in phylogenomics

Overall we find that our phylogenomic analysis based on a fairly small number of taxa provides good resolution and high support for many of the higher-level relationships within Calyptratae for which the previous multi-gene datasets from the Sanger era failed to provide high support. On the MPT, 30 out of 39 nodes have support values of 100, five nodes have support values in the range 60–99, and only four nodes have values < 60. Not surprisingly, the support values are even higher on the likelihood tree where 31 of the 39 nodes have support values of 100, six nodes have support values of 65–99, and two nodes have support values of < 60. Of course, support values of 100% have to be interpreted in phylogenomic analyses as high support values can be achieved based on large numbers of characters and systematic biases in the dataset (Yeates et al., 2016). We therefore analysed our dataset using multiple strategies because some of these biases will be restricted to DNA or amino acid data. For example, we assessed the concatenation size

Table 2
Comparison of node supports across different datasets and analysis strategies. (AA_MP = amino acid dataset+parsimony, AA_ML = amino acid dataset+maximum likelihood, NT_MP = nucleotide dataset+maximum parsimony, NT_12_MP = reduced nucleotide dataset with codon position 1 and 2 only+maximum parsimony, NT_2 _MP = reduced nucleotide dataset codon with position 2 only+maximum parsimony, ASTRAL = coalescent-based species tree analysis, AA_recoded_MP = amino acid recoded+MP, AA_recoded_ML = amino acid recoded+ML). Nodes with support < 100 on the amino acid parsimony tree (see nodes A-I on Fig. 1) and the corresponding supports on trees from the other analyses are shown.

Node support	Node on Fig. 1		AA_MP	AA_ML	NT_MP	NT_12_MP	NT_2 _MP	ASTRAL	AA_recodedMP	AA_recodedML
100			30	30	29	29	28	28	28	28
80-99	1.	**A**	99	100	100	100	93	n.a	94	97
	2.	**B**	88	n.a	n.a	n.a	<50	n.a	n.a	n.a
60-79	1.	**C**	79	78	n.a	76	70	100	n.a	n.a
	2.	**D**	70	n.a	100	68	n.a	98	n.a	n.a
	3.	**E**	69	55	n.a	88	80	n.a	n.a	n.a
< 60	1.	**F**	56	68	n.a	n.a	92	n.a	63	80
	2.	**G**	57	71	n.a	n.a	98	n.a	69	85
	3.	**H**	57	67	n.a	n.a	97	n.a	60	70
	4.	**I**	50	52	n.a	n.a	n.a	99	78	85

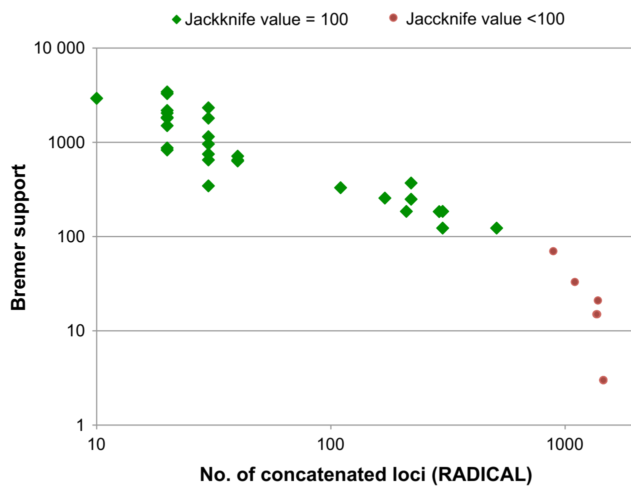

Fig. 3. True support for nodes with full jackknife values (100) varies by more than one order of magnitude when assessed with RADICAL or Bremer. Higher Bremer support values are observed for nodes that are recovered from a smaller number of concatenated loci.

or the number of loci required to confidently recover a node as carried out in the RADICAL analysis as a measure for relative support. This revealed that full jackknife support (100%) requires 20–510 (1–35%) of the 1456 genes. Similarly, Bremer support values for nodes with full jackknife support also vary by more than one order of magnitude (123–3427; Fig. 3 and Table S8). This highlights the need for using clade support measures in phylogenomic analyses that are more sensitive than jackknife/bootstrap support values (Brower, 2018). Another relative support measure could be obtained by using higher deletion proportions during jackknife analyses (Farris et al., 1996).

One major source of error in phylogenomic analyses can be compositional heterogeneity at the nucleotide or amino acid level. We therefore also analysed our data at both levels and also used Dayhoff-6 recoding to explore if our findings are influenced by compositional biases at the amino acid level. Regardless of coding, both datasets exhibit among-lineage compositional heterogeneity, but the heterogeneity was remarkably reduced using recoding based on the Dayhoff-6 matrix. (Fig. S5). However, we find that topology and support values are very similar to those that are obtained with datasets that had higher levels of among-lineage compositional heterogeneity. Indeed, most of the highly supported nodes in one analysis are also recovered with strong support in all other analyses and vice versa (Fig. S7). There are only two exceptions. The recoded data recover a monophyletic Mydaeinae with moderate to high support. Similarly, the anthomyiid *Eustalomyia* is recovered as the sister group to the scathophagids when recoding is used, although support is low. Additionally, the strongly supported sister-group relationship between Oestridae and Mystacinobiidae was affected in the likelihood analysis, where node support dropped from 100 to < 80. However, this relationship hypothesis remains favoured given the signal identified in the FcLM analysis. The quartet permutation tests also revealed that confounding signal coming from model violation due to among-lineage heterogeneity and missing data distribution can be excluded. Congruence between MP topologies of the full nucleotide dataset and datasets excluding 3rd codon position also reveals that long-branch attraction artefacts are not influencing relationships.

Phylogenomics in calyptrates: the success stories

The impact of improved gene sampling is evident when the calyptrate relationships from this study are compared to two previously published higher-level

calyptrate phylogenetic studies: the eight-gene parsimony analysis of 275 species by Kutty et al. (2010) and the three-gene likelihood analysis by Cerretti et al. (2017) (Fig. 4) based on 89 species. The trees obtained in both analyses suffered from poor node support in several areas of the topologies. For example, some families were not recovered as monophyletic in Kutty et al. (2010), including Tachinidae and Rhinophoridae, both of which are now strongly supported based on our phylogenomic dataset comprising 1456 molecular loci. Oestroidea were not monophyletic in Cerretti et al. (2017) due to the placement of *Mystacinobia* as sister to the Anthomyiidae–Scathophagidae clade. However, with increased phylogenomic gene sampling, this superfamily is now monophyletic with strong support. We also find that many relationships that previously only had moderate to high support based on Sanger data now have full support (= 100) in the phylogenomic analysis (Fig. 4). For example, in Kutty et al. (2010) the calyptrate backbone relationships between Hippoboscoidea, the families of the muscoid grade and the Oestroidea previously had node supports < 80, but these relationships are now highly supported. Node support for the monophyly of Sarcophagidae and Tachinidae ranged from < 50 to 100 in studies based on Sanger data, but again these families now have increased or full support. Within Muscidae, all subfamily relationships have higher node support (Fig. 3), except for the relationships within the Phaoniinae + Mydaeinae clade, which still cannot be resolved with confidence. There are two prominent cases where well-supported nodes in Kutty et al. (2010; BS = 88, 100) are now contradicted by the phylogenomic tree. These are *Mystacinobia*+*Ulurumyia* and Bengaliinae+Chrysomyinae, which are now placed differently with moderately to high support.

The relationships among the non-oestroid calyptrate families are congruent with the previous molecular hypotheses by Kutty et al. (2010). Scathophagidae are nested within a paraphyletic Anthomyiidae, which conflicts with morphological evidence that recovers Anthomyiidae to be monophyletic (Michelsen, 1991, 1996; Michelsen and Pape, 2017). Within the monophyletic Muscidae, the monophyly of the subfamilies Azeliinae, Muscinae and Coenosiinae is corroborated and a close relationship between Mydaeinae and Phaoniinae is in agreement with the findings in Kutty et al. (2014) and Haseyama et al. (2015).

Phylogenomics in calyptrates: partial successes and failures

The relationships within the main oestroid lineages remain the main challenge in calyptrate phylogenetics. *Ulurumyia* (as McAlpine's fly) and *Mystacinobia* were sister taxa in Kutty et al.'s (2010) study, and together they were sister to Sarcophagidae, although with such low support that both species could only be assigned with confidence to the oestroid clade. Our study now reduces the number of supported relationship hypotheses to two. *Ulurumyia* is either the sister group to Mesembrinellidae or to a clade consisting of Mesembrinellidae plus the remaining Oestroidea, with the latter placement having support from morphology (Michelsen and Pape, 2017). Another major step towards understanding the relationships within Oestroidea is finding strong support for a sister-group relationship between *Mystacinobia* and Oestridae. It is interesting to note that both taxa are mammal-associated. Various mammals serve as hosts of Oestridae, and larvae of different species develop under the skin, in the nasal passages or in the digestive tract. Larvae of *Cuterebra austeni*, the species of Oestridae included in this study, are subcutaneous parasites of woodrats (Baird, 1997). *Mystacinobia* is associated with bat guano with adults being phoretic on the bats (Holloway, 1976).

Mesembrinellidae used to be treated as a calliphorid subfamily but they are here again confirmed as a separate lineage from the other calliphorids. Junqueira et al. (2016) found support for a sister-group relationship between Mesembrinellidae and Tachinidae, while Marinho et al. (2017) were in favour of a sister-group relationship to a clade consisting of *Pollenia* and Sarcophagidae, in both cases with poor support and without the inclusion of *Ulurumyia*. Cerretti et al. (2017) found strong molecular and modest morphological support for Mesembrinellidae being sister to *Ulurumyia*. Both taxa are macrolarviparous, but their natural breeding habits have remained elusive. Our study is the first to provide strong support for a placement of Mesembrinellidae as the sister group of Oestroidea; either Mesembrinellidae+Ulurumyiidae form the sister clade to the remaining Oestroidea, or Mesembrinellidae are the sister group to all Oestroidea excluding Ulurumyiidae. Ulurumyiidae are restricted to Australia, and Mesembrinellidae are restricted to the Neotropics, suggesting a biogeographical signal of a transantarctic dispersal event at or near the base of Oestroidea.

Within the monophyletic Tachinidae our study includes species from three or four subfamilies depending on which position of *Euthera* is accepted (this genus has been classified in both Dexiinae and Phasiinae and is sometimes even hypothesized to be the sister group of Dexiinae+Phasiinae: Cerretti et al., 2014b). Molecular data appear to support inclusion in Dexiinae (Winkler et al., 2015; Blaschke et al., 2018), which is in agreement with the present study. However, while the monophyly of Phasiinae is recovered here, support is low. Subfamily classification in the Tachinidae is still controversial, particularly given the

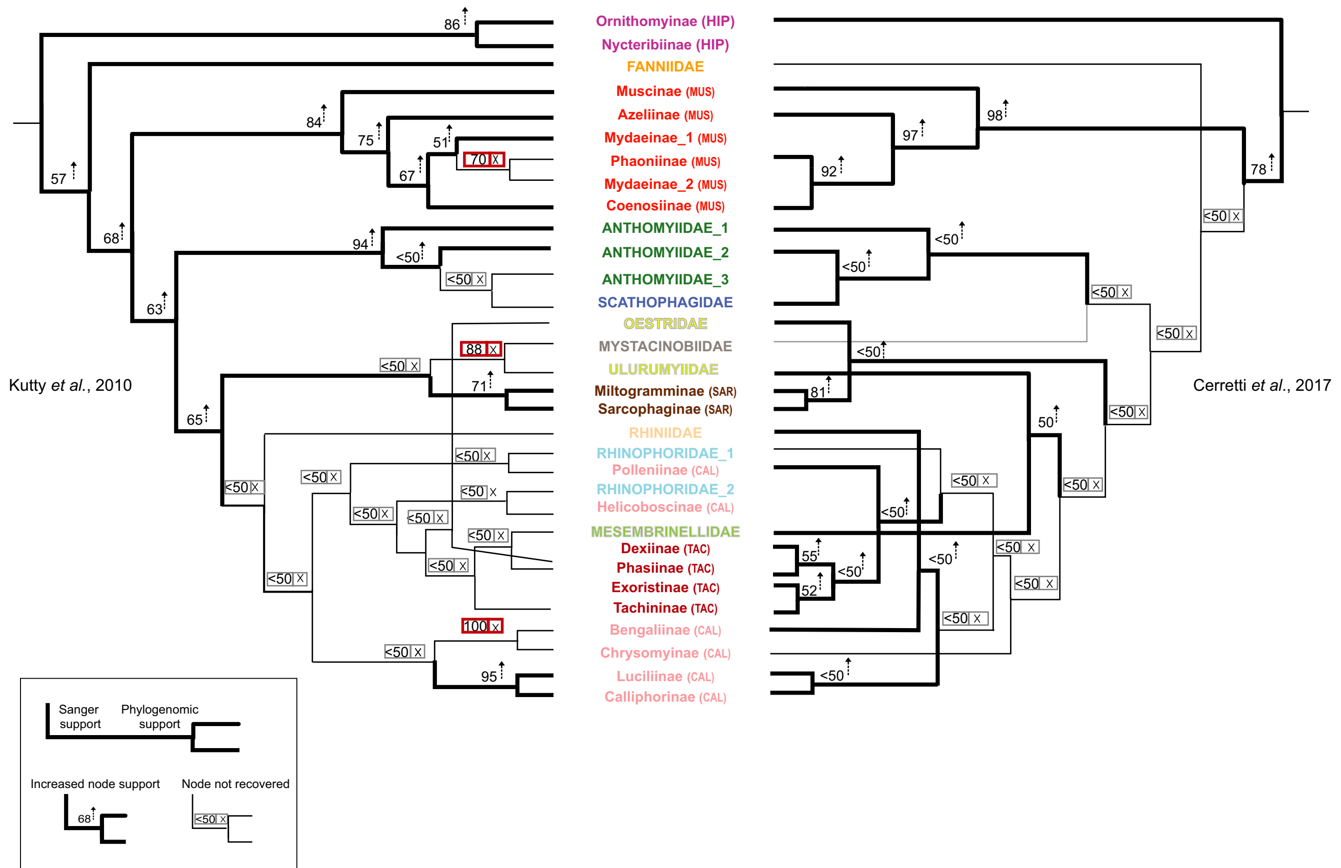

Fig. 4. Phylogenomic topology compared with multi-gene-studies: Kutty et al.'s (2010) parsimony topology (left) and Cerretti et al.'s (2017) likelihood topology (right). The phylogenomic topology confirms all but three supported nodes from previous studies (see red boxes) but increases support for many poorly supported nodes (see arrows).

potential non-monophyly of the Dexiinae and Tachininae (Cerretti et al., 2014b). The relationships obtained in the present study are congruent with Cerretti et al. (2014b), Winkler et al. (2015) and Blaschke et al. (2018) in showing a sister-group relationship between the Dexiinae–Phasiinae clade and the Tachininae–Exoristinae clade. Polleniinae are recovered as the sister taxon to Tachinidae, which is consistent with other recent molecular analyses that included both groups (Winkler et al., 2015; Zhang et al., 2016), except for Marinho et al. (2017) where these taxa are not closely related. Larvae of both Polleniinae and Tachinidae are parasitic but on disparate taxa: earthworms are the hosts for polleniines while tachinids attack a range of arthropod hosts (mainly insects).

Calliphoridae, the remaining stumbling block

Overall, the phylogenomic data help in resolving and supporting some relationships within Oestroidea, but many challenges remain. Several relationships are poorly supported although consistently recovered across a range of different analyses. Other poorly supported nodes are unstable and topologies change between analyses. It is apparent that the 1456 single-copy protein-encoding genes analysed in this study do not contain sufficient signal for resolving the phylogenetic position of these taxa. The main challenge is the relationships of the subfamilies of Calliphoridae (other than Polleniinae), some of which have never been analysed based on molecular data. A number of these subfamilies have previously been accorded family rank, but this helps little with resolving the relationships between them. RADICAL reveals that the nodes pertaining to these (sub)families are particularly unstable and emerge only with support when almost all genes are concatenated. This is likely to be partially due to the short branch lengths within the Oestroidea (Fig. 2).

The Rhiniidae, formerly a calliphorid subfamily that was raised to family level by Kutty et al. (2010), may be the sister group to Bengaliinae. This hypothesis matches the results of Cerretti et al. (2017), and the Chrysomyinae+Bengaliinae+Rhiniidae clade is also congruent with the findings of Singh and Wells (2013), although support values were very low in both studies. *Calliphora* and *Lucilia*, representing the subfamilies

Calliphorinae and Luciliinae, are strongly supported as sister groups, but the node joining this clade to the Chrysomyinae+Bengaliinae+Rhiniidae clade is again poorly supported. A clade comprising Helicoboscinae, Ameniinae and Rhinophoridae receives high support. An association between Helicoboscinae and a rhinophorid species was recovered previously (Kutty et al., 2010), but Ameniinae were not included. Species of Helicoboscinae and Ameniinae are macrolarviparous and breed in dead/dying or living snails, respectively (see Cerretti et al., 2017 for references), while the Rhinophoridae are exclusively parasitoids of woodlice (Cerretti et al., 2014a). Based on the natural history data, the ML topology in which Ameniinae and Helicoboscinae are sister taxa is more compelling. The node subtending these snail- and woodlouse-associated taxa with the other calliphorid subfamilies and the Rhiniidae is poorly supported. The relationships between these calliphorid lineages, and the monophyly of the group overall, are awaiting better supported resolution.

Conclusions

We overall obtain a well-supported phylogeny based on transcriptome data. The main outstanding challenges are the following: identifying the basal split within the Oestroidea, identifying well-supported taxonomic units to replace the traditional "Calliphoridae", and obtaining a well-supported phylogeny at the subfamily level free from systematic bias. Our results are not unusual in the sense that many phylogenomic studies based on sparse taxon sampling resolve some but not all higher-level relationships (Johnson et al., 2013; Bond et al., 2014). Increasing the taxon density within "problematic" clades and including data from additional character systems (e.g. morphology, physiology, genome architecture) is the next logical step for addressing the questions that remain. With broader taxon and gene coverage, and perhaps other metacharacters not based on primary sequence data, we will hopefully be able to illuminate evolutionary patterns regarding morphology, geographical ancestry and the remarkably diverse life histories across this group of flies. In this sense, by increasing the number of relationships that are now well understood, although it highlights which questions require more attention, our study sets the stage for future work.

Acknowledgements

The project was supported by funding from SEABIG (NUS grant nos. R-154-000-648-646 and R-154-000-648-733 to RM) and by the US National Science Foundation (DEB-1257960 to BMW, DKY and RM). DKY's research is supported by the Schlinger Endowment to the Australian National Insect Collection. We acknowledge Alexander Donath, Lars Jermiin and Michael Ott from 1KITE for help with bioinformatics. We also thank Ralph S. Peters, Rolf Beutel and Michelle Trautwein from the 1KITE Antliophora group, and furthermore Andrew McKenzie, James E. O'Hara and Jens Trasberger for collecting efforts. Material collected after October 2014 has been handled in agreement with the Nagoya Protocol on Access and Benefit Sharing (ABS).

References

Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Baird, C.R., 1997. Bionomics of *Cuterebra austeni* (Diptera: Cuterebridae) and its association with *Neotoma albigula* (Rodentia: Cricetidae) in the southwestern United States. J. Med. Entomol. 34, 690–695.

Bernasconi, M.V., Pawlowski, J., Valsangiacomo, C., Piffaretti, J.-C., Ward, P.I., 2000a. Phylogeny of the Scathophagidae (Diptera, Calyptratae) based on mitochondrial DNA sequences. Mol. Phylogenet. Evol. 16, 308–315.

Bernasconi, M.V., Valsangiacomo, C., Piffaretti, J.-C., Ward, P.I., 2000b. Phylogenetic relationships among Muscoidea (Diptera: Calyptratae) based on mitochondrial DNA sequences. Insect Mol. Biol. 9, 67–74.

Blaschke, J.D., Stireman, J.O.III., O'Hara, J.E., Cerretti, P., Moulton, J.K., 2018. Molecular phylogeny and the evolution of piercers in the bug-killing flies (Diptera: Tachinidae: Phasiinae). Syst. Entomol. 43, 218–238.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.

Bond, J.E., Garrison, N.L., Hamilton, C.A., Godwin, R.L., Hedin, M., Agnarsson, I., 2014. Phylogenomics resolves a spider backbone phylogeny and rejects a prevailing paradigm for orb web evolution. Curr. Biol. 24, 1765–1771.

Bowker, A.H., 1948. A test for symmetry in contingency tables. J. Am. Stat. Assoc. 43, 572–574.

Bremer, K., 1994. Branch support and tree stability. Cladistics 10, 295–304.

Brower, A. V. Z., 2018. "Maximum support" = 100% BS. Cladistics. https://doi.org/10.1111/cla.12356.

Buenaventura, E., Pape, T., 2017. Multilocus and multiregional phylogeny reconstruction of the genus *Sarcophaga* (Diptera, Sarcophagidae). Mol. Phylogenet. Evol. 107, 619–629.

Buenaventura, E., Whitmore, D., Pape, T., 2017. Molecular phylogeny of the hyperdiverse genus *Sarcophaga* (Diptera: Sarcophagidae), and comparison between algorithms for identification of rogue taxa. Cladistics 33, 109–133.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinformatics 10, 421.

Carvalho, C.J.B.D., Couri, M.S., 2002. Cladistic and biogeographic analyses of *Apsil* Malloch and *Reynoldsia* Malloch (Diptera : Muscidae) of southern South America. Proc. Entomol. Soc. Wash. 104, 309–317.

Cerretti, P., Lo Giudice, G., Pape, T., 2014a. Remarkable Rhinophoridae in a growing generic genealogy (Diptera: Calyptratae, Oestroidea). Syst. Entomol. 39, 660–690.

Cerretti, P., O'Hara, J.E., Wood, D.M., Shima, H., Inclan, D.J., Stireman, J.O.III., 2014b. Signal through the noise? Phylogeny of

the Tachinidae (Diptera) as inferred from morphological evidence Syst. Entomol. 39, 335–353.

Cerretti, P., Stireman, J.O. III, Pape, T., O'Hara, J.E., Marinho, M.A.T., Rognes, K., Grimaldi, D.A., 2017. First fossil of an oestroid fly (Diptera: Calyptratae: Oestroidea) and the dating of oestroid divergences. PLoS ONE 12, e0182101.

Crosskey, R.W., 1977. A review of the Rhinophoridae (Diptera) and a revision of the Afrotropical species. Bull. br. Mus. nat. Hist. Entomol. 36, 1–66.

Dayhoff, M.O., Schwartz, R.M., Orcutt, B.C., 1978. A model of evolutionary change in proteins. In: Dayhoff, M.O. (Ed.), Atlas of Protein Sequence and Structure. National Biomedical Research Foundation, Washington DC, pp. 345–352.

Dell'Ampio, E., Meusemann, K., Szucsich, N.U., Peters, R.S., Meyer, B., Borner, J., Petersen, M., Aberer, A.J., Stamatakis, A., Walzl, M.G., Minh, B.Q., von Haeseler, A., Ebersberger, I., Pass, G., Misof, B., 2014. Decisive data sets in phylogenomics: lessons from studies on the phylogenetic relationships of primarily wingless insects. Mol. Biol. Evol. 31, 239–249.

Ding, S., Li, X., Wang, N., Cameron, S.L., Mao, M., Wang, Y., Xi, Y., Yang, D., 2015. The phylogeny and evolutionary timescale of Muscoidea (Diptera: Brachycera: Calyptratae) inferred from mitochondrial genomes. PLoS ONE 10, e0134170.

Dittmar, K., Porter, M.L., Murray, S., Whiting, M.F., 2006. Molecular phylogenetic analysis of nycteribiid and streblid bat flies (Diptera : Brachycera, Calyptratae): implications for host associations and phylogeographic origins. Mol. Phylogenet. Evol. 38, 155–170.

Dittmar, K., Morse, S., Dick, C.W., Patterson, B.D., 2015. Bat fly evolution from the Eocene to the Present (Hippoboscoidea, Streblidae and Nycteribiidae). In: Morand, S., Krasnov, B., Littlewood, T. (Eds.), Parasite Diversity and Diversification: Evolutionary Ecology Meets Phylogenetics. Cambridge University Press, Cambridge, pp. 246–264.

Domínguez, M.C., Roig-Juñent, S.A., 2008. A phylogeny of the family Fanniidae Schnabl (Insecta : Diptera: Calyptratae) based on adult morphological characters, with special reference to the Austral species of the genus *Fannia*. Invertebr. Syst. 22, 563–587.

Embley, T.M., van der Giezen, M., Horner, D.S., Dyal, P.L., Foster, P., 2003. Mitochondria and hydrogenosomes are two forms of the same fundamental organelle. Philos. Trans. R. Soc. Lond. B Biol. Sci. 358, 191–201.

Farris, J.S., Albert, V.A., Källersjö, M., Lipscomb, D., Kluge, A.G., 1996. Parsimony jackknifing outperforms neighbor-joining. Cladistics 12, 99–124.

Ferrar, P., 1979. Immature stages of dung-breeding muscoid flies in Australia, with notes on the species, and keys to larvae and puparia. Aust. J. Zool. Suppl. Ser. 27, 1–106.

Gleeson, D.M., Howitt, R.L.J., Newcomb, R.D., 2000. The phylogenetic position of the New Zealand batfly, *Mystacinobia zelandica* (Mystacinobiidae; Oestroidea) inferred from mitochondrial 16S ribosomal DNA sequence data. J. R. Soc. N. Z. 30, 155–168.

Goloboff, P.A., Catalano, S.A., 2016. TNT version 1.5, including a full implementation of phylogenetic morphometrics. Cladistics 32, 221–238.

Gorodkov, K.B., 1986. Scathophagidae. In: Soós, A., Papp, L. (Eds.), Catalogue of Palaearctic Diptera, Scathophagidae—Hypodermatidae. Elsevier, Amsterdam, pp. 11–41.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotech. 29, 644–652.

Griffiths, G.C.D., 1972. The phylogenetic classification of Diptera Cyclorrhapha, with special reference to the male postabdomen. Ser. Entomol. 8, 1–340.

Griffiths, G.C.D., 1982. On the systematic position of *Mystacinobia* (Diptera: Calliphoridae). Mem. Entomol. Soc. Wash. 10, 70–77.

Grimaldi, D., Engel, M.S., 2005. Evolution of the Insects. Cambridge University Press, Cambridge, UK.

Guimarães, J.H., 1977. A systematic revision of the Mesembrinellidae, stat. nov. (Diptera, Cyclorrhapha). Arq. Zool. 29, 1–109.

Hackman, W., Väisänen, R., 1985. The evolution and phylogenetic significance of the costal chaetotaxy in the Diptera. Ann. Zool. Fenn. 22, 169–203.

Haseyama, K.L.F., Wiegmann, B.M., Almeida, E.A.B., Carvalho, C.J.B.d, 2015. Say goodbye to tribes in the new house fly classification: a new molecular phylogenetic analysis and an updated biogeographical narrative for the Muscidae (Diptera). Mol. Phylogenet. Evol. 89, 1–12.

Hennig, W., 1973. Ordnung Diptera (Zweiflügler). Handbuch der Zoologie. Eine Naturgeschichte der Stämme des Tierreiches. IV. Band: Arthropoda – 2 Hälfte: Insecta. 2. Teil: Spezielles. De Gruyter, Berlin, pp. 1–337.

Herting, B., 1957. Das weibliche Postabdomen der calyptraten Fliegen (Diptera) und sein Merkmalswert für die Systematik der Gruppe. Z. Morphol. Okol. Tiere 45, 429–461.

Herting, B., 1984. Catalogue of Palearctic Tachinidae (Diptera). Stuttgarter Beiträge zur Naturkunde (A) 369, 1–228.

Holloway, B.A., 1976. A new bat-fly family from New Zealand (Diptera: Mystacinobiidae). New Zeal. J. Zool. 3, 279–301.

Hrdy, I., Hirt, R.P., Dolezal, P., Bardonová, L., Foster, P.G., Tachezy, J., Martin Embley, T., 2004. *Trichomonas* hydrogenosomes contain the NADH dehydrogenase module of mitochondrial complex I. Nature 432, 618.

Jermiin, L., Ott, M., 2017. SymTest version 2.0.47. Available online at: https://github.com/ottmi/symtest.

Jermiin, L.S., Ho, S.Y.W., Ababneh, F., Robinson, J., Larkum, A.W.D., 2004. The biasing effect of compositional heterogeneity on phylogenetic estimates may be underestimated. Syst. Biol. 53, 638–643.

Johnson, B. R., Borowiec Marek, L., Chiu Joanna, C., Lee Ernest, K., Atallah, J., Ward Philip, S., 2013. Phylogenomics resolves evolutionary relationships among ants, bees, and wasps. Curr. Biol. 23, 2058–2062.

Junqueira, A.C.M., Azeredo-Espin, A.M.L., Paulo, D.F., Marinho, M.A.T., Tomsho, L.P., Drautz-Moses, D.I., Purbojati, R.W., Ratan, A., Schuster, S.C., 2016. Large-scale mitogenomics enables insights into *Schizophora* (Diptera) radiation and population diversity. Sci. Rep. 6, 21762.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.

Kück, P., Longo, G.C., 2014. FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. Front. Zool. 11, 81.

Kück, P., Meusemann, K., Raupach, M., von Reumont, B., Wägele, W., Misof, B., 2010. Parametric and non-parametric masking of randomness in sequence alignments can be improved and leads to better resolved trees. Front. Zool. 7, 10.

Kutty, S.N., Bernasconi, M.V., Sifner, F., Meier, R., 2007. Sensitivity analysis, molecular systematics and natural history evolution of Scathophagidae (Diptera: Cyclorrhapha: Calyptratae). Cladistics 23, 64–83.

Kutty, S.N., Pape, T., Pont, A., Wiegmann, B.M., Meier, R., 2008. The Muscoidea (Diptera: Calyptratae) are paraphyletic: evidence from four mitochondrial and four nuclear genes. Mol. Phylogenet. Evol. 49, 639–652.

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. Syst. Entomol. 35, 614–635.

Kutty, S.N., Pont, A.C., Meier, R., Pape, T., 2014. Complete tribal sampling reveals basal split in Muscidae (Diptera), confirms saprophagy as ancestral feeding mode, and reveals an evolutionary correlation between instar numbers and carnivory. Mol. Phylogenet. Evol. 78, 349–364.

Kutty, S.N., Wong, W.H., Meusemann, K., Meier, R., Cranston, P.S., 2018. A phylogenomic analysis of Culicomorpha (Diptera) resolves the relationships among the eight constituent families. Syst. Entomol. 43, 434–446.

Lambkin, C.L., Sinclair, B.J., Pape, T., Courtney, G.W., Skevington, J.H., Meier, R., Yeates, D.K., Blagoderov, V., Wiegmann, B.M., 2013. The phylogenetic relationships among infraorders and superfamilies of Diptera based on morphological evidence. Syst. Entomol. 38, 164–179.

Lanfear, R., Calcott, B., Kainer, D., Mayer, C., Stamatakis, A., 2014. Selecting optimal partitioning schemes for phylogenomic datasets. BMC Evol. Biol. 14, 82.

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2017. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol. Biol. Evol. 34, 772–773.

Macleod, N., Rawson, P.F., Forey, P.L., Banner, F.T., Boudagher-Fadel, M.K., Bown, P.R., Burnett, J.A., Chambers, P., Culver, S., Evans, S.E., Jeffery, C., Kaminski, M.A., Lord, A.R., Milner, A.C., Milner, A.R., Morris, N., Owen, E., Rosen, B.R., Smith, A.B., Taylor, P.D., Urquhart, E., Young, J.R., 1997. The Cretaceous–Tertiary biotic transition. J. Geol. Soc. 154, 265–292.

Marinho, M.A.T., Junqueira, A.C.M., Paulo, D.F., Esposito, M.C., Villet, M.H., Azeredo-Espin, A.M.L., 2012. Molecular phylogenetics of Oestroidea (Diptera: Calyptratae) with emphasis on Calliphoridae: insights into the inter-familial relationships and additional evidence for paraphyly among blowflies. Mol. Phylogenet. Evol. 65, 840–854.

Marinho, M.A.T., Wolff, M., Ramos-Pastrana, Y., de Azeredo-Espin, A.M.L., Amorim, D.d.S., 2017. The first phylogenetic study of Mesembrinellidae (Diptera: Oestroidea) based on molecular data: clades and congruence with morphological characters. Cladistics 33, 134–152.

McAlpine, J.F., 1989. Phylogeny and classification of the Muscomorpha. In: McAlpine, J.F., Wood, D.M. (Eds.), Manual of Nearctic Diptera. Research Branch, Agriculture Canada, Monograph, Ottawa, pp. 1397–1518.

Michelsen, V., 1991. Revision of the aberrant new world genus *Coenosopsia* Diptera Anthomyiidae with a discussion of anthomyiid relationships. Syst. Entomol. 16, 85–104.

Michelsen, V., 1996. First reliable record of a fossil species of Anthomyiidae (Diptera), with comments on the definition of recent and fossil clades in phylogenetic classification. Biol. J. Linn. Soc. 58, 441–451.

Michelsen, V., Pape, T., 2017. Ulurumyiidae – a new family of calyptrate flies (Diptera). Syst. Entomol. 42, 826–836.

Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE), New Orleans, LA, pp. 1–8

Mirarab, S., Warnow, T., 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics 31, 44–52.

Misof, B., Misof, K.A., 2009. Monte Carlo approach successfully identifies randomness of multiple sequence alignments: a more objective approach of data exclusion. Syst. Biol. 58, 21–34.

Misof, B., Meyer, B., von Reumont, B.R.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., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., Frandsen, P.B., Ware, J., Flouri, T., Beutel, R.G., Niehuis, O., Petersen, M., Izquierdo-Carrasco, F., Wappler, T., Rust, J., Aberer, A.J., Aspöck, U., Aspöck, H., Bartel, D., Blanke, A., Berger, S., Böhm, A., Buckley, T.R.Calcott, B., Chen, J., Friedrich, F., Fukui, M., Fujita, M., Greve, C., Grobe, P., Gu, S., Huang, Y., Jermiin, L.S., Kawahara, A.Y., Krogmann, L., Kubiak, M., Lanfear, R., Letsch, H., Li, Y., Li, Z., Li, J., Lu, H., Machida, R., Mashimo, Y., Kapli, P., McKenna, D.D., Meng, G., Nakagaki, Y., Navarrete-Heredia, J.L., Ott, M., Ou, Y., Pass, G., Podsiadlowski, L., Pohl, H., von Reumont, B.M., Schütte, K., Sekiya, K., Shimizu, S., Slipinski, A., Stamatakis, A., Song, W., Su, X., Szucsich, N.U., Tan, M., Tan, X., Tang, M., Tang, J., Timelthaler, G., Tomizuka, S., Trautwein, M., Tong, X., Uchifune, T., Walzl, M.G., Wiegmann, B.M., Wilbrandt, J., Wipfler, B., Wong, T.K.F., Wu, Q., Wu, G., Xie, Y., Yang, S., Yang, Q., Yeates, D.K., Yoshizawa, K., Zhang, Q., Zhang, R., Zhang, W., Zhang, Y., Zhao, J., Zhou, C., Zhou, L., Ziesmann, T., Zou, S., Li, Y., Xu, X., Zhang, Y., Yang, H., Wang, J., Wang, J., Kjer, K.M., Zhou, X., 2014. Phylogenomics resolves the timing and pattern of insect evolution. Science 346, 763–767.

Moulton, J.K., Wiegmann, B.M., 2007. The phylogenetic relationships of flies in the superfamily Empidoidea (Insecta : Diptera). Mol. Phylogenet. Evol. 43, 701–713.

Narechania, A., Baker, R.H., Sit, R., Kolokotronis, S.-O., DeSalle, R., Planet, P.J., 2012. Random addition concatenation analysis: a novel approach to the exploration of phylogenomic signal reveals strong agreement between core and shell genomic partitions in the cyanobacteria. Genome Biol. Evol. 4, 30–43.

Nelson, L.A., Lambkin, C.L., Batterham, P., Wallman, J.F., Dowton, M., Whiting, M.F., Yeates, D.K., Cameron, S.L., 2012. Beyond barcoding: a mitochondrial genomics approach to molecular phylogenetics and diagnostics of blowflies (Diptera: Calliphoridae). Gene 511, 131–142.

Nichols, D.J., Johnson, K.R., 2008. Plants and the K–T Boundary. Cambridge University Press, Cambridge, UK.

Nihei, S.S., Carvalho, C.J.B.d., 2004. Taxonomy, cladistics and biogeography of *Coenosopsia* Malloch (Diptera, Anthomyiidae) and its significance to the evolution of anthomyiids in the Neotropics. Syst. Entomol. 29, 260–275.

Nihei, S.S., Carvalho, C.J.B.d., 2007. Phylogeny and classification of Muscini (Diptera, Muscidae). Zool. J. Linn. Soc. 149, 493–532.

Nirmala, X., Hypsa, V., Zurovec, M., 2001. Molecular phylogeny of calyptratae (Diptera: Brachycera): the evolution of 18S and 16S ribosomal rDNAs in higher dipterans and their use in phylogenetic inference. Insect Mol. Biol. 10, 475–485.

Norris, K. R., 1999. Establishment of a subfamily Aphyssurinae for the Australian genus *Aphyssura* Hardy (Diptera : Calliphoridae), with a review of known forms and descriptions of new species. Invertebr. Taxon. 13, 511–628.

O'Hara, J.E., 1985. Oviposition strategies of the Tachinidae, a family of beneficial parasitic flies. Agriculture and Foresty Bulletin, University of Alberta, 8, 31–34.

O'Hara, J.E., Raper, C., Pont, A.C., Whitmore, D., 2013. Reassessment of *Paleotachina* Townsend and *Electrotachina* Townsend and their removal from the Tachinidae (Diptera). ZooKeys 361, 27–36.

Pape, T., 1986. A phylogenetic analysis of the woodlouse flies (Diptera, Rhinophoridae). Tijdschr. Entomol. 129, 15–34.

Pape, T., 1992. Phylogeny of the Tachinidae Family-Group Diptera Calyptratae. Tijdschr. Entomol. 135, 43–86.

Pape, T., 1996. Catalogue of the Sarcophagidae of the world (Insecta: Diptera). Mem. Entomol. Int. 8, 1–558.

Pape, T., 1998. Rhinophoridae. In: Papp, L., Darvas, B. (Eds.), Contributions to a Manual of Palaearctic Diptera. Science Herald, Budapest, pp. 679–689.

Pape, T., 2001. Phylogeny of Oestridae (Insecta: Diptera). Syst. Entomol. 26, 133–171.

Pape, T., 2006. Phylogeny and evolution of the bot flies. In: Colwell, D., Scholl, P., Hall, M. (Eds.), The Oestrid Flies: Biology, Host-Parasite Relationships, Impact and Management. CABI Publishers, Wallingford, UK, pp. 20–50.

Pape, T., Arnaud, P.H., 2001. *Bezzimyia* – a genus of native New World Rhinophoridae (Insecta, Diptera). Zool. Scr. 30, 257–297.

Pape, T., Blagoderov, V., Mostovski, M.B., 2011. Order Diptera Linnaeus, 1758. In: Zhang, Z.-Q. (Ed.), Animal Biodiversity: An Outline of Higher-Level Classification and Survey of Taxonomic Richness. Zootaxa 3148, 222–229.

Pattengale, N.D., Alipour, M., Bininda-Emonds, O.R., Moret, B.M., Stamatakis, A., 2010. How many bootstrap replicates are necessary? J. Comput. Biol. 17, 337–354.

Pauli, T., Burt, T.O., Meusemann, K., Bayless, K., Donath, A., Podsiadlowski, L., Mayer, C., Kozlov, A., Vasilikopoulos, A., Liu, S., Zhou, X.I.N., Yeates, D., Misof, B., Peters, R.S., Mengual, X., 2018. New data, same story: phylogenomics does not support Syrphoidea (Diptera: Syrphidae, Pipunculidae). Syst. Entomol. 43, 447–459.
Peters, R.S., Krogmann, L., Mayer, C., Donath, A., Gunkel, S., Meusemann, K., Kozlov, A., Podsiadlowski, L., Petersen, M., Lanfear, R., Diez, P.A., Heraty, J., Kjer, K.M., Klopfstein, S., Meier, R., Polidori, C., Schmitt, T., Liu, S., Zhou, X., Wappler, T., Rust, J., Misof, B., Niehuis, O., 2017. Evolutionary history of the Hymenoptera. Curr. Biol. 27, 1013–1018.
Petersen, F.T., Meier, R., Kutty, S.N., Wiegmann, B.M., 2007. The phylogeny and evolution of host choice in the Hippoboscoidea (Diptera) as reconstructed using four molecular markers. Mol. Phylogenet. Evol. 45, 111–122.
Petersen, M., Meusemann, K., Donath, A., Dowling, D., Liu, S., Peters, R.S., Podsiadlowski, L., Vasilikopoulos, A., Zhou, X., Misof, B., Niehuis, O., 2017. Orthograph: a versatile tool for mapping coding nucleotide sequences to clusters of orthologous genes. BMC Bioinformatics 18, 111.
Piwczyński, M., Pape, T., Deja-Sikora, E., Sikora, M., Akbarzadeh, K., Szpila, K., 2017. Molecular phylogeny of Miltogramminae (Diptera: Sarcophagidae): implications for classification, systematics and evolution of larval feeding strategies. Mol. Phylogenet. Evol. 116, 49–60.
Rognes, K., 1991. Blowflies (Diptera, Calliphoridae) of Fennoscandia and Denmark. Fauna Entomol. Scand. 24, 1–272.
Rognes, K., 1997. The Calliphoridae (blowflies) (Diptera: Oestroidea) are not a monophyletic group. Cladistics 13, 27–66.
Savage, J., Wheeler, T.A., 2004. Phylogeny of the Azeliini (Diptera: Muscidae). Stud. Dipterol. 11, 259–299.
Savage, J., Wheeler, T.A., Wiegmann, B.M., 2004. Phylogenetic analysis of the genus *Thricops* Rondani (Diptera : Muscidae) based on molecular and morphological characters. Syst. Entomol. 29, 395–414.
Sayyari, E., Mirarab, S., 2016. Fast coalescent-based computation of local branch support from quartet frequencies. Mol. Biol. Evol. 33, 1654–1668.
Schuehli, G.S., Carvalho, C.J.B.d, Wiegmann, B.M., 2007. Molecular phylogenetics of the Muscidae (Diptera : Calyptratae): new ideas in a congruence context. Invertebr. Syst. 21, 263–278.
Singh, B., Wells, J.D., 2013. Molecular systematics of the Calliphoridae (Diptera: Oestroidea): evidence from one mitochondrial and three nuclear genes. J. Med. Entomol. 50, 15–23.
Song, Z.K., Wang, X.Z., Liang, G.Q., 2008. Molecular evolution and phylogenetic utility of the internal transcribed Spacer 2 (ITS2) in Calyptratae (Diptera: Brachycera). J. Mol. Evol. 67, 448–464.
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.
Stevens, J.R., 2003. The evolution of myiasis in blowflies (Calliphoridae). Int. J. Parasitol. 33, 1105–1113.
Stireman, J.O., 2002. Phylogenetic relationships of tachinid flies in subfamily Exoristinae (Tachinidae: Diptera) based on 28S rDNA and elongation factor-1 alpha. Syst. Entomol. 27, 409–435.
Stireman, J.O. III, O'Hara, J.E., Wood, D.M., 2006. Tachinidae: evolution, behavior, and ecology. Annu. Rev. Entomol. 51, 525–555.
Strimmer, K., von Haeseler, A., 1997. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. Proc. Natl Acad. Sci. USA 94, 6815–6819.
Su, K.F.Y., Kutty, S.N., Meier, R., 2008. Morphology versus molecules: the phylogenetic relationships of Sepsidae (Diptera: Cyclorrhapha) based on morphology and DNA sequence data from ten genes. Cladistics 24, 902–916.
Susko, E., Roger, A.J., 2007. On reduced amino acid alphabets for phylogenetic inference. Mol. Biol. Evol. 24, 2139–2150.
Suyama, M., Torrents, D., Bork, P., 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 34, W609–W612.
Swofford, D.L. 2000. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer Associates, Sunderland, MA.
Tachi, T., Shima, H., 2010. Molecular phylogeny of the subfamily Exoristinae (Diptera, Tachinidae), with discussions on the evolutionary history of female oviposition strategy. Syst. Entomol. 35, 148–163.
Tanabe, A.S., 2008. Phylogears version 2, software distributed by the author at https://www.fifthdimension.jp/.
Tóthová, A., Rozkošný, R., Knutson, L., Kutty, S.N., Wiegmann, B.M., Meier, R., 2013. A phylogenetic analysis of Sciomyzidae (Diptera) and some related genera. Cladistics 29, 404–415.
von Tschirnhaus, M., Hoffeins, C., 2009. Fossil flies in Baltic amber – insights in the diversity of Tertiary Acalyptratae (Diptera, Schizophora), with new morphological characters and a key based on 1,000 collected inclusions. Denisia 26, 171–212.
Tschorsnig, H.-P., 1985. Taxonomie forstlich wichtiger Parasiten: Untersuchungen zur Struktur des männlichen Postabdomens der Raupenfliegen (Diptera, Tachinidae). Stuttgarter Beiträge zur Naturkunde, Serie A (Biologie) 383, 1–137.
Tschorsnig, H.-P., Richter, V.A., 1998. Family Tachinidae. In: Papp, L., Darvas, B. (Eds.), Contributions to a Manual of Palaearctic Diptera (With Special Reference to Flies of Economic Importance), Higher Brachycera. Science Herald, Budapest, pp. 691–827.
Verves, Y.G., 1989. The phylogenetic systematics of the miltogrammatine flies (Diptera, Sarcophagidae) of the world. Jpn. J. Med. Sci. Biol. 42, 111–126.
Vockeroth, J.R., 1989. Scathophagidae. In: McAlpine, J.F., Wood, D.M. (Eds.), Manual of Nearctic Diptera, Vol. 2. Agriculture Canada Research Branch, Ottawa, pp. 1085–1097.
Vossbrinck, C.R., Friedman, S., 1989. A 28S ribosomal-RNA phylogeny of certain cyclorrhaphous Diptera based upon a hypervariable region. Syst. Entomol. 14, 417–431.
Waterhouse, R.M., Zdobnov, E.M., Tegenfeldt, F., Li, J., Kriventseva, E.V., 2011. OrthoDB: the hierarchical catalog of eukaryotic orthologs in 2011. Nucleic Acids Res. 39, D283–D288.
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. Proc. Natl Acad. Sci. USA 108, 5690–5695.
Winkler, I.S., Blaschke, J.D., Davis, D.J., Stireman, J.O. III, O'Hara, J.E., Cerretti, P., Moulton, J.K., 2015. Explosive radiation or uninformative genes? Origin and early diversification of tachinid flies (Diptera: Tachinidae) Mol. Phylogenet. Evol. 88, 38–54.
Winterton, S.L., Hardy, N.B., Wiegmann, B.M., 2010. On wings of lace: phylogeny and Bayesian divergence time estimates of Neuropterida (Insecta) based on morphological and molecular data. Syst. Entomol. 35, 349–378.
Wood, D.M., 1986. Are Cuterebridae, Gasterophilidae, Hypodermatidae and Oestridae a monophyletic group? In: Darvas, B., Papp, L. (Eds.), Abstracts of the First International Congress of Dipterology (Budapest, 17th–24th August, 1986). Hungarian Academy of Sciences, Budapest, pp. 261.
Wood, D.M., 1987. Tachinidae. In: McAlpine, J. F. (Ed.), Manual of Nearctic Diptera. Research Branch, Agriculture Canada, Ottawa, pp. 1193–1269.
Xia, X., 2017. DAMBE6: new tools for microbial genomics, phylogenetics, and molecular evolution. J. Hered. 108, 431–437.
Xie, Y., Wu, G., Tang, J., Luo, R., Patterson, J., Liu, S., Huang, W., He, G., Gu, S., Li, S., Zhou, X., Lam, T. W., Li, Y., Xu, X., Wong, G.K., Wang, J., 2014. SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. Bioinformatics 30, 1660–1666.
Yeates, D. K., Meusemann, K., Trautwein, M., Wiegmann, B., Zwick, A., 2016. Power, resolution and bias: recent advances in

insect phylogeny driven by the genomic revolution. Curr. Opin. Insect. Sci. 13, 16–23.
Young, A.D., Lemmon, A.R., Skevington, J.H., Mengual, X., Ståhls, G., Reemer, M., Jordaens, K., Kelso, S., Lemmon, E.M., Hauser, M., De Meyer, M., Misof, B., Wiegmann, B.M., 2016. Anchored enrichment dataset for true flies (order Diptera) reveals insights into the phylogeny of flower flies (family Syrphidae). BMC Evol. Biol. 16, 143.
Zhang, D., Yan, L., Zhang, M., Chu, H., Cao, J., Li, K., Hu, D., Pape, T., 2016. Phylogenetic inference of calyptrates, with the first mitogenomes for Gasterophilinae (Diptera: Oestridae) and Paramacronychiinae (Diptera: Sarcophagidae). Int. J. Biol. Sci. 12, 489–504.
Zhao, L., Ang, S.H.A., Srivathsan, A., Su, K.F.Y., Meier, R., 2013. Does better taxon sampling help? A new phylogenetic hypothesis for Sepsidae (Diptera: Cyclorrhapha) based on 50 new taxa and the same old mitochondrial and nuclear markers. Mol. Phylogen. Evol. 69, 153–164.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. Plots of third codon positions from the nucleotide dataset from DAMBE (Xia, 2017) showing accumulation of transitions (s) and transversions (v) over GTR distances indicating saturation of this position.

Fig. S2. Heat map showing species-pairwise amino acid site coverage for the dataset from AliStat.

Fig. S3. Comparison of amino acid (left) and nucleotide (right) MPT topologies.

Fig. S4. Species tree estimated in ASTRAL v.5.5.6 (Mirarab and Warnow, 2015) from ML-based amino acid gene trees.

Fig. S5. Heat map showing pairwise Bowker's tests on the (I) aligned amino acid matrix and (II) aligned Dayhoff-6 recoded amino acid matrix ("aminoacid_recoded").

Fig. S6. Four-cluster likelihood mapping and permutation tests on the sister-group relationship between Oestridae and *Mystacinobia zelandica* represented here as 2D simplex graphs.

Fig. S7. Relationships and node support inferred from the "aminoacid_recoded" dataset.

Fig. S8. Most parsimonious tree (MPT) inferred from the amino acid dataset with Bremer branch support values above nodes.

Table S1. NCBI accession numbers for newly generated data

Table S2. Collection information for species in the study

Table S3. List of species obtained from the NCBI database

Table S4. Information on sequences removed during various contamination filtering steps of 1KITE data.

Table S5. Reference data for the orthologue set

Table S6. Models tested in PartitionFinder 2.1.1

Table S7. Four cluster likelihood mapping and permutations test

Table S8. Comparison of Jackknife support, RADICAL values and Bremer branch support at nodes on the MPT.