



Molecular Phylogenetics, Phylogenomics, and Phylogeography

Mining Ultraconserved Elements From Transcriptome and Genome Data to Explore the Phylogenomics of the Free-living Lice Suborder Psocomorpha (Insecta: Psocodea)

Oscar Fernando Saenz Manchola,^{1,4,✉} Ernesto Samacá Sáenz,¹
Stephany Virrueta Herrera,^{2,3} Lorenzo Mario D'Alessio,^{2,3} Alfonso Nerí García Aldrete,¹
and Kevin P. Johnson²

¹Instituto de Biología, Universidad Nacional Autónoma de México, Mexico City, Mexico, ²Illinois Natural History Survey, Prairie Research Institute, University of Illinois, Champaign, IL, USA, ³Program in Ecology, Evolution, and Conservation Biology, University of Illinois at Urbana-Champaign, Champaign, IL, USA, and ⁴Corresponding author, e-mail: oscar.saenz@st.ib.unam.mx

Subject Editor: Kazunori Yoshizawa

Received 27 January, 2022; Editorial decision 10 April, 2022

Abstract

The order Psocodea includes the two historically recognized groups Psocoptera (free-living bark lice) and Phthiraptera (parasitic lice) that were once considered separate orders. Psocodea is divided in three suborders: Trogiomorpha, Troctomorpha, and Psocomorpha, the latter being the largest within the free-living groups. Despite the increasing number of transcriptomes and whole genome sequence (WGS) data available for this group, the relationships among the six known infraorders within Psocomorpha remain unclear. Here, we evaluated the utility of a bait set designed specifically for parasitic lice belonging to suborder Troctomorpha to extract UCE loci from transcriptome and WGS data of 55 bark louse species and explored the phylogenetic relationships within Psocomorpha using these UCE loci markers. Taxon sampling was heavily focused on the families Lachesillidae and Elipsocidae, whose relationships have been problematic in prior phylogenetic studies. We successfully recovered a total of 2,622 UCE loci, with a 40% completeness matrix containing 2,081 UCE loci and an 80% completeness matrix containing 178 UCE loci. The average number of UCE loci recovered for the 55 species was 1,401. The WGS data sets produced a larger number of UCE loci (1,495) on average than the transcriptome data sets (972). Phylogenetic relationships reconstructed with Maximum Likelihood and coalescent-based analysis were concordant regarding the paraphyly of Lachesillidae and Elipsocidae. Branch support values were generally lower in analyses that used a fewer number of loci, even though they had higher matrix completeness.

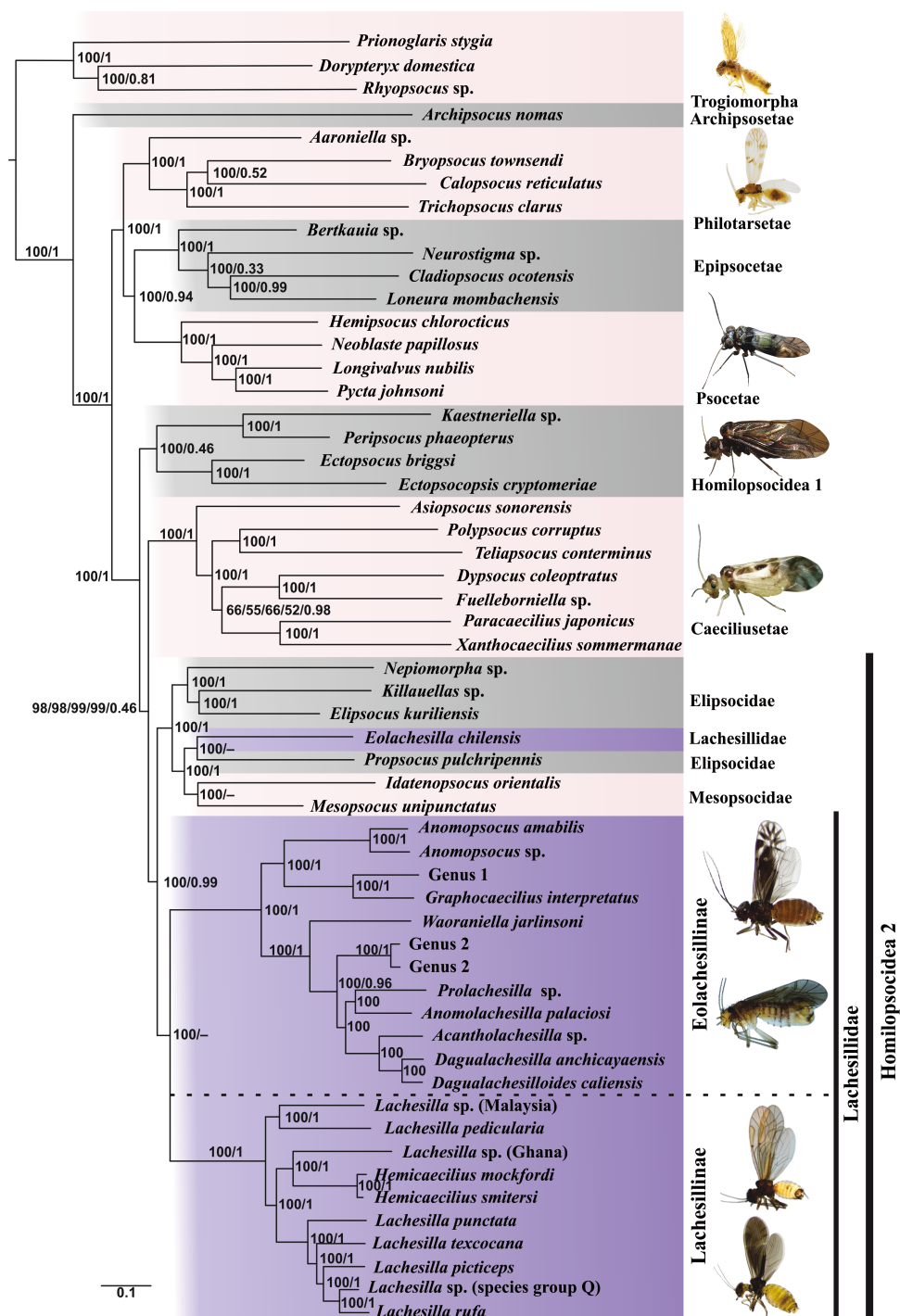
Resumen

El orden Psocodea incluye actualmente a dos grupos históricamente reconocidos y que una vez fueron considerados órdenes separados, Psocoptera (piojos de vida libre o de las cortezas) y Phthiraptera (piojos verdaderos). Psocodea está dividido en tres subórdenes: Trogiomorpha, Troctomorpha y Psocomorpha, este último siendo el más grande entre los piojos de vida libre. A pesar de que la cantidad de información disponible sobre transcriptomas y secuenciación del genoma completo (WGS) para este grupo se ha incrementado notablemente en los últimos años, las relaciones filogenéticas dentro de Psocomorpha permanecen poco claras. En este estudio, evaluamos la utilidad de un conjunto de sondas diseñadas específicamente a partir de especies de piojos verdaderos del suborden Troctomorpha, para capturar elementos ultra-conservados (UCE) a partir de las secuencias de transcriptomas y WGS de 55 especies de piojos de las cortezas. Igualmente, exploramos las relaciones filogenéticas dentro de Psocomorpha usando estos marcadores de UCE. El muestreo taxonómico estuvo fuertemente enfocado en las familias Lachesillidae y Elipsocidae, ya que sus relaciones han demostrado ser problemáticas en estudios filogenéticos previos. Como resultado, logramos recuperar exitosamente un total de 2,622 marcadores de UCE, con las matrices de completitud del 40% y 80% conteniendo 2,081 y 178 marcadores de UCE respectivamente. El número promedio de UCE recuperados para las 55 especies fue de 1,401. En promedio, el conjunto de datos de WGS produjo

un mayor número de loci de UCE (1,495) que las secuencias de transcriptomas (972). Las relaciones filogenéticas reconstruidas a partir de análisis de máxima verosimilitud y métodos coalescentes fueron concordantes respecto a la parafilia de Lachesillidae y Elipsocidae, mientras que los valores de soporte de ramas fueron generalmente más bajos en los análisis que incluyeron menor número de loci aún cuando la matriz de completitud era más grande.

Palabras claves: Psocidos, Homilopsocidea, Lachesillidae, relaciones filogenéticas, transcriptomas

Graphical Abstract



Key words: Psocids, Homilopsocidea, Lachesillidae, phylogenetic relationship, transcriptome

High throughput DNA sequencing technologies, coupled with improvements in high performance computing, have contributed to substantial advances in the field of evolutionary biology, leading to an era of big data, and improving our understanding of the tree of life (Andermann et al. 2020). Genome-scale studies are rapidly supplanting Sanger-based sequencing, dwarfing previous approaches in the sheer scale of data they generate. However, sometimes the cost and scale of these genome-scale approaches are prohibitive for many researchers, leading to so-called genome reduction approaches as an effective alternative to generate datasets with thousands of loci, at relatively low cost, for model and nonmodel taxa alike (Zhang et al. 2019b). To date, most of the approaches using next-generation sequencing technology to develop insect nuclear phylogenomic datasets have focused on genome reduction techniques rather than sequencing the entire genome (Johnson 2019).

Several genome reduction techniques have been widely used to study phylogenetic relationships at both deep and shallow scales, including transcriptome sequencing (Misof et al. 2014; Peters et al. 2014, 2017; Zhang et al. 2016, Johnson et al. 2018, Pauli et al. 2018, Simon et al. 2018, Vasilakopoulos et al. 2019, Wang et al. 2019, Wipfler et al. 2019), ultraconserved elements (UCE: Faircloth et al. 2012), and anchored hybrid enrichment (AHE: Lemmon et al. 2012). Recently Johnson (2019) compared these phylogenomic approaches, providing several examples of the use of AHE and UCE approaches across several insect groups ranging from Hemiptera, Hymenoptera, Lepidoptera, Coleoptera, Neuroptera, to Diptera. The main differences between AHE and UCE approaches is that AHE focuses on capturing exonic regions and fewer loci with a methodology in part proprietary, while UCs may capture both exonic and intronic regions for more loci with a fully open source bioinformatics pipeline to design and extract targeted regions (Zhang et al. 2019b). UCE markers have been developed for a large number of insect groups, so there is increasing interest in using these markers for arthropod phylogenomics (Zhang et al. 2019b). In addition, UCE markers have proved to be useful in resolving phylogenies at multiple phylogenetic scales (Van Dam et al. 2018).

Given the number of different approaches for developing phylogenomic datasets, the capability of combining large datasets across various sources will become increasingly important to generate a complete tree of life (Kieran et al. 2019). Combination of transcriptome and UCE data has been recently applied in several insect groups, including Hemiptera (Kieran et al. 2019), Hymenoptera (Bossert et al. 2019; Apidae), Diptera (Cohen et al. 2021; Asilidae; Buenaventura, 2021; Sarcophagidae, Buenaventura et al. 2021; Oestroidea), and other arthropods (Hedin et al. 2019; Araneae, Mygalomorphae). Regarding the order Psocodea (true and bark/book lice), there is only a single phylogenomic study that has successfully combined transcriptomes with single copy orthologous genes derived from data mining of whole genome sequence (WGS) (de Moya et al. 2020). Moreover, although a bait set designed to capture UCE loci has been proved *in silico* within the infraorder Phthiraptera (parasitic lice) (Zhang et al. 2019a), it has not been used to date with free living lice (bark lice).

The order Psocodea encompasses the two historically recognized groups, Psocoptera (free-living bark lice) and Phthiraptera (parasitic lice), that were once considered separate orders. Species in these groups have a range of feeding preferences, from detritus, plant material (i.e., pollen, decaying leaves), and microflora (i.e., cyanobacteria films, fungal and lichen) in nonparasitic members; to obligate ectoparasitism on birds and mammals (i.e., skin debris, feathers, blood and skin secretions; Clayton et al. 2016). Psocodea is divided into three suborders:

Trogiomorpha, Troctomorpha, and Psocomorpha, the latter being the largest within the free-living groups, with over 3,600 species in 25 families (Lienhard & Smithers, 2002; Yoshizawa and Johnson 2014). To date, six infraorders have been recognized within Psocomorpha: Archipsocetae, Caeciliusetae, Epipsocetae, Homilopsocidea, Philotarsetae, and Psocetae (Yoshizawa and Johnson 2014). However the relationships among infraorders remain unresolved, since neither morphology (Yoshizawa 2002) nor an extensive taxon sampling using Sanger data (Yoshizawa and Johnson 2014) have resolved deep relationships. More recently, a phylogenomic study based on 2,370 loci (de Moya et al. 2020) recovered generally stable relationships within Psocomorpha, although the monophyly of the infraorder Homilopsocidea was not supported and the relationships among this infraorder and the infraorder Caeciliusetae were unstable.

Given the increasing number of transcriptomes and WGS data available for the species of free-living bark lice, plus the recent development of a bait set targeting 2,832 UCE loci designed for parasitic lice (Phthiraptera) (Zhang et al. 2019a), here we evaluated the utility of this bait set to capture UCE loci from transcriptomes and WGS for species of the suborder Psocomorpha. Similarly, we explored the phylogenetic relationships within Psocomorpha across a variety of taxonomic scales (within genera to among infraorders) as a result of combining UCE loci extracted from both transcriptomes and WGS. This study constitutes the first within Psocodea that successfully combines different sources of genomic data, expanding the diversity of loci available for future analysis, which may help to reduce sequencing costs and computing resources to perform phylogenomic analyses within this group of insects.

Material and Methods

Taxon Sampling and Data Generation

For the UCE phylogenomic analyses, a total 55 species were available for DNA extraction. WGS data from 35 species (22 from the family Lachesillidae, representing two subfamilies and 13 genera, Table 1) were generated by extracting total genomic DNA using a Qiagen DNeasy Microkit. Library preparation and Illumina sequencing were conducted at the Roy J. Carver Biotechnology Center at the University of Illinois. A Covaris M220 machine was used to sonicate DNA fragments to approximately 300–500 bp. Libraries were prepared using a Hyper Library construction kit from Kapa Biosystems. Libraries were quantified by qPCR and pooled for sequencing using Illumina HiSeq2500 or NovaSeq6000 S4 lanes for 151 cycles. The bcl2fastq v2.20 Conversion Software was used to demultiplex and generate FASTQ files. Raw reads were deposited in the NCBI Sequence Read Archive (SRA, Accession release pending acceptance, Table 1).

Additionally, WGS data from 10 species (belonging to nine families and 10 genera) and transcriptomes sequences from 10 species (representing eight families and 10 genera) were downloaded from NCBI (Table 1). The total number of samples included species from the suborders Trogiomorpha (three outgroup species) and Psocomorpha (bark lice), representing 19 families and 36 genera (two undescribed). Given the large number of reads, raw data were subsampled at 5 million reads with reformat.sh (Bushnell, 2014). WGS data were trimmed with Trimmomatic (Bolger et al. 2014) implemented in PHYLUCE v1.7.1 (Faircloth 2016), while transcriptome data were trimmed with a pipeline designed by Zhang et al. (2019a, b) which implements BBTools, including remove duplicates (clumpify.sh), both sides trim to Q15 (bbduk.sh), discard reads

Table 1. Taxa used in this study

Suborder	Infraorder	Family	Species	Accession
Psocomorpha	Homilopsocidea	Lachesillidae	<i>Eolachesilla chilensis</i> Badonnel, 1967	PRJNA555288
			<i>Acantholachesilla</i> sp.	PRJNA555306
			<i>Anomolachesilla palaciosi</i> García Aldrete, González & Carrejo, 2012	PRJNA555303
			<i>Anomopsocus amabilis</i> Walsh, 1862	SRR5308259**
			<i>Anomopsocus</i> sp.	PRJNA555296
			<i>Dagualachesilla anchicayaensis</i> García Aldrete, González & Carrejo, 2013	PRJNA555301
			<i>Dagualachesilloides caliensis</i> García Aldrete, González & Carrejo, 2013	PRJNA555308
			<i>Graphocaeciliini</i> gen. nov. 2.	PRJNA555309
			<i>Graphocaeciliini</i> gen. nov. 1.	PRJNA555300
			<i>Graphocaeciliini</i> gen. nov. 2.	PRJNA555305
			<i>Graphocaecilius interpretatus</i> Roesler, 1940	PRJNA555307
			<i>Prolachesilla</i> sp.	PRJNA555299
			<i>Waoraniella jarlinsoni</i> Saenz Manchola, González & García Aldrete, 2018	PRJNA555304
			<i>Hemicaecilius mockfordi</i> García Aldrete, González & Carrejo, 2012	PRJNA555302
			<i>Hemicaecilius smithersi</i> García Aldrete, González & Carrejo, 2012	PRJNA555310
			<i>Lachesilla pedicularia</i> Linnaeus, 1758	PRJNA555268
			<i>Lachesilla picticeps</i> Mockford, 1986	PRJNA555295
			<i>Lachesilla punctata</i> Banks, 1905	PRJNA555297
			<i>Lachesilla rufa</i> Walsh, 1863	PRJNA555292
			<i>Lachesilla</i> sp. Ghana	PRJNA555272
			<i>Lachesilla</i> sp. Malasya	PRJNA555289
			<i>Lachesilla</i> sp. 7	PRJNA555293
			<i>Lachesilla texcocana</i> García Aldrete, 1972	PRJNA555294
		Ectopsocidae	<i>Ectopsocopsis cryptomeriae</i> Enderlein, 1907	PRJNA555282
		Peripsocidae	<i>Kaestneriella</i> sp.	PRJNA555265
		Elipsocidae	<i>Kilauella</i> sp.	SRR5308272**
			<i>Nepiomorpha</i> sp.	SRR5308276**
			<i>Propocus pulchripennis</i> Perkins, 1899	SRR5308281**
		Mesopsocidae	<i>Idatenopsocus orientalis</i> Vishnyakova, 1986	SRR5308271**
	Caeciliusetae	Asiopsocidae	<i>Asiopsocus sonorensis</i> Mockford & García Aldrete, 1976	SRR5308261**
		Caeciliusidae	<i>Dypsocus coleoptratus</i> Hagen, 1858	PRJNA555285
			<i>Fuelleborniella</i> sp.	PRJNA555311
		Paracaeciliidae	<i>Paracaecilius japonicus</i> Enderlein, 1906	PRJNA555290
			<i>Xanthocaecilius sommermanae</i> Mockford, 1955	SRR5308288**
		Amphipsocidae	<i>Polypsocus corruptus</i> Hagen, 1861	PRJNA555266
		Dasydemellidae	<i>Teliapsocus cotermis</i> Walsh, 1863	PRJNA555312
	Archipsocetae	Archipsocidae	<i>Archipsocus nomas</i> Gurney, 1939	SRR5308260**
	Epipsocetae	Cladiopsocidae	<i>Cladiopsocus ocotensis</i> García Aldrete, 1996	SRR5308265**
		Ptiloneuridae	<i>Loneura mombachensis</i> Garcia Aldrete, 2000	SRR5308274**
		Epipsocidae	<i>Bertkauia</i> sp.	SRR2051473*
			<i>Neurostigma</i> sp.	SRR5308277**
	Homilopsocidea	Ectopsocidae	<i>Ectopsocus briggsi</i> McLachlan, 1899	SRR645929*
		Elipsocidae	<i>Elipsocus kuriliensis</i> Vishnyakova, 1986	SRR2051485*
		Mesopsocidae	<i>Mesopsocus unipunctatus</i> Müller, 1764	SRR2051502*
		Peripsocidae	<i>Peripsocus phaeopterus</i> Stephens, 1836	SRR2051507*
	Philotarsetae	Philotarsidae	<i>Aaroniella</i> sp.	SRR2051465*
		Pseudocaeciliidae	<i>Calopsocus reticulatus</i> Thornton & Smithers, 1984	SRR5308264**
			<i>Bryopsocus townsendi</i> Smithers, 1969	SRR5308263**
		Trichopsocidae	<i>Trichopsocus clarus</i> Banks, 1908	SRR5308287**
	Psocetae	Hemipsocidae	<i>Hemipsocus chloroticus</i> Hagen, 1858	SRR2051492*
		Psocidae	<i>Neoblaste papillosus</i> Thornton, 1960	SRR2051505*
			<i>Longivalvus nubilus</i> Enderlein, 1906	SRR2051498*
			<i>Ptycta johnsoni</i> Bess & Yoshizawa, 2007	SRR1821962*
Trogomorpha	Prionoglaridetae	Prionoglarididae	<i>Prionoglaris stygia</i> Enderlein, 1909	SRR5308282**
	Psyllipsocetae	Psyllipsocidae	<i>Dorypteryx domestica</i> Smithers, 1958	SRR5308267**
	Atropetae	Psoquillidae	<i>Rhyopsocus</i> sp.	SRR5308284**

Accession numbers marked with (*) refer to transcriptomes and (**) refers to previously sequenced data sources.

shorter than 15 bp or with more than 5 Ns, trim poly-A or poly-T tails of at least 10 bp, overlapping paired reads correction and depth normalization at 10× (bbnorm.sh). Finally, FastQC version 0.11.9 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) was used to check quality, GC content, over-represented sequences, and duplications levels of the sequences.

Data Assembly and UCE Extraction

Assemblies for WGS data were obtained with ABySS version 1.3.6 (Simpson et al. 2009) implemented in PHYLUC v1.7.1 (Faircloth 2016), while transcriptomes were assembled with rnaSPAdes version 3.9.0 (Bushmanova et al. 2019) in Galaxy server (<https://usegalaxy.org/>). The UCE probe set Phthiraptera-2.8Kv1 (containing 55,030

baits targeting 2,832 loci) which was primarily designed by Zhang et al. (2019a, b) for parasitic lice, was used to find UCE loci and test their utility for resolving family-level relationships within the sub-order Psocomorpha.

We followed the Tutorial I from PHYLUCE web server (<https://phyluce.readthedocs.io/en/latest/tutorials/tutorial-1.html#>) to match the resulting contigs to the probes using LASTZ (Harris 2007) and the script 'phyluce_assembly_match_contigs_to_probes' with parameters *-min_coverage* and *-min_identity* set to 65, additionally we explored the effect of minimum coverage and minimum identity using default parameters (*-min_coverage* and *-min_identity* set to 80). UCE loci extraction, paralogues removed, and individual UCE loci FASTA format were generated using the scripts 'phyluce_assembly_get_match_counts' and 'phyluce_assembly_get_fastas_from_match_counts' respectively. Individual loci were aligned with MAFFT version 7.130 b (Katoh and Standley 2013) and only internal trimming was performed with 'phyluce_align_seqcap_align' and *-no-trim* flag on. Finally, the resulting alignments were filtered and trimmed with Gblocks version 0.91 b (Castresana 2000) and default parameters (0.5, 0.85, 8, and 10 for the *b1-b4* options).

Data Matrices Generation and Phylogenetic Analysis

Phylogenomic analysis was conducted with five datasets, each corresponding to a concatenated data matrix generated with the script 'phyluce_align_get_only_loci_with_min_taxa' (Supplementary Data 1–5 [online only]), with completeness that corresponds to 80–70–60–50 and 40% of the total taxa (this means that, in this study of 55 taxa, 80% complete matrix, will contain at least 44 of these 55 taxa). Given that matrix with 90% loci completeness only contains an alignment with seven loci, this data set was not used in the subsequent phylogenetic analysis. To improve model-fit and parameter estimates, all data matrices were partitioned with the Sliding-Window Site Characteristics (SWSC) and UCE Site Position (UCESP) methods (Tagliacollo and Lanfear 2018) using default parameters. These partition scheme methods account for within-UCE heterogeneity. SWSC uses proxies of rates and patterns of molecular evolution and a sliding-window approach to determine whether a central region of a UCE evolves in a different way than the two flanking regions, which basically splits each UCE into three data blocks corresponding to the conserved cores and more variable flanking regions. UCESP groups nucleotide sites across UCEs using their physical location within

the UCE, such as grouping all of the central sites of each UCE into a single partition (Tagliacollo and Lanfear 2018, Zhang et al. 2019a, b).

For each concatenated supermatrix, a maximum likelihood (ML) phylogenetic analysis was conducted with IQ-TREE 2 v2.1.3 (Minh et al. 2020). The best fit models for each partition obtained with SWSC and UCESP methods, were estimated with ModelFinder (Kalyaanamoorthy et al. 2017), followed by tree reconstruction (Nguyen et al. 2015) using the best partitioning scheme founded (*-m* TESTNEWMERGE) and the fast relaxed clustering algorithm (*-rclusterf* 10) (Lanfear et al. 2017) parameters. Tree support was estimated using ultrafast bootstrapping with UFBoot2 (*-bb* 1000) (Minh et al. 2013, Hoang et al. 2018). Additionally, a coalescent species-tree analysis in ASTRAL-III v5.7.7 (Zhang et al. 2018) was performed for each of the completeness datasets, using as input the gene trees generated from the UCE loci with IQ-TREE 2 v2.1.3 (Minh et al. 2020) for each of the five datasets, with local posterior probabilities (LPP) computed for branch support (Sayyari and Mirarab 2016).

Results

Data Assembly and UCE Features

The ABySS assembly for the WGS data produced an average of 649,934 (1,154,557 max; 39,031 min) contigs, while the RNAspades assembly for transcripts produced an average of 70,609 (108,242 max; 31,419 min) contigs (Tables 2 and 3), with a total average 350 bp contig length. From the 2,832 loci targeted by the probe set Phthiraptera-2.8Kv1, a total of 2,622 UCE loci were recovered in the incomplete matrix, with an average of 1,401 UCE loci across all taxa. This number of loci was larger for genomic data (1,495) than from transcriptomes (972) (Table 2). Within UCE data sets derived from genomic sequencing data, the lowest and the highest numbers of UCE loci captured were *Hemicaecilius mockfordi* Garcia Aldrete, González & Carrejo (Psocodea: Lachesillidae) and *Trichopsocus clarus* Banks (Psocodea: Trichopsocidae), with 1,906 and 740 UCE loci respectively. For transcriptome data, the lowest number of UCE loci recovered belongs to *Elipsocus kuriliensis* Vishnyakova (Psocodea: Elipsocidae) and the highest number belongs to *Ectopsocus briggsi* McLachlan (Psocodea: Ectopsocidae), with 593 and 1,214 loci respectively (Supp Table 1 [online only]).

Starting from the matrix containing 2,622 loci, the five data matrices analyzed (80–40% completeness, Supplementary Data 1–5 [online only]) contained between 178 and 2,081 UCE loci, between

Table 2. Contigs statistics resulting from each source of genetic data used in this study

Sequence source	No. species	Contigs bp	Contigs bp mean	UCE mean	UCE avg. length
WGS	46	649934 (1154557-39031)	466 (1741-157)	1495 (1906-740)	1286 bp
Transcriptomes	10	70609 (108242-31419)	233 (272-157)	972 (1214-593)	634 bp

Parenthetical numbers are maximum–minimum values (max–min).

Table 3. Nucleotide characteristics for each matrix of varying UCE loci completeness

Completeness matrix	No. UCE	No. characters	% Missing data	Informative sites	GC content
80%	178	71,029	35.51	44,276	43.8
70%	653	250,092	39.69	154,622	43.7
60%	1248	449,633	43.35	273,504	43.9
50%	1650	570,852	45.96	341,266	44
40%	2081	690,689	48.95	406,981	44
Incomplete matrix	2622	852,725	18.54	470,719	43.8

71,022 and 690,689 aligned base pairs and the percent of missing data ranged from 35.81% to 48.95% (Table 3). GC content was 43.8% (44% for 40–50 completeness matrix and 43.7% for 80 completeness matrix, Table 3), with *Eolachesilla chilensis* Badonnel (Psocodea: Lachesillidae) being the species with the highest GC content in all the five completeness matrices analyzed, ranged from 52% (40 completeness matrix) to 51.5% (80 completeness matrix) (Supplementary Table 2 [online only]).

Phylogenetic Analysis

Phylogenetic relationships resulting from the maximum likelihood (ML) analyses showed topologies generally consistent across all the matrices, but ultra-fast bootstrap support values (UFB) were lower for the matrices with fewer number of loci included in the analysis (e.g., 80% completeness matrix: 178 loci vs 40% completeness matrix: 2081 loci) (Fig. 1; Supp Figs. 1–4 [online only]). The infraorder Archipsocetae (represented by *Archipsocus nomas* Gurney [Psocodea: Archipsocidae]) was recovered as sister to all the remaining infraorders of Psocomorpha. Similarly, all ML analyses recovered the infraorders Elipsocetae and Psocetae as sister taxa, together in a clade with Philotarsetae. On the other hand, Homilopsocidea was not recovered as monophyletic, forming two major separated clades: Homilopsocidea 1 (Homilo 1), grouping the monophyletic families Peripsocidae and Ectopsocidae (UFB = 100) and Homilopsocidea 2 (Homilo 2), containing the families Elipsocidae, Mesopsocidae, and Lachesillidae. The infraorder Caeciliusetae was recovered as monophyletic (UFB = 100) and sister to Homilo 2 with moderate–high bootstrap support values (UFB = 89–100, for 80% to 40% completeness matrices, Fig. 1 and Supp Figs. 1–4 [online only]).

Relationships within the infraorder Caeciliusetae indicated Asiopsocidae (*Asiopsocus sonorensis* Mockford & Garcia Aldrete, Psocodea: Asiopsocidae) as sister to the remaining taxa in this group. Within this latter clade, the families Dasydemellidae and Amphipsocidae were sister taxa and sister to a clade containing Caeciliusidae and Paracaciliidae. However, within these latter group relationships were somewhat unstable across data sets (Fig. 1, Supp Figs. 1–4 [online only]).

Within the group Homilo 2, relationships were unstable, and monophyly of Lachesillidae and Elipsocidae was not recovered (Figs. 1 and 2; Supp Figs. 1–4 [online only]). Despite this, a highly supported and consistent group formed by *E. chilensis* and species belonging to Mesopsocidae and Elipsocidae was recovered by all analysis and datasets (Subclade E). Within subclade E, matrices that included more UCE loci (40–60%, 2081–1248 loci; Fig. 1; Supp Figs. 1–2 [online only]), always grouped *Prospocus pulchripennis* Perkins (Psocodea: Elipsocidae) and *Eolachesilla chilensis* (Lachesillidae: Eolachesillinae) as sister to Mesopsocidae with high branch support (UFB = 100), while matrices with fewer UCE loci (70–80%, 653–178 loci), showed *P. pulchripennis* as sister to the *E. chilensis* + Mesopsocidae species clade, but poorly supported (Supp Figs. 3–4 [online only]). Similarly, relationships within Elipsocidae (excluding *P. pulchripennis*) were unstable and poorly supported with these datasets.

The remaining species of Lachesillidae (ex. *E. chilensis*) were always grouped together in a clade, but matrices with fewer UCE loci (70–80%) resulted in lower branch support values (UFB = 97–78, respectively). Subfamilies Eolachesillinae and Lachesillinae were recovered as monophyletic (UFB = 100 with all datasets). However, within Eolachesillinae, the monophyly of tribe Graphocaeciliini was not recovered, with the tribe Waoraniellini (*Waoraniella jarlinsoni* Saenz, García Aldrete & González [Psocodea: Lachesillidae]) sister to a clade grouping five monospecific genera and *Prolachesilla*

sp. (Psocodea: Lachesillidae), while *Anomopsocus* Roesler (Psocodea: Lachesillidae) formed a cluster with an undescribed genus (Genus 1) and *Graphocaecilius interpretatus* Roesler (Psocodea: Lachesillidae). Within Lachesillinae, the genus *Lachesilla* Westwood (Psocodea: Lachesillidae) was not recovered as monophyletic, with the *pedicularia* species group (*Lachesilla* sp. from Malaysia and *Lachesilla pedicularia* Linnaeus) sister to a clade that includes the genus *Hemicaecilius* Enderlein (Psocodea: Lachesillidae) plus the remaining *Lachesilla* species groups.

The ML trees and the coalescent trees estimated for each dataset were concordant regarding the paraphyly of Lachesillidae and Elipsocidae. Similarly, relationships within the subfamilies Eolachesillinae and Lachesillinae, and between the infraorders Philotarsetae, Epipsocetae, and Psocetae were concordant. The LPP values became lower as fewer UCE loci were included (Fig. 1; Supp Figs. 1–4 [online only]). Relationships within the Homilo 2 clade were highly discordant between each data set under the coalescent tree approach, but generally the monophyly of Eolachesillinae + Lachesillinae was not supported or supported with low LPP (50%: LPP = 0.66, 80%: LPP = 0.34; Fig. 2A). Monophyly of Mesopsocidae was not recovered under the coalescent analyses, clustering *Mesopsocus unipunctatus* Muller (Psocodea: Mesopsocidae) with *E. chilensis* (Fig. 2A). Additionally, the phylogenetic position of the Elipsocidae + Mesopsocidae + *E. chilensis* (subclade E) was highly unstable, sometimes clustered with Eolachesillinae (40–60–70% matrices), sister to Caeciliusetae (50% matrix), or grouped with a monophyletic Eolachesillinae + Lachesillinae clade (80% matrix), with low LPP values.

Discussion

Data Assembly and UCE Features

This study constitutes the first phylogeny for bark lice (Psocodea: Psocomorpha) based entirely on UCE loci extracted in silico from WGS and transcriptome data. These results also provide additional evidence regarding the utility of combining UCE and transcriptomes in phylogenomics as has been showed for several insects groups including hemipterans (Kieran et al. 2019), bees (Apidae, Bossert et al. 2019), assassin flies (Diptera: Asilidae, Cohen et al. 2021), flesh flies (Buenaventura 2021), Oestroidea flies (Buenaventura et al. 2021), and other arthropods as mygalomorph spiders (Araneae: Mygalomorphae, Hedin et al. 2019). However, it should be noted that UCE loci extracted from transcriptomes may be missing sequence data associated with genomic regions that are not translated, such as promoter regions or introns (Cohen et al. 2021).

Surprisingly, our analysis recovered on average 1,404 UCE loci for the psocomorphan species, even though we used a bait set designed entirely based on parasitic louse species (Zhang et al. 2019a, b, recovered between 2,125–2,272 UCE loci for the target species for which baits were designed), a group for which the common ancestor with Psocomorpha probably lived around 175mya (de Moya et al. 2020). This relatively high number of UCE loci for bark lice, could be influenced by the fact that we used a minimum coverage and identity of 65 for matching contigs to probes. These relaxed values had an impact on the final data matrices, increasing the number of UCE loci recovered compared to results obtained with default parameters (min. coverage and identity = 80, results not showed).

Phylogenetic Analysis

Relationships recovered with each of the data matrices and analyses were congruent with previous studies (at both deep and shallow phylogenetic levels). Here, with the only exception of

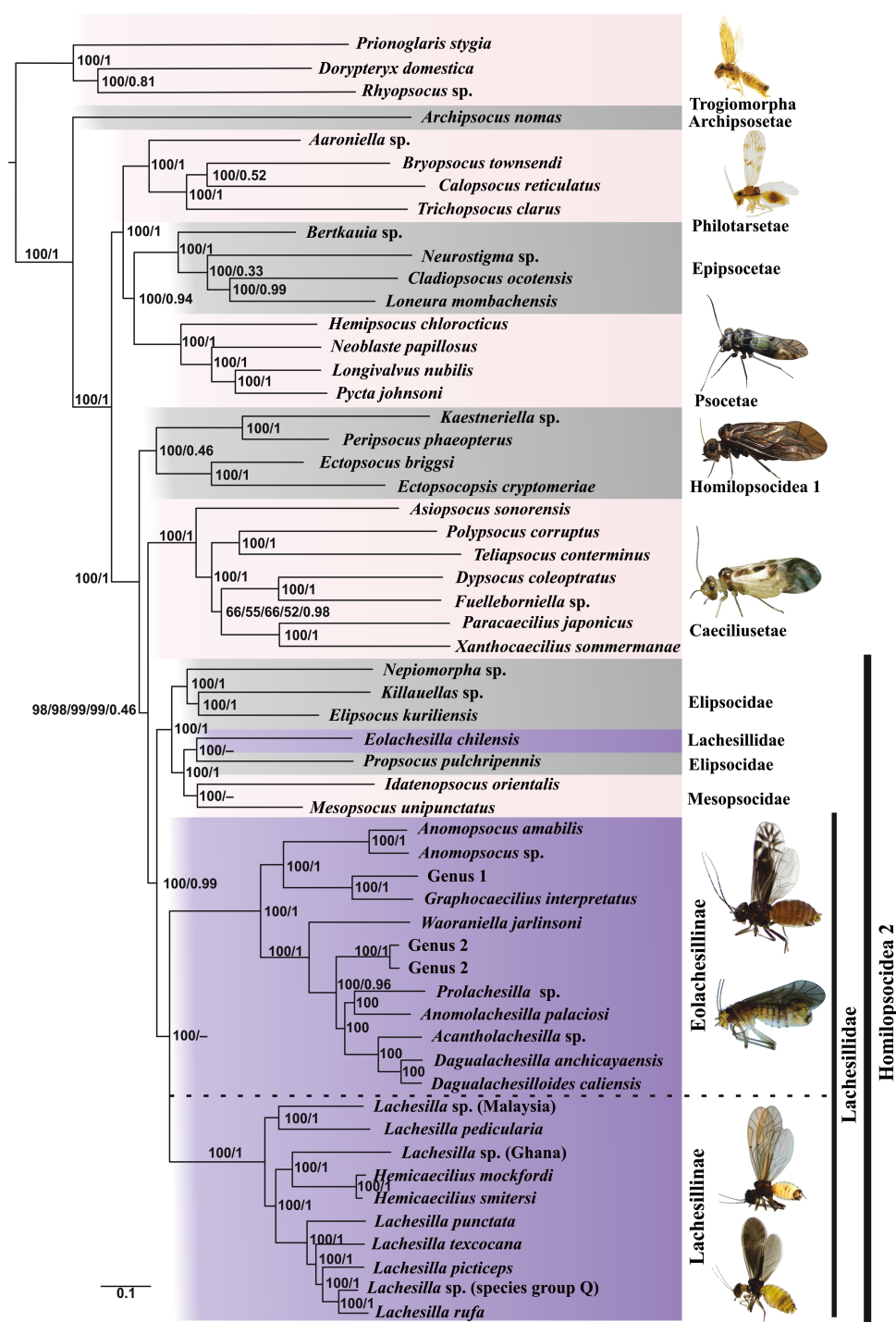


Fig. 1. Maximum Likelihood tree of Psocomorpha inferred from the 40% completeness matrix using iq-tree. Numbers associated with branches correspond to UFB support/LPP. If some difference in UFB were obtained depending the partition scheme, numbers are displayed as follows: SWSC-EN/SWSC-GC/SWSC-Multi/UCESP/LPP – indicate no UFB/LPP support. Dotted line separates Eolachesillinae versus Lachesillinae subfamilies.

Homilopsocidea, we recovered the monophyly for each of the infraorders of Psocomorpha analyzed. The relationships among most of the infraorders were also generally stable across analyses. Previous studies based on Sanger sequencing (Yoshizawa and Johnson 2014) and phylogenomics combining nuclear orthologous genes and transcriptome data (de Moya et al. 2020), showed that the infraorders Psocetidae and Epipsocetidae are sister groups, clustering together with Philotarsetae. Here, both ML and coalescent trees based on UCE loci recovered the same phylogenetic relationships. Within Epipsocetidae, similar to the 2,370 orthologous genes analysis (de Moya et al. 2020),

the monophyly of Epipsocidae was not recovered, with *Neurostigma* sp. (Epipsocidae) closest to the Cladiopsocidae + Ptiloneuridae clade, but with low LPP for all coalescent trees.

Within Caeciliusetae, relationships were generally stable, agreeing with previous Sanger analysis (e.g., monophyly of Caeciliusidae and Paracaeliidae) (Yoshizawa et al. 2014). We recovered Asiopsocidae as sister to all other Caeciliusetae species, and this is partially congruent with recent phylogenomic studies that recovered Asiopsocidae as sister to Paracaeliidae (de Moya et al. 2020). The ML analyses of all concatenated datasets always recovered Caeciliusetae as sister

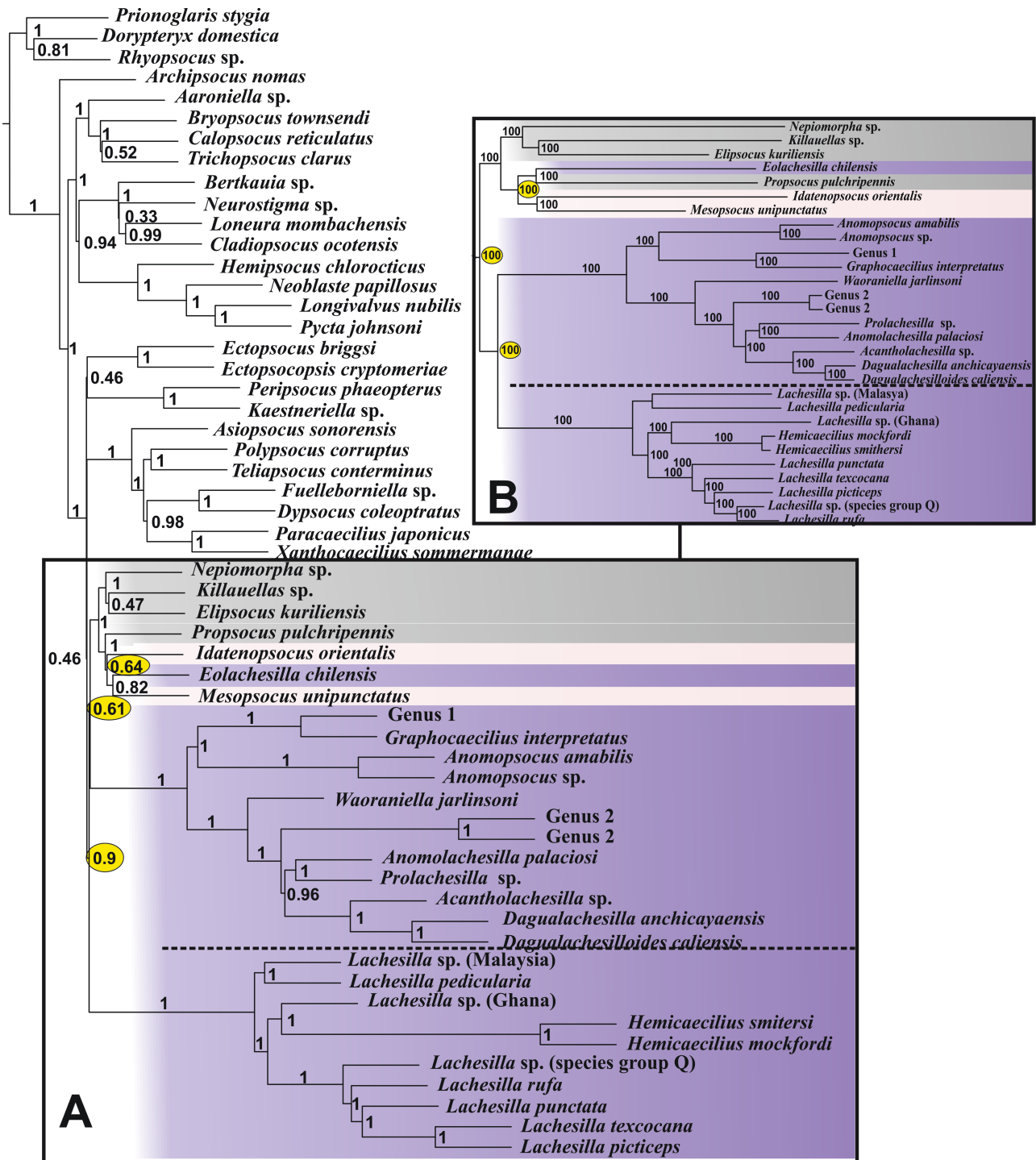


Fig. 2. Coalescent-based tree of Psocomorpha inferred using ASTRAL-III. (A) Topology obtained from the 40% completeness matrix. (B) Alternative topology obtained for 40% completeness matrix under ML approach. Numbers associated with branches correspond to LPP (A) or UFB (B). Yellow markers correspond to difference in tree topologies between (A) and (B). Dotted line separates Eolachesillinae versus Lachesillinae subfamilies.

to the Homilo 2 clade with high UFB. However, coalescent analyses showed some instability, with Caeciliusetae sometimes clustered with the Elipsocidae + Mesopsocidae + *E. chilensis* clade (50% completeness matrix, Supp Fig. 1 [online only]) or with Peripsocidae (80% completeness matrix, Supp Fig. 4 [online only]). These results were similar to the relationships obtained by the de Moya et al. (2020) phylogenomic study and by a recent study based on mitochondrial genomes (Saenz Manchola et al. 2021). In the first study, the position of Caeciliusetae within Psocomorpha is always associated

with Homilopsocidea either as sister to the families Ectopsocidae or Peripsocidae (here Homilo 1 clade); whereas in the study based on mitochondrial genomes Caeciliusetae was recovered closest to the subfamily Eolachesillinae, rendering Homilopsocidea paraphyletic.

Since the publication of the Yoshizawa and Johnson's ((2014) phylogenetic study based on Sanger data, the monophyly of Homilopsocidea has continued to be in question. Their analysis recognized an additional infraorder (Philotarsetae), to include species previously included within Homilopsocidea. Philotarsetae includes

species from the families Philotarsidae, Pseudocaeciliidae, and Trichopsocidae. Yoshizawa and Johnson (2014) also provided support for recognizing Archipsocetidae as a separate infraorder, as had been previously suggested by studies of morphological (Yoshizawa 2002) and molecular data (Johnson and Mockford 2003) that included species belonging to the family Archipsocidae. The remaining families of Homilopsocidea (Lachesillidae, Ectopsocidae, Peripsocidae, Mesopsocidae, and Elipsocidae) had phylogenetic relationships that were highly unstable, especially when *Lachesilla* (Lachesillidae) was included. This genus clustered with Peripsocidae as sister to Caeciliusetae and rendering Homilopsocidea paraphyletic. Our findings also support a paraphyletic Homilopsocidea and, as it was previously found by de Moya et al. (2020), the families Ectopsocidae and Peripsocidae (Homilo 1 clade) were recovered as sister to the Caeciliusetae + Homilo 2 clade in all ML concatenated analyses and datasets.

Within the Homilo 2 clade, *E. chilensis* always was recovered either as sister to the species of Mesopsocidae, or to *Protopsocus pulcripennis* (Elipsocidae) rendering both Lachesillidae and Elipsocidae paraphyletic. Paraphyly of these two families was also found by previous phylogenomic studies (Saenz Manchola et al. 2021, de Moya et al. 2020). Although the remaining species of Lachesillidae (Lachesillinae and Eolachesillinae subfamilies) clustered in all ML concatenated analyses and datasets with high UFB support, the coalescent trees showed instability, and recovered Eolachesillinae as sister to the Elipsocidae + Mesopsocidae

+ *E. chilensis* subclade E for some datasets (40–60–70% matrices), with low LPP (Fig. 2A). In general, it appears that relationships within Homilopsocidea continue to be unstable and may be highly sensitive to taxon sampling.

Although the ML concatenated analyses were consistent across all data sets, some discordance between concatenated and coalescent topologies were found. This discordance was particularly centered around the relationships within and between species of Homilopsocidea, but also minor differences in the infraorders Caeciliusetae, Epipsocetidae, and Philotarsetae. These phylogenetic conflicts could be explained by bias caused by GC content, which has been demonstrated to have negative impact in phylogenetic inferences using UCE loci (Bossert et al. 2017, Cohen et al. 2021). Caeciliusetae had the highest GC content (45.7% avg.), followed by Homilopsocidea (45% avg), Epipsocetidae (44.2% avg), and Philotarsetae (44.1% avg) (Fig. 3; Supp Table 2 [online only]). In particular, 10 species had GC content that exceeded 46% (Supp Table 2 [online only]): *E. chilensis* (52% GC), *M. unipunctatus* (49.2% GC), and *E. kuriliensis* (48.5% GC) belonging to Homilo 2 (subclade E) clade; *Bryopsocus townsendi* Smithers (Psocoda: Pseudocaeciliidae) (49.8% GC) belonging to infraorder Philotarsetae; *Polypsocus corruptus* Hagen (Psocoda: Amphipsocidae) (46.3% GC), *Teliapsocus conterminus* Walsh (Psocoda: Dasydemellidae) (46.8% GC), and *Paracaecilius japonicus* Enderlein (Psocoda: Paracaeciliidae) (46.9% GC) belonging to infraorder Caeciliusetae; *E. briggsi* (47.2% GC) belonging to Homilo 1 clade; and *Lachesilla picticeps* Mockford (47.1% GC) and *Lachesilla*

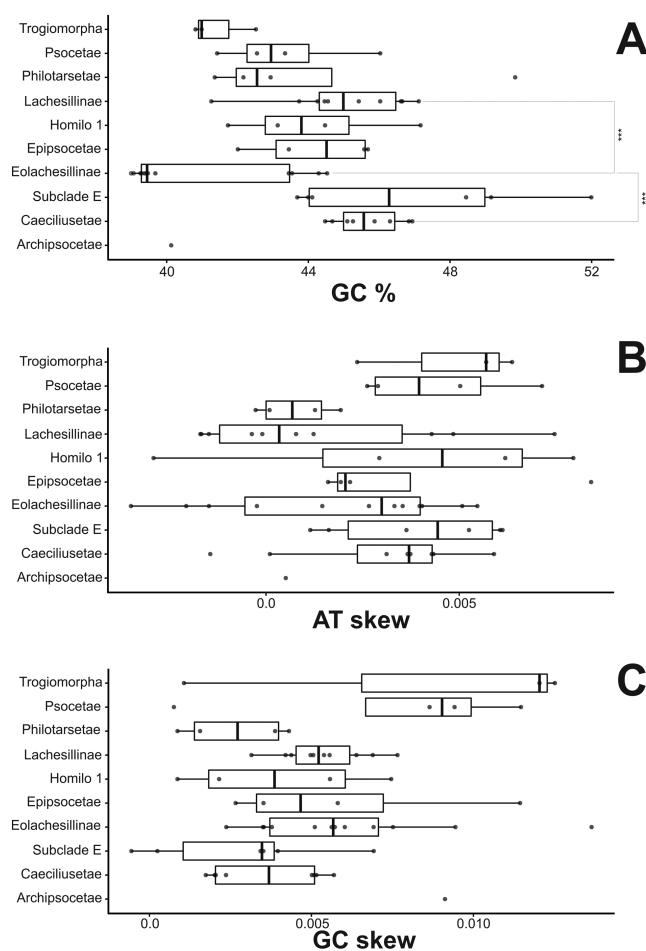


Fig. 3. Nucleotide composition box-and-whisker plots for groups within Psocomorpha. (A) refers to GC%. (B) and (C) refer to AT–GC skew. Asterisks indicate *P* values < 0.05 (***).

sp. Ghana (46.6%) belonging to Lachesillinae. All clades containing these species showed phylogenetic instability for the coalescent-based topologies (Figs. 1 and 2; Supp Figs. 1–4 [online only]). Similar results were obtained by Espeland et al. (2018), since some of the loci recovered in that study with high GC bias showed high gene tree – species tree discordance. Considering this, it is possible that the GC biases found in our study could have negatively impacted the phylogenetic relationships and caused discordant topologies.

Supplementary Material

Supplementary data are available at *Insect Systematics and Diversity* online.

Supplementary Figure 1. Maximum Likelihood tree of Psocomorpha inferred from the 50% completeness matrix using iq-tree. Number associated to branches correspond to UFB support/LPP, if some difference in UFB were obtained depending the partition scheme, numbers were displayed as follows: SWSC-EN/ SWSC-GC/ SWSC-Multi/ UCESP/LPP. - indicate no UFB/LPP support. Grey scale shades depicted major clades, blue shades depicted Lachesillidae species. Dotted line separated Eolachesillinae – Lachesillinae subfamilies.

Supplementary Figure 2. Maximum Likelihood tree of Psocomorpha inferred from the 60% completeness matrix using iq-tree. Number associated to branches correspond to UFB support/LPP, if some difference in UFB were obtained depending the partition scheme, numbers were displayed as follows: SWSC-EN/ SWSC-GC/ SWSC-Multi/ UCESP/LPP. - indicate no UFB/LPP support. Grey scale shades depicted major clades, blue shades depicted Lachesillidae species. Dotted line separated Eolachesillinae – Lachesillinae subfamilies.

Supplementary Figure 3. Maximum Likelihood tree of Psocomorpha inferred from the 70% completeness matrix using iq-tree. Number associated to branches correspond to UFB support/LPP, if some difference in UFB were obtained depending the partition scheme, numbers were displayed as follows: SWSC-EN/ SWSC-GC/ SWSC-Multi/ UCESP/LPP. - indicate no UFB/LPP support. Grey scale shades depicted major clades, blue shades depicted Lachesillidae species. Dotted line separated Eolachesillinae – Lachesillinae subfamilies.

Supplementary Figure 4. Maximum Likelihood tree of Psocomorpha inferred from the 80% completeness matrix using iq-tree. Number associated to branches correspond to UFB support/LPP, if some difference in UFB were obtained depending the partition scheme, numbers were displayed as follows: SWSC-EN/ SWSC-GC/ SWSC-Multi/ UCESP/LPP. - indicate no UFB/LPP support. Grey scale shades depicted major clades, blue shades depicted Lachesillidae species. Dotted line separated Eolachesillinae – Lachesillinae subfamilies.

Supplementary data 1. 40% completeness data matrix used in the analysis.

Supplementary data 2. 50% completeness data matrix used in the analysis.

Supplementary data 3. 60% completeness data matrix used in the analysis.

Supplementary data 4. 70% completeness data matrix used in the analysis.

Supplementary data 5. 80% completeness data matrix used in the analysis.

Supplementary table 1. General information, number of contigs, and number of UCE loci for each of the taxa used in this study. Accession numbers marked with (*) refers to transcriptomes and (**) to previously sequenced data source.

Supplementary table 2. Base composition for each species and the concatenated data matrices. Bold species indicate GC content above 46%.

Acknowledgments

We thank Alvaro Hernandez, Chris Wright and the Roy J. Carver Biotechnology Center at the University of Illinois for assistance with Illumina sequencing. We thank Kim Walden for assistance in submission of sequence files to NCBI. We thank R. de Moya, E. L. Mockford, T. Muroi, and D. Percy for assistance in obtaining specimens for this study. OFSM, ESS and ANGA thank Instituto de Biología, Universidad Nacional Autónoma de México, for continuous research support. This work is in part product of the Ph.D. thesis presented by OSFM. This work was supported by US National Science Foundation Awards DEB-1239788, DEB-1855812, DEB-1925487 and DEB-1926919 granted to K.P.J. The author declares that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Author Contributions

Oscar Fernando Saenz Manchola: Conceptualization, Methodology, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – Review and Editing, Visualization, Software. Ernesto Samacá Sáenz: Formal analysis, Writing – original draft, Writing- Review and Editing, Software. Stephany Virrueta-Herrera: Data Curation, Writing – original draft, Writing – Review and Editing. Lorenzo Mario D'Alessio: Data Curation, Writing – original draft, Writing – Review and Editing. Alfonso Nerí García Aldrete: Conceptualization, Resources, Writing – original draft, Writing – Review and Editing, Supervision. Kevin P. Johnson: Conceptualization, Methodology, Resources, Validation, Writing – original draft, Writing – Review and Editing, Funding acquisition.

Data Availability

Supplementary material, including figures, data matrices and tables are available from the figshare repository <https://doi.org/10.6084/m9.figshare.18737075.v1>

References Cited

- Andermann, T., M. F. Torres Jiménez, P. Matos-Maraví, R. Batista, J. L. Blanco-Pastor, A. L. S. Gustafsson, L. Kistler, I. M. Liberal, B. Oxelman, C. D. Bacon, et al. 2020. A guide to carrying out a phylogenomic target sequence capture project. *Front. Genet.* 10: 1–20.
- Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 30: 2114–2120.
- Bossert, S., E. A. Murray, B. B. Blaimer, and B. N. Danforth. 2017. The impact of GC bias on phylogenetic accuracy using targeted enrichment phylogenomic data. *Mol. Phylogenet. Evol.* 111: 149–157.
- Bossert, S., E. A. Murray, E. A. B. Almeida, S. G. Brady, B. B. Blaimer, and B. N. Danforth. 2019. Combining transcriptomes and ultraconserved elements to illuminate the phylogeny of Apidae. *Mol. Phylogenet. Evol.* 130: 121–131.
- Buenaventura, E. 2021. Museomics and phylogenomics with protein-encoding ultraconserved elements illuminate the evolution of life history and phallic morphology of flesh flies (Diptera: Sarcophagidae). *BMC Ecol. Evol.* 21: 1–28.
- Buenaventura, E., M. W. Lloyd, J. M. Perilla López, V. L. González, A. Thomas-Cabianca, and T. Dikow. 2021. Protein-encoding ultraconserved elements provide a new phylogenomic perspective of Oestroidea flies (Diptera: Calyptratae). *Syst. Entomol.* 46: 5–27.
- Bushmanova, E., D. Antipov, A. Lapidus, and A. D. Pribelski. 2019. RnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. *GigaScience.* 8: 1–13.
- Bushnell, B. 2014. BBTools software package. <https://sourceforge.net/projects/bbmap/>
- Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17: 540–552.
- Clayton, D. H., S. E. Bush, and K. P. Johnson. 2016. *Coevolution of life on hosts: integrating ecology and history.* University of Chicago Press, Chicago, USA.
- Cohen, C. M., K. Noble, T. Jeffrey Cole, and M. S. Brewer. 2021. The phylogeny of robber flies (Asilidae) inferred from ultraconserved elements. *Syst. Entomol.* 46: 812–826.

- Espeland, M., J. Breinholt, K. R. Willmott, A. D. Warren, R. Vila, E. F. A. Toussaint, S. C. Maunsell, K. Aduse-Poku, G. Talavera, et al. 2018. A comprehensive and dated phylogenomic analysis of Butterflies. *Curr. Biol.* 28: 770–778.
- Faircloth, B. C. 2016. PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics.* 32: 786–788.
- Faircloth, B. C., J. E. McCormack, N. G. Crawford, M. G. Harvey, R. T. Brumfield, and T. C. Glenn. 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Syst. Biol.* 61: 717–726.
- Harris, B. 2007. Improved pairwise alignment of genomic DNA. Ph.D. dissertation, The Pennsylvania State University, Philadelphia.
- Hedin, M., S. Derkarabetian, A. Alfaro, M. J. Ramírez, and J. E. Bond. 2019. Phylogenomic analysis and revised classification of atypoid mygalomorph spiders (Araneae, Mygalomorphae), with notes on arachnid ultraconserved element loci. *PeerJ.* 7: e6864.
- Hoang, D. T., O. Chernomor, A. Von Haeseler, B. Q. Minh, and L. S. Vinh. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35: 518–522.
- Johnson, K. P. 2019. Putting the genome in insect phylogenomics. *Curr. Opin. Insect Sci.* 36: 111–117.
- Johnson, K. P., and E. L. Mockford. 2003. Molecular systematics of Psocomorpha (Psocoptera). *Syst. Entomol.* 28: 409–416.
- Johnson, K. P., C. H. Dietrich, F. Friedrich, R. G. Beutel, B. Wipfler, R. S. Peters, J. M. Allen, M. Petersen, A. Donath, K. K. O. Walden, et al. 2018. Phylogenomics and the evolution of hemipteroid insects. *Proc. Natl. Acad. Sci. U.S.A.* 115: 12775–12780.
- Kalyaanamoorthy, S., B. Q. Minh, T. K. F. Wong, A. Von Haeseler, and L. S. Jermini. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods.* 14: 587–589.
- Katoh, K., and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30: 772–780.
- Kieran, T. J., E. R. L. Gordon, M. Forthman, R. Hoey-Chamberlain, R. T. Kimball, B. C. Faircloth, C. Weirauch, and T. C. Glenn. 2019. Insight from an ultraconserved element bait set designed for hemipteran phylogenetics integrated with genomic resources. *Mol. Phylogenet. Evol.* 130: 297–303.
- Lanfear, R., P. B. Frandsen, A. M. Wright, T. Senfeld, and B. Calcott. 2017. Partitionfinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* 34: 772–773.
- Lemmon, A. R., S. A. Emme, and E. M. Lemmon. 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. *Syst. Biol.* 61: 727–744.
- Lienhard, C., and C. N. Smithers. 2002. Psocoptera (Insecta): world catalogue and bibliography. *Instrumenta Biodiversitatis* 5: xli+745 pp. Muséum d'histoire naturelle, Genève.
- Minh, B. Q., M. A. T. Nguyen, and A. Von Haeseler. 2013. Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* 30: 1188–1195.
- Minh, B. Q., H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. Von Haeseler, R. Lanfear, and E. Teeling. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37: 1530–1534.
- Misof, B., S. Liu, K. Meusemann, R. S. Peters, A. Donath, C. Mayer, P. B. Frandsen, J. Ware, T. Flouri, R. G. Beutel, et al. 2014. Phylogenomics resolves the timing and pattern of insect evolution. *Science.* 346: 763–767.
- de Moya, R. S., K. Yoshizawa, K. K. O. Walden, C. H. Dietrich, and K. P. Johnson. 2020. Phylogenomics of parasitic and non-parasitic lice (Insecta: Psocodea): combining sequence data and exploring compositional bias solutions in next generation datasets. *Syst. Biol.* 53: 1689–1699.
- Nguyen, L. T., H. A. Schmidt, A. Von Haeseler, and B. Q. Minh. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32: 268–274.
- Pauli, T., T. O. Burt, K. Meusemann, K. Bayless, A. Donath, L. Podsiadlowski, C. Mayer, A. Kozlov, A. Vasilakopoulos, S. Liu, et al. 2018. New data, same story: phylogenomics does not support Syrphoidea (Diptera: Syrphidae, Pipunculidae). *Syst. Entomol.* 43: 447–459.
- Peters, R. S., K. Meusemann, M. Petersen, C. Mayer, J. Wilbrandt, T. Ziesmann, A. Donath, K. M. Kjer, U. Aspöck, H. Aspöck, et al. 2014. The evolutionary history of holometabolous insects inferred from transcriptome-based phylogeny and comprehensive morphological data. *BMC Evol. Biol.* 14: 1–16.
- Peters, R. S., L. Krogmann, C. Mayer, A. Donath, S. Gunkel, K. Meusemann, A. Kozlov, L. Podsiadlowski, M. Petersen, R. Lanfear, et al. 2017. Evolutionary history of the hymenoptera. *Curr. Biol.* 27: 1013–1018.
- Saenz Manchola, O. F., S. Virrueta Herrera, L. M. D'Alessio, K. Yoshizawa, A. N. García Aldrete, and K. P. Johnson. 2021. Mitochondrial genomes within bark lice (Insecta: Psocodea: Psocomorpha) reveal novel gene rearrangements containing phylogenetic signal. *Syst. Entomol.* 46: 938–951.
- Sayyari, E., and S. Mirarab. 2016. Fast coalescent-based computation of local branch support from quartet frequencies. *Mol. Biol. Evol.* 33: 1654–1668.
- Simon, S., A. Blanke, and K. Meusemann. 2018. Reanalyzing the Palaeoptera problem – the origin of insect flight remains obscure. *Arthropod Struct. Dev.* 47: 328–338.
- Simpson, J. T., K. Wong, S. D. Jackman, J. E. Schein, S. J. M. Jones, and I. Birol. 2009. ABySS: a parallel assembler for short read sequence data. *Genome Res.* 19: 1117–1123.
- Tagliacollo, V. A., and R. Lanfear. 2018. Estimating improved partitioning schemes for ultraconserved elements. *Mol. Biol. Evol.* 35: 1798–1811.
- Van Dam, M. H., A. W. Lam, K. Sagata, B. Gewa, R. Laufa, M. Balke, B. C. Faircloth, and A. Riedel. 2018. Correction: ultraconserved elements (UCEs) resolve the phylogeny of Australasian smurfweevils (PLoS ONE (2017) 12: 11 (e0188044) DOI: 10.1371/journal.pone.0188044). *PLoS One.* 13: 1–21.
- Vasilakopoulos, A., M. Balke, R. G. Beutel, A. Donath, L. Podsiadlowski, J. M. Pflug, R. M. Waterhouse, K. Meusemann, R. S. Peters, H. E. Escalona, et al. 2019. Phylogenomics of the superfamily Dytiscoidea (Coleoptera: Adephaga) with an evaluation of phylogenetic conflict and systematic error. *Mol. Phylogenet. Evol.* 135: 270–285.
- Wang, Y. H., H. Y. Wu, D. Rédei, Q. Xie, Y. Chen, P. P. Chen, Z. E. Dong, K. Dang, J. Damgaard, P. Štys, et al. 2019. When did the ancestor of true bugs become stinky? Disentangling the phylogenomics of Hemiptera–Heteroptera. *Cladistics.* 35: 42–66.
- Wipfler, B., H. Letsch, P. B. Frandsen, P. Kapli, C. Mayer, D. Bartel, T. R. Buckley, A. Donath, J. S. Edgerly-Rooks, M. Fujita, et al. 2019. Evolutionary history of Polyneoptera and its implications for our understanding of early winged insects. *Proc. Natl. Acad. Sci. U.S.A.* 116: 3024–3029.
- Yoshizawa, K. 2002. Phylogeny and higher classification of suborder Psocomorpha (Insecta: Psocodea: 'Psocoptera'). *Zool. J. Linn. Soc.* 136: 371–400.
- Yoshizawa, K., and K. P. Johnson. 2014. Phylogeny of the suborder Psocomorpha: congruence and incongruence between morphology and molecular data (Insecta: Psocodea: 'Psocoptera'). *Zool. J. Linn. Soc.* 171: 716–731.
- Yoshizawa, K., E. L. Mockford, and K. P. Johnson. 2014. Molecular systematics of the bark lice infraorder Caeciliusetae (Insecta: Psocodea). *Syst. Entomol.* 39: 279–285.
- Zhang, J. E., R. L. Gordon, M. Forthman, W. S. Hwang, K. Walden, D. R. Swanson, K. P. Johnson, R. Meier, and C. Weirauch. 2016. Evolution of the assassin's arms: insights from a phylogeny of combined transcriptomic and ribosomal DNA data (Heteroptera: Reduviidae). *Sci. Rep.* 6: 1–8.
- Zhang, C., M. Rabiee, E. Sayyari, and S. Mirarab. 2018. ASTRAL-III: poly-nomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinf.* 19: 15–30.
- Zhang, F., Y. Ding, C. D. Zhu, X. Zhou, M. C. Orr, S. Scheu, and Y. X. Luan. 2019a. Phylogenomics from low-coverage whole-genome sequencing. *Methods Ecol. Evol.* 10: 507–517.
- Zhang, M. Y., J. L. Williams, and A. Lucky. 2019b. Understanding UCEs: a comprehensive primer on using ultraconserved elements for arthropod phylogenomics. *Insect Syst. Divers.* 3(5): 3; 1–12.