Systematic Entomology

Systematic Entomology (2021), DOI: 10.1111/syen.12529

Molecular phylogeny of *Allodia* (Diptera: Mycetophilidae) constructed using genome skimming

TRUDE MAGNUSSEN[1], ARILD JOHNSEN[1], JOSTEIN KJÆRANDSEN[2], TORSTEN H. STRUCK[1] and GEIR E. E. SØLI[1]

[1]Natural History Museum, University of Oslo, Oslo, Norway and [2]The Arctic University Museum of Norway, The Arctic University of Norway, Tromsø, Norway

Abstract. In this study, we use low-coverage shotgun sequencing of genomic DNA, commonly referred to as genome skimming, to investigate the phylogenetic relationships of the fungus gnat genus *Allodia* Winnertz (Mycetophilidae, Mycetophilinae, Exechiini). Nineteen specimens, representing 16 in-group and three outgroup taxa, were successfully sequenced and molecular markers of both mitochondrial and nuclear origin were retrieved. The phylogenetic analyses of 13 protein-coding mitochondrial genes, two ribosomal mitochondrial genes and the nuclear ribosomal 18S and 28S strongly support the monophyly of its two subgenera *Allodia* s.s. and *Brachycampta*. Complete mitochondrial genomes of 15 species were assembled, which enables further comparisons with the mitochondrial genomes of other Diptera. Overall, the methodology used in this study proved successful and promising for other dipteran groups. In addition to the phylogenetic reconstruction, the morphological characters previously used to separate the two subgenera were evaluated and re-examined. Together with the composite structure of the male genitalia, we consider details of the scutal bristles appropriate for separating the two groups. Based on the achieved results, we reinstate *Brachycampta* **stat. rev.** as a separate genus.

Introduction

High throughput sequencing (HTS) is increasingly being used in insect phylogenetic studies, especially through reduced-representation techniques such as target capture or transcriptomics (Misof *et al.*, 2014; Young *et al.*, 2016; Peters *et al.*, 2017). Reduced-representation techniques provide only a snapshot or specifically subsampled representation of the genome instead of the entire genome and the retrieved genomic information is thus reduced. However, the amount of information represented is a larger part of the genome than in traditional marker-based approaches. One of these reduced-representation methods is genome skimming. Genome skimming equals random sequencing of genomic DNA at low coverage (Straub *et al.*, 2012). Even though the coverage of the genome is low, repetitive parts of the genome with high copy numbers will have higher coverage. Therefore, organellar DNA such as the mitochondrial genome (mitogenome) and genes from the nuclear ribosomal DNA cluster (18S and 28S) will have high coverage and recovery rates in skims of animal genomes (Grandjean *et al.*, 2017; Ren *et al.*, 2017; Trevisan *et al.*, 2019). Genome skimming has been used to reconstruct phylogenies in several insect taxa (Ren *et al.*, 2017; Zhang *et al.*, 2019) and has been used for DNA barcoding and DNA sequencing of bulk samples for biodiversity studies (Papadopoulou *et al.*, 2015; Coissac *et al.*, 2016; Linard *et al.*, 2016).

Genome skimming is considered a relatively simple method in terms of bioinformatics and has the advantages that no *a priori* knowledge of the investigated genomes, genomic resources or optimization steps are required (Lemmon & Lemmon, 2013; Knyshov *et al.*, 2021). In addition, the method demands less laboratory work and the quality requirements of the samples are lower than for many other HTS methods, such as whole-genome sequencing, transcriptomics or target-capture methods. On the other hand, genome skimming is most efficient for sequencing organellar DNA, which might not be suitable for resolving deep

Correspondence: Trude Magnussen, Natural History Museum, University of Oslo, P.O. Box 1172 Blindern, NO-0318 Oslo, Norway. E-mail: trude.magnussen@nhm.uio.no

phylogenies. Nevertheless, mitogenomic data have been used in higher-level systematics in Diptera (Cameron *et al.*, 2006; Zhang *et al.*, 2016; Zhang *et al.*, 2019), and also recently in Mycetophilidae (Wang *et al.*, 2021).

The fungus gnat tribe Exechiini (Diptera: Mycetophilidae: Mycetophilinae) contains approximately 700 described species in 20 genera (Ševčík & Kjærandsen, 2012; Burdíková *et al.*, 2019). Members of the tribe are known from all continents except Antarctica and its highest species diversity is in the Holarctic region, although they are understudied in tropical and subtropical areas. The tribe is estimated to have originated in the Paleogene, with a rapid radiation during the Eocene (Rindal *et al.*, 2007; Burdíková *et al.*, 2019; Kaspřák *et al.*, 2019). This young age has proven to be a challenge in trying to reconstruct the intergeneric phylogeny of the tribe (Kjærandsen, 2006; Rindal & Søli, 2006; Rindal *et al.*, 2007; Burdíková *et al.*, 2019; Camier & Nel, 2020). The recent radiation may also explain the overall uniform morphology within the tribe as several genera can only be diagnosed by subtle differences in characters (Rindal & Søli, 2006). As a result, the outline of the male genitalia is not only used in the identification of species, but often also hold the most defining characters of some genera. Undoubtedly, Tuomikoski (1966) has given the most comprehensive and detailed presentation of the systematics within the tribe Exechiini. He clearly accounted for the many ambiguities linked to the conception of its included genera, a debate that has been going on ever since.

Allodia Winnertz contains 96 described species in two subgenera; *Allodia* s.s. and *Brachycampta* Winnertz. The genus is distributed in the Holarctic (Zaitzev, 1983, 1984), the Afrotropical (Matile, 1978; Magnussen *et al.*, 2018), the Oriental (Senior-White, 1922; Edwards, 1928; Magnussen *et al.*, 2019) and the Oceanian regions (Colless, 1966). The majority of the species, however, has been described from the Holarctic region. The larva of *Allodia* have been found in fruiting bodies of agaric fungi, mainly in the orders Boletales and Russulales; several species in subgenus *Brachycampta* are mainly or exclusively associated with the order Pezizales (Jakovlev, 2012). Although many rearing records exist for the genus (Chandler, 1993; Ševčík, 2010; Jakovlev, 2011), the biology of the larval stages of most species is unknown. Based on commonly applied collecting methods, like interception traps and sweep netting, *Allodia* s.s. is very common and abundant in boreal forests with bimodal flight periods in spring and autumn, while *Brachycampta* appears sporadic and rare (personal observation). Individuals of *Allodia* s.s. tend to congregate in shady and humid places, often several species together and in large numbers. Species of *Brachycampta*, however, are usually collected in low numbers.

Winnertz (1864) originally described *Allodia* and *Brachycampta* as separate genera. He included five species in *Allodia* and eight species in *Brachycampta*. Of these, one species is retained in each of the two subgenera, while the others have been regarded as invalid, synonymized or ascribed to other genera. Johannsen (1909) merged the two subgenera, and this practice was followed by Coquillett (1911), and later by Edwards (1925) who also included *Synplasta* Skuse and *Brevicornu* Marshall. Tuomikoski (1966), in his review of Exechiini, maintained the sub-generic status of *Allodia* and *Brachycampta*, but reinstated *Brevicornu* as a separate genus. This view has been followed by subsequent authors, including the most recent revisions of the subgenera by Zaitzev (1983, 1984). In the most recent molecular phylogenetic study of Exechiini (Burdíková *et al.*, 2019) *Allodia* and *Brachycampta* exhibit a sister-group relationship with high support, but only two species from each subgenus were included.

Members of *Allodia* and *Brachycampta* are small to medium-sized with slender body and long legs; they are very similar in general appearance, and are typical representatives of Exechiini (Fig. 1). Typical characters to separate them from most other genera in the tribe include the following: wings with point of furcation of posterior before level of point of furcation of anterior fork; branches of both forks without setulae; wing with subcostal vein ending in vein radial vein (R1); thorax with a bare anepisternum and scutum with bristles, but not evenly clothed by setae (for a full key to genera, see e.g. Søli et al. (2000)) Characters to distinguish between the two subgenera exist in the chaetotaxy of mesonotum, wing venation and abdominal coloration. However, these characters are imprecisely defined and do not necessarily apply to all species of the two subgenera. Johannsen (1909) claimed that the characters given by Winnertz (1864), for example, the outline of the anal vein and the position of the cubital fork, were far from adequate to separate between the North American species, and he consequently combined the two. Later, the most conspicuous and consistent morphological differences between the two subgenera were found in the structures of the male terminalia, as first pointed out by Tuomikoski (1966). In general, the shaping of the various elements that constitute the male terminalia is more consistent and comparable across species in *Allodia* s.s., than in *Brachycampta*, where the outline of one or more elements sometimes may be highly divergent. Based on the morphological differences, Kjærandsen (2007) in a study of the *A.* (*B.*) *barbata*-group, suggested that the two subgenera should be regarded as separate genera, but before doing so, he called for a rigorous test of their reciprocal monophyly.

In this study, we will test the utility of genome skimming for obtaining molecular markers for phylogenetic reconstruction of the Exechiini genus *Allodia*. We use genome skimming to target the complete mitogenomes and the nuclear ribosomal markers 18S and 28S, representing both more rapidly evolving mitochondrial markers and more slowly evolving nuclear markers. In addition, we will re-examine and evaluate morphological characters used in previous attempts to delimit and define the subgenera.

Materials and methods

Taxon sampling and DNA extraction

We initially extracted genomic DNA from 43 specimens, representing 25 species, including one from each of the genera *Allodiopsis* Tuomikoski, *Brevicornu*, *Cordyla* Meigen and *Pseudexechia* Tuomikoski as outgroup taxa. The choice of outgroup taxa

 doi: 10.1111/syen.12529

Fig. 1. General morphology of the two subgenera, represented by (A) *Allodia lugens* (Wiedemann); (B) *Allodia Brachycampta barbata* Lundstrom. Photo: Karsten Sund (NHM, Oslo).

is based on previous hypotheses on phylogenetic relationships between the Exechiini genera. *Brevicornu* and *Cordyla* were included in the analyses because they originally were merged with *Allodia* by Edwards (1925), and share several morphological features with *Allodia*. Tuomikoski (1966) considered *Allodia* to have a close relationship to *Pseudexechia* and *Allodiopsis*, we, therefore, also included them as outgroups. Too late for our study, Burdíková *et al.* (2019) hypothesized a previously unrecognized sister-group relationship between *Allodia* and *Anatella* Winnertz. We did include two specimens of *Anatella lenis* Dziedzicki in the DNA extractions, but the measured DNA concentration was too low to proceed with these samples (see Table S1).

All specimens were collected between 2010 and 2017 from various localities in Norway; some were pinned, but most specimens were stored in ethanol (>70%). The full list of specimens and localities is given in Table S1. For each specimen thorax, legs and head were used for DNA extraction, while the abdomen including the terminalia was kept as voucher in the collection at the Natural History Museum in Oslo (NHMO), stored in micro vials with glycerol (Col. Numbers: NHMO 442461 – 442498). The ethanol-preserved specimens were incubated in a heating chamber for 20 min, 50°C, to remove excess ethanol, before DNA extraction. DNA was extracted using the tissue DNA protocol of the E.Z.N.A Tissue DNA kit, following the manufacturer's protocol with minor modifications to optimize the DNA yield and concentration: in the first step, a pestle was used to crush the exoskeleton effectively, the lysis was done overnight and the elution once with 60 µL Elution Buffer. The DNA concentration was measured using Qubit Spectrophotometer (ThermoFisherScientific) and subsequently checked for fragmentation of the product on a 2% agarose gel to ensure a high quality and a minimum concentration (>10 ng) of genomic DNA. The specimens stored in ethanol worked better with regard to DNA quality (i.e. concentration and fragmentation), compared to the pinned specimens.

To reduce costs, the limit was set to approximately 20 taxa altogether, as we were sequencing all on one lane. Based on the measurements of DNA concentration and evaluation of the fragmentation, we therefore chose 20 samples for subsequent sequencing, one sample per species (highlighted in Table S1). The selected samples represented 16 in-group species; nine of the species from *Allodia* s.s., seven species from *Brachycampta* and four outgroup species. Our main purpose when choosing in-group taxa was to cover as many morphologically distinct groups as possible, primarily assessed by variation in their male terminalia. Due to greater variation inside the group, this was easier to accomplish for *Brachycampta* than for *Allodia* s.s. From *Brachycampta*, it was important for us to include *A.* (*B.*) *czernyi* (Landrock) and *A.* (*B.*) *alternans* (Zetterstedt), as they are morphologically similar to *Allodia* s.s. One of the included species represents an undescribed species, and is referred to with the interim name *Allodia* JKJ-spA.

Morphology

For some specimens, the terminalia were temporarily slide mounted as described in Magnussen *et al.* (2018) and photographed using Zeiss Axio imager M2, fitted with the camera Axiocam 506 colour. The photos were stacked with Zerene Stacker version 1.04 (ZereneSystems LLC), and edited in Adobe Photoshop and assembled using Adobe Illustrator. In addition, scanning electron microscope (SEM) images were taken of pinned, noncoated, specimens, in a Hitachi S-3600N SEM at the Geo Lab at the Natural History Museum, University of Oslo. The instrument was operated at low vacuum (20 Pa), 15.0 kV

 doi: 10.1111/syen.12529

accelerating voltage and 70% beam current. To obtain detailed images of the specimen's surface, the backscattered detector in 3D mode was used.

Library construction and sequencing

The library preparation and sequencing were performed at the Norwegian Sequencing Centre (NSC). Library preparations were performed with the Nextera DNA Flex Library Preparation Kit (Illumina), with 5–12 cycles of amplification depending on the amount of input DNA of each sample. The samples were subsequently sequenced paired-end on one lane with Illumina HiSeq 4000, with a desired insert size of 350 bp and read length of 150 bp.

Sequence assembly, target gene search and annotation

The paired end libraries were first assembled *denovo* using St. Petersburg genome assembler (SPAdes) (Bankevich *et al.*, 2012), with k-mer sizes 21, 33 and 55. The reads were corrected by BayesHammer (Nikolenko *et al.*, 2013) included in SPAdes. The quality of each assembly was subsequently evaluated with Quast (Gurevich *et al.*, 2013). The Basic Local Alignment Search Tool version 2.8.1 (BLAST+) (Altschul *et al.*, 1990) was used locally to screen for the whole mitogenome, specific mitochondrial genes and the nuclear ribosomal genes (18S and 28S) using 'tblastn' and 'blastn' in combination with different bait sequences as queries. The ∼650 bp barcoding region of cytochrome oxidase subunit 1 (cox1) was used to find long contigs representing the complete or parts of the mitogenome. The DNA barcodes used as queries are available through the publicly available dataset 'Allodia Phylogeny Queries' (DS-ALPHYL) in Barcode of Life Datasystems (BOLD) (Ratnasingham & Hebert, 2007). As queries for the nuclear ribosomal genes, we used the partial 18S sequence from *Allodia sp*. (GenBank: DQ787912.1, in Rindal *et al.* (2007)), and for 28S, we used the partial 28S of *Allodia sp*. (GenBank: EU219584.1, in Rindal *et al.* (2009)).

Using short sequences, such as cox1, did not work to obtain long contigs of the mitogenome. Therefore, we tested using complete mitochondrial sequences from other dipterans as queries, namely; *Tipula cockerelliana* (GenBank: KT970065.1, in Zhang *et al.* (2016)) and *Arachnocampa flava* (GenBank: NC_016204.1, in Beckenbach (2012)). Using the complete mitochondrial references aided in getting some longer contigs in the blast searches, but the mitogenomes were still broken into short fragments (∼1000 bp). In most cases, only short and incomplete sequences were also retrieved for 18S and 28S.

An alternative strategy was therefore implemented, which consisted of a reference-based method with a nucleotide sequence as a seed in MITObim version 1.8 (Hahn *et al.*, 2012). First, the raw reads were trimmed with Trim Galore! V.0.3.3 (Krueger, 2013), using the 'paired' option and other options as default. Due to the high copy number for both the mitogenome and the rRNA cluster, the method was used for both the mitochondrial genes/genomes and the nuclear ribosomal genes. MITObim was run with the parameter 'quick' and the 'pair' option, other settings were left unchanged. Initially, the mitogenome of all 20 species was attempted to be reconstructed with MITObim using the barcode queries as a reference. This resulted in most cases in short sequences (∼4000 bp), because the iteration process ended before the complete circular mitogenome sequence was assembled. Still, the sequences were longer than what was retrieved with BLAST+ using the SPAdes assemblies. To be able to assemble longer and almost complete mitogenomes we therefore used *Arachnocampa flava* (Acc.no NC_016204.1) as a reference for the assembly of *A*. JKJ-spA. Then, *A*. JKJ-spA was used as a reference for the rest of the species, which resulted in longer mitochondrial sequences and several complete mitogenomes.

The mitochondrial sequences obtained through the two approaches were annotated using MITOS web server (Bernt *et al.*, 2013), using default settings and the invertebrate genetic code. The short sequences were challenging to annotate, and often contained incomplete genes. If the genes annotated with MITOS were of approximate expected length, each sequence was blasted against NCBI to verify the identity. The nuclear 18S were annotated with RNAmmer 1.2 Server (Lagesen *et al.*, 2007), and the partial 28S were not annotated, but checked with manual BLAST searches.

Coverage plots and alignments of the assemblies from MITObim were inspected with the software TABLET v 1.17.08.17 (Milne *et al.*, 2013). Most of the complete mitogenomes had areas of low quality in the assembly, where the coverage dropped and contained ambiguous characters. These regions were likely due to long stretches of repetitive bases. The low-quality regions were reassembled with a few iterations and compared against the initial sequence. For those species where we could not obtain the complete or almost complete mitogenome with any of the methods, we assembled separate genes or regions using MITObim with a few iterations, in order to get a complete set of mitochondrial genes.

Genes and alignment

Even though we could not determine the complete mitogenome for all our species, sequence information of all 13 protein-coding mitochondrial genes, two ribosomal mitochondrial genes and two nuclear rRNA genes 18S and partial 28S could still be retrieved, except for mitochondrial genes of *Cordyla* sp., which turned out to be difficult to assemble probably due to poor quality of the retrieved sequence data. All alignments are available in the Figshare database (https://doi.org/10.6084/m9.figshare.16565994).

Each gene was aligned separately using MAFFT version 7.300 (Katoh & Standley, 2013) with max iterations set to 1000, using the 'globalpair' and 'reorder' options for the input file. In several of the species, the gene NADH dehydrogenase subunit 5 (nad5) was split into two parts (∼500 and ∼1000 bp long) given the MITOS annotation. In this case, both parts were aligned together with the complete nad5 sequences of

 doi: 10.1111/syen.12529

the other species, and the parts were subsequently merged into one sequence based on the alignment. The reason for the splitting was a low-quality region with a stretch of T's of different length introducing shifts in the reading frame. After alignment, ambiguous regions (i.e. containing gaps or poorly aligned) in nad5 (amino acid), 12S, 16S, 18S and 28S were removed with Gblocks version 0.91b (Castresana, 2000) using the following parameters: minimum length of a block after gap cleaning set to 10, no gap positions were allowed in the final alignment, all segments with contiguous nonconserved positions bigger than eight were rejected, minimum number of sequences for a flank position were set to 85%. The alignments were concatenated into five different datasets using FASconCAT-G (Kück & Longo, 2014) and the following criteria:

1 Only mitochondrial protein-coding genes (nucleotide data).
2 Only mitochondrial protein-coding genes (amino acid data).
3 All mitochondrial protein-coding and rRNA genes (nucleotide data).
4 Only nuclear genes, 18S and partial 28S (nucleotide data).
5 All mitochondrial and nuclear genes used above combined (nucleotide data).

Phylogenetic analyses

ModelFinder (Kalyaanamoorthy *et al.*, 2017) was used to find the best fitting substitution model for each of the separate gene alignments for RaxML analyses (Table S2). Phylogenetic analyses of the datasets were conducted with maximum likelihood (ML) using RAxML and ultrafast bootstrap was done in IQ-TREE multicore version 1.6.8 for Linux (Nguyen *et al.*, 2015). Each analysis was set up with 1000 initial parsimony trees, 15 trees were maintained during the ML tree search and the ultrafast bootstrap replicates were set to 1000 iterations. Each dataset was analysed with each partition (i.e. gene) allowed to have its own evolutionary rate, using the '-spp' option.

Bayesian analyses were carried out on dataset 5 using BEAST2 v2.6.5 (Bouckaert *et al.*, 2019). The protein-coding genes were partitioned by codon position (1 + 2 + 3), the mitochondrial rRNA were treated as one partition and the nuclear data as one. The substitution models were selected automatically using the add-on package bModelTest (Bouckaert & Drummond, 2017) in BEAST2 for each of the five partitions, using the 'namedExtend' set of substitution models. The analysis was run with linked trees and linked clock models, while the site models were unlinked. We applied a birth death model and selected a strict, log normal clock with the offset set to 16 million years with a soft maximum boundary, covering the Eocene. This age is based on the fossil of an *Allodia* sp. found in Miocene Ethiopian amber, which to our knowledge represent the youngest fossil of the genus (Bouju *et al.*, 2021). The MCMC chain was set to 20 million iterations and with a 10% burn-in. The log files were inspected in Tracer v1.7.2 (Rambaut *et al.*, 2018) to confirm convergence and inspect the effective sample size (ESS). The ESS values were at least 200 for all relevant parameters, with the posterior, prior and likelihood all above 1400. A maximum credibility consensus tree was obtained using TreeAnnotator v2.6.4 (Bouckaert *et al.*, 2019) with a 25% burn-in. The phylogenetic trees from the RaxML and Bayesian analyses were visualized and edited in FigTree version 1.4.3 (Rambaut, 2009) and Adobe Illustrator. All tree files are available in the Figshare database (https://doi.org/10.6084/m9.figshare.16565997).

Results and discussion

Sequencing results and assembly

The number of obtained paired-end reads for the 20 specimens varied from 20 313 314 to 42 434 194 reads (Table S3), the average read number per specimen was 31 981 741, which gave ∼4.8 gb per specimen. We were able to assemble complete mitogenomes (i.e. with all genes present) for 15 of the included species. For better readability, we refer to *Allodia* and *Brachycampta* as genera throughout this section, representing the taxa as newly revised by our work (see below).

Of the complete mitogenomes, nine were from *Allodia* s.s. [*A*. JKJ-spA, *Allodia anglofennica* Edwards, *Allodia zaitzevi* Kurina, *Allodia tuomikoskii* Hackman, *Allodia pyxidiiformis* Zaitzev, *Allodia truncata* Edwards, *Allodia lundstroemi* Lundstroemi, *Allodia lugens* (Wiedemann), *Allodia ornaticollis* (Meigen)], five from *Brachycampta* (*Brachycampta angulata* Lundstrom, *Brachycampta protenta* Laštovka and Matile, *Brachycampta foliifera* (Strobl), *Brachycampta neglecta* Edwards and *Brachycampta czernyi* (Landrock)) and one outgroup species (*Brevicornu bipartitum*). For the remaining species *Brachycampta alternans* (Zetterstedt), *Brachycampta barbata* (Lundstrom), *Allodiopsis domestica* and *Pseudexechia tuomikoskii*, the mitogenome had to be assembled in several pieces and all mitochondrial genes used for the phylogenetic inference could be obtained. Only for one outgroup species (*Cordyla* sp.), we were unable to retrieve most of the mitochondrial genes, and therefore, excluded this species from further analyses involving the mitochondrial genes. The nuclear ribosomal genes 18S (∼2017 bp) and partial 28S (∼954 bp) were successfully assembled with MITObim for all of the 20 species. The coverage of 18S and 28S were approximately four times higher than for the mitogenome assemblies. All mitogenomes and genes obtained, with GenBank accession numbers, are listed in Table S3.

Consequently, we were able to base our following analyses on a total of 17 genes, which together comprised 16 072 positions in the alignment after trimming. Of these, 3772 were parsimony informative. Hence, substantially more data could be used than in traditional Sanger sequencing studies that include relatively few loci (Rindal *et al.*, 2007; Burdíková *et al.*, 2019). Compared to genome skimming, RADseq and target-capture usually provide a more even representation of the genome, which at present mostly recovers sequence data present in high-copy numbers within the genome. However, bioinformatic methods are being developed to increase the value of low coverage genome data by retrieving more genes, including

 doi: 10.1111/syen.12529

single-copy ones, and hence increasing the representation of the genome (Tan *et al.*, 2021). In addition, it is worth pointing out that methods like target-capture or RADseq are tailored towards conservative parts of the genomes to reduce the amount of allelic dropouts due to substitutions (Cerca *et al.*, 2021; Lemmon & Lemmon, 2013) and hence also are a biased representation of the genome. As mentioned in the introduction, a strong advantage of genome skimming is that it does not require *a priori* knowledge of the investigated genomes, while target-capture or RADseq, for example, require some prior information about the genome (Lemmon & Lemmon, 2013). While this can be circumvented by using information from other species in case this is lacking for the targeted species, this has some caveats the more distantly related the other species is to the targeted species. The chance of substitutions in the target of the capture or the restriction site increases with genetic distance and hence the chance of allelic dropout (Cerca *et al.*, 2021; Lemmon & Lemmon, 2013). This results in increased level of missing data. In contrast, genome skimming, similar to whole-genome sequencing, has a much lower chance of being affected by substitutions and allelic dropout (Lemmon & Lemmon, 2013).

In our analyses, we have a low percentage of missing data [4.3% in the complete dataset (Dataset 5)] and only in one of the outgroup species (*Cordyla* sp.), we had problems with assembly of several mitochondrial genes. However, in this case, this is not related to the more distant relationship of this outgroup species to the ingroup species, but due to the poor quality of the retrieved sequence data. An additional factor, which can affect the recovery of genes in genome skimming is the genome size. Naturally, when the genome size increases one has to generate more sequencing reads to obtain the same level of coverage than for a smaller genome. However, in genome skimming each individual sample (and hence genome) is in principal targeted with the same number of reads, but as can be seen in our result the actual difference can be two-fold ranging from about 20 to 40 million reads because the samples were pooled prior to sequencing. Furthermore, when the nuclear genome size increases the proportion of the mitochondrial genome to the whole-genome becomes also smaller and accordingly the sequencing output will have less reads of the mitochondrial genome. All of these can reduce the number of genes recovered by genome skimming, but this can easily be amended by increasing the sequencing depth and hence increase the coverage, while for the other methods just sequencing more will not solve the underlying problem of allelic dropout.

Mitochondrial genomes

The gene order was identical for all species for which we obtained the complete mitogenome (see Fig. 2), except for a missing trnV between the small ribosomal RNA subunit (rrnS) and large ribosomal RNA subunit (rrnL) in *A. truncata* and *B. protenta*. The gene order found corresponds to a typical ancestral insect gene order (Cameron, 2014) and to the results in Wang *et al.* (2021) for Mycetophilidae. The mitochondrial genes are of approximately the same length in the studied species, with the largest length differences observed between in-group and outgroup species. The control region (AT-rich region) of the species was difficult to assemble due to repetitive regions; this area was therefore excluded from the mitogenome sequences and subsequent analyses.

Phylogenetic inference

Dataset 1, based on the nucleotide sequences of all the mitochondrial protein-coding genes consisted of 11 686 positions and of these 3286 were parsimony informative. Dataset 2, containing the amino acid sequences of all the mitochondrial protein-coding genes, had in total 3639 sites, of which 583 were parsimony informative. Dataset 3, containing the nucleotide sequences of all 15 mitochondrial protein-coding and rRNA genes, which comprised 13 101 positions with 3613 parsimony-informative ones. Dataset 4, of the nuclear ribosomal genes 18S and partial 28S, consisted of 2971 positions in total, of which 124 were parsimony informative. Dataset 5 included all 17 genes mentioned above with a total of 16 072 positions, of which 3772 sites were parsimony informative.

The results of the phylogenetic analyses of all five datasets support the monophyly of the two subgenera *Allodia* and *Brachycampta* as well as of the genus *Allodia* with maximal bootstrap support of 99–100 for all three clades in all RaxML analyses (Fig. 3, Figs S1–S4). In addition, in the Bayesian analysis these three clades have strong support with a posterior probability of 1. The RaxML and Bayesian analyses of dataset 5 (with all genes included) are congruent and with the exception of *A. lugens* and *A. truncata*, which switched place in the Bayesian analysis (Fig. 3, Fig S5). Within each subgenus, incongruences among the datasets occur, and several nodes, especially in *Allodia* have low branch support. Within *Brachycampta*, there are only minor incongruences between the different datasets and several phylogenetic relationships are strongly supported. In general, the resolution is better in the datasets containing mitochondrial data, with longer branches and higher bootstrap support. In dataset 4, with only nuclear data, the branches are shorter and the internal topology of both *Allodia* and *Brachycampta* is largely unresolved. The topological conflict between the analyses of the mitochondrial data (Figs S1–S3) and the nuclear ribosomal data (Fig S4) is therefore most likely due to differences in resolution, leading to branches with low statistical support.

The resolution of the mitochondrial data in our analyses suggests that it is applicable also at the genus and to a certain degree at the species level. Furthermore, by combining and comparing the mitochondrial data with the nuclear markers, as discussed in Caravas & Friedrich (2013), we could independently evaluate the results given by the mitogenome analyses alone. Our results also indicate that the phylogenetic signal of the nuclear ribosomal genes seems well suited for genus-level relationships, but these genes did not give adequate resolution to resolve between-species relationships (i.e. with regard to branch lengths or support). To resolve the species-level relationships, we would

 doi: 10.1111/syen.12529

Fig. 2. A diagram showing the mitochondrial gene order of *Allodia* JKJ-spA. The inner-circle represents the light strand, and the outer circle the heavy strand. The transferRNA genes are indicated by the abbreviation letter of the amino acid. The gene sizes are approximately proportional to the nucleotide length of the genes. Abbreviations: ATP synthase subunit 6 (=atp6) and 8 (=atp8); control region (=CR); cytochrome oxidase b (=cob); cytochrome oxidase c subunit 1 (=cox1), 2 (=cox2) and 3 (=cox3); NADH dehydrogenase subunit 1 (=nad1), 2, (=nad2), 3 (=nad3), 4 (=nad4), 4L (=nad4L), 5 (=nad5) and 6 (=nad6); small ribosomal RNA subunit (rrnS) and large ribosomal RNA subunit (rrnL).

need to include data from more species and also include samples from other biogeographic regions.

Our study aimed to solve the phylogenetic relationship between the two subgenera, and not the intraspecific relationships. Accordingly, we will only briefly discuss the internal topology of the two clades in the following two sections.

The internal topology of Brachycampta

The resolution within *Brachycampta* is higher than what is observed within *Allodia* and the phylogenetic relationships are largely supported (Fig. 3) also in the other datasets, with some exceptions. Likewise, the analysis of the amino acid data results in a much better resolution (i.e. longer branches) in *Brachycampta* compared to *Allodia*, where the topology has no support (Fig S2). Several species pairs in *Brachycampta* exhibit strong affinity in the analyses, and are revealed as sister groups with high statistical support in all but the nuclear dataset (Fig S4). These species pairs also correspondingly show a high degree of morphological similarity in the shape and structure of the male gonostyle. The clade consisting of *B. protenta*, *B. foliifera*, *B. angulata* and *B. barbata* is recovered as monophyletic in the analyses of all datasets, although in dataset 4, the branches are very short, and without support (Fig S4). Within the clade, the species pairs of *B. protenta* and *B. foliifera*, and *B. angulata* and *B. barbata* are recovered in all analyses, except in dataset 4. These relationships are supported by morphology. *Brachycampta protenta* and *B. foliifera* both have a very large median lobe of the gonostyle, while the dorsal lobe only forms a very narrow comb-like structure (Fig. 4). *Brachycampta angulata* and *B. barbata* both have highly reduced and slender dorsal and

 doi: 10.1111/syen.12529

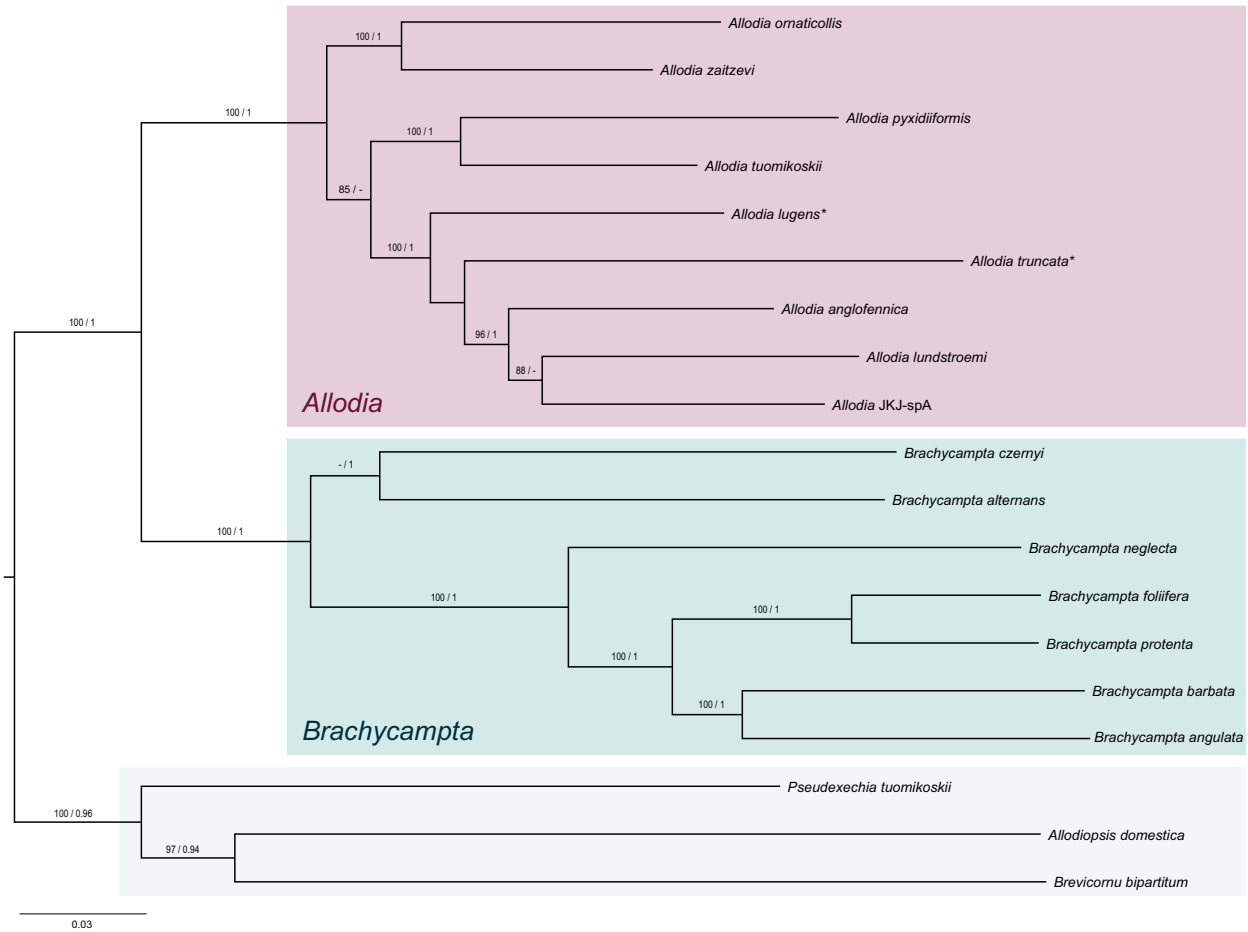

Fig. 3. Maximum likelihood analysis of dataset 5; all genes included. Brach support refers to Bootstrap support values above 85, together with the posterior probabilities above 0.90 from the Bayesian analysis of the dataset. The topology of the Maximum likelihood analysis and the Bayesian analysis of the dataset is congruent with the exception of the placement of *Allodia lugens* (Wiedemann) and *Allodia truncata* Edwards, indicated by an asterisk.

median lobes of the gonostyle (Fig. 4). Moreover, *B. neglecta* has a stable position as a sister to the latter clade *B. protenta*, *B. foliifera*, *B. angulata* and *B. barbata* with strong nodal support (>89) in all analyses. The two species *B. alternans* and *B. czernyi* are either placed as sister species to each other (Figs S1, S3 and Fig. 3) or as distinct nodes (Figs S2 and S4) branching out basally within *Brachycampta*. The three species *B. alternans*, *B. neglecta* and, in particular, *B. czernyi* are morphologically more similar to *Allodia* than any other of the included *Brachycampta* species, as they all have simple outlines of the male gonostyli.

The internal topology of Allodia

In contrast to *Brachycampta*, the relationships in subgenus *Allodia* s.s. vary between datasets, and the support of the different clades is correspondingly low. In fact, none of the internal nodes in *Allodia* are present in all the analyses of the different datasets. A similar pattern is found in the morphological characters, probably caused by the more restrained outline of the composite male terminalia (see Fig. 4). The relationships between the species are completely incongruent between the nuclear and mitochondrial data (Figs S1–S4), but also among the different analyses of the mitochondrial data. Moreover, the inclusion of mitochondrial rRNA in the analyses (Fig S3), has a large influence on the topology within *Allodia*. The differences occur, in particular, with respect to the positions of *A. truncata*, *A. lugens*, *A. tuomikoskii* and *A. pyxidiiformis*. As already mentioned, *A. truncata* and *A. lugens* also have a different placement in the Bayesian analysis of the dataset 5, compared to the RaxML topology (Fig S5). The group consisting of *A. anglofennica*, *A.* JKJ-spA, *A. lundstroemi* and *A. truncata* shows up in datasets 1 and 3 (Figs S1 and S3), with relatively high statistical support (i.e. bootstrap > 90). The position of *A. lugens* is very unstable, either it groups together with *A. anglofennica*, *A.* JKJ-spA, *A. lundstroemi* and *A. truncata* (Figs S1, S2 and S4), or together with *A. tuomikoskii* and *A. pyxidiiformis*, but neither of these positions have strong bootstrap support. There is also a tendency for *A. zaitzevi* and *A. ornaticollis* to group

 doi: 10.1111/syen.12529

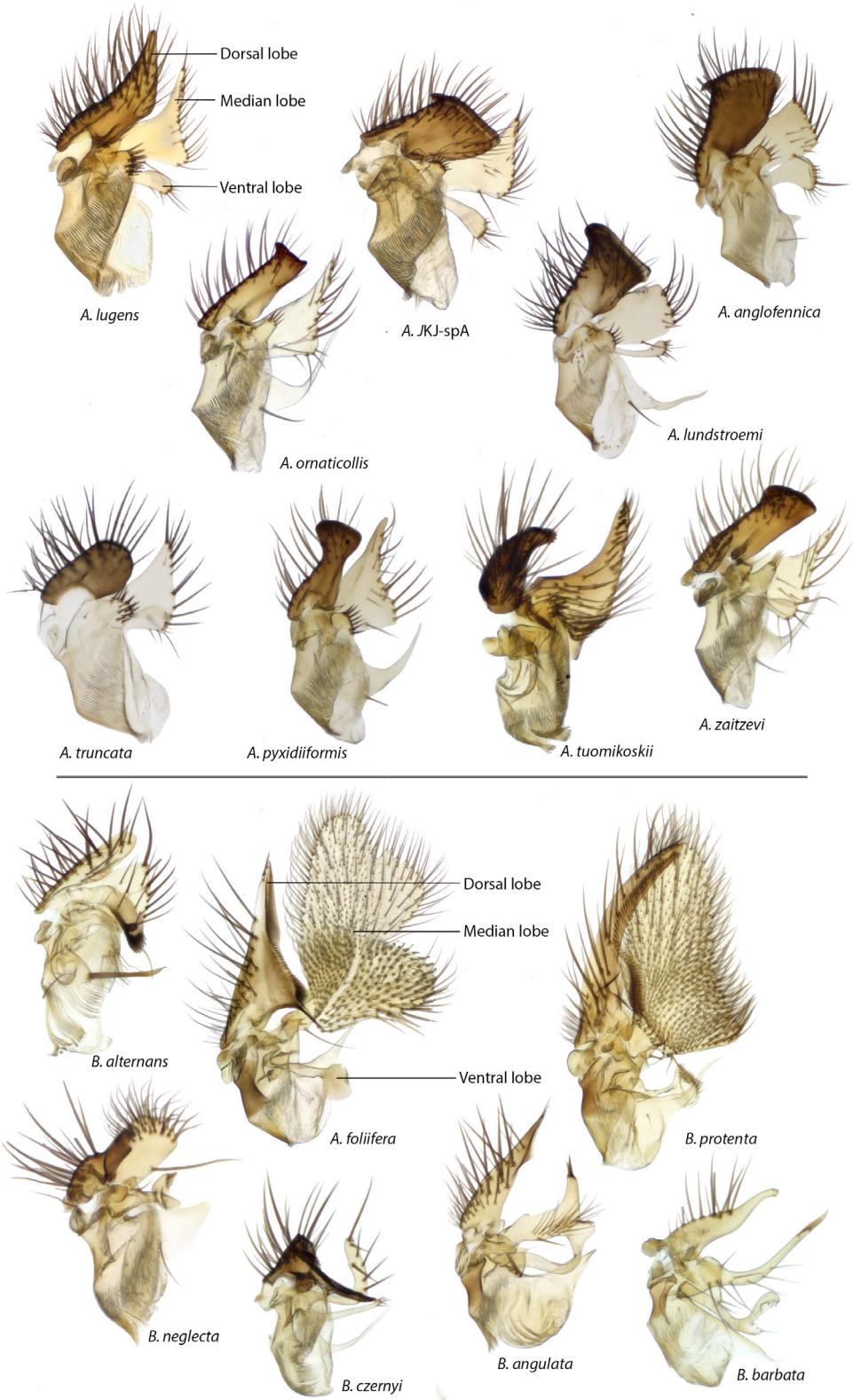

Fig. 4. The male gonostylus from the inner side of all the species included in the study.

 doi: 10.1111/syen.12529

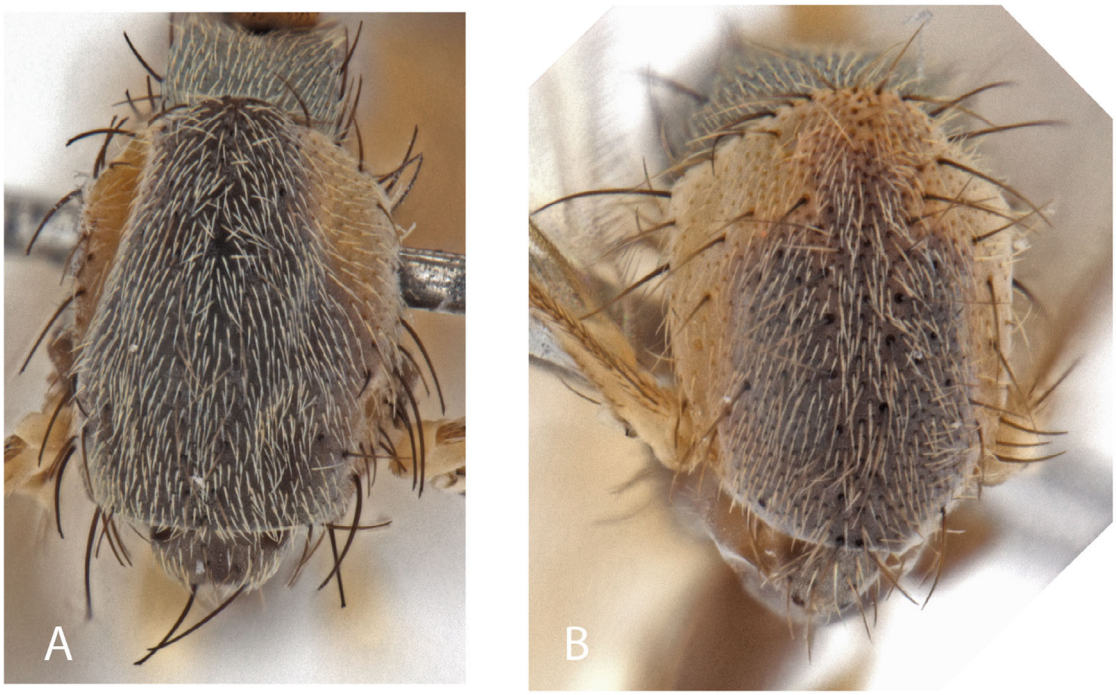

Fig. 5. Scutum, dorsal view, showing the position of bristles in *Allodia* and *Brachycampta*. (A) *Allodia pyxidiiformis* Zaitzev, 1983; (B) *Brachycampta alternans* (Zetterstedt, 1838).

together (Figs S1, S2 and Fig. 3), but with variable statistical support. Morphologically these species are not very similar, and *A. zaitzevi* has much more in common with *A. pyxidiiformis*, as also pointed out in the original description by Kurina (1997). In the analysis of dataset 4, with 18S and 28S, these two species do show a strong phylogenetic relationship, together with *A. tuomikoskii* (Fig S4).

Morphological characters

The original characters used by Winnertz (1864) to describe *Allodia* and *Brachycampta* as separate genera were the outline of the anal vein and the position of the posterior fork in relation to the base of the R-M crossvein. Nevertheless, as already commented on by Johannsen (1909), a large degree of variation exists. At the time of Winnertz (1864), the notion of the genera included in the tribe Exechiini was very different from today, and most species described at that time have later been transferred to other genera. It is illustrative that of the 13 species originally included in *Allodia* and *Brachycampta* by Winnertz (1864), 11 have been ascribed to other genera.

Edwards (1925), in his concept of *Allodia*, also included *Brevicornu* and *Cordyla*, and mainly used the chaetotaxy to separate between the different species groups. Edwards was the first to use the presence/absence of discal bristles at the anterior part of scutum to separate the *lugens*- and *grata*-group (roughly corresponding to the subgenera *Allodia* and *Brachycampta*, respectively). We have found this particular character to be important to separate the two genera, due to its overall consistency in both groups. Edwards also illustrated the genital structures of species in *Allodia* s.l., and used the characters of the male genitalia and coloration to identify the species. The convergent pattern in the outline of the genitalia within each of the two groups, however, has not been used as diagnostic character. The clearest differences can be seen in the individual lobes of the male gonostyle (see Fig. 4), but differences can also be found in the shape of the genital capsule and the hypandrial lobe, particularly in *Brachycampta*. In *Brachycampta*, with large interspecific variation, it may be difficult to see a common pattern (Fig. 4), while in *Allodia*, such a pattern is much more evident (Fig. 4).

A careful study of the microstructure of the large bristles on the notum, revealed an interesting and seemingly consistent variation between the two subgenera (Fig. 5). In *Allodia*, these bristles have an apical split, one tip longer than the other (Fig. 6B–D); in *Brachycampta*, the bristles are pointy (Fig. 6E, F), more or less clear-cut, or with somewhat splintery tips (Fig. 6G). The microstructure of the two long bristles on the scutellum display the same variation (Fig. 6B, E). This character has been confirmed for all 12 *Allodia* and 15 *Brachycampta* species studied: *A. anglofennica*, *Allodia embla* Hackman, 1971, *A.* JKJ-spA, *Allodia laccariae* Sasakawa and Ishizaki, 2003, *A. lugens*, *A. lundstroemi*, *A. ornaticollis*, *A. pyxidiiformis*, *Allodia septentrionalis* Hackman, 1971, *A. truncata*, *A. tuomikoskii*, *A. zaitzevi*, *Brachycampta adunca* Zaitzev, 1992, *B. alternans*, *B. angulata*, *B. barbata*, *B. czernyi*, *B. foliifera*, *Brachycampta grata* (Meigen, 1830), *Brachycampta huggerti* Kjærandsen, 2007, *B. neglecta*, *Brachycampta penicillata* (Lundstrom,

 doi: 10.1111/syen.12529

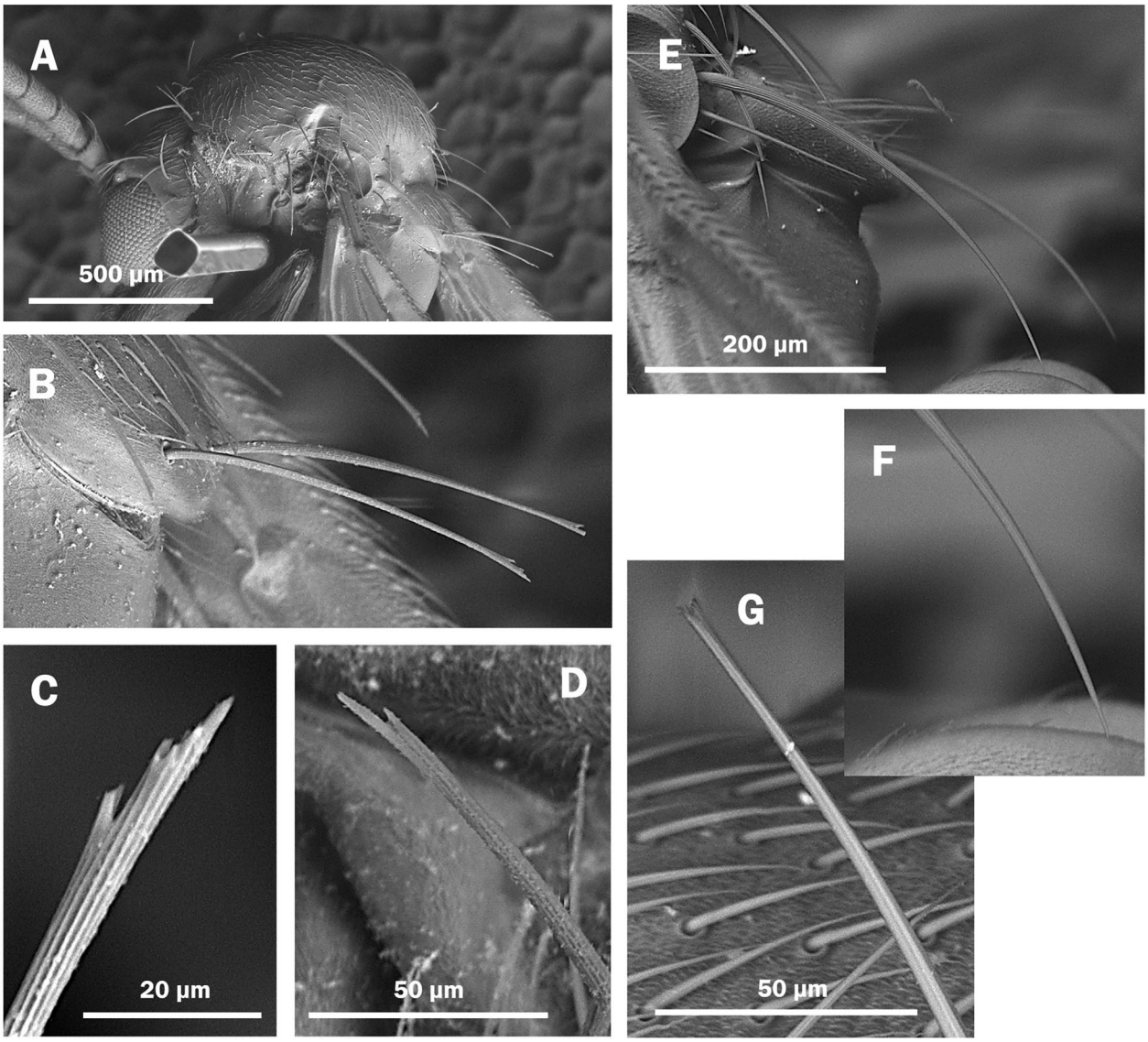

Fig. 6. Microstructure of bristles on scutum. Scanning electron microscope images of setae on scutum and scutellum of in the two genera *Allodia* and *Brachycampta*. (A) Notum of *Allodia pyxidiiformis* Zaitzev, 1983; (B) Scutellum of *Allodia pyxidiiformis*; (C) Detail, scutellar bristles of *Allodia pyxidiiformis*; (D) Scutellar bristles of *Allodia lugens* (Wiedemann, 1817); (E) Scutellum of *Brachycampta foliifera* (Strobl, 1910); (F) Detail scutellar bristle of *Brachycampta foliifera*; (G) Tip of setae in *Brachycampta adunca* Zaitzev, 1992.

1912), *Brachycampta pistillata* (Lundstrom, 1911), *B. protenta*, *Brachycampta rindeni* Kjærandsen, 2007, *Brachycampta subspillata* Ševčík, 1999 and *Brachycampta triangularis* (Strobl, 1895).

Taxonomy

Based on the results of our molecular phylogenetic analyses, in combination with the evaluation and re-examination of morphological characters, we find it justified to raise the subgenus *Brachycampta* to the genus level and consequently redefine *Allodia* to comprise the current species of the subgenus *Allodia s.s*. only.

***Allodia* Winnertz, 1864**

(Figs 5A, 6B–D)

Allodia Winnertz, 1864: 826.

Type species: *Mycetophila ornaticollis* Winnertz, 1864, by designation of Johannsen, 1909: 104. [Misidentification = *lugens* Wiedemann (1817)].

= *Parallodia* Plassmann, 1969

Amended diagnosis. The genus differs from *Brachycampta* in having the anterior half of scutum devoid of strong discal bristles, or with very short bristles only; smaller flat-lying setae may

 doi: 10.1111/syen.12529

be present (Fig. 5A). Furthermore, the bristles on scutum and scutellum have an apical split, one tip distinctly longer than the other (Fig. 6B–D). The male terminalia of the genus are rather homogenous and the gonostyli are uniformly shaped. Typical for the gonostylus is a prominent, heavily sclerotized dorsal lobe, a more flattened, often sub-triangular median lobe and a club-shaped ventral lobe (Fig. 4). The hypandrial lobe of the species in this genus is also rather uniformly shaped, elongated and sclerotized. Additional characters, helpful for identification, but not completely reliable, are the pale abdominal markings, which, when present, are broader towards the hind margins of the tergites; moreover, the base of the posterior fork is usually located after or opposite the base of the R-M crossvein.

***Brachycampta* Winnertz, 1864 stat. rev.**
(Figs 5B, 6E–G)
Brachycampta Winnertz, 1864: 833.

Type species: *Mycetophila alternans*, Winnertz, 1864, by designation of Coquillett, 1910: 515. [Misidentification = *grata* (Meigen, 1830)] (Further details about any possible uncertainties on the designated type species is given by Tuomikoski (1966)).

Diagnosis. The genus differs from *Allodia* by the presence of well-developed discal bristles on the anterior part of scutum, in addition to smaller flat-lying setae, the bristles can be arranged in two or three rows (Fig. 5B). The apices of the bristles on scutum and scutellum are pointed (Fig. 6E, F), or somewhat splintery (Fig. 6G), but never with splits of different lengths. The male terminalia of the genus is highly variable, with diversely shaped gonostyli, often with a less intricate dorsal lobe and more elaborate medial and ventral lobes (Fig. 4). The hypandrial lobe is prominent and has a complex outline that often varies between species and can be of diagnostic importance. Additional characters, helpful for identification, but not completely reliable are the pale abdominal markings, which, when present are broader towards the anterior margin of the tergites; moreover, the base of the posterior fork is usually located before or opposite the base of the R-M crossvein.

Conclusion

In summary, genome skimming was successfully applied to retrieve molecular markers to construct the phylogeny of a Mycetophilid genus. The methodology used is promising for resolving deeper relationships within the tribe Exechiini, which to date has been challenging due to the young age of the radiation. The genome skimming approach also seems promising for studies focusing on resolving problems at shallow phylogenetic levels in Diptera in general, and we believe that the method is a transition from Sanger sequencing to HTS that is within reach also for scientists without strong background in bioinformatics.

As accounted for, the perception of *Allodia* and *Brachycampta* has varied through time, originally described as two separate genera (Winnertz, 1864), but as a consequence of inadequacies in their suggested diagnoses Johannsen (1911) suggested to treat them as subgenera of *Allodia* s.s. This practice was followed by most subsequent authors, including Tuomikoski (1966) in his highly acknowledged and foresighted analysis of the Exechiini, above all because of the lack of good synapomorphies to separate the two taxa. Despite the absence of good diagnostic morphological characters to separate the two subgenera, most workers have been conscious of distinct dissimilarities in the outline of their male genitalia, as well as in different behaviour and biology. Nevertheless, despite strong indications for the contrary (Kjærandsen, 2007), the two have still been treated as one genus by most workers. As the results of our phylogenetic analyses strongly support the monophyly of both *Allodia* and *Brachycampta*, and moreover, we are able to present morphological characters to diagnose and separate them, we find it justified to treat *Allodia* and *Brachycampta* as separate genera. Furthermore, we argue that distinguishing two genera will make it easier to recognize their diversity and facilitate future taxonomic studies.

Supporting Information

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

Figure S1. Maximum likelihood analysis of Dataset 1, mitochondrial protein coding genes (nucleotide data). Bootstrap support values above 85 are listed next to the branches.

Figure S2. Maximum likelihood analysis of Dataset 2, mitochondrial protein coding genes (amino acid data). Bootstrap support values above 85 are listed next to the branches.

Figure S3. Maximum likelihood analysis of Dataset 3, all mitochondrial protein-coding and rRNA genes (nucleotide data). Bootstrap support values above 85 are listed next to the branches.

Figure S4. Maximum likelihood analysis of Dataset 4, only nyclear genes, 18S and partial 28S (nucleotide data). Bootstrap support values above 85 are listed next to the branches.

Figure S5. Bayesian analysis of Dataset 5, mitochondrial and nuclear genes (nucleotide data). Posterior probabilities above 0.95 are listed next to the branches.

Table S1. Full list of specimens with DNA extracts, id. information, DNA concentration, storage medium and collection data. The sequenced specimens highlighted. Abbreviation: NHMO = Natural History Museum, University of Oslo.

Table S2. Alignment information and substitution models. Substitution models listed for each gene alignment estimated with ModelFinder (Kalyaanamoorthy *et al.*, 2017) in IQ-TREE (Nguyen *et al.*, 2015). The best fit model is found for nucleotide- and amino acid data according to the Bayesian Information Criterion (BIC) score. The information in parentheses represents the number of

 doi: 10.1111/syen.12529

positions before Gblocks version 0.91b (Castresana, 2000). Abbreviations: ASC = ascertainment bias correction model (Lewis 2001), F = Empirical base frequencies, GTR = General time reversible (Tavare 1986), G = discrete Gamma model (Yang 1994), I = allowing for a proportion of invariable sites, K3Pu = three substitution types model and unequal base freq (Kimura 1981), mtART = Mitochondrial Arthropoda (Abascal et al. 2007), mtMAM = Mitochondrial Mammalia (Yang et al. 1998), mtInv = Mitochondrial Inverterbrate (Vinh et al. 2017), mtZOA = Mitochondrial Metazoa (Rota-Stabelli et al. 2009), R = FreeRate model (Yang 1995; Soubrier et al. 2012), TIM = Transition model, AC = GT, AT = CG and unequal base freq, TIM2 = AC = AT, CG = GT and unequal base freq, TIM3 = AC=CG, AT = GT and unequal base freq, TN = Unequal transition/transversion rates and unequal purine/pyrimidine rates (Tamura and Nei 1993). TPM3u = AC=CG, AG = CT, AT = GT and unequal base freq, TVM = Transversion model, AG = CT and unequal base freq.

Table S3. Library statistics and assembly information. The GenBank accession numbers are listed for each of the mitogenomes/parts of the mitogenomes, 18S and 28S.

Appendix S1. References_Supplementary.

Acknowledgements

T.M. is grateful to José Cerca and Lutz Bachmann for invaluable help with the bioinformatics, and to Michael Matschiner and Jon Peder Lindemann for valuable advice and discussions about the analyses. We also want to thank the staff at the DNA lab and Geo Lab at the Natural History Museum, University of Oslo, the help is highly appreciated. A special thanks to Nélia Castro for help with the SEM pictures. The valuable comments and suggestions from the reviewers Emma Wahlberg, Peter Kerr and one anonymous reviewer greatly improved the manuscript. The sequencing service was provided by the NSC (www.sequencing.uio.no), a national technology platform hosted by the University of Oslo and supported by the 'Functional Genomics' and 'Infrastructure' programs of the Research Council of Norway and the Southeastern Regional Health Authorities. We acknowledge the Norwegian national e-infrastructure for high-performance computing and storage via the project NN9605K. Most of the specimens included in this study were collected as part of the two projects; 'Soppmygg i nordøstlige skogområder av Norge' (grant number 45-14) and 'Insekter på rikmyrer i Hedmark' (grant number 70184235) financed by the Norwegian Species Information Centre. The authors declare no conflict of interest.

Data availability statement

The data obtained and analysed in this project is available in genbank in the figshare database: https://doi.org/10.6084/m9.figshare.16565994.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

Bankevich, A., Nurk, S., Antipov, D. *et al.* (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, **19**, 455–477. https://doi.org/10.1089/cmb.2012.0021.

Beckenbach, A.T. (2012) Mitochondrial genome sequences of Nematocera (lower Diptera): evidence of rearrangement following a complete genome duplication in a winter crane fly. *Genome Biology and Evolution*, **4**, 89–101. https://doi.org/10.1093/gbe/evr131.

Bernt, M., Donath, A., Jühling, F. *et al.* (2013) MITOS: improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution*, **69**, 313–319.

Bouckaert, R.R. & Drummond, A.J. (2017) bModelTest: Bayesian phylogenetic site model averaging and model comparison. *BMC Evolutionary Biology*, **17**(42), 1–11. https://doi.org/10.1186/s12862-017-0890-6.

Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A. *et al.* (2019) BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, **15**, e1006650. https://doi.org/10.1371/journal.pcbi.1006650.

Bouju, V., Rosse-Guillevic, S., Griffon, M., Bojarski, B., Szwedo, J. & Perrichot, V. (2021) The genus Allodia (Diptera: Mycetophilidae) in Miocene Ethiopian amber. *Fossil Record*, **24**, 339–346. https://doi.org/10.5194/fr-24-339-2021

Burdíková, N., Kjærandsen, J., Lindemann, J.P., Kaspřák, D., Tóthová, A. & Ševčík, J. (2019) Molecular phylogeny of the Paleogene fungus gnat tribe Exechiini (Diptera: Mycetophilidae) revisited: monophyly of genera established and rapid radiation confirmed. *Journal of Zoological Systematics and Evolutionary Research*, **57**(4), 806–821. https://doi.org/10.1111/jzs.12287.

Cameron, S.L. (2014) Insect mitochondrial genomics: implications for evolution and phylogeny. *Annual Review of Entomology*, **59**, 95–117. https://doi.org/10.1146/annurev-ento-011613-162007.

Cameron, S.L., Lambkin, C.L., Barker, S.C. & Whiting, M.F. (2006) A mitochondrial genome phylogeny of Diptera: whole genome sequence data accurately resolve relationships over broad timescales with high precision. *Systematic Entomology*, **32**, 40–59. https://doi.org/10.1111/j.1365-3113.2006.00355.x.

Camier, M. & Nel, A. (2020) The oldest fungus gnat of the tribe Exechiini in the lowermost Eocene Oise amber (Diptera: Mycetophilidae). *Zootaxa*, **4722**, 91–98 10.11646/zootaxa.4722.1.9.

Caravas, J. & Friedrich, M. (2013) Shaking the Diptera tree of life: performance analysis of nuclear and mitochondrial sequence data partitions. *Systematic Entomology*, **38**, 93–103. https://doi.org/10.1111/j.1365-3113.2012.00657.x.

Castresana, J. (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, **17**, 540–552. https://doi.org/10.1093/oxfordjournals.molbev.a026334.

Cerca, J., Maurstad, M.F., Rochette, N.C., Rivera-Colón, A.G., Rayamajhi, N., Catchen, J.M. & Struck, T.H. (2021) Removing the bad apples: a simple bioinformatic method to improve loci-recovery in de novo RADseq data for non-model organisms. *Methods in Ecology and Evolution*, **12**, 805–817. https://doi.org/10.1111/2041-210X.13562.

Chandler, P.J. (1993) New rearing records of fungus gnats (Diptera: Mycetophilidae and allied families). *Dipterists Digest (1st Series)*, **13**, 29–35.

 doi: 10.1111/syen.12529

Coissac, E., Hollingsworth, P.M., Lavergne, S. & Taberlet, P. (2016) From barcodes to genomes: extending the concept of DNA barcoding. *Molecular Ecology*, **25**, 1423–1428. https://doi.org/10.1111/mec.13549.

Colless, D.H. (1966) Diptera: Mycetophilidae – insects of Micronesia. *Bernice P Bishop Museum, Honolulu*, **12**, 637–667.

Coquillett, D.W. (1911) The type-species of the North American genera of Diptera. *Proceedings U. S. National Museum*, **37**, 499–647.

Edwards, F.W. (1925) British fungus-gnats (Diptera, Mycetophilidae) with a revised generic classification of the family. *Transactions of the Royal Entomological Society of London*, **72**, 505–670. https://doi.org/10.1111/j.1365-2311.1925.tb03369.x.

Edwards, F.W. (1928) Diptera Nematocera from the federated Malay states museum. *Journal of the Federated Malay States Museum*, **14**, 1–139.

Grandjean, F., Tan, M.H., Gan, H.M. *et al.* (2017) Rapid recovery of nuclear and mitochondrial genes by genome skimming from Northern Hemisphere freshwater crayfish. *Zoologica Scripta*, **46**, 718–728. https://doi.org/10.1111/zsc.12247.

Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. (2013) QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, **29**, 1072–1075. https://doi.org/10.1093/bioinformatics/btt086.

Hahn, C., Bachmann, L. & Chevreux, B. (2012) Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads – a baiting and iterative mapping approach. *Nucleic Acids Research*, **41**, e129. https://doi.org/10.1093/nar/gkt371.

Jakovlev, J. (2011) Fungus gnats (Diptera: Sciaroidea) associated with dead wood and wood growing fungi: new rearing data from Finland and Russian Karelia and general analysis of known larval microhabitats in Europe. *Entomologica Fennica*, **22**, 157–189.

Jakovlev, J. (2012) Fungal hosts of mycetophilids (Diptera: Sciaroidea excluding Sciaridae): a review. *Mycology*, **3**, 11–23. https://doi.org/10.1080/21501203.2012.662533.

Johannsen, O.A. (1909) Diptera, fam. Mycetophilidae. *Genera Insectorum* (ed. by P. Wytsman), pp. 1–141. Brussels.

Johannsen, O.A. (1911) The Mycetophilidae of North America. Part III, The Mycetophilinae. *Maine Agricultural Experiment Station Bulletin*, **196**, 249–327. https://doi.org/10.5962/bhl.title.86513

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., Von Haeseler, A. & Jermiin, L.S. (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, **14**, 587–589.

Kaspřák, D., Kerr, P., Sýkora, V., Tóthová, A. & Ševčík, J. (2019) Molecular phylogeny of the fungus gnat subfamilies Gnoristinae and Mycomyinae, and their position within Mycetophilidae (Diptera). *Systematic Entomology*, **44**, 128–138. https://doi.org/10.1111/syen.12312.

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. https://doi.org/10.1093/molbev/mst010.

Kjærandsen, K. (2006) Review of fungus gnats of the genus *Tarnania* Tuomikoski, with a phylogeny of the *Rymosia* s.l. genus group (Diptera: Mycetophilidae). *Insect Systematics and Evolution*, **37**, 121–148.

Kjærandsen, J. (2007) Two new species of *Allodia* subgenus *Brachycampta* Winnertz from Norway and Sweden (Diptera: Mycetophilidae). *Entomologica Fennica*, **18**, 17–23.

Knyshov, A., Gordon, E.R.L. & Weirauch, C. (2021) New alignment-based sequence extraction software (AliBaSeq) and its utility for deep level phylogenetics. *PeerJ*, **9**, e11019. https://doi.org/10.7717/peerj.11019.

Krueger, F. (2013). *Trim Galore!: A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files*. V.0.3.3 URL https://github.com/FelixKrueger/TrimGalore [accessed on XXX].

Kück, P. & Longo, G.C. (2014) FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Frontiers in Zoology*, **11**, 81. https://doi.org/10.1186/s12983-014-0081-x.

Kurina, O. (1997) A new species of fungus gnats of the genus *Allodia* Winnertz, 1863 (Diptera, Mycetophilidae) from Estonia. *Studia Dipterologica*, **4**, 275–279.

Lagesen, K., Hallin, P., Rødland, E.E., Staerfeldt, H.H., Rognes, T. & Ussery, D.W. (2007) RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research*, **35**, 3100–3108. https://doi.org/10.1093/nar/gkm160.

Lemmon, E.M. & Lemmon, A.R. (2013) High-throughput genomic data in systematics and phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, **44**, 99–121. https://doi.org/10.1146/annurev-ecolsys-110512-135822.

Linard, B., Arribas, P., Andújar, C., Crampton-Platt, A. & Vogler, A.P. (2016) Lessons from genome skimming of arthropod-preserving ethanol. *Molecular Ecology Resources*, **16**, 1365–1377. https://doi.org/10.1111/1755-0998.12539.

Magnussen, T., Kjærandsen, J., Johnsen, A. & Søli, G.E.E. (2018) Six new species of Afrotropical Allodia (Diptera: Mycetophilidae): DNA barcodes indicate recent diversification with a single origin. *Zootaxa*, **4407**, 301–320 10.11646/zootaxa.4407.3.1.

Magnussen, T., Søli, G.E.E. & Kjærandsen, J. (2019) *Allodia* Winnertz from the Himalayas, with nine species new to science (Diptera, Mycetophilidae). *ZooKeys*, **820**, 119–138. https://doi.org/10.3897/zookeys.820.31618.

Matile, L. (1978) Diptères Mycetophilidae de l'Archipel des Comores. *Mémoires du Museum national d'Histoire naturelle, Sèrie A, Zoologie*, **109**, 247–306.

Milne, I., Stephen, G., Bayer, M. *et al.* (2013) Using Tablet for visual exploration of second-generation sequencing data. *Briefings in Bioinformatics*, **14**, 193–202. https://doi.org/10.1093/bib/bbs012.

Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C. *et al.* (2014) Phylogenomics resolves the timing and pattern of insect evolution. *Science*, **346**, 763–767. https://doi.org/10.1126/science.1257570.

Nguyen, L.T., Schmidt, H.A., Von Haeseler, A. & Minh, B.Q. (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Molecular Biology and Evolution*, **32**, 268–274. https://doi.org/10.1093/molbev/msu300.

Nikolenko, S., Korobeynikov, A. & Alekseyev, M.A. (2013) BayesHammer: Bayesian clustering for error correction in single-cell sequencing. *BMC Genomics*, **14**(Suppl. 1), S7. https://doi.org/10.1186/1471-2164-14-S1-S7.

Papadopoulou, A., Taberlet, P. & Zinger, L. (2015) Metagenome skimming for phylogenetic community ecology: a new era in biodiversity research. *Molecular Ecology*, **24**, 3515–3517. https://doi.org/10.1111/mec.13263.

Peters, R.S., Krogmann, L., Mayer, C., Donath, A., Gunkel, S., Meusemann, K. *et al.* (2017) Evolutionary history of the hymenoptera. *Current Biology*, **27**, 1013–1018. https://doi.org/10.1016/j.cub.2017.01.027.

Rambaut, A. (2009). *FigTree version 1.4.3*. URL http://tree.bio.ed.ac.uk/software/figtree/ [accessed on XXX].

Rambaut, A., Drummond, A.J., Xie, D., Baele, G. & Suchard, M.A. (2018) Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, **67**, 901–904. https://doi.org/10.1093/sysbio/syy032.

Ratnasingham, S. & Hebert, P.D.N. (2007) BOLD: the barcode of life Datasystem (http://www.barcodinglife.org). *Molecular*

Ecology Notes, **7**, 355–364. https://doi.org/10.1111/j.1471-8286.2007.01678.x.

Ren, Z., Harris, A.J., Dikow, R.B., Ma, E., Zhong, Y. & Wen, J. (2017) Another look at the phylogenetic relationships and intercontinental biogeography of eastern Asian – North American Rhus gall aphids (Hemiptera: Aphididae: Eriosomatinae): evidence from mitogenome sequences via genome skimming. *Molecular Phylogenetics and Evolution*, **117**, 102–110. https://doi.org/10.1016/j.ympev.2017.05.017.

Rindal, E. & Søli, G.E.E. (2006) Phylogeny of the subfamily Mycetophilinae (Diptera: Mycetophilidae). *Zootaxa*, **1302**, 43–59 10.11646/zootaxa.1302.1.4.

Rindal, E., Søli, G.E.E., Kjærandsen, J. & Bachmann, L. (2007) Molecular phylogeny of the fungus gnat tribe Exechiini (Mycetophilidae, Diptera). *Zoologica Scripta*, **36**, 327–335. https://doi.org/10.1111/j.1463-6409.2007.00285.x.

Rindal, R., Søli, G.E.E. & Bachmann, L. (2009) On the systematics of the fungus gnat subfamily Mycetophilinae (Diptera): a combined morphological and molecular approach. *Journal of Zoological Systematics and Evolutionary Research*, **47**, 227–233. https://doi.org/10.1111/j.1439-0469.2008.00498.x.

Senior-White, R.A. (1922) New Ceylon Diptera (part II). *Spolia Zeylanica*, **12**, 195–206. https://doi.org/10.5962/bhl.part.12654.

Ševčík, J. (2010) *Czech and Slovak Diptera Associated with Fungi*, p. 112. Slezské zemské muzeum, Opava.

Ševčík, J. & Kjærandsen, J. (2012) *Brachyradia*, a new genus of the tribe Exechiini (Diptera: Mycetophilidae) from the oriental and Australasian regions. *The Raffles Bulletin of Zoology*, **60**, 117–127.

Søli, G.E.E., Vockeroth, J.R. & Matile, L. (2000) Families of Sciaroidea. Papp, L. & Darvas, B., Contributions to a Manual of Palaearctic Diptera, Appendix. Budapest: Science Herald, 49–92.

Straub, S.C., Parks, M., Weitemier, K., Fishbein, M., Cronn, R.C. & Liston, A. (2012) Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. *American Journal of Botany*, **99**, 349–364. https://doi.org/10.3732/ajb.1100335.

Tan, M.H., Gan, H.M., Bracken-Grissom, H., Chan, T., Grandjean, F. & Austin, C.M. (2021) More from less: genome skimming for nuclear markers for animal phylogenomics, a case study using decapod crustaceans. *Journal of Crustacean Biology*, **41**, ruab009. https://doi.org/10.1093/jcbiol/ruab009.

Trevisan, B., Alcantara, D.M.C., Machado, D.J., Marques, F.P.L. & Lahr, D.J.G. (2019) Genome skimming is a low-cost and robust strategy to assemble complete mitochondrial genomes from ethanol preserved specimens in biodiversity studies. *PeerJ*, **7**, e7543. https://doi.org/10.7717/peerj.7543.

Tuomikoski, R. (1966) Generic taxonomy of the Exechiini (Diptera, Mycetophilidae). *Annales Entomologici Fennici*, **32**, 159–194.

Wang, Q., Huang, J. & Wu, H. (2021) Mitogenomes provide insights into the phylogeny of Mycetophilidae (Diptera: Sciaroidea). *Gene*, **783**, 145564. https://doi.org/10.1016/j.gene.2021.145564.

Winnertz, J. (1864) *Beitrag zu einer monographie der pilzmücken*, p. 332. Braumüller, Wien. https://doi.org/10.5962/bhl.title.9961.

Young, A.D., Lemmon, A.R., Skevington, J.H., Mengual, X., Ståhls, G., Reemer, M. *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. https://doi.org/10.1186/s12862-016-0714-0.

Zaitzev, A. (1983) A review of Holarctic species of the subgenus *Allodia* s. str. (Diptera, Mycetophilidae). *Zoologicheskiĭ Zhurnal*, **62**, 1915–1920.

Zaitzev, A. (1984) A review of species of the subgenus *Brachycampta* (Diptera, Mycetophilidae) of the Holarctic fauna. *Zoologicheskiĭ Zhurnal*, **63**, 1504–1515.

Zhang, X., Kang, Z., Mao, M. *et al.* (2016) Comparative Mt genomics of the Tipuloidea (Diptera: Nematocera: Tipulomorpha) and its implications for the phylogeny of the Tipulomorpha. *PLoS One*, **11**, e0158167. https://doi.org/10.1371/journal.pone.0158167.

Zhang, X., Kang, Z., Ding, S., Wang, Y., Borkent, C., Saigusa, T. & Yang, D. (2019) Mitochondrial genomes provide insights into the phylogeny of Culicomorpha (Insecta: Diptera). *International Journal of Molecular Sciences*, **20**, 747. https://doi.org/10.3390/ijms20030747.

Accepted 18 October 2021