Systematic Entomology

Systematic Entomology (2018), DOI: 10.1111/syen.12302

Resolving the psyllid tree of life: phylogenomic analyses of the superfamily Psylloidea (Hemiptera)

DIANA M. PERCY[1,2], ALEX CRAMPTON-PLATT[1,3], SAEMUNDUR SVEINSSON[4], ALAN R. LEMMON[5], EMILY MORIARTY LEMMON[6], DAVID OUVRARD[1] and DANIEL BURCKHARDT[7]

[1]Department of Life Sciences, Natural History Museum, London, U.K., [2]Department of Botany, Faculty of Science, University of British Columbia, Vancouver, Canada, [3]Department of Genetics, Evolution and Environment, University College London, London, U.K., [4]Agricultural University of Iceland, Hvanneyri, Iceland, [5]Department of Scientific Computing, Florida State University, Tallahassee, FL, U.S.A., [6]Department of Biological Science, Florida State University, Tallahassee, FL, U.S.A. and [7]Naturhistorisches Museum, Basel, Switzerland

Abstract. Understanding evolutionary relationships in the superfamily Psylloidea is challenging due to the lack of clear morphological synapomorphies for many groups. Some families and many of the genera, including the two largest, *Cacopsylla* Ossiannilsson and *Trioza* Foerster, have long been acknowledged as nonmonophyletic and the circumscription of natural groups has remained fluid. We present the best phylogenetic hypothesis to date for Psylloidea and provide a working systematic framework to better reflect evolutionary relationships. A shotgun sequencing approach using mixed pool DNAs for more than 400 species resulted in recovery from de novo assemblies of near-complete mitogenomes (≥10 kb) for 359 species, and partial genomes (5–10 kb) for an additional 40 species. The resulting phylogeny improves and clarifies the family classification and resolves some of the longstanding uncertainties in relationships within and between genera. A whole-nuclear-genome scan approach (yielding data from an estimated 373 nuclear genes) using the anchored hybrid enrichment method for a representative subset of taxa confirms the placement of major groupings and overall tree topology recovered with the mitochondrial data. The data generated represent a major increase in molecular resources for this superfamily. In addition, we highlight areas of remaining uncertainty that require further sampling and/or additional sources of data. The phylogeny provides new insights for both evolutionary and applied research, and a backbone constraint tree allows the placement of taxa of particular interest or concern (e.g. pest taxa) with only small fragments of sequence available (e.g. DNA barcodes).

Introduction

Assembling the tree of life is a huge challenge and construction must necessarily be done piecemeal. Contributions from each branch will ultimately allow meta-analysis at multiple scales, helping to understand the immense diversity of life as well as common evolutionary processes (Watson *et al.,* 2014; Hinchliff *et al.,* 2015; Nelson-Sathi *et al.,* 2015).

Correspondence: Diana M. Percy, Department of Life Sciences, Natural History Museum, Cromwell Road, London, SW7 5BD, U.K. E-mail: d.percy@nhm.ac.uk

Psyllids (Psylloidea), or jumping plant-lice, are arguably the least well-known superfamily in the suborder Sternorrhyncha (Hemiptera). They are a relatively small group (~4000 described species in eight families) compared with sister groups aphids and scale insects. All Sternorrhyncha are plant sap-sucking insects with many well-known pests mostly found among aphids, scales and whiteflies. Damage to the host plant is caused directly via feeding or gall development, and/or indirectly via transmission of pathogens (Perilla-Henao & Casteel, 2016). Although relatively few members of Psylloidea are serious pests, those that do cause damage to crops can be considered serious

enough agricultural threats to require proactive control (Boina & Bloomquist, 2015; Kistner *et al.*, 2016). Nevertheless, the lower profile in agricultural/horticultural systems has resulted in psyllids receiving less attention in evolutionary and systematic studies. Our understanding of psyllid systematics comes mainly from early cladistic analyses (e.g. White & Hodkinson, 1985) and more recent revisions to the superfamily classification (Burckhardt & Ouvrard, 2012). Although some interesting recent applications of molecular methods have been applied to particular taxon groups (e.g. Hall *et al.*, 2016; Martoni *et al.*, 2017; Wu *et al.*, 2017), there have been no comprehensive molecular systematic studies.

Psylloidea classification has some notable uncertainties and the superfamily has long been recognized as needing better resolution at multiple taxonomic scales due to the lack of clear morphological synapomorphies for many groups (White & Hodkinson, 1985). Several families, particularly Aphalaridae, Liviidae and Calophyidae, have been difficult to circumscribe, with the last being acknowledged as polyphyletic. In addition, many genera, particularly the two largest genera, *Cacopsylla* and *Trioza*, in the two largest families, Psyllidae and Triozidae, have long been acknowledged as polyphyletic (Burckhardt, 1987, 1988). Other affiliations within Psyllidae and Triozidae have also been problematic, and the circumscription of natural groups has remained fluid. In contrast, some families (e.g. Carsidaridae, Homotomidae and Triozidae) and subfamilies (e.g. Aphalarinae, Calophyinae, Liviinae and Spondyliaspidinae) are more clearly defined morphologically, but relationships among them remain uncertain. A robust phylogeny is needed to define taxa and resolve relationships within and amongst higher-level groupings. A classification reflecting natural groupings will enable more relevant interpretation of evolutionary processes such as potential codiversification with host plant groups (Ouvrard *et al.*, 2015), bacterial endosymbionts (Hall *et al.*, 2016), and even prediction of disease vectoring potential.

Approaches allowing both rapid and cost-efficient analyses of large datasets are greatly improving the realization of the tree of life; in particular, next-generation phylogenomics is increasingly being used to resolve large-scale phylogenies (Misof *et al.*, 2014; Prum *et al.*, 2015; Stout *et al.*, 2016; Young *et al.*, 2016; Feng *et al.*, 2017). A valuable tool in the phylogenomics era is a meta-sample approach whereby large quantities of species-diverse sample pools yield considerable amounts of data at marginal cost per sample. The use of individual index labelling for each taxon sampled within a pool provides a more time-efficient and guaranteed identification of resulting sequences, but indexing hundreds of samples can be costly. A nonindexed species pool approach is typically referred to as metagenomics, or voucher metagenomics when each sample in a pool is linked to a voucher specimen (Taberlet *et al.*, 2012; Crampton-Platt *et al.*, 2016). The proven success of a metagenomic approach in systematics (Taberlet *et al.*, 2012; Crampton-Platt *et al.*, 2016) and its cost-effectiveness make it currently a viable alternative to the use of sample indexing. For resolving systematic relationships, both mitogenomic and meta-mitogenomic approaches have proven effective with insect lineages at multiple taxonomic scales (Cui *et al.*, 2013; Gillett *et al.*, 2014; Tang *et al.*, 2014; Crampton-Platt *et al.*, 2015; Choo *et al.*, 2017). However, despite the proven phylogenetic utility of the mitogenome, there remain justified doubts in relying solely on maternally inherited haploid genomes to reconstruct evolutionary histories (Marlétaz *et al.*, 2017; Rodríguez *et al.*, 2017). Therefore, combining a mitogenomic approach with an amenable nuclear data approach, such as genome-wide targeted hybrid enrichment methods (Prum *et al.*, 2015; Hamilton *et al.*, 2016; Young *et al.*, 2016), should improve resolution as well as confidence in phylogenetic reconstructions from mitochondria alone. An advantage of both genomic approaches implemented here is that they can be used with little or no previous sequence reference. The anchored hybrid enrichment (AHE; Lemmon *et al.*, 2012) approach, in particular, has been designed to recover hundreds of low or single-copy, phylogenetically informative markers from across the nuclear genome, and it is a useful approach for resolving evolutionary relationships in nonmodel lineages for which a good reference genome is lacking (Lemmon & Lemmon, 2013). Furthermore, both the mitogenome and nuclear AHE methods are particularly effective for small organisms (such as psyllids: ∼1–7 mm) that typically yield low-quantity DNAs from single individuals.

The primary goal of this study is to apply next-generation sequencing approaches to generate the first large-scale phylogenomic framework for the superfamily that will provide both the required framework to study evolutionary processes, and a guide phylogeny for future classifications of this taxonomically challenging group.

Materials and methods

Mitogenomics: sampling and sample preparation

More than 400 taxa were selected to provide representation of all eight currently recognized psyllid families (according to Burckhardt & Ouvrard, 2012) (Table 1, Table S1). Representatives of each family are shown in Fig. 1. A nondestructive DNA extraction protocol using whole individual specimens was performed using a Qiagen Blood and Tissue Kit (Qiagen, U.K.) with initial incubation for ∼12–15 h; the voucher specimens were then placed in 70% ethanol before the remainder of the extraction protocol was completed. Final elution was in 200 μL. Voucher specimens are deposited in the Natural History Museum, London, UK (BMNH). Each specimen voucher was examined and identified where possible. The mitogenome sampling employed a vouchered metagenomic approach whereby mixed DNA pools of vouchered but unindexed specimens were sequenced with Illumina technology (U.K.). Each individual DNAs sample was also Sanger-sequenced for two short regions of mitochondrial DNA (mtDNA) which were then used downstream as the 'bait' sequences in the identification of assemblies from mixed pool data (Gillett *et al.*, 2014). PCR amplification of mtDNA baits was performed following protocols in Percy (2003): cytochrome oxidase 1 (cox1, length 472 bp) was amplified with primers mtd6 and mtd9 (equivalent to C1-J-1718 and C1-N-2191, respectively, in Simon *et al.*, 1994) with annealing

Table 1. Number of taxa included in the final phylogenetic analyses (see Methods), according to the family classification of Burckhardt & Ouvrard (2012) (see Table S1 for complete species list and Table S1 and S2 for GenBank numbers).

Family (no. of genera)	No. of genera sampled (% named genera)	No. of species sampled (Mito)	No. of species sampled (AHE)
Triozidae (70)	27 (39%)	179	16
Psyllidae (69)	28 (40%)	125	2
Aphalaridae (68)	20 (29%)	48	3
Liviidae (29)	11 (38%)	26	4
Calophyidae (11)	3 (27%)	10	1
Homotomidae (11)	3 (27%)	3	2
Carsidaridae (8)	5 (63%)	5	2
Phacopteronidae (5)	2 (40%)	3	1
Total	99	399	31

Mito, mitogenome dataset; AHE, anchored hybrid enrichment dataset.

temperature 50°C; and cytochrome B (cytB, length 385 bp) with primers cytBF and cytBR [equivalent to Sytb-F and Sytb-R, respectively, in Timmermans *et al.* (2010), with minor modifications: cytBF 5′-TGAGGNCAAATATCHTTYTGA-3′, cytBR 5′-GCAAATARRAARTATCATTCDG-3′] and annealing temperature 56°C.

We partitioned taxa into three sample pools and avoided including closely related taxa in the same pools; each pool included ∼20–30 taxa shared with one other pool. Each individual DNA sample was quantified (using Qubit; Thermo Fisher Scientific, U.K.) and pools were then prepared by calculating equimolar concentrations per taxon per pool (Table S2). Pool 1 included 164 taxa (of which 21 were also sampled in other pools); pool 2 had 137 taxa (of which 19 were also sampled in other pools); and pool 3 had 178 taxa (of which 29 were also sampled in other pools). Shared taxa across pools served as a quality check of both pools and recovered data.

Twelve additional mitogenomes obtained from GenBank were also included in the analyses: four ingroup Psylloidea and eight outgroup taxa in Hemiptera and Thysanoptera (see supplementary information Table S3).

Mitogenome sequencing, assembly and bioinformatics

The three sample pools were run on an Illumina MiSeq using a TruSeq library preparation (300 bp, paired-end) which yielded 14, 11 and 13 Gb of raw data for the pools. A quality (Q) assessment of raw FASTQ files for each library was made using FASTQC v.0.10.1 (www.bioinformatics.babraham.ac.uk/projects/fastqc) prior to the removal of adapter sequences with TRIMMOMATIC v.0.30 (Lohse *et al.*, 2012). Following removal of adapters, the proportions of paired reads retained for analysis were 83%, 90% and 84% of raw data pairs for the three pools. The data were then further filtered with BLASTN (BLAST v.2.2.27+; Altschup *et al.*, 1990) against a database of four identified Psylloidea mitogenomes obtained from GenBank (Table S3).

Assembly of mitochondrial reads from the three libraries was conducted using three different assembly protocols to maximize data capture: CELERA ASSEMBLER (CA, v.8.0; Myers *et al.*, 2000), IDBA-UD (Peng *et al.*, 2012), and NEWBLER (v.2.7, Margulies *et al.*, 2005). For CA, we used largely default settings as described in Crampton-Platt *et al.* (2015). Prior to assembly with IDBA-UD the reads were passed through a quality control step with PRINSEQ-LITE v.0.19.2 (Schmieder & Edwards, 2011) to trim low-quality bases and remove short sequences. Pairs of reads where both reads passed quality control were extracted using a CDBFASTA pipeline (see Crampton-Platt *et al.*, 2015 for details). Each assembly method behaves differently and the benefits of using multiple protocols are described in Crampton-Platt *et al.* (2016), including the option of merging assemblies from different protocols to increase sequence contiguity; this was done for 59 taxa that otherwise would have been excluded from our final dataset, which required a minimum of five loci to be represented.

Removal of redundant contigs, concatenation and attempted circularization of assembled contigs (≥ 14 kb) was performed in GENEIOUS R9 (Biomatters Ltd., Kearse *et al.*, 2012). Identification of assembled contigs was conducted using megablast in GENEIOUS against the bait sequences (98% identity, maximum target sequence 1), and identification of partial assemblies was sometimes improved by using previously sequenced 12S fragments as supplementary baits (Table S1). Protein-coding gene (PCG) sequences were identified by mapping to the mitogenome references obtained from GenBank. The PCGs were aligned with TRANSALIGN (invertebrate mitochondrial code; Bininda-Emonds, 2005) and MAFFT v.7 (Katoh & Standley, 2013). After removal of duplicates (samples in multiple pools) and short contigs with only partial or poorly aligning gene fragments, the total number of aligned contigs contributing to the phylogenetic analysis was 399 (385 identified and 14 unidentified). With the inclusion of 12 additional taxa obtained from GenBank (Table S3), the complete mitogenome dataset included 411 taxa. The alignment has been deposited in the Natural History Museum Data Portal as dataset (https://doi.org/10.5519/0023421): 31 complete mitogenomes, representing at least two species from each family, have been annotated and deposited in GenBank (Table S4), and the bait sequences for all taxa have been deposited in GenBank (Table S1).

Anchored phylogenomics: sampling, probe design and assembly

To obtain comparative nuclear data, we used an AHE approach (Lemmon *et al.*, 2012) requiring only 60 ng of genomic DNAs per sample. We sampled 40 psyllid species with at least one representative for each family, and we show results for 31 species here (Table S5) (only excluding multiple taxa sampled from a single genus); for rooting, we used a single aphid outgroup (*Tamalia coweni*; Aphididae). Genomic samples (treated with RNAse) were sent to the Center for Anchored Phylogenomics (www.anchoredphylogeny.com) at Florida State University for library preparation and enrichment. Capture of relevant loci for

a given taxonomic group and level of resolution are dependent on the similarity between probe and target. Our data were obtained from probes designed using Auchenorrhyncha and Sternorrhyncha lineages, and referred to as the 'Paraneoptera 2' kit (described in Dietrich *et al.*, 2017). General methodology for locus selection and probe design are described in Lemmon *et al.* (2012), Prum *et al.* (2015), and Hamilton *et al.* (2016). Library preparation and indexing were performed following a protocol modified from Meyer & Kircher (2010) (described in Lemmon *et al.*, 2012). Enrichment reactions were performed using Agilent Custom SureSelect kits (Agilent Technologies, Santa Clara, CA, U.S.A), and sequencing was performed with an Illumina HiSeq2500. After merging following Rokyta *et al.* (2012), the raw reads were trimmed, filtered and assembled into contigs following the methods described in Prum *et al.* (2015). The alignment was deposited in the Natural History Museum Data Portal as a dataset (https://doi.org/10.5519/0023421).

Phylogenetic analyses

Maximum likelihood (ML) analyses were conducted using RAXML (v. 8.2.4) with Gamma estimation of model parameters for all partitions (mitogenome coding regions: first/second/third position, rRNA regions, and AHE loci) and Gamma optimization of tree space. Multiple RAXML analyses (GTRCAT, 1000 rapid bootstraps) were run on the CIPRES Science Gateway (Miller *et al.*, 2010; Stamatakis, 2014). For the mitogenome dataset, two reduced character sets were employed to explore the effect on topology and node support of using all nucleotides versus more conserved nucleotides/regions. For the first reduced set, third codon positions were removed and RY coding employed for first and second positions to reduce compositional biases and homoplastic substitutions (Hassanin, 2006); and in the second set the rRNA regions were additionally excluded. In addition, a reduced taxon set was used to explore the effect of reducing taxon sampling by approximately half (from 411 to 202 taxa) but still retaining representatives of generic groupings for all families. There are 14 assemblies not identified with baits but represented by >5 kb length included in the phylogenetic analyses (Table S1), thus enabling assignment to family, and in some cases genus (identification can be narrowed down taxonomically by this process and additional mtDNA/bait regions sequenced for future confirmation).

For the AHE data, we explored the effect on tree topology and node support of excluding the aphid outgroup, using mid-point rooting, and changing the basal grouping to reflect that in the mitogenome conserved codon dataset. We also performed a combined analysis of mitochondrial and nuclear data with representation of all families and including the majority of major groups. In order to increase taxon representation in the combined analysis, we utilized previously generated 18S data for 37 taxa providing a combined mitochondrial–nuclear dataset of 62 taxa (Table S6). To compare tree building methods, we also performed Bayesian analyses on the combined dataset using MRBAYES (v. 3.2.6) run on CIPRES with two independent runs with four coupled Markov chain Monte Carlo analyses run for 20 million generations, sampling every 1000th generation and applying uniform priors to tree topologies and an exponential prior to branch lengths. Results were visualized using a 50% majority-rule consensus tree with 25% of topologies discarded as burn-in (Ronquist & Huelsenbeck, 2003).

Constraint tree

We explored incorporating four taxa into the phylogeny for which we had short bait sequences (cox1 and cytB). The taxa selected were two *Trioza* species (*T. apicalis* Foerster and *T. flixiana* Burckhardt & Lauterer); *T. apicalis* is considered a serious pest of carrot crops in northern and central Europe and is also suspected of vectoring pathogens (Nissinen *et al.*, 2012); more recently the sister taxon *T. anthrisci* Burckhardt was a confirmed carrier and potential vector of the proteobacterium '*Candidatus* Liberibacter solanacearum' (Sjölund *et al.*, 2017). Secondly, a taxon in the genus *Diclidophlebia* Crawford [*D. fremontiae* (Klyver)] is of particular interest as the single North American species in an otherwise mostly tropical and subtropical group (distributed worldwide) with a complex taxonomic history (Ouvrard, 2016). Lastly, a rare and highly unusual palm-feeding species from Malaysia has proved difficult to assign taxonomically, even to family level.

To explore the placement of these taxa, we implemented a binary backbone constraint specification with RAXML (according to Stamatakis *et al.*, 2008) using the best ML tree obtained with the all-nucleotide/all-taxa minimum five loci dataset. The backbone constraint method adds the taxon/taxa of interest initially using the maximum parsimony criterion, and once a comprehensive tree with all taxa has been obtained, it then optimizes under ML respecting the restrictions of the backbone constraint tree (Stamatakis *et al.*, 2008).

Results

Mitogenome assemblies

The three sample pools yielded somewhat different sequence outputs despite using similar sample density and sequencing protocol, in particular a smaller average fragment length was recovered from pool 3 (Table S2). It is not clear how this impacted downstream assemblies, but comparing assembly results, pool 3 generated a larger number of contigs with two of the protocols (IDBA and NEWBLER) (Table S7), but these were of comparatively shorter length, and notably pool 3 yielded the fewest full-length mitogenome assemblies (Table S8). From all three pools, a total of 638 nonredundant contigs were recovered with 472 identified with one or more bait sequences. There were 117 assemblies that could not be identified with baits ($101 <$ five loci, $14 \geq$ five loci; Table S9), mostly as a result of the bait region not being recovered in the assembly, but occasionally due to a failure to obtain bait sequences for a particular taxon. Complete circular genomes were obtained for 192 species with a genome size variation of 14.7–16.8 kb.

Anchored phylogenomics: assembly, locus performance and analysis

The AHE method generated an average of 4.8 million raw reads per sample, with ~1.9 million successfully merged with overlapping reads (Table S10, Fig. S1). After trimming, quality filtering and assembly, a total of 425 loci (average length 291 bases) from an estimated 373 genes resulted in an aligned dataset, with the aphid outgroup included, of over 100 000 bases (108 484 bp) with 50 811 variable and 43 181 informative sites; and excluding the aphid, alignment length was 90 045 bases, with 42 833 variable and 37 050 informative sites.

Phylogenetic analyses

As comparative Bayesian analyses did not yield notably different results from the ML analyses, the ML results are reported here (Tables 2–3, Figs 2–4). Both mitochondrial and nuclear data indicated a strongly supported crown clade of taxa comprising Psyllidae, Triozidae, Calophyinae and *Diaphorina* Löw (currently in Liviidae, Euphyllurinae). The mitogenome topology groups Psyllidae + *Diaphorina* and Triozidae as sister to Calophyinae, and we refer to this terminal clade as the 'PTCD clade'. Although the placement of the remaining families differs somewhat in the all nucleotide and conserved codon (using only first and second codon positions) character sets, both datasets cast doubt on the monophyly of Aphalaridae, recover Liviidae as polyphyletic, and strongly support the monophyly of Carsidaridae, Homotomidae and Phacopteronidae. The placement of these families in relationship to one another still remains uncertain due to weak support in key basal nodes. The different data and taxon partitions we implemented affected only the weakly supported basal portion of the tree. For instance, the all-taxa/all-nucleotides dataset recovered the topology in Fig. 2 (left), as did the conserved codon with rRNA dataset; only datasets using conserved codons only (i.e. without third codon positions and without rRNA regions), with both all-taxa and reduced taxon sets, recovered the topology on the right in Fig. 2. The reduced-taxon analyses using all PCG nucleotides, either with or without rRNA regions, recovered Phacopteronidae as the basal most group, but with unsupported/collapsing basal nodes; using conserved codon positions only, either with or without rRNA regions, recovered the same topologies as those shown in Fig. 2 left (for the all-taxa dataset).

Despite the impressive quantity of nuclear data obtained with the AHE method, there remains similar topological uncertainty around the basal groupings of the Psylloidea to that found in the mitogenome data; in particular, weakly supported basal nodes in the rooted AHE analysis (Fig. 3A, left). However, the basal uncertainty is partly caused by the inclusion of the single aphid outgroup in the alignment (Fig. S3 shows strengthened basal node support for the same topology from analysis without the aphid). All other nodes representing major groups in the AHE analysis were strongly supported, and the overall topology of the nuclear data either clearly supports or was consistent with the mitogenome results (Fig. 3A).

The combined mitochondrial–nuclear analysis also did not resolve the basal node uncertainty (Fig. 3B). It did, nevertheless, corroborate the major groupings and lend additional support to the placement of a paraphyletic Aphalaridae as sister to the remaining Psylloidea. Interestingly *Diaphorina* and Calophyinae were placed in a nested position higher up within the PTCD clade (sister to Triozidae as in the AHE analyis), and Psyllidae was paraphyletic (also found in the AHE analysis). However, the combined dataset had a number of taxa with either AHE or 18S data, but not both (Table S6).

Summary of taxon group arrangements in the mitogenome analyses

In many cases, the mitogenome data confirm the current classification based on morphology. Table 3 gives the taxa sampled in relation to the phylogenetic groups recovered with moderate to strong bootstrap support.

With over 1000 species in 70 genera, Triozidae is the second largest family. We identified 14 groups within Triozidae (Groups A–N; Table 3, Fig. 2). Notably, in our analyses, the highly polyphyletic genus *Trioza* (~430 species) has taxa currently distributed in eight of these groups. *Trioza urticae* (Linné), the type species of *Trioza* is in Group M, and therefore we refer to this *Trioza* subclade as *Trioza* 'sensu stricto'. Other problematic genera include *Kuwayama* Crawford and *Megatrioza* Crawford. Triozidae can be separated into three major clades: the first includes Groups A–L, the second Group M, and the third is the basal triozid clade represented by *Triozoida* Crawford (Group N) as sister to the remaining Triozidae.

With well over 1100 species in 69 genera, Psyllidae is the largest family. We identified 13 groups (Groups O–AA) within the family Psyllidae (sensu Burckhardt & Ouvrard, 2012), including *Katacephala* Crawford (currently in Group Z). Group BB representing *Diaphorina* is placed as sister to Psyllidae (but see alternative placements in the AHE and combined analyses below) (Table 3, Figs 2, 3). The Psyllidae groupings are partly consistent with current subfamily classification with a few notable exceptions. The subfamily Psyllinae is polyphyletic (Groups O and P, which include the vast majority of Psyllinae, and Group Y with the single genus *Platycorypha* Tuthill), and taxa currently in subfamilies Ciriacreminae and Aphalaroidinae are also recovered in multiple groups (Table 3). For example, *Limbopsylla* Brown & Hodkinson was erected as a polyphyletic genus for various members of Psyllidae, the type species is related to *Platycorypha*, and it was assigned to Psyllinae by Burckhardt & Ouvrard (2012); however, *Limbopsylla nigrivenis* Brown & Hodkinson (sampled here) has characteristically ciriacremine immatures. *Cacopsylla* is the largest psyllid genus (~460 species) and separates into two primary clades in Group O, with clade 1 containing the type species and including Old and New World taxa on a variety of plant groups. *Cacopsylla* clade 2 comprises North American taxa [and includes *Ceanothia aculeata* (Crawford) and *Pexopsylla* Jensen] on host plants in *Cercocarpus* and *Purshia* (both Rosaceae). *Baeopelma* Enderlein is not monophyletic but clusters, along with *Chamaepsylla*,

Fig. 1. Images of psyllids (adults and immatures) with representatives from all eight families: (A) *Trioza urticae* (Triozidae), (B) *Eryngiofaga mesomela* (Triozidae), (C) *Cacopsylla pulchella* (Psyllidae), (D) *Cacopsylla hippophaes* (Psyllidae), (E) *Arytainilla spartiophila* (Psyllidae), (F) *Arytaina genistae* (Psyllidae), (G) *Cacopsylla parvipennis* (Psyllidae), (H) *Livilla radiata* (Psyllidae), (I) *Psylla buxi* (Psyllidae), (J) *Calophya rhois* (Calophyidae), (K) *Livia junci* (Liviidae), (L) *Euphyllura phillyreae* (Liviidae), (M) *Mesohomotoma hibisci* (Carsidaridae), (N, O) *Homotoma ficus* (immatures and adult) (Homotomidae), (P) *Pseudophacopteron alstonium* (Phacopteronidae), (Q) *Aphalara sauteri* (Aphalaridae), (R) *Rhinocola aceris* (Aphalaridae). Photo credits: (A–D, F–I, K, L, N, O, Q, R) Gernot Kunz; (E, M, P) David Ouvrard; (J) Gabrijel Seljak. [Colour figure can be viewed at wileyonlinelibrary.com].

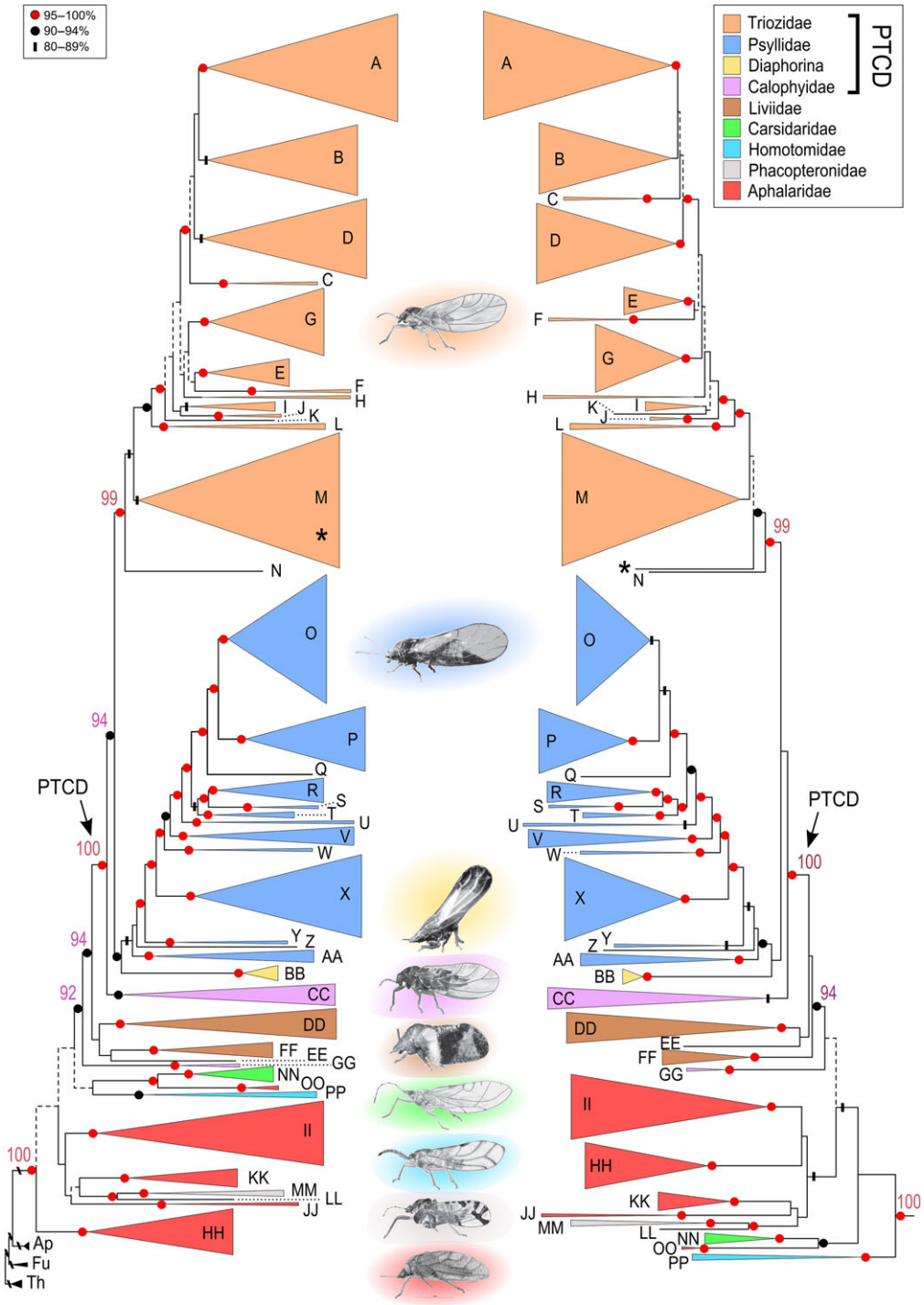

Fig. 2. Best maximum likelihood trees generated using the all-taxa/all-nucleotide mitogenome dataset (left) and the all-taxa/conserved-codon mitogenome dataset (right). Indicated in capital letters are the groups referred to in the text and Table 3 (species given in Table S1); triangle size indicates number of taxa (species names are shown for the left tree in Fig. S2). Node support is indicated by shape and colour of node symbols, and bootstrap numbers are shown only for critical backbone nodes; nodes with < 50% support are indicated by dotted lines. An asterix indicates a single species (*Trioza tabebuiea*) which is included in Group M (left tree) or falls outside (right tree); taxon group assignments are otherwise identical. Both trees are rooted with the same outgroups (indicated on the left tree; branch lengths not to scale for: Ap, Aphidoidea, Fu, Fulgoroidea, Th, Thysanoptera; outgroup species given in Table S3). Images of psyllids, top to bottom represent: *Trioza* (Triozidae), *Cacopsylla* (Psyllidae), *Diaphorina* (Liviidae), *Calophya* (Calophyidae), *Euphyllura* (Liviidae), *Mesohomotoma* (Carsidaridae), *Homotoma* (Homotomidae), *Phacopteron* (Phacopteronidae), *Rhinocola* (Aphalaridae). [Colour figure can be viewed at wileyonlinelibrary.com].

 Systematic Entomology, doi: 10.1111/syen.12302

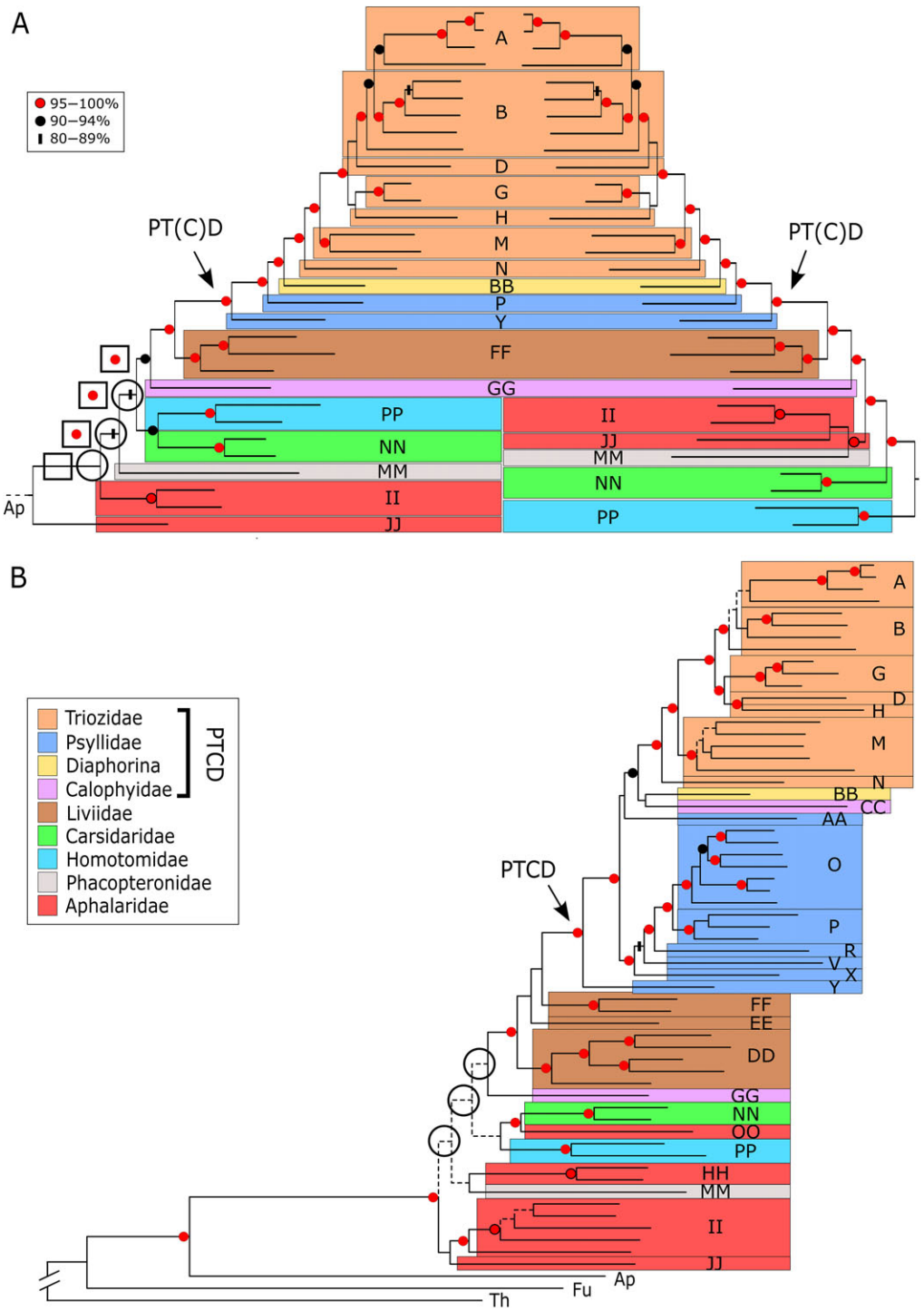

Fig. 3. (A) Best maximum likelihood (ML) trees generated using nuclear anchored hybrid enrichment (AHE) dataset (species names are given in Table S5 and Fig. S3). Left: Tree rooted with a single aphid outgroup; right: tree excluding aphid from analysis and rooted with Homotomidae to reflect topology in Fig. 2 (right). (B) Best ML tree generated using the combined mitochondrial–nuclear dataset (species names are given in Table S6 and Fig. S4). Indicated in capital letters are taxon groups represented relative to the mitogenome analysis (the 'C' in PTCD is in parentheses in A because there is no representative of Calophyidae sensu stricto, i.e. subfamily Calophyinae, in the AHE dataset). Node support is indicated by shape and colour of node symbols, and circles indicate weakly supported basal nodes referred to in the text (support for the same nodes using analysis excluding aphid outgroup is shown within squares in A, left); stronger support on the right tree also reflects analysis run without the aphid outgroup (see Fig. S3 for topology rooted with Aphalaridae, Group JJ, reflecting aphid outgroup rooted topology). [Colour figure can be viewed at wileyonlinelibrary.com].

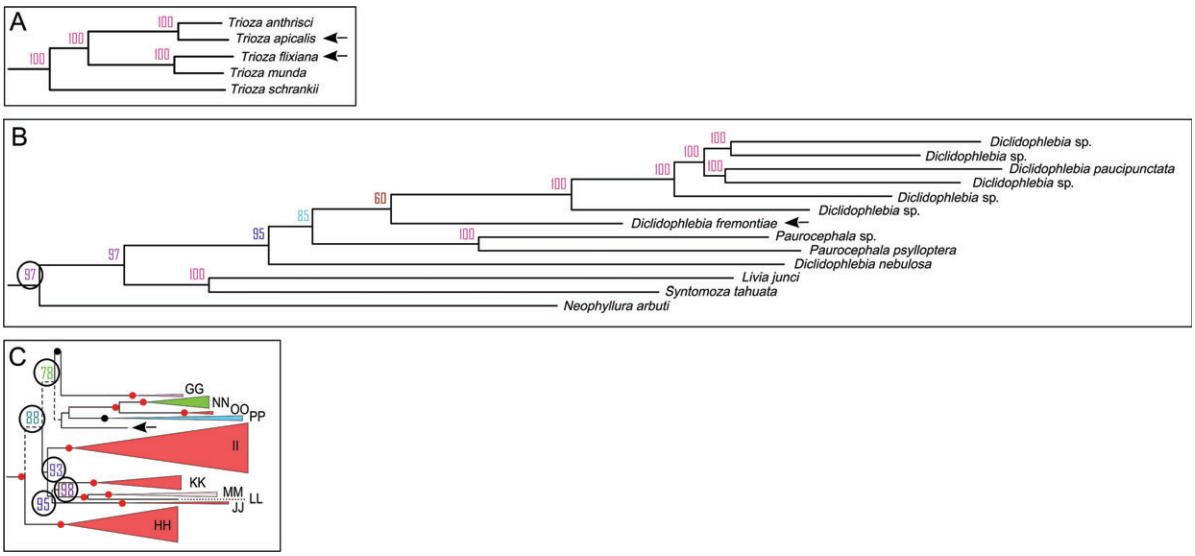

Fig. 4. Placement of taxa using a backbone constraint tree (in each example, arrows indicate the taxa added with only a short sequence): (A) *Trioza apicalis* and *Trioza flixiana*; (B) *Diclidophlebia fremontiae*; (C) an unnamed genus. In examples (B) and (C), topology and/or support values are impacted by the addition of taxa, with critical nodes circled, and with C otherwise showing the topology and node symbols in Fig. 2 (left). Colour legend as for Fig. 2. [Colour figure can be viewed at wileyonlinelibrary.com].

Table 2. Family level classification and modified major groupings in the mitogenome analysis showing support from maximum likelihood analyses with all taxa.

Group	BS
Triozidae	99/99
Psyllidae (including *Katacephala*)	80/94
Psyllidae (including *Katacephala*) + *Diaphorina*	94/73
Calophyidae (excluding *Mastigimas*)	94/85
PTCD clade	100/100
Liviidae (excluding *Katacephala* + *Diaphorina*)	77/*
Aphalaridae (excluding *Pachypsylla*) + Phacopteronidae	*/87
Phacopteronidae	100/100
Carsidaridae	100/100
Carsidaridae + *Pachypsylla*	100/94
Homotomidae	93/99

BS indicates ML bootstrap support for datasets: all-nucleotide/conserved-codon positions (i.e. without third codon positions and without rRNA regions); asterisk indicates not recovered as monophyletic. The 'PTCD clade' refers to a major clade incorporating three families, Psyllidae, Triozidae and Calophyidae sensu stricto (i.e. excluding *Mastigimas*), and the genus *Diaphorina*.

Ossiannilsson, with *Psylla* Geoffroy; while *Psylla buxi* (Linné) clusters with *Spanioneura* Foerster. *Ceanothia* Heslop-Harrison and *Euglyptoneura* Heslop-Harrison cluster together, and, with *Nyctiphalerus* Bliven, form a well-supported subgroup. In Group P, which are all legume feeders, *Arytainilla* Loginova, *Arytinnis* Percy and *Livilla* Curtis cluster together as sister to *Arytaina* Foerster. In Group V, *Colophorina* Capener, *Epiacizzia* Li and *Paraphyllura* Yang cluster together as sister to *Euphalerus* Schwarz. Group Q represents an Asian species currently assigned to *Cacopsylla*.

Diaphorina is a reasonably large genus (~75 described and many undescribed species) that is notable for its atypically wide host plant associations (Ouvrard *et al.*, 2015). The placement of this genus has been problematic with past and current placements in families Aphalaridae, Psyllidae and Liviidae most recently. Our analysis clearly indicates a position grouping it with Psyllidae, Triozidae and Calophyinae as part of the 'PTCD clade'. There is reasonable support for *Diaphorina* as sister to Psyllidae in the mitogenome analysis, but nuclear data – both AHE data and preliminary transcriptome data (K. Johnson, personal communication) – and the combined mitochondrial–nuclear analyses (Fig. 3B) place it as sister to Triozidae. The mitogenome analysis also clearly indicates the polyphyly of Diaphorininae/Diaphorinini (Groups Z, BB and FF).

Calophyidae is a small family (~100 species in 11 genera) that is poorly circumscribed morphologically and is currently polyphyletic (e.g. Groups CC, GG) but includes five probably monophyletic subfamilies. In all our analyses, *Mastigimas* Enderlein (Group GG), a small Neotropical genus (eight species), is placed outside of the 'PTCD clade'.

Aphalaridae, Liviidae, and Phacopteronidae: Aphalaridae and Liviidae are medium to moderately large-sized families with historically fluid taxonomies. Both these families may have diversified early in the evolution of Psylloidea given their position relative to the 'PTCD clade' and fossil evidence for Aphalaridae (see Discussion). Aphalaridae (~700 species in 68 genera) is only recovered as monophyletic (excluding *Pachypsylla* Riley and assuming inclusion of a nested Phacopteronidae) using the conserved codon position dataset (Fig. 2, right); in all other analyses it is recovered as paraphyletic and sister to all other families.

In the all-taxa analyses (Fig. 2), the small but highly distinct family Phacopteronidae is nested within Aphalaridae and is

Table 3. Family/subfamily placement in current classification; generic group placement as indicted in the trees (Figs 2–4).

Family	Group	BS	Taxa/clades
Triozidae	Group A	100/94	*Hemitrioza**, *Pariaconus*, *Pauropsylla**, *Trioza* spp. (two clades, including *T. anthrisci*, *T. galii*, *T. munda*, *T. remota*, *T. schrankii*, *T. tatrensis*)
Triozidae	Group B	88/64	*Crawforda**, *Hevaheva*, *Megatrioza* (including *M. kauaiensis*), *Trioza* spp. (including *T. alacris*, *T. anceps*, *T. uniqua*), genus unnamed (Austro-Pacific)
Triozidae	Group C	100/97	*Leuronota*
Triozidae	Group D	83/95	*Aacanthocnema*, *Anomocephala**, *Casuarinicola**, *Cerotrioza*, *Kuwayama* spp. (including *K. minutura*, *K. pisonia*, *K. tipicola*), *Megatrioza* (including *M. zanthoxyli*), *Trioza* spp. (including *T. chenopodii*, *T. erytreae*, *T. percyae*, *T. tricornuta*), Genus unnamed (Australia)
Triozidae	Group E	100/97	*Calinda*, *Kuwayama* sp., *Powellia**
Triozidae	Group F	100/100	*Trioza* spp. (including *T. malloticola*, *T. pitformis*)
Triozidae	Group G	100/100	*Schedotrioza*, *Trioza* spp. (including *T. alipellucida*, *T. eugeniae*, *T. obunca*, *T. outeiensis*, *T. vitiensis*, *T. zimmermani*)
Triozidae	Group H	79/59	*Leptynoptera**, *Trioza* spp. (including *T. incrustata*)
Triozidae	Group I	83/79	*Stenopsylla*, *Trichochermes**, *Trioza* spp. (including *T. rhamni*)
Triozidae	Group J	100/100	Genus unnamed (South America)
Triozidae	Group K	S	Genus unnamed (Africa)
Triozidae	Group L	98/97	*Ceropsylla**, Genus unnamed (Singapore)
Triozidae	Group M	80/65	*Bactericera*, *Baeoalitriozus*, *Hemischizocranium**, *Leptotrioza*, *Phylloplecta**, *Schedoneolithus**, *Swezeyana**, *Trioza**, Genus unnamed (Hawaii)
Triozidae	Group N	S	*Triozoida**
Psyllidae, Psyllinae	Group O	99/86	*Baeopelma**, *Cacopsylla* clade 1*, *Cacopsylla* clade 2 (includes *Ceanothia aculeata*), *Ceanothia**, *Chamaepsylla**, *Euglyptoneura**, *Nyctiphalerus*, *Pexopsylla**, *Psylla**, *Spanioneura**
Psyllidae, Psyllinae	Group P	100/100	*Arytainilla*, *Arytaina*, *Arytinnis*, *Livilla**
Psyllidae, Psyllinae	Group Q	S	*Cacopsylla eriobotryae* (Taiwan)
Psyllidae, Ciriacreminae	Group R	100/97	*Auchmerina**, *Euceropsylla**, *Heteropsylla**
Psyllidae, Aphalaroidinae and Psyllidae, incertae sedis	Group S	100/100	*Telmapsylla**, *Limbopsylla laguncularia*
Psyllidae, Ciriacreminae and Psyllidae, incertae sedis	Group T	97/100	*Mitrapsylla*, *Limbopsylla nigrivenis*
Psyllidae, Psyllinae	Group U	97/85	*Amorphicola**, Genus unnamed
Psyllidae, Macrocorsinae	Group V	100/100	*Colophorina*, *Epiacizzia*, *Euphalerus*, *Paraphyllura**
Psyllidae, cf. Macrocorsinae	Group W	100/99	Genera unnamed (Madagascar)
Psyllidae, Acizziinae	Group X	100/100	*Acizzia*
Psyllidae, Psyllinae	Group Y	97/81	*Platycorypha*
Liviidae, Euphyllurinae	Group Z	S	*Katacephala*
Aphalaroidinae	Group AA	100/100	*Aphalaroida*, *Freysuila*, *Russelliana**
Liviidae, Euphyllurinae	Group BB	100/100	*Diaphorina*
Calophyidae, Calophyinae	Group CC	94/85	*Calophya*, *Strogylocephala*
Liviidae, Liviinae	Group DD	100/100	*Diclidophlebia*, *Livia*, *Paurocephala*, *Syntomoza*
Liviidae, Euphyllurinae	Group EE	S	*Neophyllura*
Liviidae, Euphyllurinae	Group FF	100/100	*Euphyllura*, *Psyllopsis*, *Strophingia*
Calophyidae, Mastigimatinae	Group GG	100/100	*Mastigimas*
Aphalaridae, Aphalarinae	Group HH	100/100	*Aphalara*, *Colposcenia*, *Craspedolepta*, *Lanthanaphalara*, *Limataphalara*, *Neaphalara*
Aphalaridae, Spondyliaspidinae	Group II	100/100	*Anoeconeossa*, *Australopsylla*, *Blastopsylla*, *Boreioglycaspis**, *Cardiaspina*, *Creiis*, *Cryptoneossa*, *Ctenarytaina*, *Glycaspis*, *Lasiopsylla*, *Platyobria*
Aphalaridae, Rhinocolinae	Group JJ	100/100	*Apsylla*, *Rhinocola*
cf. Aphalaridae	Group KK	100/100	Genus unnamed [New Caledonia]
Aphalaridae, Rhinocolinae	Group LL	S	'*Parapaurocephala*' (= '*Paurocephala longicella* group' in Burckhardt & Basset, 2000)
Phacopteronidae	Group MM	100/100	*Cornegenapsylla*, *Pseudophacopteron*
Carsidaridae	Group NN	100/100	*Allocarsidara*, *Mesohomotoma*, *Paracarsidara*, *Protyora*, *Tenaphalara*
Aphalaridae, Pachypsyllinae	Group OO	100/100	*Pachypsylla*
Homotomidae	Group PP	100/99	*Homotoma*, *Macrohomotoma*, *Mycopsylla*

BS indicates ML bootstrap support for datasets: all nucleotide/conserved codon position; 'S' indicates single taxon sampled for that group. Taxa included in the mitogenome analysis are listed, and an asterisk indicates that the type species was sampled. See text for details of suprageneric and subgeneric clades. See Table S1 for a list of all species sampled.

strongly supported as sister to an Asian Pacific genus referred to here as '*Parapaurocephala*' (current name invalid; Burckhardt & Ouvrard, 2012). Monophyly of Aphalaridae is not recovered in the mitogenome all-nucleotide dataset (Fig. 2, left), or the outgroup rooted AHE analysis (Fig. 3A, left), or the combined data analysis (Fig. 3B), and there is further indication of instability in the basal portion of the tree, with the reduced-taxon set recovering Phacopteronidae (albeit unsupported) as sister to the rest of Psylloidea.

Liviidae (~330 species in 29 genera) is not recovered as monophyletic in the conserved codon dataset (Fig. 2, right) or the combined data (Fig. 3B), and is recovered with only weak support in the all-nucleotide dataset (Fig. 2, left). However, both subfamilies, Euphyllurinae (excluding *Diaphorina* and *Katacephala*) and Liviinae, are always recovered with strong support, with the exception of *Neophyllura* Loginova (Euphyllurinae) whose position is uncertain as it clusters as sister taxon to one or other subfamily with weak support (but see constraint analysis results later, which implicate taxon sampling again in impacting both topology and support values).

Carsidaridae, *Pachypsylla*, and Homotomidae: The small families Carsidaridae (~40 species in eight genera) and Homotomidae (~80 species in 11 genera) are well supported as monophyla by a series of morphological characters (Hollis, 1987; Hollis & Broomfield, 1989), and Hollis (1987) suggested that these families are sister taxa. Only in the conserved codon position dataset are Homotomidae and Carsidaridae recovered as sister to the remaining Psylloidea (Fig. 2, right). In the all-nucleotide and outgroup rooted AHE analyses, Homotomidae and Carsidaridae group as sister families (but this is only strongly supported in the AHE analysis). Due to the weak resolution at the base of the Psylloidea, a conservative interpretation of our data is that these families are in an effectively unresolved basal position with a paraphyletic Aphalaridae. However, both our combined data analysis (Fig. 3B) and preliminary transcriptome data analysis (K. Johnson, personal communication) place a paraphyletic Aphalaridae at the base of the Psylloidea, followed by Carsidaridae and Homotomidae as sister to the remaining Psylloidea, and this lends further confidence to the mitogenome topology using all-taxa/all-data (Fig. 2 left), as well as the rooted AHE analysis (see below). The position of the genus *Pachypsylla* (Aphalaridae, Pachypsyllinae) as sister to Carsidaridae is strongly supported in all mitogenome analyses (Fig. 2) and the combined data analysis (Fig. 3B). Interestingly, confirmation of a close relationship between Carsidaridae and *Pachypsylla* was also found in a phylogenetic analysis of the bacterial endosymbionts (Hall *et al.*, 2016).

Comparison of mitogenome, AHE and combined analyses

The anchored phylogenomic analysis provides a nuclear data comparison and confirms the overall topology of major groupings found in the mitogenome analysis (Figs 3A, S3). In particular, within Triozidae, where the majority of the AHE sampling is focused, the topology confirms that of the mitogenome data but with placement shifts among Groups A, B and D (where node support is less than 95% in Fig. 3) depending on whether the aphid outgroup is included or excluded from the analysis (see Fig. S3). The nuclear data do not fully resolve the ambiguity around the base of Psylloidea, with the most basal nodes lacking strong support in the outgroup rooted topology (Fig. 3A, left; see also combined analysis in Fig. 3B). Nevertheless, this topology (i.e. with a basal, paraphyletic Aphalaridae) is the same as that shown on the left in Fig. 2 using the all-nucleotide/all-taxa mitogenome dataset, and also recovered in the combined data analysis (Fig. 3B), and therefore currently represents the best working hypothesis for the Psylloidea. In addition to recovering similar basal ambiguity, the AHE and combined analyses confirm the 'PTCD clade'. The most notable difference from the mitogenome analysis is the placement of *Diaphorina* (and Calophyinae in the combined data ML analysis) as sister to Triozidae, not Psyllidae (Bayesian analysis grouped Calophyinae with *Aphalaroida* Crawford), and the paraphyly of Psyllidae due to the placement of *Platycorypha* (Group Y, Fig. 3). The tree on the right in Fig. 3A is manually rooted with Homotomidae, and the resulting topology can be seen to reflect that recovered using the conserved codon analysis (Fig. 2, right), with a monophyletic Aphalaridae + Phacopteronidae. When the aphid is excluded from the AHE analysis (shown in Fig. S3 rooted with Aphalaridae, Group JJ, to reflect outgroup rooted topology), bootstrap support is generally improved, notably in the basal nodes (Fig. 3A). Interpreting the nuclear data is done with caution given the limited taxon sampling in the AHE analysis (including outgroup rooting with a single aphid), and missing data in the combined analysis where a number of taxa are represented by either AHE or 18S data but not both (Table S6).

Taxon groups relative to host preference and biogeography

Our analysis recovers numerous novel biological and biogeographic findings although we mention only a few here. There are a number of clades notable because taxa in two or more genera group together according to host plant association rather than current taxonomy. Examples include: (i) *Psylla buxi* (Linné) and *Spanioneura fonscolombii* Foerster on *Buxus*; (ii) *Psylla alni* (Linné) and *Baeopelma foersteri* (Flor) on *Alnus*; (iii) *Pexopsylla cercocarpi* Jensen, *Ceanothia aculeata* and *Cacopsylla brevistigmata* (Patch) on *Cercocarpus*; and (iv) *Trioza rhamni* (Schrank) and *Trichochermes walkeri* (Foerster) on *Rhamnus*; and a subgroup in Group O includes taxa that are all on host plants in Betulaceae but are currently placed in four different psyllid genera. Additionally, notable groupings in our data reflecting host association are also evident within genera such as *Cacopsylla*, where there is a subgroup in clade 1 of taxa on Salicaceae from both Nearctic and Palaearctic regions, and the remaining subgroups are predominantly on Rosaceae. *Cacopsylla* species feeding on *Salix* (Salicaceae) are geographically mixed (Old and New World) but cluster together, and these are sister to an Asian species feeding on *Akebia* and *Stauntonia* (Lardizabalaceae) (Ouvrard, 2016). The large, species-rich genus *Acizzia* has a clade representing African/European taxa that is nested within taxa from Asia, Austro-Pacific, and Polynesia.

Several subclades in Triozidae provide potential insights into the evolution of galling in this family. For instance, a strongly supported terminal subclade, Groups A–D (Fig. 2), is global in distribution and includes genera with both galling and non-galling taxa. Each group (A–D) has moderate to strong support (Table 3) and includes a number of unnamed genera from the Austro-Pacific region. Group G is an exclusively galling clade, with all taxa feeding on Myrtaceae in Austro-Pacific and Oriental regions; Groups E–F, H and J are also tropical and southern hemisphere groups; Group L currently includes only galling taxa on Sapotaceae. The second major clade, Group M, is widely distributed but appears to be predominantly nongalling. Notably, Group M does not include *Leptynoptera* Crawford (in Group H); previous authors had proposed a sister relationship of this genus with *Baeoalitriozus* Li based on the reduced hindwing (Martin & Hollis, 1992), and *Pauropsylla* Rübsaamen (Group A) based on the broad, short forewing (White & Hodkinson, 1985), but these characters are homoplasious in this instance. Lastly, Group N (represented by *Triozoida*) is always recovered as sister to the remaining triozids, and *Triozoida* is a galling group on host plants in Myrtaceae in Central and South America.

For a comparison of taxa on remote islands, Madagascar (a continental island) has morphologically disparate endemic genera that group together, implying a single ancient radiation, whereas endemic Hawaiian genera (oceanic islands) are generally more closely related to non-Hawaiian genera, suggesting multiple recent colonizations. Similar biogeographic patterns are found in other animal and plant groups (Valente *et al.*, 2014).

Constraint tree placements

We provide three examples of how a backbone constraint tree can be used to place taxa of interest for which only short DNA sequences are available. First, the two *Trioza* species (*T. apicalis* and *T. flixiana*) were placed convincingly in the phylogeny (Fig. 4A), and we confirm subgroup placement of *T. apicalis* and *T. flixiana* within Group A. Secondly, the backbone constraint analysis places *Diclidophlebia fremontiae* as basal and sister to a group representing *Diclidophlebia* from Central America (Fig. 4B). Interestingly, the addition of *D. fremontiae* serves to resolve, with relatively strong support, *Neophyllura* (Euphyllurinae) as sister to Liviinae rather than the remaining members of Euphyllurinae. Lastly, the unusual palm-feeding species (undescribed) from Malaysia is recovered as closest to Homotomidae and Carsidaridae, but is placed at the collapsing node for these two families (Fig. 4C). Again, it is notable that the addition of this last taxon serves to strengthen the basal node support values for the topology shown in Fig. 2 (left), i.e. a paraphyletic Aphalaridae as sister to remaining Psylloidea (Fig. 4C).

Discussion

We present the first molecular phylogeny to sample across all families of Psylloidea. Almost 100 genera and ~385 species represent 37% and 10% of the described genera and species, respectively. The resulting phylogeny supports the monophyly of many taxa previously defined (summarized in Burckhardt & Ouvrard, 2012) and provides definitive clarification for many long disputed relationships, in addition to a number of novel and unexpected findings. At the suprafamilial level, the previously postulated sister-group relationship of Psyllidae + Triozidae, together with Calophyinae is recovered in the 'PTCD clade'. Other groupings of Carsidaridae + Homotomidae and Rhinocolinae + Spondyliaspidinae are not, or are not strongly, supported. At the family and subfamily levels, the monophyly of Carsidaridae, Homotomidae, Phacopteronidae, Triozidae, Aphalarinae, Calophyinae, Liviinae and Spondyliaspidinae are well supported, as well as, to a large extent (albeit with less support), Psyllidae, Aphalaroidinae, Ciriacreminae, Macrocorsinae and Rhinocolinae. Our analyses also confirm the artificial nature of families/subfamilies Aphalaridae, Calophyidae, Psyllinae and the two most species-rich genera, *Cacopsylla* and *Trioza*.

The phylogenetic reconstructions presented here differ in several aspects from the classification of Burckhardt & Ouvrard (2012). One of the most notable differences is the polyphyly of Euphyllurinae. Our analyses strongly support a monophyletic clade including the genera *Euphyllura* Foerster, *Psyllopsis* Löw and *Strophingia* Enderlein, with *Neophyllura* either as a weakly supported sister group or grouped more strongly with Liviinae (in the backbone constraint analysis); two other genera, *Diaphorina* and *Katacephala*, previously thought to be closely related to *Psyllopsis* (Hollis, 1985; Hodkinson, 1991), are here included within the strongly supported 'PTCD clade'. Within this clade, *Diaphorina* is sister to Psyllidae (mitogenome analysis) or sister to Triozidae (AHE and combined analyses), and *Katacephala* is consistently recovered in Psyllidae but in a basal position together with Aphalaroidinae (mitogenome analysis). This suggests that the morphological characters previously used to group these taxa are homoplasious (i.e. metatibial spurs forming an open crown, short antennae, inner face of paramere densely covered in long bristles, immatures lacking capitate setae).

A number of other differences from the current classification concern several small taxonomic units previously assigned to particular groups, often using vaguely characterized justifications. For example, Pachypsyllinae, previously referred to Aphalaridae, here constitute a well-supported sister group to Carsidaridae. Phacopteronidae may be nested within a probably paraphyletic Aphalaridae. Subfamily Mastigimatinae (currently in Calophyidae) is not closely related to Calophyinae, but instead is recovered as a reasonably well-supported sister group to Liviidae groups (Euphyllurinae *sensu stricto* and Liviinae) + PTCD clade. *Platycorypha* and *Amorphicola* Heslop-Harrison were previously assigned to Psyllinae, but the former constitutes a moderately supported sister group to a 'core' Psyllidae (i.e. excluding *Diaphorina*, Aphalaroidinae and *Katacephala*), and the latter is a strongly supported sister group to Ciriacreminae + Psyllinae. Despite these notable differences from the current classification, the monophyly of many of the remaining genera is confirmed in this study. Most examples of nonmonophyletic genera are found in Psyllinae

(e.g. *Cacopsylla*, *Ceanothia*, *Livilla* and *Psylla*), and Triozidae (e.g. *Bactericera* Puton, *Kuwayama*, *Megatrioza*, *Trioza*), which is not surprising as most of these genera are already suspected of being polyphyletic or are presently vaguely defined. There are only five subfamilies not included in our analysis, and of these, Togepsyllinae (Aphalaridae) and Dynopsyllinae (Homotomidae) are probably classified correctly, while the placement of Atmetocraniinae, Metapsyllinae and Symphorosinae (Calophyidae) remains uncertain given the polyphyly of the family found here. Regarding the earliest diversification of Psylloidea, the most likely scenario and best working hypothesis suggested by our study is a paraphyletic Aphalaridae as the sister group to the remaining Psylloidea. The 'soft' backbone nodes at the base of the phylogeny may indicate a period of rapid diversification early in the evolution of extant Psylloidea, e.g. the ancient rapid radiation model (Rothfels *et al.*, 2012).

Like most Sternorrhyncha, psyllids may exhibit high levels of morphological divergence between relatively closely related species, often influenced by ecology (e.g. Percy, 2017). The potential for rapid morphological divergence coupled with relatively simplified body plan and wing venation results in the absence of reliable synapomorphies for interpreting early divergence in the group. A lack of morphological synapomorphies makes the contribution of molecular data critical to reconstructing psyllid evolutionary history. In the majority of cases where morphology does provide an unambiguous indication of relationships, the phylogenomic data provide added support and clarify the placement of groups in relation to one another. For example, the differentiation of major Triozidae clades provides insights into evolution in this large, often homogeneous family (which is particularly lacking a clear morphological signal) that would not be possible using morphology alone. One example would be the interpretation of galling biology: an ancestral galling state (Group N) and subsequent splits into major clades that are either predominantly nongalling (Group M) or predominantly galling (Group G), with yet other groups with mixed biologies, suggests switches back and forth (e.g. in Group A), with the loss of galling occurring many times. Less clear is whether galling, once lost, has re-evolved; a mostly nongalling clade (including *Trioza* sensu stricto and *Bactericera*) includes many taxa that cause plant tissue cell death and distortion [e.g. *Bactericera cockerelli* (Šulc), *Baeoalitriozus diospyri* (Ashmead) and *Phylloplecta* Riley spp.], but without production of fully formed galls, and therefore the mechanisms to induce gall formation (e.g. Bailey *et al.*, 2015; Nabity, 2016) could potentially be lost permanently in this clade.

In summary, we provide the first comprehensively sampled molecular phylogenetic framework for the superfamily Psylloidea, providing a working hypothesis of the psyllid tree of life to guide further analyses. In many instances, support for the traditional classification is confirmed; in others, unusual convergences and/or unexpected biological and host associations have come to light. We highlight the uses of this study for guiding future classification and research, and in particular as a reference point for further evolutionary studies in this group.

Supporting Information

Additional Supporting Information may be found in the online version of this article under the DOI reference: 10.1111/syen.12302

Table S1. Taxa (411 in all) included in the mitogenome analysis listed alphabetically using current family classification with group placement from this study.

Table S2. MiSeq taxon sampling per pool and sequencing results.

Table S3. Additional taxa obtained from GenBank and included in the mitogenome analysis.

Table S4. The 31 taxa for which annotated mitogenomes have been submitted to GenBank.

Table S5. Taxa included in the anchored hybrid enrichment (AHE) analysis.

Table S6. Taxa included in the combined mitochondrial–nuclear analysis.

Table S7. Number of contigs generated per sample pool for each of three different assembly protocols applied to the mitogenome data.

Table S8. Breakdown of assembled contig lengths generated per sample pool with three different assembly protocols applied to the mitogenome data.

Table S9. Recovery rates for bait identified mitogenome assemblies for each sample pool.

Table S10. Raw read and assembly details for the anchored hybrid enrichment (AHE) data.

Figure S1. Assembly results for the nuclear genome loci using the AHE method.

Figure S2. Best ML tree generated using the all-taxa/all-nucleotide mitogenome data.

Figure S3. Best ML trees generated using the nuclear anchored hybrid enrichment (AHE) data.

Figure S4. Best ML tree generated using the combined mitochondrial–nuclear dataset.

Data and supplementary tables and figures are available from the Natural History Museum Data Portal, dataset Psylloidea phylogenomics: https://doi.org/10.5519/0023421.

Acknowledgements

We thank Shilo Dickens at Cambridge University Department of Biochemistry for the TruSeq library preparation and Illumina sequencing, and the Natural History Museum (NHM) sequencing service for Sanger Sequencing of the baits. We thank the

Vogler Lab (NHM), particularly Martijn Timmermans and Conrad Gillett, for advice on sample preparation. From the Lemmon lab (FSU), we thank Michelle Kortyna, Alyssa Bigelow Hassinger, Sean Holland, Kirby Birch and Ameer Jalal for support with AHE data generation and analysis. We thank those who supplied additional psyllid material, including Philip Butterill, Elin Claridge, Lyn Cook, Quentin Cronk, Rosie Gillespie, Dan Gruner, Mark Hoddle, Kevin Johnson, Jon Martin, Steve Montgomery, Laurence Mound, John Noyes, George Roderick and Alice Wells. We acknowledge the late Gilbert Hodebert (illustrator at the Muséum national d'Histoire naturelle, Paris) for the habitus illustrations in Fig. 2. For helpful comments on earlier drafts we are grateful to Quentin Cronk. Financial support that assisted in fieldwork is gratefully acknowledged from Natural Environment Research Council (grant number NERC GT04/97/109/TS studentship to DMP), Smithsonian Institution Postdoctoral Fellowship (2004–2005 to DMP), an NSF Dimensions of Biodiversity award (grant number DEB 1241253 to the University of California, Berkeley); additional sample collecting and the molecular work were supported by NHM (Molecular Taxonomy and DIF funds awarded to DMP). The authors are unaware of any conflict of interest.

References

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

Bailey, S., Percy, D.M., Hefer, C.A. & Cronk, Q.C.B. (2015) The transcriptional landscape of insect galls: psyllid (Hemiptera) gall formation in Hawaiian *Metrosideros polymorpha* (Myrtaceae). *BMC Genomics*, **16**, 943.

Bininda-Emonds, O.R.P. (2005) transAlign: using amino acids to facilitate the multiple alignment of protein-coding DNA sequences. *BMC Bioinformatics*, **6**, 156.

Boina, D.R. & Bloomquist, J.R. (2015) Chemical control of the Asian citrus psyllid and of huanglongbing disease in citrus. *Pest Management Science*, **71**, 808–823.

Burckhardt, D. (1987) Jumping plant lice (Homoptera: Psylloidea) of the temperate neotropical region. Part 2: Psyllidae (subfamilies Diaphorininae, Acizziinae, Ciriacreminae and Psyllinae). *Zoological Journal of the Linnean Society*, **90**, 145–205.

Burckhardt, D. (1988) Jumping plant lice (Homoptera: Psylloidea) of the temperate neotropical region. Part 3: Calophyidae and Triozidae. *Zoological Journal of the Linnean Society*, **92**, 115–191.

Burckhardt, D. & Basset, Y. (2000) The jumping plant-lice (Hemiptera, Psylloidea) associated with *Schinus* (Anacardiaceae); systematics, biogeography and host plant relationships. *Journal of Natural History*, **34**, 57–155.

Burckhardt, D. & Ouvrard, D. (2012) A revised classification of the jumping plant-lice (Hemiptera: Psylloidea). *Zootaxa*, **3509**, 1–34.

Choo, L.Q., Crampton-Platt, A. & Vogler, A.P. (2017) Shotgun mitogenomics across body size classes in a local assemblage of tropical Diptera: phylogeny, species diversity and mitochondrial abundance spectrum. *Molecular Ecology*, **26**, 5886–5098. https://doi.org/10.1111/mec.14258.

Crampton-Platt, A., Timmermans, M.J.T.N., Gimmel, M.L. *et al.* (2015) Soup to tree: the phylogeny of beetles inferred by mitochondrial metagenomics of a Bornean rainforest sample. *Molecular Biology and Evolution*, **32**, 2302–2316.

Crampton-Platt, A., Yu, D.W., Zhou, X. & Vogler, A.P. (2016) Mitochondrial metagenomics: letting the genes out of the bottle. *GigaScience*, **5**, 15.

Cui, Y., Xie, Q., Hua, J. *et al.* (2013) Phylogenomics of Hemiptera (Insecta: Paraneoptera) based on mitochondrial genomes. *Systematic Entomology*, **38**, 233–245.

Dietrich, C.H., Allen, J.M., Lemmon, A.R. *et al.* (2017) Anchored hybrid enrichment-based phylogenomics of leafhoppers and treehoppers (Hemiptera: Cicadomorpha: Membracoidea). *Insect Systematics and Diversity*, **1**, 57–72.

Feng, Y.J., Blackburn, D.C., Liang, D. *et al.* (2017) Phylogenomics reveals rapid, simultaneous diversification of three major clades of Gondwanan frogs at the Cretaceous-Paleogene boundary. *Proceedings of the National Academy of Sciences of the United States of America*, **114**, E5864–E5870.

Gillett, C.P.D.T., Crampton-Platt, A., Timmermans, M.J.T.N., Jordal, B., Emerson, B.C. & Vogler, A.P. (2014) Bulk de novo mitogenome assembly from pooled total DNA elucidates the phylogeny of weevils (Coleoptera: Curculionoidea). *Molecular Biology and Evolution*, **31**, 2223–2237.

Hall, A.A.G., Morrow, J.L., Fromont, C. *et al.* (2016) Codivergence of the primary bacterial endosymbiont of psyllids versus host switches and replacement of their secondary bacterial endosymbionts. *Environmental Microbiology*, **18**, 2591–2603.

Hamilton, C.A., Lemmon, A.R., Lemmon, E.M. & Bond, J.E. (2016) Expanding anchored hybrid enrichment to resolve both deep and shallow relationships within the spider tree of life. *BMC Evolutionary Biology*, **16**, 212.

Hassanin, A. (2006) Phylogeny of Arthropoda inferred from mitochondrial sequences: strategies for limiting the misleading effects of multiple changes in pattern and rates of substitution. *Molecular Phylogenetics and Evolution*, **38**, 100–116.

Hinchliff, C.E., Smith, S.A., Allman, J.F. *et al.* (2015) Synthesis of phylogeny and taxonomy into a comprehensive tree of life. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 12764–12769.

Hodkinson, I.D. (1991) A review of *Katacephala* Crawford with the description of an allied genus from South America (Insecta, Homoptera, Psylloidea). *Zoologica Scripta*, **20**, 77–87.

Hollis, D. (1985) *Parapsylla*, a Gondwanan element in the psyllid fauna of southern Africa (Homoptera). *Zoological Journal of the Linnean Society*, **83**, 325–342.

Hollis, D. (1987) A review of the Malvales-feeding psyllid family Carsidaridae (Homoptera). *Bulletin of the British Museum (Natural History) Entomology*, **56**, 87–127.

Hollis, D. & Broomfield, P.S. (1989) *Ficus*-feeding psyllids (Homoptera), with special reference to the Homotomidae. *Bulletin of the British Museum (Natural History) Entomology*, **58**, 131–183.

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.

Kearse, M., Moir, R., Wilson, A. *et al.* (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**, 1647–1649.

Kistner, E.J., Amrich, R., Castillo, M., Strode, V. & Hoddle, M.S. (2016) Phenology of Asian citrus psyllid (Hemiptera: Liviidae), with special reference to biological control by *Tamarixia radiata*, in the residential landscape of Southern California. *Journal of Economic Entomology*, **109**, 1047–1057.

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.

Lemmon, A.R., Emme, S.A. & Lemmon, E.M. (2012) Anchored Hybrid Enrichment for massively high-throughput phylogenomics. *Systematic Biology*, **61**, 717–726.

Lohse, M., Bolger, A.M., Nagel, A. *et al.* (2012) RobiNA: a user-friendly, integrated software solution for RNA-Seqbased transcriptomics. *Nucleic Acids Research*, **40**, W622–W627.

Margulies, M., Egholm, M., Altman, W.E. *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, **437**, 376–380.

Marlétaz, F., Le Parco, Y., Liu, S. & Peijnenburg, K.T.C.A. (2017) Extreme mitogenomic variation in natural populations of chaetognaths. *Genome Biology and Evolution*, **9**, 1374–1384.

Martin, J.H. & Hollis, D. (1992) The *Calophyllum*-feeding triozid genus *Leptynoptera* (Hemiptera: Psylloidea). *Journal of Natural History*, **26**, 555–585.

Martoni, F., Bulman, S.R., Pitman, A. & Armstrong, K.F. (2017) Elongation Factor-1α accurately reconstructs relationships amongst psyllid families (Hemiptera: Psylloidea), with possible diagnostic implications. *Journal of Economic Entomology*, **110**, 2618–2622.

Meyer, M. & Kircher, M. (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, **6**, 1–10.

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

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

Myers, E.W., Sutton, G.G., Delcher, A.L. *et al.* (2000) A whole-genome assembly of *Drosophila*. *Science*, **287**, 2196–2204.

Nabity, P.D. (2016) Insect-induced plant phenotypes: revealing mechanisms through comparative genomics of galling insects and their hosts. *American Journal of Botany*, **103**, 1–3.

Nelson-Sathi, S., Sousa, F.L., Roettger, M. *et al.* (2015) Origins of major archaeal clades correspond to gene acquisitions from bacteria. *Nature*, **517**, 77–80.

Nissinen, A.I., Lemmetty, A., Pihlava, J.-M., Jauhiainen, L., Munyaneza, J.E. & Vanhala, P. (2012) Effects of carrot psyllid (*Trioza apicalis*) feeding on carrot yield and content of sugars and phenolic compounds. *Annals of Applied Biology*, **161**, 68–80.

Ouvrard, D. (2016) *Psyl'list: The World Psylloidea Database* [WWW document]. URL http://www.hemiptera-databases.com/psyllist/ [accessed on 20 September 2016].

Ouvrard, D., Chalise, P. & Percy, D.M. (2015) Host-plant leaps versus host-plant shuffle: a global survey reveals contrasting patterns in an oligophagous insect group (Hemiptera, Psylloidea). *Systematics and Biodiversity*, **13**, 434–454.

Peng, Y., Leung, H.C.M., Yiu, S.M. & Chin, F.Y.L. (2012) IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*, **28**, 1420–1428.

Percy, D.M. (2003) Radiation, diversity, and host-plant interactions among island and continental legume-feeding psyllids. *Evolution*, **57**, 2540–2556.

Percy, D.M. (2017) Making the most of your host: the *Metrosideros*-feeding psyllids (Hemiptera, Psylloidea) of the Hawaiian Islands. *ZooKeys*, **649**, 1–163.

Perilla-Henao, L.M. & Casteel, C.L. (2016) Vector-borne bacterial plant pathogens: interactions with hemipteran insects and plants. *Frontiers in Plant Science*, **7**, 1163.

Prum, R.O., Berv, J.S., Dornburg, A. *et al.* (2015) A comprehensive phylogeny of birds (Aves) using targeted nextgeneration DNA sequencing. *Nature*, **526**, 569–573.

Rodríguez, A., Burgon, J.D., Lyra, M. *et al.* (2017) Inferring the shallow phylogeny of true salamanders (*Salamandra*) by multiple phylogenomic approaches. *Molecular Phylogenetics and Evolution*, **115**, 16–26.

Rokyta, D.R., Lemmon, A.R., Margres, M.J. & Arnow, K. (2012) The venom-gland transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). *BMC Genomics*, **13**, 312.

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

Rothfels, C.J., Larsson, A., Kuo, L.-Y., Korall, P., Chiou, W.-L. & Pryer, K.M. (2012) Overcoming deep roots, fast rates, and short internodes to resolve the ancient rapid radiation of eupolypod II ferns. *Systematic Biology*, **61**, 490–509.

Schmieder, R. & Edwards, R. (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, **27**, 863–864.

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

Sjölund, M.J., Clark, M., Carnegie, M. *et al.* (2017) First report of '*Candidatus* Liberibacter solanacearum' in the United Kingdom in the psyllid *Trioza anthrisci*. *New Disease Reports*, **36**, 4.

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

Stamatakis, A., Hoover, P. & Rougemont, J. (2008) A rapid bootstrap algorithm for the RAxML web-servers. *Systematic Biology*, **75**, 758–771.

Stout, C.C., Tan, M., Lemmon, A.R., Lemmon, E.M. & Armbruster, J.W. (2016) Resolving Cypriniformes relationships using an anchored enrichment approach. *BMC Evolutionary Biology*, **16**, 244.

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. & Willerslev, E. (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, **21**, 2045–2050.

Tang, M., Tan, M., Meng, G. *et al.* (2014) Multiplex sequencing of pooled mitochondrial genomes - a crucial step toward biodiversity analysis using mito-metagenomics. *Nucleic Acids Research*, **42**, e166.

Timmermans, M.J., Dodsworth, S., Culverwell, C.L. *et al.* (2010) Why barcode? High-throughput multiplex sequencing of mitochondrial genomes for molecular systematics. *Nucleic Acids Research*, **38**, e197.

Valente, L.M., Etienne, R.S. & Phillimore, A.B. (2014) The effects of island ontogeny on species diversity and phylogeny. *Proceedings of the Royal Society B: Biological Sciences*, **281**, 20133227.

Watson, R.A., Wagner, G.P., Pavlicev, M., Weinreich, D.M. & Mills, R. (2014) The evolution of phenotypic correlations and "developmental memory". *Evolution*, **68**, 1124–1138.

White, I.M. & Hodkinson, I.D. (1985) Nymphal taxonomy and systematics of the Psylloidea (Homoptera). *Bulletin of the British Museum (Natural History) Entomology*, **50**, 123–301.

Wu, F., Kumagai, L., Cen, Y. *et al.* (2017) Analyses of mitogenome sequences revealed that Asian Citrus Psyllids (*Diaphorina citri*) from California were related to those from Florida. *Scientific Reports*, **7**, 10154.

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

Accepted 15 March 2018

