Systematic Entomology

Systematic Entomology (2021), **46**, 508–525 DOI: 10.1111/syen.12473

Towards a new classification of Muscidae (Diptera): a comparison of hypotheses based on multiple molecular phylogenetic approaches

ANDRZEJ GRZYWACZ[1,2], PAULINA TRZECIAK[1], BRIAN M. WIEGMANN[3], BRIAN K. CASSEL[3], THOMAS PAPE[4], KINGA WALCZAK[1,2], CEZARY BYSTROWSKI[5], LEANNE NELSON[6] and MARCIN PIWCZYŃSKI[1]

[1]Department of Ecology and Biogeography, Nicolaus Copernicus University in Toruń, Toruń, Poland, [2]Centre for Modern Interdisciplinary Technologies, Nicolaus Copernicus University in Toruń, [3]Department of Entomology and Plant Pathology, North Carolina State University, Raleigh, NC, U.S.A., [4]Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark, [5]Forest Research Institute, Raszyn, Poland and [6]Biosecurity Sciences Laboratory, Department of Agriculture and Fisheries, Health and Food Science Precinct, Archerfield, QLD, Australia

Abstract. Muscidae are a megadiverse dipteran family that exhibits extraordinary diversity in morphology and life history as both immatures and adults. The classification of Muscidae has been long debated, and most higher-level relationships remain unknown. In this study, we used multilocus Sanger sequencing (mS-seq), anchored hybrid enrichment (AHE) and restriction-site associated DNA sequencing (RAD-seq) approaches to examine relationships within Muscidae. The results from AHE and RAD-seq largely correspond to those obtained from mS-seq in terms of overall topology, yet phylogenomic approaches received much higher nodal support. The results from all molecular approaches contradict the traditional classification based predominantly on adult morphology, but provide an opportunity to re-interpret the morphology of immature stages. Rearrangements in Muscidae classification are proposed as follows: (i) *Mesembrina* Meigen and *Polietes* Rondani are transferred from Muscinae to Azeliinae; (ii) Reinwardtiinae **stat. rev.** is resurrected as a subfamily distinct from Azeliinae; (iii) *Eginia* Robineau-Desvoidy, *Neohelina* Malloch, *Syngamoptera* Schnabl and *Xenotachina* Malloch are transferred to Reinwardtiinae **stat. rev.**

Introduction

Muscidae represent the most species-rich family within the muscoid grade of Calyptratae with some 6000 known species in *ca*. 180 genera (Pape *et al.*, 2011; Pont, unpublished data). This dipteran family exhibits extraordinary diversity in morphology (Fig. 1), life history, and behaviour at both immature and adult stages (Skidmore, 1985). Some muscids are of great agricultural, medical or veterinary significance and their morphology and natural history have been subjects of detailed examination (Pohl *et al.*, 2012; Tiusanen *et al.*, 2016; Grzywacz *et al.*, 2017a). At the same time, considerable work has been conducted on muscid taxonomy and their higher-level classification has changed over time (see Kutty *et al.*, 2014). Early classifications differed significantly from the current circumscription of Muscidae as a family, and apart from including Fanniidae, Anthomyiidae and Scathophagidae (Karl, 1928; van Emden, 1951), even taxa currently recognized as Hippoboscoidea and Oestroidea were occasionally included (Séguy, 1937). Muscid higher-level classification improved with the adoption of rigorous morphological data analyses applying explicit phylogenetic reasoning (Hennig, 1965; Carvalho, 1989; Couri & Pont, 2000; Couri & Carvalho, 2003; Savage & Wheeler, 2004; Nihei & Carvalho, 2007). The majority of studies using this approach were based exclusively on data from adult morphology and characters drawn from immature stages were used comprehensively in only one

Correspondence: Andrzej Grzywacz, Department of Ecology and Biogeography, Nicolaus Copernicus University, 87-100 Toruń, Poland. E-mail: hydrotaea@gmail.com

Fig. 1. Representative taxa of Muscidae: (A) *Atherigona varia*, male; (B) *Azelia cilipes*, male; (C) *Thricops simplex*, male; (D) *Hydrotaea dentipes*, male; (E) *Hydrotaea ignava*, male; (F) *Australophyra rostrata*, male; (G) *Polietes lardarius*, female; (H) *Mesembrina meridiana*, female; (I) *Musca domestica*, male; (J) *Pyrellia rapax*, male; (K) *Haematobosca stimulans*, male; (L) *Passeromyia steinii*, female; (M) *Synthesiomyia nudiseta*, female; (N) *Eginia ocypterata*, male; (O) *Dichaetomyia varia*, female; (P) *Helina reversio*, male; (R) *Phaonia valida*, female; (S) *Graphomya maculata*, male; (T) *Gymnodia humilis*, male; (U) *Lispe tentaculata*, male. Scale bar 2 mm. [Colour figure can be viewed at wileyonlinelibrary.com].

case, but without a thorough formal phylogenetic analysis (Skidmore, 1985). Furthermore, Skidmore (1985) refrained from decisive classificatory changes when morphological data from immatures and adults were in conflict (e.g. Achanthipterinae).

The modern concept of Muscidae (Roback, 1951; Pont, 1986; Michelsen, 1991) has been corroborated by multilocus Sanger sequencing (mS-seq) (Kutty *et al.*, 2008, 2014; Haseyama *et al.*, 2015) as well as transcriptomic data (Kutty *et al.*, 2019), which have also provided major breakthroughs in reconstructing relationships within the family. Several studies have split Muscidae into a basal dichotomy: Muscinae + Azeliini versus all other Muscidae (Schuehli *et al.*, 2007; Kutty *et al.*, 2008, 2010, 2014; Haseyama *et al.*, 2015). However, the use of mS-seq was unable to robustly resolve the deeper splits within these two groups, and some highly supported clades were at odds with classifications based on adult morphology (Kutty *et al.*, 2014; Haseyama *et al.*, 2015; Grzywacz *et al.*, 2017b). For example, *Azelia* Robineau-Desvoidy emerged as the sister group to a clade containing the remaining Azeliini + Muscinae (Grzywacz *et al.*, 2017b) instead of being closely related to *Thricops* Rondani as predicted based on adult morphology (Savage & Wheeler, 2004). The traditionally recognized subfamilies Coenosiinae, Cyrtoneurininae, Muscinae, Mydaeinae and Phaoniinae were either nonmonophyletic (Kutty *et al.*, 2014; Haseyama *et al.*, 2015) or relationships were strongly affected by different analytical approaches (fig. 3 in Kutty *et al.*, 2014; fig. 1 in Haseyama *et al.*, 2015); *Eginia* Robineau-Desvoidy unexpectedly emerged as the sister taxon of *Passeromyia* Rodhain & Villeneuve within Reinwardtiini (Kutty *et al.*, 2014; Haseyama *et al.*, 2015). Alternative higher-level classifications emerged from these analyses (Haseyama *et al.*, 2015), however these hypotheses lacked robustness due to topological incongruence and low statistical support. Furthermore, recent studies on Schizophora (Junqueira *et al.*, 2016) and Calyptratae (Zhang *et al.*, 2016; Kutty *et al.*, 2019) incorporating mitochondrial genomes and transcriptomic data, were unable to provide new insights into the systematics of Muscidae due to limited taxon sampling within the family.

The application of mS-seq prompted numerous ground-breaking systematic discoveries (Hedges, 1994; Springer *et al.*, 1997; APG, 1998; Kutty *et al.*, 2010; Mutanen *et al.*, 2010; Wiegmann *et al.*, 2011; Piwczyński *et al.*, 2017). However, limited phylogenetic signal when using single or several markers often resulted in poor topological resolution, low statistical support and varying results based on the method employed. Next-generation sequencing (NGS) has opened up new possibilities in phylogenetics by allowing sampling of multiple loci scattered throughout the genome in nonmodel organisms (Lemmon *et al.*, 2012; McCormack *et al.*, 2013; Kjer *et al.*, 2016). However, adding more data may not always increase phylogenetic resolution. Moreover, higher statistical support obtained for nodes does not always provide greater accuracy, as it may be susceptible to systematic errors due to misspecification of model parameters, or incongruences due to the stochasticity of the lineage sorting. One way to overcome these problems is to generate different genomic datasets and employ various phylogenetic approaches. Congruence among these results can be used for building confidence in certain phylogenetic hypotheses (Kjer *et al.*, 2016).

The primary objective of this study was to incorporate mS-seq, anchored hybrid enrichment (AHE) and restriction site associated DNA sequencing (RAD-seq) approaches to examine relationships within the dipteran family Muscidae. To achieve this, we explored the influence of bioinformatic processing of RAD-seq data on the obtained phylogenetic hypotheses, and examine congruence among relationships inferred from these molecular approaches. We also incorporated morphological data from immature stages to examine morphological support for nodes on phylogenetic trees that are in conflict with traditional concepts. Based on our results, we propose changes to higher-level classification within Muscidae.

Materials and methods

Taxon sampling and DNA isolation

We sampled 33 species of Muscidae from 27 genera representing six of seven subfamilies (i.e. all except Cyrtoneurininae) (Table S1). Adult flies were identified by the first author using keys in Hennig (1955–1964), Pont (1969, 1973) and Gregor *et al.* (2002). Voucher specimens, when available, were deposited at the Department of Ecology and Biogeography, Nicolaus Copernicus University (Table S1). Species identifications were also verified by comparing cytochrome oxidase subunit I (COI) barcode sequences, obtained according to Grzywacz *et al.* (2017b), against sequences in the NCBI database (National Center for Biotechnology Information, MD, U.S.A.) using the Basic Local Alignment Search Tool (BLAST).

Prior to DNA extraction, adult flies stored in ethanol were rinsed for a few minutes in distilled water and air-dried on tissue paper. Entire specimens, *Atherigona varia* (Meigen) and *Azelia* spp., or thoracic muscle samples (other taxa) were subsequently homogenized in 1.5 mL Eppendorf tubes and dried on a thermoblock at 65°C for several minutes. Total genomic DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA, U.S.A.) following the manufacturer's instructions. Isolated DNA was quantified with a Qubit 3.0 fluorometer using dsDNA High Sensitivity Assay Kit (Life Technologies, Inc., Carlsbad, CA, U.S.A.) following the manufacturer's instructions. Samples with low DNA yield were subjected to whole genome amplification with REPLI-g Mini Kit (Qiagen) to increase DNA concentration.

mS-seq data matrix preparation

Mitochondrial COI, cytB (cytochrome b) and nuclear Ef1α (elongation factor 1-alpha) markers were obtained *de novo* and combined with nuclear CAD (carbamoyl-phosphate synthetase 2) markers retrieved from GenBank (Table S1). The details of the PCR amplifications and sequencing are provided in Grzywacz *et al.* (2017b); the final products of sequencing were resolved using an automated DNA sequencer at the Laboratory of Molecular Biology Techniques, UAM (Poznań, Poland). Both forward and reverse strands were edited and then assembled using *SeqMan II* ver. 4.0 (DNASTAR, Lasergene, Madison, WI, U.S.A.). All sequences were deposited in GenBank (Table S1).

AHE library preparation and data assembly

For each sample 7.9–110 ng/μL (47 ng/μL mean) DNA in 50 μL total volume was sheared to approximately 300 bp by sonication with a Covaris E220 Focused-ultrasonicator using Covaris microTUBES (Covaris, Inc., Woburn, MA, U.S.A.). The sheared DNA was used as input for genomic DNA library preparation and indexing using the protocol of Meyer & Kircher (2010), but modified to include a size-selection step after blunt-end repair using SPRIselect beads (Beckman Coulter, Inc., Brea, CA, U.S.A.; 0.9 × ratio of bead to sample volume). Each sample was indexed and pooled together in groups of 48 samples. We enriched each 48-sample pool using the 57 681 tiled, custom-designed probes contained in the Diptera AHE kit (Young *et al.*, 2016), an Agilent Custom SureSelect Kit (Agilent Technologies, Santa Clara, CA, U.S.A.) that targets 559 unique loci. The Diptera probe kit design is detailed in Young *et al.* (2016) and is based on comparison and selection of conserved 150 bp gene regions found among 7 diverse fly genomes and 14 transcriptomes. We sequenced the pooled libraries using two lanes of an Illumina HiSeq 2500 (Illumina, San Diego, CA, U.S.A.) run (single read, 100 bp). All AHE laboratory procedures and sequencing were conducted in laboratory facilities of the North Carolina State University (NCSU), Department of Entomology & Plant Pathology (Wiegmann Lab).

AHE data were assembled, processed and analysed using methods described in Buenaventura *et al.* (2020). We demultiplexed raw reads using *cassava* 1.8.2 at the NCSU Genomic Sciences Laboratory and trimmed of adapters and low-quality sequences using *Trimmomatic* v.0.36 (Bolger *et al.*, 2014). This included a locus-by-locus cleaning step to remove non-Diptera sequences and low quality reads based on E-values reported by BLAST search against the NCBI database. We used *trinity* v.2.4 (Grabherr *et al.*, 2011) to assemble the clean reads. We also rechecked files containing very low numbers of reads for contamination against a custom database of microbial sequences, and NCBI databases. Single-copy orthologs were confirmed using the program *Orthograph* v.0.5.14 (Petersen *et al.*, 2017), which uses Hidden Markov Model-based fit to assign orthology to known gene models. We used the reciprocal blast hit criterion in *Orthograph* and all other default settings to assign loci using the 'Mecopterida' ortholog set of 3145 single-copy nuclear gene models identified in comparison of two Lepidoptera and three Diptera genomes (Pauli *et al.*, 2018; available for download on Mendeley.com). Multiple sequence alignments (MSAs) were carried out using *MAFFT* v.7.273 with the *L-INS-I* algorithm (Katoh & Standley, 2013) on FASTA files of amino acid sequences from each orthologous gene set. We followed the procedure of Misof *et al.* (2014) to assess alignment quality by using the *addfragments* algorithm in *MAFFT* to refine alignments through identification of outlier sequences and removing outliers from both amino acid MSAs and nucleotide sequences (Evangelista *et al.*, 2019). Orthologous sequences from reference species were removed from all MSAs, as well as all empty or X-only data columns in each alignment. We used a modified version of *Pal2nal* v.14 (Suyama *et al.*, 2006; Misof *et al.*, 2014) to generate the corresponding nucleotide MSAs with the amino acid MSAs as a guide. We checked the amino acid MSAs of each orthologous gene set for ambiguously aligned regions with the software *Aliscore* v.1.2 (Misof & Misof, 2009; Kück *et al.*, 2010; Misof *et al.*, 2014) and subsequently removed from the nucleotide MSAs at the amino acid level. We used the -e option for gap-prone datasets and forced a comparison of all sequence pairs. These refined nucleotide sequences were realigned using the amino acid alignment as a guide in *Pal2Nal* as modified by Misof *et al.* (2014). Alignments were concatenated to construct phylogenetic datasets in *FasConCat-G* v1.0 (Kück & Longo, 2014).

RAD-seq library preparation and data processing

Genomic DNA for each species was individually barcoded and processed as a reduced complexity library based on the traditional RAD-seq protocol described by Ali *et al.* (2016) with the following modifications: (i) for each species, four samples of 50–75 ng of DNA each were separately digested using SbfI-HF restriction enzyme to avoid unexpected reaction failure; (ii) P1 adapter-ligated fragments were sheared for 26 s to a peak target of 700 bp using an ultrasonicator Covaris M220 (Covaris, Inc.); (iii) Pippin Prep (Sage Science, Beverly, MA, U.S.A.) was used to select fragments between 550 and 700 bp with prior library cleaning with AMPure XP (Beckman Coulter, Carlsbad, CA, U.S.A.; 1 × ratio of bead to sample volume); (iv) three independent PCRs (15 cycles) were carried out and subsequently pooled; (v) pooled PCR products were purified twice with AMPure XP (1 × ratio of bead to sample volume) to completely remove the remaining primers. The final library quantification was carried out using a Qubit 3.0 fluorometer and 2100 Bioanalyzer with the High Sensitivity DNA Analysis Kit (Agilent Technologies). Before sequencing, we added 10% of PhiX DNA (i.e. a library of fragments from a well-characterized GC neutral genome; PhiX Control v3, Illumina) to the denatured library to mitigate potential problems with an unbalanced and low diversity library. The library was sequenced at the Molecular Ecology Lab, Institute of Environmental Sciences, Jagiellonian University (Poland) on the MiSeq platform using MiSeq Reagent Kit v3 (600-cycle) (Illumina). Four libraries were created using this preparation scheme. The first two included the same set of 11 pooled species (Table S1) because the number of reads from the first attempt was too low. The third and fourth comprised 8 and 14 species, respectively (Table S1).

We processed the raw RAD-seq data (Illumina FASTQ output files) for phylogenetic analysis using the *ipyrad* v.0.7.28 pipeline (Eaton, 2014). Before running the analyses, we trimmed each read to 200 bp after inspecting raw data using the *fastx_trimmer* script from *FASTX-Toolkit* (http://hannonlab.cshl.edu/fastx_ toolkit/index.html; accessed 25 May 2018), because the quality of sequences dropped rapidly above this value. We excluded R2 reads, because the *ipyrad* pipeline does not assemble paired end data for standard RAD-seq. Moreover, R2 reads had very low quality terminal part, especially in one of the libraries. According to *ipyrad* developers [see discussion on Gitter 25 September 2020 (https://gitter.im/dereneaton/ipyrad)] and from our experience (Piwczyński *et al.*, 2021), the exclusion of R2 reads usually do not influence the final result. Typical RAD-seq dataset generated by HiSeq platforms contains reads that are 50–100 bp long. Increasing the sequence length beyond this value may increase phylogenetic signal for alternative gene tree topologies, deteriorating support for species tree obtained by concatenation method (Rivers *et al.*, 2016). Taking advantage of MiSeq generated reads, we tested this prediction analysing full R1 reads (200 bp) as well as reads trimmed to 150, 100 and 50 bp.

We used *ipyrad* to demultiplex and assign reads to species based on sequence barcodes (allowing one mismatch). Further, we performed *de novo* read assembly using five clustering thresholds: 0.70, 0.75, 0.80, 0.85 and 0.90 for full R1 (200 bp). An increase in clustering threshold reduces the number of recovered loci because more variable loci are split into multiple loci that are not present in enough taxa to pass the coverage filter. The highest number of parsimony informative sites (PIS) in alignments, a proxy of phylogenetic signal strength, was obtained under 0.70 and 0.75 thresholds (Table 1; *AMAS*, Borowiec, 2016). Thus, we applied these two thresholds for the trimmed R1 reads. We performed two additional analyses to check whether read assembly with respect to the *Musca domestica* Linnaeus reference genome would increase the number of recovered loci. We performed: (i) reference read assembly with *M. domestica* 2.0.2 (GCA_000371365.1) genome (Scott *et al.*, 2014) as a reference, and (ii) reference + *de novo* read assembly. Other parameters used for filtering out poor-quality reads and reducing base-calling error were the same for all analyses. We implemented a minimum Phred quality score (= 33), minimum sequence coverage (= 4 ×), minimum read length (= 100 bp, except analyses where reads were trimmed to 50 bp), the maximum number of uncalled bases (= 10), maximum proportion of heterozygous sites per locus (= 0.5) and maximum number of heterozygous individuals per locus (= 8) while ensuring that variable sites had no more than two alleles.

Phylogenetic inference

The mS-seq alignment was analysed by a maximum likelihood (ML) approach as implemented in *RAxML* 8.2.6 (Stamatakis, 2014). The partitioning scheme was selected by *PartitionFinder* 1.1.1 (Lanfear *et al.*, 2012) using the Bayesian Information Criterion (BIC) (Table S2). Subsequently, the GTR + Γ model of nucleotide substitution was applied for each partition. The phylogenetic tree was rooted using Anthomyiidae species: *Delia platura* (Meigen), *Hydrophoria lancifer* (Harris), *Lasiomma latipenne* (Zetterstedt) and *Paregle coerulescens* (Strobl) as outgroups.

AHE data were analysed using the multispecies coalescent and concatenation approaches. The SVDquartets method (Chifman & Kubatko, 2014, 2015) implemented in the most recent version of *PAUP** (Swofford, 2002) was used to analyse datasets under the multispecies coalescent model. We evaluated all possible quartets and treated ambiguities as missing data. We applied the *QFM* algorithm to assemble quartets into a species tree. Confidence in the groupings on the tree was measured using the standard nonparametric bootstrap (BS) procedure with 1000 repetitions. For the concatenation analysis, AHE loci were analysed by ML in *RAxML* (Stamatakis, 2014) and by the Bayesian method (BM) implemented in *ExaBayes* 1.5 (Aberer *et al.*, 2014), with one GTR + Γ model of nucleotide substitution for the whole alignment. In the ML analysis, support for each node was assessed by standard 1000 nonparametric bootstrap replicates, followed by a search for the best scoring ML tree. For the BM, we ran four independent runs of 1.0 million generations each. We visually assessed runs for convergence in *Tracer* v. 1.7 (Rambaut *et al.*, 2018) by checking for effective sample sizes that were >200 for all model parameters. We also calculated the standard deviation of the split frequencies using the *sdsf* programme included in the *ExaBayes* package. We summarized tree samples from each run on a consensus tree using the *consense* programme from the *ExaBayes* package. Phylogenetic trees resulting from the AHE dataset were rooted using species from the family Sarcophagidae: *Sarcophaga* (*Pandelleisca*) *similis* Meade, *Sarcophaga* (*Liopygia*) *crassipalpis* Macquart, *Paramacronychia flavipalpis* (Girschner), *Brachicoma devia* (Fallén), *Miltogramma rohdendorfi* (Tscharykuliev), *Sphecapatodes inornata* Rohdendorf.

All RAD-seq-based alignments were analysed using the concatenation approach using ML and BM. All parameters for both types of analyses were the same as in the case of AHE. Since we did not include any outgroup taxa in RAD-seq analyses to reduce the effect of locus dropout, we used the midpoint rooting as implemented in R package *phytools* (Revell, 2012). To investigate possible influence of midpoint rooting on the obtained topologies we used the midpoint rooting method for rooting mS-seq and AHE trees (without outgroup taxa). Outgroup and midpoint rooting approaches were congruent both for mS-seq and AHE trees and revealed the same basal dichotomy in rooted trees.

To quantify differences between all obtained trees, we calculated Robinson-Foulds topological distance (TD) and Kuhner and Felsenstein branch length scores (BLS) for unrooted phylogenetic trees (without outgroup taxa) using *ape* package (Popescu *et al.*, 2012) in R (R Core Team, 2019).

Results

mS-seq phylogeny

The alignment obtained from four molecular markers consisted of 4145 base pairs (bp) and 1342 PIS (Table 1). Similar to previous studies (Kutty *et al.*, 2014; Haseyama

Table 1. Summary statistics of analysed data and summary of bootstrap support values for phylogenies inferred from multilocus Sanger sequencing (mS-seq), anchored hybrid enrichment (AHE) and restriction site associated DNA sequencing (RAD-seq) alignments with maximum likelihood approach. RAD-seq data (R1 reads with 200, 150, 100 or 50 nucleotides positions) were either *de novo* assembled, mapped to reference genome or clustered using combination of both schemes (reference + *de novo*) under 0.70, 075, 0.80, 0.85 or 0.90 clustering threshold. Topologies are provided in Figure S2.

Analysed data		Alignment (bp)	Loci	Missing data (%)	PIS	PIS proportion (%)	Bootstrap support			Topology
							Mean	Median	CV	
mS-seq		4145	4	15.21	1342	32.37	52.97	60.00	0.52	A
AHE	Concatenated	365 931	489	60.49	71 527	19.55	93.16	100	0.15	B
	Coalescent						85.53	94.00	0.23	D
RAD-seq										
Assembling scheme	Threshold									
de novo										
R1 200 bp	0.70	106 434	538	83.54	7502	7.05	83.87	97.00	0.25	E
	0.75	107 089	538	83.38	7390	6.90	77.40	84.50	0.29	F
	0.80	93 242	462	83.10	4962	5.32	73.00	78.50	0.34	G
	0.85	76 279	369	82.93	2155	2.83	51.97	50.00	0.56	H
	0.90	58 301	274	82.29	645	1.11	32.77	26.50	0.70	I
R1 150 bp	0.70	110 119	746	82.48	8535	7.75	86.80	96.00	0.21	J
	0.75	111 375	752	82.29	8316	7.47	85.73	95.00	0.22	L
R1 100 bp	0.70	88 811	913	81.95	6020	6.78	88.40	98.00	0.19	M
	0.75	88 455	909	81.55	5742	6.49	90.20	97.50	0.17	N
R1 50 bp	0.70	51 945	1139	81.95	2256	4.34	71.63	75.00	0.38	O
	0.75	52 435	1154	81.81	2187	4.17	72.70	69.50	0.36	P
Reference										
R1 200 bp	0.70	14 447	99	80.51	1288	8.92	77.00	79.00	0.28	Q
	0.75	14 297	98	83.21	1295	9.06	74.79	79.00	0.30	R
Reference + *de novo*										
R1 200 bp	0.70	73 664	497	80.34	5268	7.15	52.90	52.00	0.70	S
	0.75	71 823	484	80.62	4747	6.61	63.10	71.50	0.52	T

Abbreviations: PIS, parsimony informative sites; CV, coefficient of variation.

et al., 2015), our mS-seq tree resulted in many poorly to moderately supported phylogenetic relationships (Fig. 2), particularly at tribal and subfamilial levels (Figure S2: topology A). We obtained 38% of nodes with poor BS, lower than 50%, 34% of nodes with moderate BS ranging from 50 to 70%, and 28% with BS higher than 70%. The tree obtained from the ML analysis of the mS-seq alignment is congruent with those of recently published studies (Kutty *et al.*, 2014; Haseyama *et al.*, 2015; Grzywacz *et al.*, 2017b). A poorly supported dichotomy (BS = 36%) split the family into one clade composed of Muscinae and Azeliini (hereafter clade 1) and a second clade containing all remaining Muscidae (hereafter clade 2) (Figure S2: topology A; for taxonomic limits of clade 1 and clade 2 see Fig. 3). Azeliinae were recovered as nonmonophyletic, with Azeliini and Reinwardtiini separated into clade 1 and clade 2, respectively. Azeliini, Muscinae and Muscini were not recovered as monophyletic; *Azelia* emerged as the sister-group of the remaining Azeliini + Muscinae. *Hydrotaea* Robineau-Desvoidy was monophyletic, yet with poor support (BS < 50%). *Huckettomyia watanabei* Pont & Shinonaga and *Thricops simplex* (Wiedemann) clustered within Muscinae as the sister of *Polietes lardarius* (Fabricius) + *Mesembrina meridiana* (Linnaeus) (BS = 63%). Within clade 2, *Atherigona varia* was the sister-taxon of the remaining muscids. Phaoniinae were polyphyletic, with *Eginia ocypterata* (Meigen) nested within Reinwardtiini. The remaining Phaoniinae emerged as the sister-group of Coenosiinae + Mydaeinae (BS = 74%).

AHE phylogeny

We obtained an alignment of 489 loci with 365 931 bp and 71 527 PIS (Table 1). Results obtained from the ML and BM analyses of the AHE concatenated dataset were topologically congruent (Table S3; compare Fig. 3 and Figure S2: topology C). The trees had 75% (24 out of 32) and 31% (10 out of 32) of nodes receiving 100% BS in ML concatenated (Fig. 2) and species tree analyses (Figure S1), respectively. In the BM concatenated analysis, only three nodes received posterior probability (PP) lower than 1.0 (Figure S2: topology C). Azeliinae were recovered as nonmonophyletic, with Azeliini clustered within clade 1 and Reinwardtiini clustered within clade 2 (Fig. 3). Within clade 1, neither Azeliini nor Muscinae were monophyletic. *Azelia* was found to be the sister-group of the remaining Azeliini + Muscinae (BS = 100% and PP = 1). *Hydrotaea* was recovered as monophyletic (BS = 100%, PP = 1), whereas Muscinae were polyphyletic with *P. lardarius* and *M. meridiana* clustered within Azeliini, as the sister-group of *H. watanabei* + *T. simplex* (BS = 100%, PP = 1). Muscini were paraphyletic with regard to Stomoxyini, herein *Haematobosca stimulans* (Meigen) and

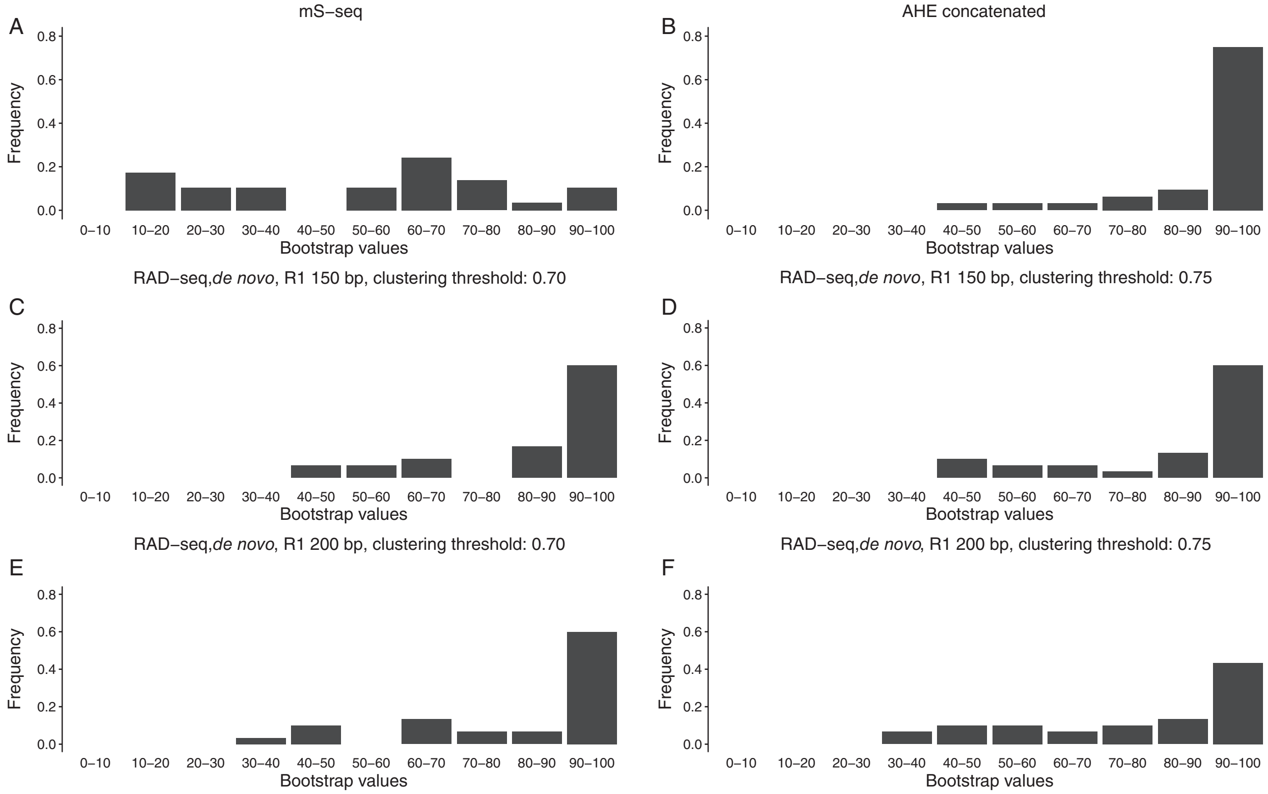

Fig. 2. Distribution of bootstrap values for selected maximum likelihood trees of: (A) mS-seq matrix; (B) concatenated AHE and (C–F) RAD-seq R1 reads with 150 (R1 150) or 200 (R1 200) nucleotides positions analysed according to *de novo* assembly schemes under 0.70 or 0.75 clustering threshold. For distribution of bootstrap values of remaining phylogenetic trees see Supporting Information, Figure S1.

Stomoxys calcitrans (Linnaeus). Within clade 2, a dichotomy of Atherigoninae + (Reinwardtiini + *E. ocypterata*) and the remaining Phaoniinae + (Mydaeinae + Coenosiinae) was recovered (BS = 100%, PP = 0.83). Phaoniinae were polyphyletic with *E. ocypterata* nested within Reinwardtiini and Phaoniini were paraphyletic with regard to Dichaetomyiini. Mydaeinae were paraphyletic with regard to a monophyletic Coenosiinae.

Concatenated and coalescent analyses were largely topologically congruent (Table S3; Fig. 3; Figure S2: topology D). In the concatenated analysis, *A. varia* was recovered within clade 2, whereas in the coalescent analysis recovered in clade 1 as sister to Azeliini + Muscinae, albeit with poor support (BS = 39%). *Musca domestica* was the sister of Stomoxyini using the concatenation method with 100% support, yet the sister of Azeliini + the remaining Muscinae using the coalescent method with 92% support. Incongruence between these approaches was observed for species-level relationships within *Hydrotaea*. *Hydrotaea meteorica* (Linnaeus) was the sister to *H. ignava* (Harris) + *H. velutina* Robineau-Desvoidy using the concatenated method, and the sister of *H. diabolus* (Harris) + *H. dentipes* (Fabricius) using the coalescent method. The major difference within clade 2 was paraphyly of Mydaeinae with regard to Phaoniinae excluding *E. ocypterata* (coalescent approach), and Coenosiinae (concatenated approach).

RAD-seq phylogeny

The alignments obtained from RAD-seq data greatly differed in the number of loci, nucleotide positions, number and proportion of PIS and bootstrap support for nodes depending on the analytical scheme (Table 1, Fig. 2). The use of different analytical schemes had a relatively low impact on the proportion of missing data in all alignments, ranging from 80.34% to 83.54% (Table 1). The number of retrieved loci, alignment length and number and proportion of PIS mostly increased with a decrease of clustering threshold. Also, the length of the assembled sequences affected the amount of data obtained. Although trimming of R1 reads to 100 bp and subsequently to 50 bp sequences led to an increase in the number of assembled loci, short sequences resulted in fewer PIS and did not increase the accuracy of phylogeny estimation (Figure S2: topologies E–T). The highest number of PIS was yielded for reads trimmed to 150 bp and analysed under 0.70 clustering threshold under *de novo* assembly scheme, and was built from 746 loci with 110 119 bp (Table 1). Data assembly with respect to the *M. domestica* reference genome resulted in a significant dropout to less than 100 recovered loci (Table 1). Furthermore, use of the *M. domestica* genome for reference reads assembly did not result in the retrieval of loci for *P. indecora* (Walker).

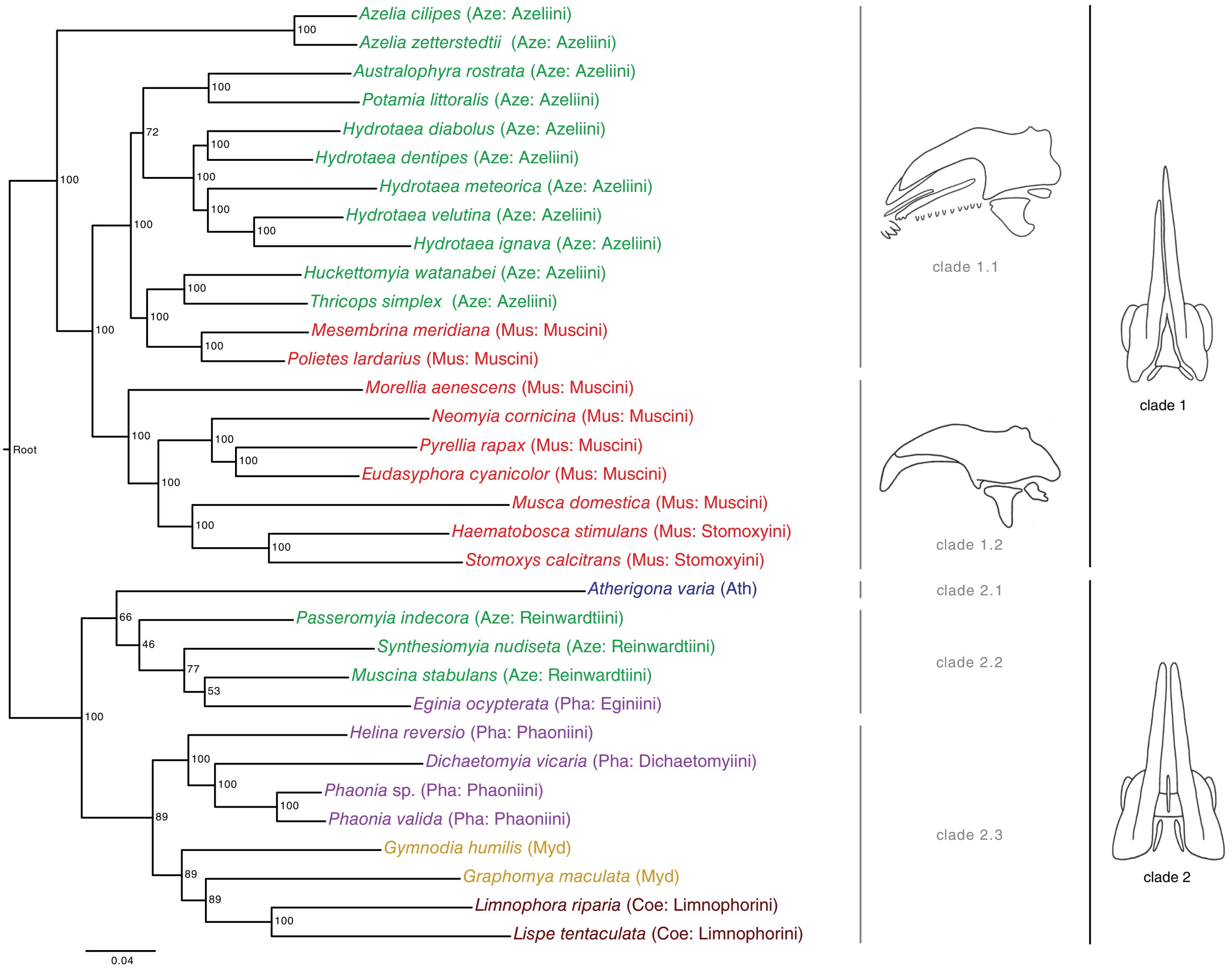

Fig. 3. Maximum likelihood phylogeny of the concatenated anchored hybrid enrichment (AHE) dataset using *RAxML* under the GTR + Γ model. The tree was rooted using midpoint rooting and bootstrap scores from 1000 nonparametric bootstrap replicates are on nodes. Clade 1 includes taxa with asymmetric mouth-hooks whereas clade 2 includes taxa with symmetric mouth-hooks in third instar larvae. Clade 1.1 is characterized by the presence of accessory oral sclerites below the apical part of the mouth-hook in third instar larvae and representatives of clade 1.2 are devoid of such sclerites. Traditional subfamilial classification: Ath, Atherigoninae; Aze, Azeliinae; Coe, Coenosiinae; Mus, Muscinae; Myd, Mydaeinae; Pha, Phaoniinae. [Colour figure can be viewed at wileyonlinelibrary.com].

Alternative datasets obtained from RAD-seq under different analytical schemes resulted in conflicting phylogenetic hypotheses, often with very high nodal support values. A major difference between the alternative topologies was the position of *P. indecora* (Figure S2: topologies E–T), which fell in either clade 1 or 2, often with high support (e.g. Figure S2: topologies E and F). Therefore, we focus on topologies derived from the RAD-seq dataset with the highest PIS, that is, R1 reads trimmed to 150 bp and *de novo* assembled under 0.70 clustering threshold (Table 1). The topologies we obtained from this dataset were fully congruent between ML and BM analyses (compare Figure S2: topologies J and K). Azeliinae were non-monophyletic, with Azeliini and Reinwardtiini separated into clade 1 and clade 2, respectively. *Azelia* was recovered as sister of the remaining Azeliini + Muscinae (BS = 99%, PP = 1). *Hydrotaea* was monophyletic with *H. diabolus* as sister of the remaining *Hydrotaea* (BS = 100%, PP = 1) and *H. ignava* was closely related to *H. velutina* (BS = 100%, PP = 1). Muscinae was polyphyletic with *P. lardarius* and *M. meridiana* clustered within Azeliini as the sister-group of *H. watanabei* + *T. simplex* (BS = 100%, PP = 1). Muscini, excluding *M. meridiana* and *P. lardarius*, were paraphyletic with regard to Stomoxyini (*H. stimulans* + *S. calcitrans*). Within clade 2, a split of Atherigoninae + (Reinwardtiini + *E. ocypterata*) and the remaining Phaoniinae + Mydaeinae + Coenosiinae was observed (BS = 99%, PP = 1). Phaoniinae were polyphyletic, with *E. ocypterata* nested within Reinwardtiini. Coenosiinae were monophyletic (BS = 87%, PP = 1), whereas Mydaeinae were paraphyletic with regard to Coenosiinae.

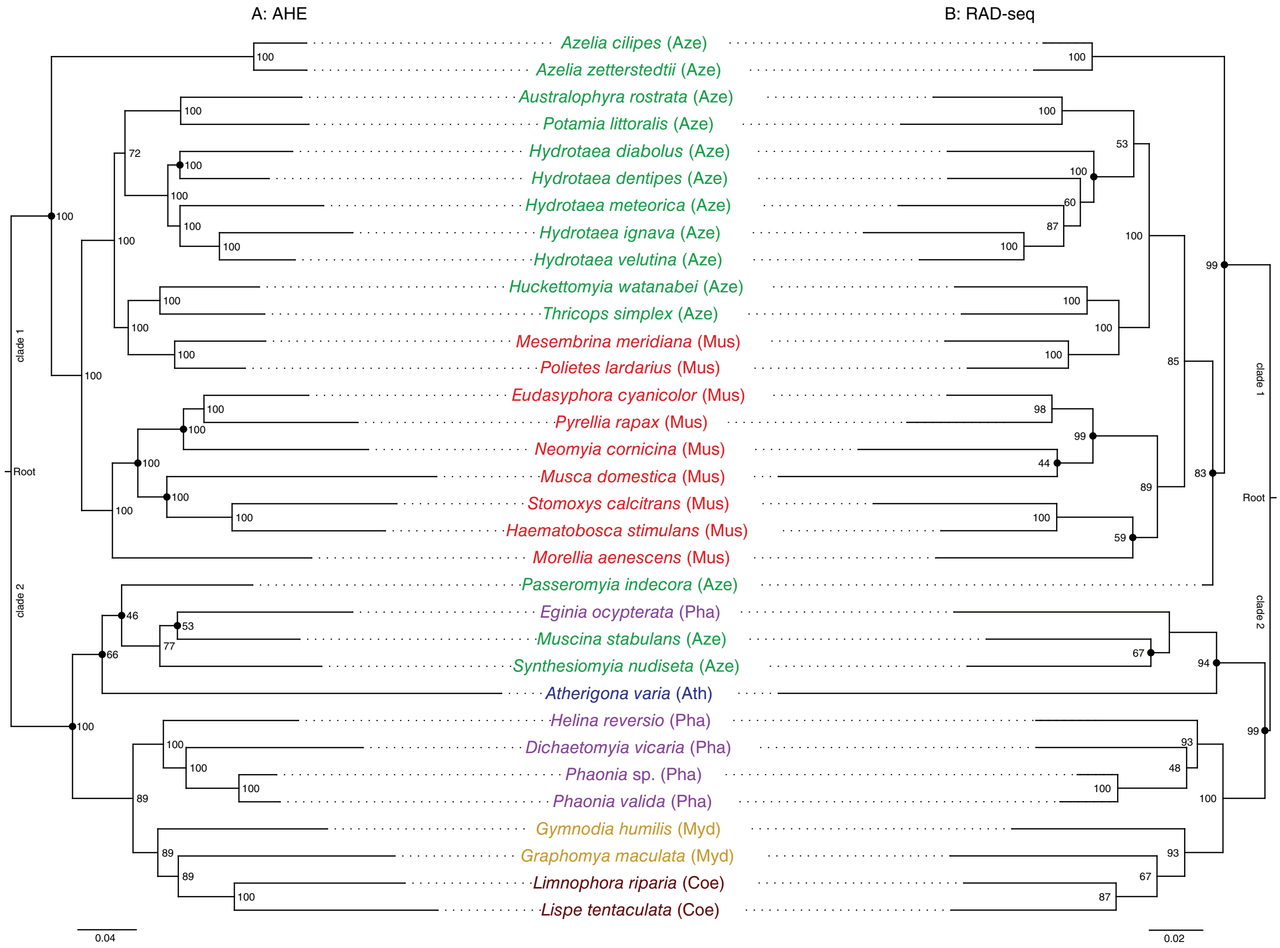

Fig. 4. Comparison of *RAxML* topologies inferred from: (A) concatenated anchored hybrid enrichment (AHE) and (B) restriction site associated DNA sequencing (RAD-seq) data. In RAD-seq, R1 reads trimmed to 150 bp were *de novo* assembled and analysed under 0.70 clustering threshold. Trees were rooted using midpoint rooting approach and basal dichotomy, a split into clade 1 and clade 2, is marked. Bootstrap scores from 1000 nonparametric bootstrap replicates are on nodes. Clades incongruent between compared trees are marked with a filled circle. Traditional subfamilial classification: Ath, Atherigoninae; Aze, Azeliinae; Coe, Coenosiinae; Mus, Muscinae; Myd, Mydaeinae; Pha, Phaoniinae. [Colour figure can be viewed at wileyonlinelibrary.com].

Comparison of AHE with RAD-seq

Phylogenetic analyses of the AHE and RAD-seq data resulted in different tree topologies (Figure S2: topologies B–S) and support values (Fig. 2; Figure S1). Congruence was greatest between the AHE concatenated tree and the RAD-seq tree derived from the highest number of PIS based on TD and BLS values (Fig. 4, Table S3). However, the position of *Passeromyia* differed between analyses and clustered within clades 1 (RAD-seq) and 2 (AHE). Aside from *Passeromyia*, relationships within clade 1 and clade 2 were similar overall. Azeliinae were nonmonophyletic, with Azeliini in clade 1 and Reinwardtiini in clade 2. Azeliini were polyphyletic with regard to Muscinae, with *Azelia* recovered as sister to the remaining Azeliini + Muscinae. Muscinae were polyphyletic with *P. lardarius* and *M. meridiana* clustered within Azeliini as the sister-group of *H. watanabei* + *T. simplex. Hydrotaea* was monophyletic, but the two methods differed in species relationships. A split in *Hydrotaea* in the AHE analysis was not revealed in the RAD-seq analysis. Different sister taxa to Stomoxyini were revealed by the two approaches, that is, *M. domestica* (AHE) and *Morellia aenescens* Robineau-Desvoidy (RAD-seq). Clade 2 presented a split into Atherigoninae + Reinwardtiini (excluding *P. indecora* in RAD-seq) + *E. ocypterata* and the remaining Phaoniinae + Mydaeinae + Coenosiinae.

Discussion

Performance of RAD-seq and AHE for phylogeny estimation

An obstacle for phylogeny reconstruction using NGS is that datasets of different origins return conflicting phylogenetic estimates, each with maximum nodal support (Jarvis *et al.*, 2014;

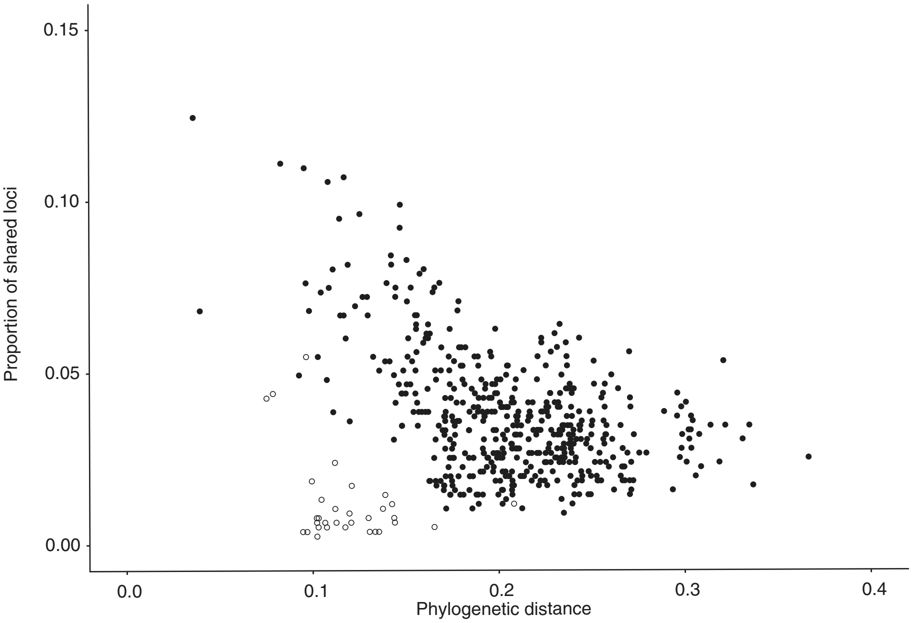

Fig. 5. Reduction in the number of shared loci with increasing phylogenetic distance among lineages observed in RAD-seq R1 reads trimmed to 150 bp and *de novo* assembled under 0.70 clustering threshold. Open circles show data for *Passeromyia indecora* and filled circles for all remaining taxa.

Prum *et al.*, 2015; Gillung *et al.*, 2018). However, congruence among trees inferred from different data sources and analytical approaches may be used for building confidence in certain phylogenetic hypotheses (Kjer *et al.*, 2016). While AHE has already been used to reveal higher-level relationships in Diptera (Young *et al.*, 2016; Gillung *et al.*, 2018; Buenaventura *et al.*, 2020), RAD-seq was used for intrageneric and intraspecific relationships (Rubin *et al.*, 2012; Suchan *et al.*, 2017) and this study is the first application of RAD-seq for reconstruction of deeper divergences within the order. Although the phylogenetic hypotheses obtained were not fully congruent between RAD-seq and AHE, the results provided new insights in our understanding of higher-level relationships within Muscidae.

RAD-seq has been shown to be most effective for lineages that diverged up to approximately 60 Mya (Rubin *et al.*, 2012; Cariou *et al.*, 2013; Leaché *et al.*, 2015; Eaton *et al.*, 2017). The recently estimated divergence time for Muscidae (Wiegmann *et al.*, 2011; Haseyama *et al.*, 2015; Cerretti *et al.*, 2017) falls within this range, and for the majority of splits RAD-seq provided phylogenetic hypotheses congruent with the AHE approach (Fig. 4). However, despite the relatively high number of loci obtained using RAD-seq, nodes with low support in ML and BM analyses (Figure S2: topologies E–T) were still observed. These nodes were also responsible for the incongruence between the RAD-seq and AHE phylogenies (Fig. 4), for example, clades *M. domestica* + *Neomyia cornicina* (Fabricius) (BS = 44%; PP = 0.87), *M. aenescens* + (*H. stimulans* + *S. calcitrans*) (BS = 59%; PP = 0.92) and *Muscina stabulans* (Fallén) + *Synthesiomyia nudiseta* (van der Wulp) (BS = 67%; PP = 0.82). Using RAD-seq, substantial amounts of missing data may affect phylogeny reconstruction. This is manifested in the reduction of the number of shared loci with increasing phylogenetic distance between lineages (Fig. 5). Furthermore, restriction sites shared across lineages with deep divergences most likely occur in slowly evolving loci and for this reason provide weaker phylogenetic signal (Rubin *et al.*, 2012; Huang & Knowles, 2016; Crotti *et al.*, 2019).

We experienced two major obstacles with our RAD-seq approach. The first, as experienced in previous studies, was bioinformatic processing of raw RAD-seq data (Cruaud *et al.*, 2014; Leaché *et al.*, 2015; Eaton *et al.*, 2017). The clustering threshold, that is, the level of sequence similarity at which two sequences are identified as being homologous (Eaton, 2014), strongly affects the number of obtained homologous sequences (Rubin *et al.*, 2012; Cariou *et al.*, 2013; Cruaud *et al.*, 2014; Harvey *et al.*, 2016). Alignments obtained under conservative clustering thresholds (above 0.80) result in reduced success in recognizing orthologous loci and therefore provided a relatively random arrangement of taxa on inferred trees (Figure S2: topologies G–I). Relaxation of the parameter to 0.70 or 0.75 allowed for the reconstruction of many relationships in congruence with mS-seq and AHE (Table S3). The second obstacle was the influence of the length of assembled reads. Read length impacted the amount of recovered data and the subsequent inferred phylogeny. For example, R1 reads trimmed to 100 and 50 bp led to an increase in the number of recovered loci, but simultaneously resulted in weaker phylogenetic signal (number of PIS), which is especially problematic

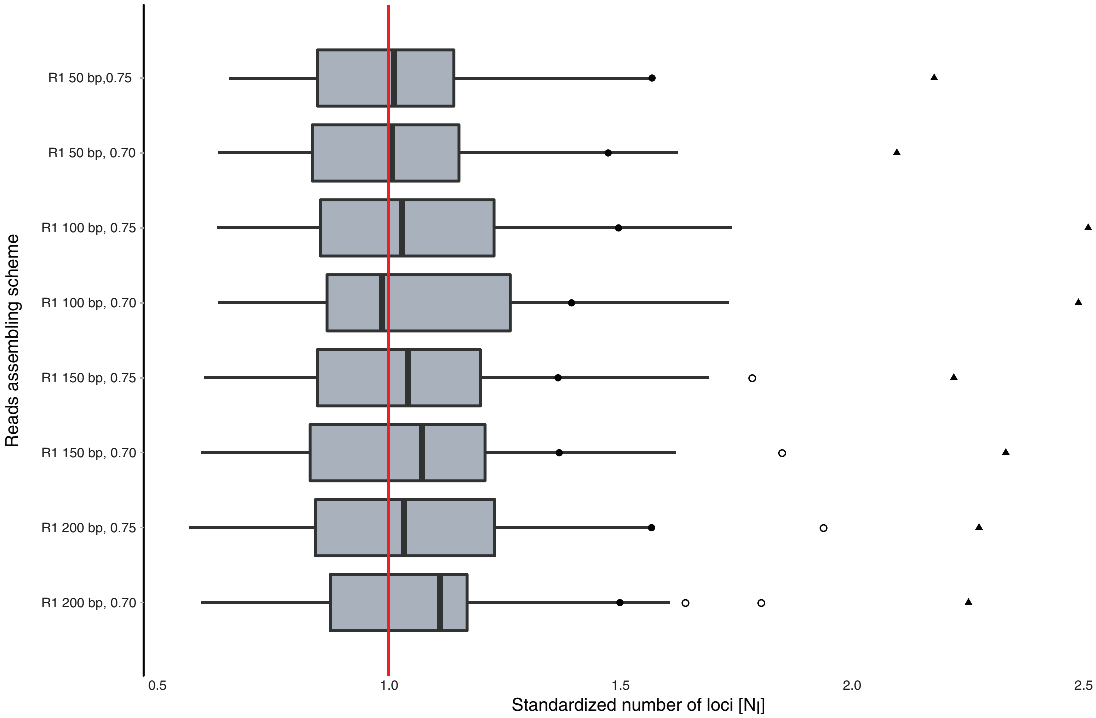

Fig. 6. Boxplots summarizing the number of loci recovered in RAD-seq for all 33 species for R1 reads with 200, 150, 100 or 50 nucleotides positions and *de novo* assembled under 0.70 or 0.75 clustering threshold. Standardized values calculated from equation $N_l = \mu/L$, where μ is the mean number of loci recovered under a specific assembling scheme and L is the number of loci recovered for a specific species under this assembling scheme. A value of 1 (red vertical line) represents the standardized mean number of loci recovered under a specific assembling scheme. For species with N_l below 1, the number of recovered loci was higher than the mean for all taxa, and for species with N_l above 1 the number of recovered loci was lower than the mean for all taxa. Filled triangles indicate the N_l value for *Passeromyia indecora*, and filled circles the N_l value for *Eginia ocypterata*. [Colour figure can be viewed at wileyonlinelibrary.com].

for deeper nodes (Cariou *et al.*, 2013). Although tree topology and bootstrap support for nodes were different between the 200 bp and 150 bp datasets, this discordance was likely caused by the number of recovered loci (e.g. 538 vs 746 under a 0.70 clustering threshold) and not by an increase in the contradictory phylogenetic signal (Rivers *et al.*, 2016). Thus, this study shows that the length of the reads together with the clustering threshold can strongly influence the topology and branch support for a phylogenetic hypothesis.

The use of reference genome for reads assembly can result in a considerably lower number of retrieved loci (Tripp *et al.*, 2017). This is explained by the fact that an increase in evolutionary distance between ingroup taxa and reference genome has a considerable effect on the quantity of assembled reads (Stetter & Schmid, 2017). Furthermore, loci with higher mutation rates have a lower chance of being mapped (Tripp *et al.*, 2017).

Passeromyia indecora significantly differed in position across a range of different RAD-seq analyses, and usually receiving high nodal support. The RAD-seq dataset providing the highest number of PIS and AHE derived phylogenies were incongruent with regard to the position of *P. indecora* (Fig. 4). Previous mS-seq studies have expressed doubts about the phylogenetic position of *P. indecora*, which is an avian parasite in the larval stage. Based on results obtained for other parasitic species (Johnson *et al.*, 2014), Kutty *et al.* (2014) hypothesized an increased rate of molecular evolution in this taxon. In this instance, higher rates of mutation within restriction sites and associated DNA fragments should affect the number of loci recovered for *P. indecora* by a higher dropout of loci shared with other taxa (Fig. 5). While the number of raw reads for *P. indecora* was slightly higher than the mean number of total raw reads (data not shown), we retrieved the lowest number of loci under all *de novo* assembling schemes for this species (Fig. 6). For example, for R1 reads trimmed to 150 bp and assembled *de novo* under the 0.70 similarity threshold we obtained only 40 loci containing 45 PIS for *P. indecora*, almost fourfold less than the mean number of loci recovered for this assembling scheme. Furthermore, loci recovered for *P. indecora* were mostly shared with few other species, mainly representatives of *Azelia* and thus frequently resulted in a close relationship with *Azelia*. Additionally, we observed lack of loci retrieved for *P. indecora* when reads for this species were mapped to *M. domestica* genome.

In this study, we subjected the AHE dataset to two alternative analytical approaches. The phylogenetic hypotheses derived from concatenated and coalescent methods were not fully congruent (Figure S2). Appropriate taxon sampling is an important consideration of coalescent-based phylogeny reconstruction at deeper timescales. The accuracy of this approach should increase with sampling more closely related taxa (Lambert *et al.*, 2015). Thus, incongruence between concatenation and coalescence approaches, particularly the unexpected positions of *A. varia* and *M. domestica*, may reflect the relatively high divergence among the taxa sampled (Lambert *et al.*, 2015; Simmons & Gatesy, 2015). However, discrepancies between concatenation and coalescence methods may be a result of using SVDquartets in the coalescence approach. The performance of SVDquartets implemented in *PAUP** depends on the amount of incomplete lineage sorting (ILS) and accuracy improves when discordance amongst gene trees is low (Chou *et al.*, 2015). Concatenation may provide greater accuracy than coalescent-based methods in predicting deeper divergences when phylogenetic signal is very weak and reconstructed gene trees are strongly divergent (Simmons & Gatesy, 2015).

Systematics of Muscidae

Despite some limitations to confidently resolve deeper splits within Muscidae, as reported by the present and previous studies (Schuehli *et al.*, 2007; Kutty *et al.*, 2014; Haseyama *et al.*, 2015; Grzywacz *et al.*, 2017b), the mS-seq approach has challenged higher-level classifications and contributed to advancement in our knowledge of Muscidae taxonomy. However, many aspects of muscid phylogeny remain unresolved since available molecular phylogenies (Kutty *et al.*, 2014; Haseyama *et al.*, 2015; Grzywacz *et al.*, 2017b) are incongruent with traditional classifications based on adult morphology. Discrepancies between molecular and morphological data are not rare; for example, subfamilial relationships within Sarcophagidae based on mS-seq data (Piwczyński *et al.*, 2014, 2017), incongruent with traditional adult morphology-based classification, have subsequently been corroborated with phylogenomic data and thereby led to novel homology assessments (Buenaventura *et al.*, 2020). On the other hand, immature stages are valuable source of data supporting higher-level taxa (see Meier & Lim, 2009), and have for example corroborated the close relationship between *Achanthiptera* Rondani and some Azeliinae (Skidmore, 1985), supported family status of Fanniidae (Roback, 1951) and some relationships within the flesh fly subfamily Miltogramminae (Piwczyński *et al.*, 2017). Thus, it may prove rewarding for muscid classification to exploit data from immature stages in order to assess conflicting topologies. Indeed, we provide evidence where morphology of immatures corroborates some aspects of the present phylogenetic hypotheses and indicate possible directions for future research. Specifically, this study supports previous mS-seq analyses (Schuehli *et al.*, 2007; Kutty *et al.*, 2014; Grzywacz *et al.*, 2017b) in recovering a dichotomy between Azeliini + Muscinae (clade 1) and the remaining Muscidae (clade 2) with very high nodal support (Fig. 4). However, this relationship has not hitherto been discussed in terms of support from adult or immature stage morphology. The most conspicuous character state we identified corroborating clade 1 was mouth-hook asymmetry in the third instar larva (Fig. 3) (Ferrar, 1979; Iwasa & Nishijima, 1984; Grzywacz, 2013; Grzywacz *et al.*, 2014, 2017a). In contrast, all taxa in clade 2 (where larvae are known) have a third instar larva equipped with symmetrical mouth-hooks; this is most likely a ground-plan feature of Cyclorrhapha (Ferrar, 1987). Mouth-hooks symmetry in clade 1 has likely been secondarily restored in some representatives of *Dasyphora* Robineau-Desvoidy, *Musca* Linnaeus, *Neomyia* Walker, *Morellia* Robineau-Desvoidy and *Pyrellia* Robineau-Desvoidy, although records in the literature concerning *Morellia* and *Pyrellia* (Ishijima, 1967; Skidmore, 1985) remain unconfirmed.

Azeliinae traditionally comprises two tribes, Azeliini and Reinwardtiini, but there is no morphological support for this relationship (Kutty *et al.*, 2014). Our results and previous mS-seq studies (Kutty *et al.*, 2014; Haseyama *et al.*, 2015) conflict with this classification, with Azeliini and Reinwardtiini placed separately in clades 1 and 2, and neither as monophyletic group. These positions are supported by immatures morphology (Fig. 3), as third instars of Azeliini are equipped with asymmetrical mouth-hooks (Grzywacz, 2013; Grzywacz *et al.*, 2014) whereas those in Reinwardtiini exhibit symmetrical mouth-hooks (Velásquez *et al.*, 2013; Grzywacz *et al.*, 2015). We propose Reinwardtiini be raised to the subfamily level, Reinwardtiinae Brauer & Bergenstamm **stat. rev.**, including all 16 genera presently contained in Reinwardtiini (Table 2) (Pont, 1986, 1989; Couri & Carvalho, 2003; Carvalho *et al.*, 2005; Haseyama *et al.*, 2019). Based on the placement of *Eginia* in this clade (Fig. 3), we also propose to transfer all genera traditionally classified within Eginiini (Pont, 1986, 1989), that is *Eginia*, *Neohelina* Malloch, *Syngamoptera* Schnabl and *Xenotachina* Malloch, from Phaoniinae to Reinwardtiinae **stat. rev.** Species relationships awaits future studies, especially the position of the nominal genus *Reinwardtia* Brauer & Bergenstamm, as it was *incertae sedis* according to mS-seq phylogeny (Haseyama *et al.*, 2015), yet a recent morphological phylogeny reconstruction revealed *Reinwardtia* as a sister-taxon of *Synthesiomyia* Brauer & Bergenstamm (Pérez *et al.*, 2020).

Relationships within clade 1: Azelia

The systematic position of *Azelia* has been debated for many years (Karl, 1928; Hennig, 1955–1964, 1965; Savage & Wheeler, 2004). For example, Schnabl & Dziedzicki (1911) classified *Azelia* within Fanniidae (as Homalomyidae). Séguy (1937) and Huckett (1965, 1975) followed this, although they classified Fanniinae as a subfamily of the Muscidae. The close morphological resemblance of *Azelia* to taxa currently classified under the family Fanniidae had considerable influence on recent studies. In a study using complete mitogenome sequences (mtDNA) by Zhang *et al.* (2016), the fanniid genus *Euryomma* Stein was nested within Muscidae. However, upon re-examination of the mtDNA sequence (GenBank accession number KP901269) and photos of the voucher specimen of *Euryomma* used by Zhang *et al.* (2016) (obtained from Ding *et al.*, 2015: fig. 1), we found this to actually be a species of *Azelia*. Karl (1928) and Hennig (1955–1964) classified *Azelia* in Mydaeinae as the sole genus in the tribe

Table 2. Summary of traditional higher-level classification or representative muscid taxa (Pont, 1986; Carvalho *et al*., 2005; Kutty *et al*., 2014; Pérez *et al*., 2020) along with proposed changes to classification marked in bold. Genera not used in this study are marked with an asterisk (*).

Traditional classification			Proposed changes		
Subfamily	Tribe	Genera	Subfamily	Tribe	Genera
Atherigoninae		*Atherigona*	Atherigoninae		*Atherigona*
Muscinae	Muscini	*Eudasyphora*, *Mesembrina*, *Morellia*, *Musca*, *Neomyia*, *Polietes*, *Pyrellia*	Muscinae	Muscini	*Eudasyphora*, *Morellia*, *Musca*, *Neomyia*, *Pyrellia*
	Stomoxyini	*Haematobosca*, *Stomoxys*		Stomoxyini	*Haematobosca*, *Stomoxys*
Azeliinae	Azeliini	*Australophyra*, *Azelia*, *Huckettomyia*, *Hydrotaea*, *Potamia*, *Thricops*	Azeliinae		*Australophyra*, *Azelia*, *Huckettomyia*, *Hydrotaea*, ***Mesembrina***, ***Polietes***, *Potamia*, *Thricops*,
	Reinwardtiini	*Balioglutum**, *Brachygasterina**, *Callainireinwardtia**, *Calliphoroides**, *Chaetagenia**, *Correntosia**, *Dalcyella**, *Fraserella**, *Itatingamyia**, *Muscina*, *Palpibracus**, *Passeromyia*, *Philornis**, *Psilochaeta**, *Reinwardtia**, *Synthesiomyia*	Reinwardtiinae **stat. rev.**		*Balioglutum**, *Brachygasterina**, *Callainireinwardtia**, *Calliphoroides**, *Chaetagenia**, *Correntosia**, *Dalcyella**, *Fraserella**, *Itatingamyia**, *Muscina*, *Palpibracus**, *Passeromyia*, *Philornis**, *Psilochaeta**, *Reinwardtia**, *Synthesiomyia*, ***Eginia***, ***Neohelina*****,** ***Syngamoptera*****,** ***Xenotachina****
Phaoniinae	Dichaetomyiini	*Dichaetomyia*	Phaoniinae	Dichaetomyiini	*Dichaetomyia*
	Eginiini	*Eginia*, *Neohelina**, *Syngamoptera**, *Xenotachina**		Phaoniini	*Helina*, *Phaonia*
	Phaoniini	*Helina*, *Phaonia*	Mydaeinae		*Graphomya*, *Gymnodia*
Mydaeinae		*Graphomya*, *Gymnodia*	Coenosiinae	Limnophorini	*Limnophora*, *Lispe*
Coenosiinae	Limnophorini	*Limnophora*, *Lispe*			

Azeliini. Hennig (1965) later considered *Azelia* within Azeliini (as Hydrotaeini), closely related to *Hydrotaea* and *Thricops*. The molecular study of Grzywacz *et al.* (2017b) questioned the position of *Azelia* and its close relationship with *Thricops*, since *Azelia* was recovered as the sister group to the remaining Azeliini + Muscinae. However, while statistical support for this relationship was high using Bayesian inference (PP = 1), it was low using a ML approach (BS = 65). All three molecular approaches utilized in our study confirm *Azelia* as the sister group to the remaining Azeliini + Muscinae with high branch support in NGS results (BS = 100%, PP = 1 in AHE and BS = 99%, PP = 1 in RAD-seq). *Azelia* was previously considered a close relative of *Thricops* due to the presence of setulae on the hind coxa (Savage & Wheeler, 2004). However, given that this feature is found also in other muscid genera (Carvalho, 2002; Pont, 2019), it is not considered a synapomorphy of this sister-group relationship. Unfortunately, the larval morphology of *Azelia* is insufficiently known. On the other hand, we refrain from classifying Azeliini under Muscinae, since a phylogenetic placement of Azeliini outside of the Muscinae is supported both in details of immature morphology and natural history as discussed below, and from the study of male terminalia musculature by Sorokina & Ovtshinnikova (2020).

Relationships within clade 1: Ophyra

The systematic position and classification of the genus *Ophyra* Robineau-Desvoidy as a synonym of *Hydrotaea* has been widely discussed (Savage & Wheeler, 2004; Schuehli *et al.*, 2004; Haseyama *et al.*, 2015). Recent molecular studies have been contradictory; according to Haseyama *et al.* (2015), *Ophyra* should be considered a valid genus whereas Grzywacz *et al.* (2017b) re-established the synonymy of Pont (1986) under *Hydrotaea* due to paraphyly. However, results obtained from both studies were not fully conclusive due to lack of statistical support (BS < 50). Here we recovered *Ophyra* (here as *Hydrotaea ignava*) nested within *Hydrotaea*, sister to *H. velutina* with very high support. Although intrageneric relationships within *Hydrotaea* (e.g. the position of *H. meteorica*) were incongruent amongst analyses, our results are in agreement with Grzywacz *et al.* (2017b) regarding the status of *Hydrotaea*.

Relationships within clade 1: Mesembrina and *Polietes*

An important incongruence between our results and traditional classification is the position of *Mesembrina* and *Polietes* (Fig. 4).

Previous mS-seq studies provided conflicting results about the position of these genera, and some phylogenetic reconstructions recovered *Mesembrina* and *Polietes* as closely related to representatives of Azeliini (see Kutty *et al.*, 2014: fig. 3; Haseyama *et al.*, 2015: fig. 1). Here we find a sister-group relationship between *Mesembrina* + *Polietes* and *Thricops* + *Huckettomyia* Pont & Shinonaga in all analyses and with high support. However, contrary to mS-seq, results from the NGS data challenged position of these genera within Muscinae, given the clade (*Mesembrina* + *Polietes*) + (*Thricops* + *Huckettomyia*) was nested within Azeliini (excluding *Azelia*) in all analyses (Fig. 4). The similarity of the larval morphology of *Polietes* and *Mesembrina* and some Azeliini was already recognized by Hennig (1965). Based on immature stage morphology and biology, we define two groups within clade 1: clade 1.1 including Azeliini, *Polietes* and *Mesembrina* and clade 1.2 including the remaining Muscinae (Fig. 3). In Azeliini, *Mesembrina* and *Polietes*, the third instar larvae are equipped with well-developed accessory oral sclerites below the apical region of the mouth-hooks (clade 1.1) whereas the remaining Muscinae (clade 1.2) lack such distinct sclerites (Fig. 3). Species of clade 1.1 are facultative or obligatory carnivores in the third larval instar, whereas those in clade 1.2 are coprophagous or saprophagous. In the second instars of clade 1.2, the distal and basal part of each mouth-hook are separated and present as two distinct sclerites, whereas members of clade 1.1 show the plesiomorphic configuration of mouth-hook, without distal and basal part separated. Previous interpretations of second instar mouth-hooks of *Mesembrina* and *Polietes* (Thomson, 1937; Skidmore, 1985) misidentified the epistomal sclerite as the basal part separated from the distal part of the mouth-hook, and both genera share the plesiomorphic mouth-hook with other Azeliini.

We propose classification changes only for *Mesembrina* and *Polietes*. However, future research should investigate whether the apparently closely related *Hennigmyia* Peris, and *Polietina* Schnabl & Dziedzicki, should also be transferred to Azeliinae. The close affinity between *Hennigmyia* and *Polietes* was initially suggested based on adult morphology (Hennig, 1965) and subsequently supported by details of the third instar larvae (Skidmore, 1985). *Polietina* was originally considered closely related to *Polietes*, but Skidmore (1985) considered the cephaloskeleton of the second instar larva of *Polietina* more similar to that of *Muscina* Robineau-Desvoidy. We determined that the latter hypothesis resulted from a misinterpretation of Albuquerque's (1956) description of a pharate second instar larva of *Polietina bicolor* Albuquerque that was misidentified as *P. flavithorax* (Stein) (Couri & Carvalho, 1997). In Albuquerque (1956: plate IV; fig. 50) the third instar mouth-hooks are asymmetric, as in other species grouped herein within clade 1. More recently, *Polietina* was classified within Muscinae (Couri & Carvalho, 1997; Nihei & Carvalho, 2007). This interpretation and the close relationship with *Polietes* and *Mesembrina* have been confirmed in recent molecular phylogenies (Kutty *et al.*, 2014; Haseyama *et al.*, 2015).

Relationships within clade 2

In all three approaches, clade 2 was basally split into one clade consisting of Atherigoninae + (former Reinwardtiini + *Eginia ocypterata*) and a clade including Coenosiinae, Mydaeinae and Phaoniinae. Immature stage morphology and biology corresponds to three distinct clades within clade 2. Members of Atherigoninae (clade 2.1) have trimorphic larvae (i.e. possess three free-living larval instars) that are facultative predators or phytophages (Skidmore, 1985; Grzywacz *et al.*, 2013; Grzywacz & Pape, 2014). Clade 2.2 comprises species with trimorphic larvae that are either facultative predators (*Synthesiomyia* and *Muscina*), parasites of vertebrates (*Passeromyia*) or invertebrates (some *Muscina*, *Eginia*), or induce secondary myiasis (*Synthesiomyia* and *Muscina*).

The position of *Eginia* within Reinwardtiinae **stat. rev.** has already been shown in previous studies (Kutty *et al.*, 2014; Haseyama *et al.*, 2015), yet changes to classification were not proposed. No adult morphological characters are known to support this placement, and the larval morphology of *Eginia* remains unknown. However, *Eginia* is a parasite of millipedes, which corresponds in terms of larval natural history, given that all other representatives of clade 2.2 consume living animal tissues to some extent. Like *Eginia*, *Syngamoptera flavipes* Coquillett was also reared from milipedes (Skidmore, 1985), whereas the biology of the remaining members is unknown.

Clade 2.3 contains Coenosiinae, Mydaeinae and Phaoniinae that encompass dimorphic (i.e. two free-living larval instars) or monomorphic (i.e. one free-living larval instar) obligatory predators. Kutty *et al.* (2014) and Haseyama *et al.* (2015) recovered a close relationship between Mydaeinae and Phaoniinae, consistent with the findings of our AHE concatenated analysis. The AHE coalescent analysis favoured a close relationship between Mydaeinae and Coenosiinae.

Changes in Muscidae higher-level classification

Previous attempts to reconstruct phylogenetic relationships within Muscidae by means of molecular data (Schuehli *et al.*, 2007; Kutty *et al.*, 2014, 2019; Haseyama *et al.*, 2015; Grzywacz *et al.*, 2017b) failed to provide well-supported phylogenetic hypotheses raising questions about several higher-level relationships. Results from our mS-seq, AHE and RAD-seq analyses, whereas at odds with adult morphology-based classifications, are partially supported by immature stage morphology and natural history. Taxon sampling in this study was relatively narrow, yet designed to address interesting taxonomic questions. The data obtained provided robust phylogenetic hypotheses, and we argue for the following changes to the higher-level classification within Muscidae (Table 2): a transfer of *Mesembrina* and *Polietes* from Muscinae to Azeliinae, defined herein as clade 1.1; resurrecting Reinwardtiinae **stat. rev.** as a subfamily distinct from Azeliinae; a transfer of *Eginia*, *Neohelina*, *Syngamoptera* and *Xenotachina* to Reinwardtiinae **stat. rev.**

Supporting Information

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

Figure S1. Distribution of bootstrap values for obtained phylogenetic trees.

Figure S2. Phylogenetic trees obtained in this study from coalescent (topology D) and concatenation (all remaining topologies) analyses. Topologies C and K are derived from Bayesian method and all remaining topologies from maximum likelihood analysis. All trees were rooted using midpoint rooting. Bootstrap scores from 1000 nonparametric bootstrap replicates and posterior probabilities are on nodes. Traditional subfamilial classification: Ath, Atherigoninae; Aze, Azeliinae; Coe, Coenosiinae; Mus, Muscinae; Myd, Mydaeinae; Pha, Phaoniinae.

Table S1. List of species used in the study.

Table S2. Partitioning scheme for the multilocus Sanger sequencing dataset.

Table S3. Comparison of obtained phylogenetic trees with Robinson-Foulds topological distance and Kuhner and Felsenstein branch length score.

Acknowledgements

We thank Mr Nikita Vikhrev (Zoological Museum of Moscow University, Moscow, Russia) for help in obtaining *Huckettomyia watanabei*, Dr Adrian C Pont (University of Oxford, Oxford, U.K.) for granting us an access to unpublished Catalogue of Muscidae (Diptera) of the World, Prof. Krzysztof Szpila (Nicolaus Copernicus University, Toruń, Poland) and anonymous reviewers for comments, which helped to improve this manuscript. We would like to express our appreciation to Prof. Wiesław Babik and Dr Katarzyna Dudek (Jagiellonian University, Kraków, Poland) for access to the MiSeq platform. This research was supported by the Polish National Science Centre grant 2017/01/X/NZ8/01119 to AG and 2015/17/B/NZ8/02453 grant to MP. BKC and BMW were supported by the Schlinger Foundation and the US National Science Foundation (DEB-1754376). The authors declare no conflict of interest relevant to the subject of their manuscript.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Aberer, A.J., Kobert, K. & Stamatakis, A. (2014) ExaBayes: massively paarallel Bayesian tree inference for the whole-genome era. *Molecular Biology and Evolution*, **31**, 2553–2556.

Albuquerque, D.O. (1956) Fauna do Distrito Federal XIII - sôbre o gênero *Polietina* Schnable & Dziedzcki, 1911, com descrições de espécies novas (Diptera-Muscidae). *Boletim do Museu Nacional do Rio de Janeiro (Zoologia)*, **1391**, 1–31.

Ali, O.A., O'Rourke, S.M., Amish, S.J., Meek, M.H., Luikart, G., Jeffres, C. & Miller, M.R. (2016) RAD capture (rapture): flexible and efficient sequence-based genotyping. *Genetics*, **202**, 389–400.

APG (1998) An ordinal classification for the families of flowering plants. *Annals of the Missouri Botanical Garden*, **85**, 531.

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

Borowiec, M.L. (2016) AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ*, **4**, e1660.

Buenaventura, E., Szpila, K., Cassel, B.K., Wiegmann, B.M. & Pape, T. (2020) Anchored hybrid enrichment challenges the traditional classification of flesh flies (Diptera: Sarcophagidae). *Systematic Entomology*, **45**, 281–301.

Cariou, M., Duret, L. & Charlat, S. (2013) Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. *Ecology and Evolution*, **3**, 846–852.

de Carvalho, C.J.B. (1989) Classificação de Muscidae (Diptera): uma proposta através da análise cladística l. *Revista Brasileira de Zoologia*, **6**, 627–648.

de Carvalho, C.J.B. (2002) *Muscidae (Diptera) of the Neotropical Region: Taxonomy*, pp. 1–287.Editora Universidade Federal do Paraná, Curitiba.

de Carvalho, C.J.B., Couri, M.S., Pont, A.C., Pamplona, D. & Lopes, S. (2005) A catalogue of the Muscidae (Diptera) of the neotropical region. *Zootaxa*, **860**, 1–282.

Cerretti, P., Stireman, J.O., 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**, 1–24.

Chifman, J. & Kubatko, L. (2014) Quartet inference from SNP data under the coalescent model. *Bioinformatics*, **30**, 3317–3324.

Chifman, J. & Kubatko, L. (2015) Identifiability of the unrooted species tree topology under the coalescent model with time-reversible substitution processes, site-specific rate variation, and invariable sites. *Journal of Theoretical Biology*, **374**, 35–47.

Chou, J., Gupta, A., Yaduvanshi, S., Davidson, R., Nute, M., Mirarab, S. & Warnow, T. (2015) A comparative study of SVDquartets and other coalescent-based species tree estimation methods. *BMC Genomics*, **16**, 1–11.

Couri, M.S. & Carvalho, C.J.B. (1997) Revision of *Polietina* Schnabl & Dziedzicki (Diptera, Muscidae) and considerations on its new systematic position. *Revista Brasileira de Zoologia*, **14**, 255–281.

Couri, M.S. & de Carvalho, C.J.B. (2003) Systematic relations among *Philornis* Meinert, *Passeromyia* Rodhain Villeneuve and allied genera (Diptera, Muscidae). *Brazilian Journal of Biology*, **63**, 223–232.

Couri, M.S. & Pont, A.C. (2000) Cladistic analysis of Coenosiini (Diptera: Muscidae: Coenosiinae). *Systematic Entomology*, **25**, 373–392.

Crotti, M., Barratt, C.D., Loader, S.P., Gower, D.J. & Streicher, J.W. (2019) Causes and analytical impacts of missing data in RADseq phylogenetics: insights from an African frog (Afrixalus). *Zoologica Scripta*, **48**, 157–167.

Cruaud, A., Gautier, M., Galan, M. *et al.* (2014) Empirical assessment of rad sequencing for interspecific phylogeny. *Molecular Biology and Evolution*, **31**, 1272–1274.

Ding, S., Li, X., Wang, N. *et al.* (2015) The phylogeny and evolutionary timescale of muscoidea (Diptera: Brachycera: Calyptratae) inferred from mitochondrial genomes. *PLoS One*, **10**, 1–17.

Eaton, D.A.R. (2014) PyRAD: assembly of *de novo* RADseq loci for phylogenetic analyses. *Bioinformatics*, **30**, 1844–1849.

Eaton, D.A.R., Spriggs, E.L., Park, B. & Donoghue, M.J. (2017) Misconceptions on missing data in RAD-seq phylogenetics with a deep-scale example from flowering plants. *Systematic Biology*, **66**, 399–412.

van Emden, F.I. (1951) Muscidae: C – scatophaginae, anthomyiinae, lispinae, fanniinae and phaoninae. *Ruwenzori Expedition 1934–5*, Vol. **2**, pp. 325–710.British Museum (Natural History), London.

Evangelista, D.A., Wipfler, B., Béthoux, O. *et al.* (2019) An integrative phylogenomic approach illuminates the evolutionary history of cockroaches and termites (Blattodea). *Proceedings of the Royal Society B: Biological Sciences*, **286**, 20182076.

Ferrar, P. (1979) The immature stages of dung-breeding muscoid flies in Australia, with notes on the species, and keys to larvae and puparia. *Australian Journal of Zoology, Supplementary Series*, **73**, 1–106.

Ferrar, P. (1987) A guide to the breeding habits and immature stages of Diptera Cyclorrhapha. *Entomonograph*, **8**, 1–907.

Gillung, J.P., Winterton, S.L., Bayless, K.M. *et al.* (2018) Anchored phylogenomics unravels the evolution of spider flies (Diptera, Acroceridae) and reveals discordance between nucleotides and amino acids. *Molecular Phylogenetics and Evolution*, **128**, 233–245.

Grabherr, M.G., Haas, B.J., Yassour, M. *et al.* (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, **29**, 644–652.

Gregor, F., Rozkošný, R., Barták, M. & Jaňhara, V. (2002) The Muscidae (Diptera) of Central Europe. *Folia Facultatis Scientiarum Naturalium Universitatis Masarykianae Brunensis, Biologia*, **107**, 1–280.

Grzywacz, A. (2013) Third instar larva morphology of *Hydrotaea cyrtoneurina* (Zetterstedt, 1845) (Diptera: Muscidae) – a species of forensic interest. *Polish Journal of Entomology*, **82**, 303–315.

Grzywacz, A., Hall, M.J.R. & Pape, T. (2015) Morphology successfully separates third instar larvae of *Muscina*. *Medical and Veterinary Entomology*, **29**, 314–329.

Grzywacz, A., Hall, M.J.R., Pape, T. & Szpila, K. (2017a) Muscidae (Diptera) of forensic importance – an identification key to third instar larvae of the western Palaearctic region and a catalogue of the muscid carrion community. *International Journal of Legal Medicine*, **131**, 855–866.

Grzywacz, A., Lindström, A. & Hall, M.J.R. (2014) *Hydrotaea similis* Meade (Diptera: Muscidae) newly reported from a human cadaver: a case report and larval morphology. *Forensic Science International*, **242**, e34–e43.

Grzywacz, A. & Pape, T. (2014) Larval morphology of *Atherigona orientalis* (Schiner) (Diptera: Muscidae) - a species of sanitary and forensic importance. *Acta Tropica*, **137**, 174–184.

Grzywacz, A., Pape, T., Hudson, W.G. & Gomez, S. (2013) Morphology of immature stages of *Atherigona reversura* (Diptera: Muscidae), with notes on the recent invasion of North America. *Journal of Natural History*, **47**, 1055–1067.

Grzywacz, A., Wallman, J.F. & Piwczyński, M. (2017b) To be or not to be a valid genus: the systematic position of *Ophyra* R.-D. revised (Diptera: Muscidae). *Systematic Entomology*, **42**, 714–723.

Harvey, M.G., Smith, B.T., Glenn, T.C., Faircloth, B.C. & Brumfield, R.T. (2016) Sequence capture versus restriction site associated DNA sequencing for shallow systematics. *Systematic Biology*, **65**, 910–924.

Haseyama, K.L.F., de Carvalho, C.J.B., Zafalon-Silva, Â. & Kirst, F.D. (2019) New systematic position of *Itatingamyia* Albuquerque (Diptera, Muscidae) based on molecular evidence, and description of the female of I. couriae. *Revista Brasileira de Entomologia*, **63**, 35–42.

Haseyama, K.L.F., Wiegmann, B.M., Almeida, E.A.B. & de Carvalho, C.J.B. (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). *Molecular Phylogenetics and Evolution*, **89**, 1–12.

Hedges, S.B. (1994) Molecular evidence for the origin of birds. *Proceedings of the National Academy of Sciences*, **91**, 2621–2624.

Hennig, W. (1955) 63 b. Muscidae. *Die Fliegen der Paläarktischen Region* (ed. by E. Lindner), pp. 1–1110.Schweizerbart'sche Verlagsbuchhandlung, Stuttgart.

Hennig, W. (1965) Vorarbeiten zu einem phylogenetischen System der Muscidae (Diptera: Cyclorrhapha). *Stuttgarter Beiträge zur Naturkunde*, **141**, 1–100.

Huang, H. & Knowles, L.L. (2016) Unforeseen consequences of excluding missing data from next-generation sequences: simulation study of rad sequences. *Systematic Biology*, **65**, 357–365.

Huckett, H.C. (1965) The Muscidae of Northern Canada, Alaska, and Greenland (Diptera). *Memoirs of the Entomological Society of Canada*, **42**, 1–369.

Huckett, H.C. (1975) The Muscidae of California: Exclusive of Subfamilies Muscinae and Stomoxyinae. *Bulletin of the California Insect Survey*, **18**, 1–148.

Ishijima, H. (1967) Revision of the third stage larvae of synanthropic flies of Japan (Diptera: Anthomyiidae, Muscidae, Calliphoridae and Sarcophagidae). *Japanese Journal of Sanitary Zoology*, **18**, 47–100.

Iwasa, M. & Nishijima, Y. (1984) Studies on the dung-breeding flies in Japan IV. The immature stages of three muscine species occurring from brown bear dung (Diptera: Muscidae, Muscinae). *Japanese Journal of Sanitary Zoology*, **35**, 381–389.

Jarvis, E.D., Mirarab, S., Aberer, A.J. *et al.* (2014) Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, **346**, 1320–1331.

Johnson, K.P., Allen, J.M., Olds, B.P., Mugisha, L., Reed, D.L., Paige, K.N. & Pittendrigh, B.R. (2014) Rates of genomic divergence in humans, chimpanzees and their lice. *Proceedings of the Royal Society B: Biological Sciences*, **281**, 20132174.

Junqueira, C.A.M., Azeredo-Espin, M.A.L., Paulo, D.F. *et al.* (2016) Large-scale mitogenomics enables insights into Schizophora (Diptera) radiation and population diversity. *Scientific Reports*, **6**, 1–13.

Karl, O. (1928) Zweiflügler oder Diptera. II: Muscidae. *Die Tierwelt Deutschlands* (ed. by F. Dahl), pp. 1–232.G. Fischer Verlag, Jena.

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.

Kjer, K., Borowiec, M.L., Frandsen, P.B. & Ware, J. (2016) Advances using molecular data in insect systematics. *Current Opinion in Insect Science*, **18**, 40–47.

Kück, P. & Longo, G.C. (2014) FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Frontiers in Zoology*, **11**, 81.

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., Meusemann, K., Bayless, K.M. *et al.* (2019) Phylogenomic analysis of Calyptratae: resolving the phylogenetic relationships within a major radiation of Diptera. *Cladistics*, **35**, 605–622.

Kutty, S.N., Pape, T., Pont, A.C., Wiegmann, B.M. & Meier, R. (2008) The Muscoidea (Diptera: Calyptratae) are paraphyletic: evidence from four mitochondrial and four nuclear genes. *Molecular Phylogenetics and Evolution*, **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. *Systematic Entomology*, **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. *Molecular Phylogenetics and Evolution*, **78**, 349–364.

Lambert, S.M., Reeder, T.W. & Wiens, J.J. (2015) When do species-tree and concatenated estimates disagree? An empirical analysis with higher-level scincid lizard phylogeny. *Molecular Phylogenetics and Evolution*, **82**, 146–155.

Lanfear, R., Calcott, B., Ho, S.Y.W. & Guindon, S. (2012) PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, **29**, 1695–1701.

Leaché, A.D., Chavez, A.S., Jones, L.N., Grummer, J.A., Gottscho, A.D. & Linkem, C.W. (2015) Phylogenomics of phrynosomatid lizards: conflicting signals from sequence capture versus restriction site associated DNA sequencing. *Genome Biology and Evolution*, **7**, 706–719.

Lemmon, A.R., Emme, S.A. & Lemmon, E.M. (2012) Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology*, **61**, 727–744.

McCormack, J.E., Hird, S.M., Zellmer, A.J., Carstens, B.C. & Brumfield, R.T. (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, **66**, 526–538.

Meier, R. & Lim, G.S. (2009) Conflict, convergent evolution, and the relative importance of immature and adult characters in Endopterygote phylogenetics. *Annual Review of Entomology*, **54**, 85–104.

Meyer, M. & Kircher, M. (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, **2010**, 1–10.

Michelsen, V. (1991) Revision of the aberrant New World genus *Coenosopsia* with a discussion of anthomyiid relationships. *Systematic Entomology*, **16**, 85–104.

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

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.

Mutanen, M., Wahlberg, N. & Kaila, L. (2010) Comprehensive gene and taxon coverage elucidates radiation patterns in moths and butterflies. *Proceedings of the Royal Society B: Biological Sciences*, **277**, 2839–2848.

Nihei, S.S. & Carvalho, C.J.B. (2007) Phylogeny and classification of Muscini (Diptera, Muscidae). *Zoological Journal of the Linnean Society*, **149**, 493–532.

Pape, T., Blagoderov, V. & Mostovski, M.B. (2011) Order Diptera Linnaeus, 1758. *Animal Biodiversity: An Outline of Higher-Level Classification and Survey of Taxonomic Richness*, Vol. **3148** (ed. by Z.-Q. Zhang), pp. 222–229, Auckland, New Zealand: Magnolia press.

Pauli, T., Burt, T.O., Meusemann, K. *et al.* (2018) New data, same story: phylogenomics does not support Syrphoidea (Diptera: Syrphidae, Pipunculidae). *Systematic Entomology*, **43**, 447–459.

Pérez, S., Fogaça, J.M., Wolff, M. & de Carvalho, C.J.B. (2020) Morphological phylogeny of *Reinwardtia* Brauer & Bergenstamm (Diptera, Muscidae), with the description of a new species from the Neotropical region. *Systematics and Biodiversity*, **0**, 1–11.

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.

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. *Molecular Phylogenetics and Evolution*, **116**, 49–60.

Piwczyński, M., Szpila, K., Grzywacz, A. & Pape, T. (2014) A large-scale molecular phylogeny of flesh flies (Diptera: Sarcophagidae). *Systematic Entomology*, **39**, 783–799.

Piwczyński, M., Trzeciak, P., Popa, M.-O., Pabijan, M., Corral, J., Spalik, K. & Grzywacz, A. (2021) Using RAD seq for reconstructing phylogenies of highly diverged taxa: a test using the tribe Scandiceae (Apiaceae). *Journal of Systematics and Evolution*, **59**, 58–72.

Pohl, D., Kühne, S., Karaca, İ. & Moll, E. (2012) Review of *Coenosia attenuata* stein and its first record as a predator of important greenhouse pests in Turkey. *Phytoparasitica*, **40**, 63–68.

Pont, A.C. (1969) Studies on Australian Muscidae (Diptera) II. A revision of the tribe Dichaetomyiini Emden. *Bulletin of the British Museum (Natural History) Entomology*, **23**, 191–286.

Pont, A.C. (1973) Studies on Australian Muscidae IV revision of the subfamilies Muscinae and stomoxyinae. *Australian Journal of Zoology Supplementary Series*, **21**, 129–296.

Pont, A.C. (1986) Family Muscidae. *Catalogue of Palaearctic Diptera. Scatophagidae - Hypodermatidae* (ed. by A. Soós and L. Papp), pp. 1–345.Elsevier, Amsterdam.

Pont, A.C. (1989) Family Muscidae. *Catalog of the Diptera of the Australasian and Oceanian Regions* (ed. by E. Evenhuis), pp. 675–699.Special Publications of the Bernice Pauahi Bishop Museum, Honolulu, Leiden.

Pont, A.C. (2019) Studies on the Australian Muscidae (Diptera). VIII. The genus *Lispe* Latreille, 1797. *Zootaxa*, **4557**, 1–232.

Popescu, A.-A., Huber, K.T. & Paradis, E. (2012) Ape 3.0: new tools for distance-based phylogenetics and evolutionary analysis in R. *Bioinformatics*, **28**, 1536–1537.

Prum, R.O., Berv, J.S., Dornburg, A., Field, D.J., Townsend, J.P., Lemmon, E.M. & Lemmon, A.R. (2015) A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature*, **526**, 569–573.

R Core Team. (2019) *R: A Language and Environment for Statistical Computing*.

Rambaut, A., Drummond, A.J., Xie, D., Baele, G. & Suchard, M.A. (2018) Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Systematic Biology*, **67**, 901–904.

Revell, L.J. (2012) Phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, **3**, 217–223.

Rivers, D.M., Darwell, C.T. & Althoff, D.M. (2016) Phylogenetic analysis of RAD-seq data: examining the influence of gene genealogy conflict on analysis of concatenated data. *Cladistics*, **32**, 672–681.

Roback, S.S. (1951) A classification of the muscoid calyptrate, Diptera. *Annals of the Entomological Society of America*, **44**, 327–361.

Rubin, B.E.R., Ree, R.H. & Moreau, C.S. (2012) Inferring phylogenies from RAD sequence data. *PLoS One*, **7**, 1–12.

Savage, J. & Wheeler, T.A. (2004) Phylogeny of the Azeliini (Diptera: Muscidae). *Studia Dipterologica*, **11**, 259–300.

Schnabl, J. & Dziedzicki, H. (1911) Die Anthomyiden. *Nova acta Academiae Caesareae Leopoldino-Carolinae*, **95**, 55–358.

Schuehli, G.S.E., de Carvalho, C.J.B. & Wiegmann, B.M. (2004) Regarding the taxonomic status of *Ophyra* Robineau-Desvoidy (Diptera: Muscidae): a molecular approach. *Zootaxa*, **712**, 1–12.

Schuehli, G.S.E., de Carvalho, C.J.B. & Wiegmann, B.M. (2007) Molecular phylogenetics of the Muscidae (Diptera: Calyptratae): new ideas in a congruence context. *Invertebrate Systematics*, **21**, 263–278.

Scott, J.G., Warren, W.C., Beukeboom, L.W. *et al.* (2014) Genome of the house fly, *Musca domestica* L., a global vector of diseases with adaptations to a septic environment. *Genome Biology*, **15**, 466.

Séguy, E. (1937) Diptera Fam. Muscidae. *Genera Insectorum*, Vol. **205** (ed. by P. Wytsman), pp. 1–604.Bruxelles, Desmet-Verteneuil.

Simmons, M.P. & Gatesy, J. (2015) Coalescence vs. concatenation: sophisticated analyses vs. first principles applied to rooting the angiosperms. *Molecular Phylogenetics and Evolution*, **91**, 98–122.

Skidmore, P. (1985) The biology of the Muscidae of the world. *Series Entomologica*, **29**, 1–550.

Sorokina, V.S. & Ovtshinnikova, O.G. (2020) The position of the Azeliinae in the Muscidae (Diptera) based on musculature of the male terminalia. *ZooKeys*, **975**, 87–110.

Springer, M.S., Cleven, G.C., Madsen, O., de Jong, W.W., Waddell, V.G., Amrine, H.M. & Stanhope, M.J. (1997) Endemic African mammals shake the phylogenetic tree. *Nature*, **388**, 61–64.

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

Stetter, M.G. & Schmid, K.J. (2017) Analysis of phylogenetic relationships and genome size evolution of the *Amaranthus* genus using GBS indicates the ancestors of an ancient crop. *Molecular Phylogenetics and Evolution*, **109**, 80–92.

Suchan, T., Espíndola, A., Rutschmann, S. *et al.* (2017) Assessing the potential of RAD-sequencing to resolve phylogenetic relationships within species radiations: the fly genus *Chiastocheta* (Diptera: Anthomyiidae) as a case study. *Molecular Phylogenetics and Evolution*, **114**, 189–198.

Suyama, M., Torrents, D. & Bork, P. (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*, **34**, W609–W612.

Swofford, D.L. (2002) *Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Sinauer Associates, Sunderland, MA.

Thomson, R.C.M. (1937) Observations on the biology and larvae of the Anthomyiidae. *Parasitology*, **29**, 273–358.

Tiusanen, M., Hebert, P.D.N., Schmidt, N.M. & Roslin, T. (2016) One fly to rule them all – muscid flies are the key pollinators in the Arctic. *Proceedings of the Royal Society B: Biological Sciences*, **283**, 20161271.

Tripp, E.A., Tsai, Y.H.E., Zhuang, Y. & Dexter, K.G. (2017) RADseq dataset with 90% missing data fully resolves recent radiation of *Petalidium* (Acanthaceae) in the ultra-arid deserts of Namibia. *Ecology and Evolution*, **7**, 7920–7936.

Velásquez, Y., Ivorra, T., Grzywacz, A., Martínez-Sánchez, A., Magaña, C., García-Rojo, A. & Rojo, S. (2013) Larval morphology, development and forensic importance of *Synthesiomyia nudiseta* (Diptera: Muscidae) in Europe: a rare species or just overlooked? *Bulletin of Entomological Research*, **103**, 98–110.

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*, **108**, 5690–5695.

Young, A.D.D., Lemmon, A.R.R., Skevington, J.H.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.

Zhang, D., Yan, L., Zhang, M. *et al.* (2016) Phylogenetic inference of calyptrates, with the first mitogenomes for Gasterophilinae (Diptera: Oestridae) and Paramacronychiinae (Diptera: Sarcophagidae). *International Journal of Biological Sciences*, **12**, 489–504.

Accepted 28 January 2021