Robert S. de Moya
Department of Entomology
University of Illinois Urbana-Champaign
505 S. Goodwin Ave.
Urbana, IL 61801
Email: Rdemoya2@illinois.edu

Phylogenomics of parasitic and non-parasitic lice (Insecta: Psocodea): Combining sequence data and Exploring compositional bias solutions in Next Generation Datasets

Robert S. de Moya[1,2], Kazunori Yoshizawa[3], Kimberly K.O. Walden[1], Andrew D. Sweet[4], Christopher H. Dietrich[2], Kevin P. Johnson[2]

[1]Department of Entomology, University of Illinois Urbana-Champaign, 505 S. Goodwin Ave., Urbana, IL 61801, USA

[2]Illinois Natural History Survey, Prairie Research Institute, University of Illinois, Champaign, IL 61820, USA

[3]Systematic Entomology, School of Agriculture, Hokkaido University, Sapporo 060-8589, Japan

[4]Department of Entomology, Purdue University, 901 W. State St., West Lafayette, IN 47907, USA

Abstract

The insect order Psocodea is a diverse lineage comprising both parasitic (Phthiraptera) and non-parasitic members (Psocoptera). The extreme age and ecological diversity of the group may be associated with major genomic changes, such as base compositional biases expected to affect phylogenetic inference. Divergent morphology between parasitic and non-parasitic members has also obscured the origins of parasitism within the order. We conducted a phylogenomic analysis on the order Psocodea utilizing both transcriptome and genome sequencing to obtain a data set of 2,370 orthologous genes. All phylogenomic analyses, including both concatenated and coalescent methods suggest a single origin of parasitism within the order Psocodea, resolving conflicting results from previous studies. This phylogeny allows us to propose a stable ordinal level classification scheme that retains significant taxonomic names present in historical scientific literature and reflects the evolution of the group as a whole. A dating analysis, with internal nodes calibrated by fossil evidence, suggests an origin of parasitism that predates the K-Pg boundary. Nucleotide compositional biases are detected in third and first codon positions and result in the anomalous placement of the Amphientometae as sister to Psocomorpha when all nucleotide sites are analyzed. Likelihood-mapping and quartet sampling methods demonstrate that base compositional biases can also have an effect on quartet-based methods.

Keywords:
Quartet sampling, Psocoptera, Phthiraptera, Illumina, Recoding methods

Introduction

The era of phylogenomic analysis has provided access to new data types originating from different genomic regions (e.g. ultra-conserved elements (UCE) and single-copy protein-coding orthologs) (Jarvis et al. 2014; Prum et al. 2015) or from post transcriptional processes (i.e. transcriptomes) (Misof et al. 2014; Johnson et al. 2018a). Availability of certain data types may be contingent on the quality or age of a specimen. For example, RNA degrades quickly, and transcriptome-based analyses are not typically feasible for old or fixed specimens (Houseley and Tollervey 2009; Bossert et al. 2019). Transcriptomes are obtainable for fresh specimens preserved in an appropriate buffer (e.g. RNA-later™) that inhibits the activity of RNases which degrade RNA (Houseley and Tollervey 2009). Transcriptomes correspond to coding gene sequences and are free of non-coding introns, and thus align well. Alternatively, most next generation methods that produce whole genome sequences include a fragmentation step prior to library construction (Alkan et al. 2011) that may mimic degradation processes that occur in specimens that are dry or old (Bossert et al. 2019). DNA-based genome sequencing is not limited by the amount of RNA present in a cell and can produce many reads across the genome (Johnson 2019). However, raw genomic DNA sequence data contain intron and non-coding data (Jarvis et al. 2014), but these can be excised prior or masked following alignment with transcriptome data. High volume sequencing technologies have often been described and implemented in phylogenomic studies; however, approaches to combine sequences derived from different next generation sequencing technologies have been less developed. A recent phylogenomic analysis of Apidae (bees) successfully combined transcriptome, genome, and UCE data to produce a

robust topology (Bossert et al. 2019). However, this is one of few studies to combine transcriptome and partial genome data in a single phylogenomic analysis.

Phylogenomic analyses have helped resolve many contentious relationships, but have also accentuated the need to test for compositional (and other) biases in molecular data sets that may be amplified by the inclusion of millions of base pairs of nucleotides and can lead to strong support for misleading hypotheses (Romiguier et al. 2016; Bossert et al. 2017, 2019; Simion et al. 2017; Laumer et al. 2018; Simon et al. 2018; Vasilikopoulos et al. 2019). Before the widespread use of phylogenomic analysis, Sanger sequencing-based phylogenetics sought to optimize the topology by testing hierarchical models of evolution (Posada and Crandall 1998). Weaknesses of these models can be exposed (Duchêne et al. 2017), due to large scale compositional biases that exist in codon positions found across thousands of loci, which violate model assumptions of stationarity (Simion et al. 2017; Laumer et al. 2018). These biases can create phylogenetic artifacts that appear well supported given traditional clade support values (i.e. bootstrap support) but is actually misleading because the large amount of data simply converges upon a stable topology due to underlying weaknesses in model assumptions.

Base compositional biases (%GC) have long been known to influence the results of phylogenetic analyses (Galtier and Gouy 1995; Jermiin et al. 2004; Bossert et al. 2017) and several methods for reducing the influence of such biases on phylogenetic inference have been proposed (Jermiin et al. 2004; Sheffield et al. 2009; Regier et al. 2010; Ishikawa et al. 2012; Zwick et al. 2012; Simmons 2017). However, most methods that incorporate time-heterogeneous approaches (Philippe et al. 2011; Roure and Philippe 2011) are extremely computationally intensive and would be difficult to apply to large phylogenomic data sets, although they have been applied in mitochondrial phylogenomics with success (Sheffield et al. 2009). Alternative

methods include recoding techniques (Simmons 2017) which use IUPAC ambiguity codes to mask variable codon positions that code for a silent mutation, such as RY recoding (Ishikawa et al. 2012) and degeneracy methods (Regier et al. 2010; Zwick et al. 2012). Another solution is to discard possible saturated data, for example removal of the third codon positions from an alignment (Breinholt and Kawahara 2013) or even the first and third codon postitions (Misof et al. 2014). These two methods are effective for concatenated datasets; however, coalescent analyses may also be influenced by compositional biases in individual genes (Romiguier et al. 2016; Bossert et al. 2017, 2019). A further solution is to analyze amino acid sequences, although it is possible that underlying base compositional biases can result in amino acid biases as well (Foster et al. 1997). In addition, molecular models for the evolution of amino acids are much more computationally intensive and may not be feasible for analysis of large genomic data sets, because a twenty-one amino acid model (two coding strategies for serine) (Zwick et al. 2012) is much more complex relative to nucleotide models based on four bases (Posada and Crandall 1998). Here we explore some of these issues using a combined genome and transcriptome data set for a group of insects (Psocodea) known to have strong variation in base compositional biases across taxa (Johnson et al. 2003; Yoshizawa and Johnson 2013).

The insect order Psocodea encompasses the two historically recognized groups Psocoptera (free-living bark lice) and Phthiraptera (parasitic lice) that were once considered separate orders. Members of Psocodea have an extensive fossil record that extends into the Lower Cretaceous (Mockford et al. 2013) and molecular divergence time estimates place their origin in the Paleozoic (~404 Ma) (Misof et al. 2014; Johnson et al. 2018a; Yoshizawa et al. 2019). The order also encompasses species with a range of feeding preferences, from detritus, plant material (i.e. pollen, decaying leaves), and microflora (i.e. cyanobacteria films, fungal, and

lichen) in non-parasitic members (Broadhead and Wapshere 1966; New 1970, 1987; Broadhead and Richards 1982); to obligate ectoparasitism on birds and mammals (i.e. skin debris, feathers, blood/skin secretions) (Price et al. 2003; Clayton et al. 2015). The ecological diversity and age of the group have likely contributed to large-scale compositional biases that have previously been detected between parasitic and non-parasitic members (Johnson et al. 2003; Yoshizawa and Johnson 2013; Johnson et al. 2018a). These known compositional biases provide an opportunity to examine the effects such biases may have on phylogenomic analyses. Other groups of organisms are also known to show such biases (Cox et al. 2014; Romiguier et al. 2016; Bossert et al. 2017; Skinner et al. 2020), thus understanding the potential effects and methods to account for such biases will have relevance to many phylogenomic studies.

The order Psocodea also represents an ideal taxon for examining the effect of combining whole genome and transcriptome derived sequence data. Parasitic lice are known to have reduced genome sizes (Pittendrigh et al. 2006; Johnston et al. 2007; Kirkness et al. 2010) and are minute insects which typically produce small amounts of RNA. *Pediculus humanus* has one of the smallest insect genomes recorded, at 108 Mbp (Kirkness at al. 2010), and coverage estimates from Illumina genome sequencing indicate small genome sizes may be a general feature of Psocodea (100 – 400 Mbp, unpublished). While it has been possible in the past to sequence transcriptome-based data for parasitic Phthiraptera based on pooling many individuals (Johnson et al. 2018b), data is more readily obtained by whole genome sequencing (Allen et al. 2017; Boyd et al. 2017; Sweet et al. 2018). The reduced genome size of parasitic lice makes it possible to produce high quality assemblies from multiplexed samples on a single sequencing lane with whole genome-based sequencing methods (Allen et al. 2017; Boyd et al. 2017; Sweet et al. 2018; Johnson 2019). In contrast, non-parasitic Psocodea (bark-lice) are typically larger in body

volume and produce higher copy transcript sequences (Johnson et al. 2018a). Less total sequence data is needed for transcriptome sequencing, thus is more economical than whole genome sequencing. (Johnson 2019). Therefore, there is a cost advantage to combining transcriptome and whole genome data in phylogenomic analyses that combine parasitic and non-parasitic Psocodea. Our study is the first to test the utility of combining these different data types in a study of Psocodea phylogeny, and this general approach should be applicable to many groups of organisms.

Although the monophyly of the lineage comprising both Phthiraptera and Psocoptera is well established based on morphological criteria (Lyal 1985; Yoshizawa and Lienhard 2010), inconsistent taxonomic treatment of the two groups continues (Emeljanov et al. 2001; Scholtz 2016; Durden 2019; Wang et al. 2019). Psocoptera traditionally consists of three recognized suborders (Trogiomorpha, Psocomorpha, and Troctomorpha) (Lienhard and Smithers 2002) and Phthiraptera has four previously recognized suborders (Amblycera, Ischnocera, Rhynchophthirina, and Anoplura) (Price et al. 2003). Based on molecular and morphological evidence, Phthiraptera is derived from within the Troctomorpha (Lyal 1985; Johnson et al. 2004; Yoshizawa and Lienhard 2010; Johnson et al. 2018a), but inconsistent use of the subordinal ranks that divide the traditional orders Psocoptera and Phthiraptera can be found in modern literature (Emeljanov et al. 2001; Scholtz 2016). Adding to the confusion, the origin of parasitism remains in question (Yoshizawa and Johnson 2010) because phylogenetic analyses of a ribosomal gene suggested that Phthiraptera could be polyphyletic (Johnson et al. 2004).

To explore the phylogenetic relationships within Psocodea, we assembled a large phylogenomic data set (2,370 orthologous genes) derived from whole genome and transcriptome sequencing using a customized pipeline. Using this dataset, comprising more than two million

base pairs of nucleotide data, we examined the effects of large base compositional biases on phylogenetic inference. We used the results from our assessment of the influence of base composition on tree topology to conduct a dating analysis accounting for these biases to explore the origins of parasitism in this group.

Materials & Methods

Taxonomic Sampling

Sampling was aimed at resolving deep level relationships between historically recognized orders or suborders that comprise the insect order Psocodea. One focus of the sampling was resolving whether or not the parasitic lice (Phthiraptera) form a monophyletic assemblage (Johnson et al. 2004; Yoshizawa and Johnson 2010). Sampling included a broad array of parasitic species and the closest non-parasitic members known as the Nanopsocetae (Mockford 1993). In total 112 individuals were sampled, encompassing all currently recognized suborders and infraorders (Table 1). Prior studies established that Trogiomorpha is monophyletic and is the sister taxon of the remainder of Psocodea (Johnson et al. 2004; Yoshizawa et al. 2006), so we used this suborder as the root. This was done because the sister taxon of Psocodea is currently unclear (Misof et al. 2014; Johnson et al. 2018a), thus we avoided outgroups that are highly divergent from the ingroup to circumvent alignment difficulties and potential long branch attraction artifacts.

Next Generation Sequencing and Orthology Inference

Given the inherent difficulties of obtaining large quantities of freshly preserved tiny insects, such as lice, we developed a pipeline to use whole genome sequencing to obtain orthologs belonging to a set included in a previous dataset derived from transcriptome data (Johnson et al. 2018a). In some cases, both genome and transcriptome sequences were available for the same species. This allowed us to verify that these two data types placed respective species in the same phylogenetic position.

Whole genome data were obtained following genomic DNA extraction procedures and using Illumina sequencing technologies. Specimens were stored in 95% ethanol at -80 ºC. From these, genomic DNA was extracted using a Qiagen DNAeasy extraction kit. The protocol was slightly modified with an extended 48-hour incubation step and use of 52 µl of elution buffer. DNA was quantified using a Qubit 3.0 fluorometer. The extractions were then sonicated with a Covaris M220 to an average size of 300-400 nt. A Kapa Library Preparation Kit (Kapa Biosystems) was used to produce pair-end libraries. The libraries were then pooled into equimolar concentrations, quantified by qPCR. Each sample was sequenced for 151-161 cycles on a Hiseq2500 (Illumina) with a TruSeq or HiSeq SBS sequencing rapid kit to produce 160 nt reads. Fastq files were produced with Casava 1.8.2 or bcl2fastq v2.17.1.14. Adaptors and low-quality bases were removed using the FASTX Toolkit 0.0.14 (Gordon and Hannon 2010). All sequencing took place at W.M. Keck Center at the University of Illinois, Urbana-Champaign. A gene set of 2,395 protein-coding orthologs previously used for phylogenomic analyses of hemipteroid insects was identified in the annotated genome of the human body louse, *P. humanus* (Johnson et al. 2018a). This ortholog set was used as a reference in aTRAM 1.0 (Allen et al. 2015) for local assembly of individual orthologs. This software uses tblastn searches to

identify reads matching the gene of interest and assembles them locally. Parameters for aTRAM loci assembly were set to three iterations, fraction one, and the ABySS de novo assembler (Simpson et al. 2009). Exon sequences assembled by aTRAM were then annotated and stitched together if needed using an Exonerate-based (Slater and Birney 2005) pipeline (Allen et al. 2017). Transcriptome assemblies and inferred ortholog transcripts were previously published (Table 1, Johnson et al. 2018a).

Phylogenomic Analyses

Nucleotide sequences inferred as being orthologous from whole genome sequence data were translated with Geneious 11.1.15 (Kearse et al. 2012). Translated whole genome and transcriptome sequences were aligned with PASTA 1.8.0 by amino acid with memory usage increased (2048 MB) and otherwise default parameters (Mirarab et al. 2014a). Nucleotide sequences were retrieved using a custom python script to produce a final nucleotide alignment based upon the amino acid alignments (Allen et al. 2017). Exonerate inserts ambiguous N's between combined exon data, therefore excess N's were recoded to gaps before subsequent masking. Multiple sequence gene alignments (MSAs) were masked on a nucleotide level with trimAl (Capella-Gutiérrez et al. 2009) using a 40% gap threshold. Subsequently, MSAs that included less than 50% of individuals were eliminated from subsequent analyses. Final concatenated supermatrices of 2,370 gene sequences were produced with SequenceMatrix 1.8 (Vaidya et al. 2011).

Several phylogenetic analyses were performed using the final nucleotide supermatrix. First, all nucleotide sites were analyzed in both partitioned and unpartitioned maximum

likelihood (ML) frameworks. Second, nucleotide sites were recoded using degeneracy coding (Regier et al. 2010; Zwick et al. 2012) for phylogenetic analyses. Finally, nucleotide supermatrices of 1) first and second codon positions and 2) second codon positions only were produced from Geneious (Kearse et al. 2012). Partitioned analyses were performed on the all nucleotide site and degeneracy recoded supermatrices. The optimal partitioning scheme was determined with PartitionFinder 2.1.1 (Lanfear et al. 2017) and the implemented version of RAxML 8.2.11 (Stamatakis 2014) with the following parameters: branch lengths linked, GTR + G model, BIC model selection, rcluster search and max set to 100.

A series of ML phylogenetic analyses were performed on the resulting supermatrices using ExaML 3.0.21 (updated: 6/4/2018) (Kozlov et al. 2015) and RAxML 8.2.11 (Stamatakis 2014). To save computation time, the ML hill-climbing algorithm was performed in ExaML with a gamma model and 100 rapid bootstrap replicates were completed with RAxML using the GTR + G model with four GAMMA categories. For each bootstrap search, we tested for bootstrap convergence using RAxML (Pattengale et al. 2009). In all cases, convergence was reached by 50 replicates, so the 100 bootstrap replicates are sufficient to provide a reliable estimator of bootstrap proportions. To ensure that the most likely topology was obtained, eight separate tree searches were performed with ExaML, each with different starting input trees derived from RAxML (four parsimony based and four random start topologies). Bootstrap support (BS) for the most likely topology obtained was then mapped using SumTrees 4.1.0 (Sukumaran and Holder 2015). These methods were used to analyze all supermatrices produced (all nucleotide sites, degeneracy recoding, second codon positions only, and third codon positions removed).

To account for possible incongruence among genes due to incomplete lineage sorting or other biases masked by concatenation, we also performed coalescent gene/species-tree analyses

using Astral 5.5.9 (Mirarab et al. 2014b; Mirarab and Warnow 2015). To infer individual gene trees as basis, each of the 2,370 MSAs were analyzed with RAxML using GTR + G and 100 rapid bootstrap replicates. Estimation of individual gene trees was performed using both 1) all sites and 2) the degeneracy recoded data for each gene. Resulting bipartition files produced from RAxML were used as in input for Astral analyses using default parameters with branch support calculated based on local posterior probability (LPP) (Sayyari and Mirarab 2016).

To evaluate support for conflicting topologies surrounding the phylogenetic position of Amphientometae (see Results) due to potential biases in the concatenated dataset, quartet sampling (Pease et al. 2018) and four cluster likelihood-mapping (quartet mapping, Strimmer and Haeseler 1997) methods were employed. Four cluster likelihood-mapping was performed in IQ-TREE 1.6.5 (Nguyen et al. 2015) testing all possible quartets, tree search skipped, GTR + G model, and the following quartets defined: Trogiomorpha, Psocomorpha, Amphientometae, and Nanopsocetae. Likelihood-mapping in IQ-TREE was performed on the nucleotide supermatrix with 1) all sites, 2) degeneracy recoding, 3) first and second positions only, and 4) second codon positions only. Four cluster likelihood-mapping is not computationally feasible across all branches in large datasets. However, Pease et al. (2018) developed a quartet sampling method which performs four cluster likelihood mapping across each node of the tree but using a random subsample of all possible quartet combinations for that node. We used quartet sampling (Pease et al. 2018) to evaluate support for conflicting topologies across all phylogenetic branches. Quartet sampling was performed using a log likelihood cutoff value of 2 and 200 replicates per branch on the supermatrices for 1) all sites, 2) degeneracy recoding, 3) first and second positions only, and 4) second codon positions only. For ease of comparison, we provide a summary of all phylogenetic analyses performed (Table 2).

Guanine and cytosine content (GC%) were calculated per gene and codon position from the masked MSAs with a custom python script (Allen et al. 2015, 2017). Following GC% calculation, biases were visualized with box and whisker plots produced from RStudio 1.1.453 (RS Team 2015). Distribution of the GC% obtained for each individual gene were arranged in ascending order per individual sampled based on the median GC% score obtained. This process was repeated for first, second, and third codon positions.

Phylogenetic dating analyses using relaxed clock methods were performed with MCMCTree in the PAML package under a correlated rates model (Yang 2007) on a topology resulting from the ML searches of the partitioned degeneracy-coded dataset. A total of nine internal calibration points with soft bounds were based on fossil evidence or previous dating analyses (Wappler et al. 2004; Mockford et al. 2013; Johnson et al. 2018a,b). The internal minimum age calibrations based on fossil evidence include the following: split of Atropetae (120 Ma), Psocomorpha (84 Ma), Caeciliusidae (33.9 Ma), Psocidae (33.9), Amphientometae (145 Ma), Liposcelididae + Phthiraptera (99 Ma), Menoponidae (44 Ma), *Pedicinus* + (*Pthirus* + *Pediculus*) (20-25 Ma), and *P. schaeffi* + *P. humanus* (5-7 Ma) (Wappler et al. 2004; Mockford et al. 2013; Johnson et al. 2018b). A maximum root age calibration for the split between Trogiomorpha and the remainder of Psocodea was set to 328 Ma based on a previous dating analysis (Johnson et al. 2018a) and was used to estimate the rate of substitution across the topology. These described calibrations are not completely independent of calibrations previously employed (Johnson et al. 2018a) but do include additional calibration points relevant to our current taxon sampling. A reversible (GTR) model was implemented for the analysis. The stationarity of two separate MCMC runs was visualized with Tracer 1.7.1 (Rambaut et al. 2018).

de Moya et al.

295 **Results**

296

297 In total 2,370 genes were successfully aligned yielding a supermatrix of 2,945,181 bp
298 including all three codon positions. Transcriptome and whole genome sequences aligned well,
299 facilitating subsequent phylogenetic analyses. On average, each individual sampled had data
300 present for 95% of genes sampled (Table 1).

301 Topologies from maximum likelihood (ML) phylogenetic analyses of the concatenated
302 sequence dataset varied depending on the methods used for coding or removing nucleotides.
303 Much of this variation centered around the placement of the Amphientometae, an infraorder of
304 Troctomorpha comprising free-living taxa. Amphientometae was recovered as sister to the
305 suborder Psocomorpha with maximum support (100% BS), which contains only non-parasitic
306 taxa, when 1) all nucleotide sites or 2) first and second codon positions were analyzed. However,
307 under degeneracy recoding or analysis of second codon positions only, Amphientometae was
308 recovered as sister to the remainder of the Troctomorpha with maximum support (100% BS), the
309 suborder into which it has traditionally been placed (Fig. 1).

310 Other than the placement of Amphientometae (i.e. monophyly of Troctomorpha),
311 relationships between the other major lineages within Psocodea were generally stable across
312 analyses. Psocomorpha was always recovered as monophyletic (100% BS). Within
313 Troctomorpha, Phthiraptera (parasitic lice) was always recovered as monophyletic regardless of
314 coding method (100% BS). The family Liposcelididae was also always recovered as
315 monophyletic and as the sister taxon of all parasitic lice (100% BS) as predicted from
316 morphology (Lyal 1985). Nanopsocetae (Pachytroctidae, Sphaeropsocidae, and Liposcelididae
317 plus Phthiraptera) was also always recovered as monophyletic (100% BS).

Among Phthiraptera, there was variation across analyses in the position of some mammal lice (Anoplura, Trichodectidae, and Rhynchophthirina). In particular, Rhynchophthirina (elephant lice) was sister to the chewing louse family Trichodectidae when all nucleotide sites were analyzed (100% BS). However, under 1) degeneracy coding, 2) first and second codon positions only, and 3) second codon positions only, Rhynchophthirina was recovered as sister to Anoplura (sucking lice) (100% BS). Either of these placements resulted in paraphyly of what is traditionally considered to be Ischnocera (one of the suborders of chewing lice): Trichodectidae (parasitizing mammals) and Philopteridae (parasitizing mainly birds). Thus, our results also support the existence of a larger mammal infesting clade comprising the Trichodectidae, Rhynchophthirina, and Anoplura, which corroborates recent analyses (Johnson et al. 2018b; de Moya et al. 2019; Song et al. 2019). The other traditional chewing louse suborder, Amblycera, was recovered as monophyletic across all analyses and sister to the remainder of Phthiraptera.

Within the Psocomorpha (bark lice, all free-living), relationships between some infraorders showed variation across analyses. In particular, the infraorder Homilopsocidea was not supported as monophyletic across all ML analyses. Two families of Homilopsocidea (Peripsocidae and Ectopsocidae) were most unstable in their placement and each of them sometimes grouped with the Caeciliusetae depending on the method of analysis. However, despite the poor support for the monophyly of the Homilopsocidea, the infraorder was consistently recovered as sister to other members of the Caeciliusetae across ML analyses (100% BS). The Psocetae and Epipsocetae were recovered as sister taxa across ML analyses (89-100% BS) and together sister to Philotarsetae (100% BS). In general, support values are higher within the Psocomorpha when all nucleotide sites are analyzed with a ML approach. The single sampled member of the Archipsocetae (Archipsocidae) was always recovered as sister to the remainder of

the Psocomorpha (100% BS), as in prior morphological (Yoshizawa 2002) and molecular studies (Yoshizawa and Johnson 2014).

Within the bark louse suborder Trogiomorpha, there was little variation in the results among analyses. The infraorder Prionoglaridetae was recovered as paraphyletic across all analyses, but this paraphyletic relationship was poorly supported in the degeneracy (6% BS) and second codon position only (33% BS) analyses. However, a paraphyletic Prionoglaridetae was supported with maximum bootstrap support (100% BS) when all nucleotide sites or first and second codon positions were analyzed. The remainder of Trogiomorpha was embedded within this paraphyletic assemblage of Prionoglaridetae. The infraorders Atropetae and Psyllipsocetae were each recovered as monophyletic and sister lineages across all ML analyses (100% BS).

Coalescent gene/species tree analyses (Astral) of individual gene trees across the 2,370 orthologous gene data set yielded similar branching patterns and measures of clade support relative to concatenated ML analyses of the same data type. Most nodes received maximum support (1.0 LPP). As in the ML analyses of all sites for the concatenated data set, coalescent analyses of gene trees from all nucleotide sites display maximum (1.0 LPP) support for a sister relationship between Psocomorpha and Amphientometae (Supplemental Fig. 2). However, when degeneracy recoded gene trees are analyzed (Supplemental Fig. 3), there is maximum support (1.0 LPP) for a monophyletic Trocotomorpha including the Amphientometae. Relationships among members of the Psocomorpha displayed some instability when degeneracy recoded data was analyzed in a coalescent context. However, a similar topology relative to the ML analyses is obtained for members of Psocomorpha when all sites are analyzed in a coalescent context.

Results of four-cluster likelihood-mapping show the distribution of discordant topologies between different methods of analyses for the placement of the Amphientometae (Fig. 2). When

all nucleotide sites are analyzed 65.2% of quartets sampled favor a sister relationship between the Amphientometae and Psocomorpha. Similarly, when first and second positions are analyzed the support declines, but 50.2% of quartets still support a sister relationship between the Amphientometae and Psocomorpha. In contrast, when the degeneracy recoded or second codon positions only data are analyzed, 43.8% and 42.5% of quartets sampled respectively, support a sister relationship between the Amphientometae and Nanopsocetae, while only 30.1% and 30.8% support Amphientometae with Psocomorpha.

Quartet sampling analyses are able to assess support from quartets (four cluster likelihood-mapping) across all nodes in the tree. Using slightly different metrics, these analyses estimate the frequency of discordant topologies across the resultant ML topology tested. When all nucleotide sites are analyzed, a weak majority of quartets support a sister Psocomorpha + Amphientometae (0.23 QC) (Fig. 3). In contrast, when using the degeneracy recoded dataset, a slight majority of quartets sampled support a monophyletic Troctomorpha, including the Amphientometae (0.01 QC) (Fig. 4). Quartet sampling also provides an estimate of which nodes are most stable given the data type and topology analyzed. For example, monophyly of the parasitic louse clade that includes Philopteridae, Trichodectidae, Anoplura, and Rhynchophthirina is supported by all quartets sampled across all nucleotide sites and degeneracy recoded analyses (1.0 QC) and no discordant topologies are detected (NA QD).

Visualization of the distribution of GC content for first, second, and third codon positions revealed substantial compositional biases at all positions between suborders or infraorders of Psocodea. Third codon positions showed the most variation in compositional biases (Fig. 5). Members of the Amphientometae possess some of the highest levels of GC content for third codon positions, similar to the pattern observed in Psocomorpha. In contrast members of the

Nanopsocetae tend to be more AT rich at third codon positions, similar to patterns observed in third codon positions of the Trogiomorpha. First and second codon positions showed similar patterns of compositional biases, but with much lower levels of variance around the medians relative to third positions (Figs. 6 and 7). First codon positions showed more variation in the medians relative to second codon positions. However, members of the Psocomorpha and Amphientometae were suggested to have the highest levels of GC content in first and second positions and the Nanopsocetae and Trogiomorpha tended to be more AT rich in first and second codon positions (Fig. 6 and 7).

Divergence time analysis indicates the main diversification of extant lineages of parasitic lice occurred approximately 60 Ma (74-123: 95% Myr) following the K-PG boundary mass extinction event. The origin of parasitism within Psocodea could have occurred a maximum of 115 Ma (95-148: 95% Myr) based on estimated divergence of the parasitic lineage from non-parasitic members of Psocodea. Divergences between suborders of Psocodea are estimated to have occurred in the lower Jurassic with the deepest split between extant suborders occurring 192 Ma (154-255: 95% Myr) (Fig. 1).

Discussion

Combining Data and Compositional Biases

Few phylogenomic studies have explored the results of combining whole genome and transcriptome-based data (Bossert et al. 2019). Higher level phylogenomic studies of insects have used transcriptome sequencing to produce large data matrices of thousands of genes (Misof

et al. 2014; Peters et al. 2017; Johnson et al. 2018a; Simon et al. 2019; Wipfler et al. 2019). Transcriptome sequencing relies on freshly preserved material, from which RNA can be extracted. Contigs can be annotated for single copy ortholog genes, and methods exist to account for splice variants to resolve these to a single gene sequence (Petersen et al. 2017). New methods have also been developed to individually assemble and annotate single copy genes from shotgun Illumina genome sequences (Allen et al. 2015, 2017). Whole genome data can include non-coding sequences, but poorly aligned regions can be masked when aligned against transcriptome sequences. Thus, it is possible to combine whole genome and transcriptome data to develop large phylogenomic datasets. Here we used a customized bioinformatic pipeline to combine transcriptome and genome data to produce a shared set of nuclear orthologs.

Large-scale compositional biases appeared to have some effect on the phylogenetic results for certain taxonomic groups within Psocodea. Most of the instability in our results centered around the placement of Amphientometae, a group of free-living bark lice traditionally placed in the suborder Troctomorpha based upon morphological synapomorphies (Mockford 1993; Lienhard and Smithers 2002; Yoshizawa and Lienhard 2010). However, in our phylogenomic analyses, large base compositional biases resulted in alternative placements of Amphientometae under different nucleotide recoding methods. This is most evident when all nucleotide sites are analyzed in which the Amphientometae are placed with 100% bootstrap support with Psocomorpha, resulting in paraphyly of Troctomorpha. Examination of GC content of third codon positions across the alignment (Fig. 5) shows that Amphientometae are similar in GC composition to Psocomorpha. This same pattern of GC biases is also seen in first and second codon positions (Figs. 6 and 7), although when second codon positions alone are analyzed, Amphientometae is recovered as sister to the remainder of the Troctomorpha, similar to the

result using degeneracy recoding. However, when third codon positions are removed, Amphientometae is still placed as the sister of Psocomorpha, suggesting modest base composition biases at first codon positions also affect the results.

Four cluster likelihood-mapping (quartet mapping) and quartet sampling analyses of the concatenated data demonstrate that quartet-based analyses may also be affected by substantial compositional biases. Likelihood-mapping analyses produce results similar to the ML tree topology itself for a given data type. For example, regarding the position of Amphientometae, degeneracy recoded, and second codon positions produce similar scores and in agreement with the phylogenetic placement of this group with Nanopsocetae (Fig. 2). Degeneracy recodes all nucleotides present in the alignment using IUPAC ambiguity codes so all codons that code for a synonymous mutation are nearly identical (Zwick et al. 2012). This recoding method helps account for the large amount of saturation that can take place at most notably third (Fig. 5) but also first (Fig. 6) codon positions, as well as accounting for some variation in base composition. However, given the similar phylogenetic results when degeneracy recoded and second codon positions only are analyzed, it appears that much of the signal supporting monophyletic Troctomorpha is present in second codon positions. Compositional biases and saturation of third and first codon positions can also skew quartet-based analyses. It appears that the relatively high percent GC base composition in both Amphientometae and Psocomorpha (blue and green in Figs. 5, 6), particularly at first and third codon positions, drives the majority of quartets to support a relationship between these two taxa (Fig. 2), similar to the full phylogeny derived from these data types. Although second base positions also show a similar pattern in the ranking of base composition frequencies (blue and green in Fig. 7), this variation spans only a few percent difference and is apparently not enough to influence the resulting tree.

Conflicting topologies in regard to the phylogenetic position of Amphientometae between recoded data types demonstrates the limitations of using a simplified model of evolution when analyzing millions of base pairs of data. Third positions are known to saturate at high rates due to degeneracy of the genetic code that allows for the emergence of silent mutations. This point is obvious when considering that a stable topology is obtained when degeneracy recoded data is analyzed (Fig. 1). Nearly identical results are obtained when second codon positions only are analyzed (Supplemental Fig. 1). Thus, when phylogenomic analyses are limited by computing power, it may be best to consider alternate methods that reduce data set sizes and recoding strategies to reduce rate heterogeneity and compositional bias that may exist at first and third codon positions between distantly related taxa.

Dating Analysis and the Origin of Parasitism

After accounting for heterogeneity by using the degeneracy recoded matrix, the dating analysis provides insight into the evolution of parasitic and non-parasitic members within Psocodea. We obtained an estimate for some of the earliest divergences within parasitic lice (i.e. shortly after the K-Pg boundary) similar to that found by a recent phylogenomic study (Johnson et al. 2018b) that used mostly calibration points from cospeciation events rather than from fossil free-living bark lice. Our estimate for the earliest divergence within parasitic lice (~100 Ma) was similar to this previous estimate (between 90 and 100 Ma) (Johnson et al. 2018b). The dating estimates completed in this study suggest generally younger origins for non-parasitic members than have been previously reported (Misof et al. 2014; Johnson et al. 2018b; Yoshizawa et al. 2019). The calibrations for analyses produced in this present study are fossil-based minimum age

479 constraints (Mockford et al. 2013) with a maximum age constraint at the root based on previous
480 Bayesian estimates (Johnson et al. 2018a). Differences in calibration methods may account for
481 some of older divergence estimates for non-parasitic members reported in other studies (Misof et
482 al. 2014; Johnson et al. 2018a; Yoshizawa et al. 2019). The dating analysis suggests the origin of
483 parasitism may have occurred a maximum of 115 Ma, predating the K-Pg boundary (66 Ma)
484 (Fig. 1). Our 95% confidence interval for the origin of parasitism also predates the K-Pg
485 boundary (95-148 Ma). Therefore, it remains a possibility that the ancient host of the first
486 parasitic louse may have been an endothermic dinosaur, although fossil evidence would be
487 needed to confirm this. However, the common ancestor of the clade that eventually evolved
488 parasitism may have been non-parasitic. Therefore, parasitism may have originated anytime
489 between 115 Ma and the initial diversification of parasitic lice (100 Ma).
490
491 *Implications for the Taxonomic Classification of Psocodea*
492
493 Our phylogenomic analyses, plus existing morphological evidence (Lyal 1985; Mockford
494 1993; Yoshizawa and Johnson 2010) help establish a stable subordinal level classification
495 scheme for the order Psocodea (Table 3). Given that previous analyses have suggested, both
496 based on morphological and molecular evidence, that free living Psocoptera and parasitic
497 Phthiraptera together form a monophyletic lineage (Lyal 1985; Johnson et al. 2004, 2018a), we
498 recognize Psocodea as a single order encompassing both traditional Psocoptera and Phthiraptera.
499 Below the level of order, the goal of this classification scheme is to reflect the higher-level
500 phylogeny, but also retain as many widely used historical names as possible. In particular, given

the widespread usage of Phthiraptera, we seek to retain this name, which necessitates changes in taxonomic rank for certain groups within Troctomorpha.

A monophyletic Phthiraptera (parasitic lice) is derived from within the Troctomorpha and sister to the family Liposcelididae across all analyses. All analyses suggest that the origin of parasitism occurred once within the Troctomorpha. Thus, it is necessary to recognize Phthiraptera at a lower taxonomic rank than the suborder Troctomorpha from which these parasites are derived. Given their widespread usage and acceptance (Lienhard and Smithers 2002), we prefer to retain the three historical suborders of bark lice (Trogiomorpha, Psocomorpha, and Troctomorpha) within Psocodea. We prefer to retain the infraordinal taxonomic ranks within Psocomorpha and Trogiomorpha given that these groups are generally supported by our analyses. Given that parasitic lice are also embedded within the traditional Infraorder Nanopsocetae, this would further reduce the rank of Phthiraptera. As a solution to this issue, we elect to divide Nanopsocetae into three infraorders (Table 3). Under this scheme, we also preserve many of the traditional subordinal names within Phthiraptera, and they are now placed at the rank of Parvorder. This scheme also allows us to retain all other existing subordinal parasitic louse names, including Amblycera, Rhynchophthirina, and Anoplura.

One final concern with regards to classification is the status of Ischnocera. Ischnocera (chewing lice) as currently defined (to include both Philopteridae and Trichodectidae) (Price et al. 2003) was paraphyletic across all analyses. This paraphyly has also been detected in previous studies (Johnson et al. 2018b; Song et al. 2019). Therefore, we suggest that Ischnocera be retained to recognize the bulk of diversity (i.e. Philopteridae, ~3,000 species) (Price et al. 2003) and that the Parvorder Trichodectera (Song et al. 2019) be recognized for the less diverse mammal infesting clade Trichodectidae (~400 species) (Price et al. 2003) (Table 3).

524 **Supplemental Material**

525 Data files and other materials, can found at the Dryad data repository at http://datadryad.org,

526 https://doi.org/10.5061/dryad.c59zw3r50

527

528 **Acknowledgments**

529 We would like to thank Rodrigo L. Ferreira for field support and supplying samples. We would

530 like to thank Charles Lienhard for information and discussion. We thank the following

531 individuals for assistance in obtaining samples: R. Wilson, D.H. Clayton, T.D. Galloway, E.

532 Osnas, T. Spradling, L. Mugisha, A. K. Saxena, T. Chesser, E. DiBlasi, F. Daunt, V. Smith, R.

533 Furness, E.L. Mockford, J. Jankowski, J.M. Allen, R. Faucett, B. O'Shea, K.G. McCracken, R.E.

534 Junge, J.D. Weckstein, A. Lawrence, K.C. Bell, M.S. Leonardi, and M. Bowser. Photo credit to

535 J. Gausas/Terrestrial Parasite Tracker Project.

536

537 **Funding**

538 This work was supported by JSPS Grants no. 15H04409 and 19H03278 to K.Y.; US NSF DEB-

539 0612938, DEB-1342604, XSEDE DEB-16002, DEB-1855812, DEB-1925487 to K.P.J.; and US

540 NSF DEB-1239788 to K.P.J. and C.H.D.

541

References

Alkan C., Sajjadian S., Eichler E.E. 2011. Limitations of next-generation genome sequence assembly. Nat Methods. 8:61–65.

Allen J.M., Boyd B., Nguyen N.-P., Vachaspati P., Warnow T., Huang D.I., Grady P.G.S., Bell K.C., Cronk Q.C.B., Mugisha L., Pittendrigh B.R., Leonardi M.S., Reed D.L., Johnson K.P. 2017. Phylogenomics from Whole Genome Sequences Using aTRAM. Syst. Biol. 66:786–798.

Allen J.M., Huang D.I., Cronk Q.C., Johnson K.P. 2015. aTRAM - automated target restricted assembly method: a fast method for assembling loci across divergent taxa from next-generation sequencing data. BMC Bioinformatics. 16.

Bossert S., Murray E.A., Almeida E.A.B., Brady S.G., Blaimer B.B., Danforth B.N. 2019. Combining transcriptomes and ultraconserved elements to illuminate the phylogeny of Apidae. Mol Phylogenet Evol. 130:121–131.

Bossert S., Murray E.A., Blaimer B.B., Danforth B.N. 2017. The impact of GC bias on phylogenetic accuracy using targeted enrichment phylogenomic data. Mol. Phylogenet Evol. 111:149–157.

Breinholt J.W., Kawahara A.Y. 2013. Phylotranscriptomics: Saturated Third Codon Positions Radically Influence the Estimation of Trees Based on Next-Gen Data. Genome Biol Evol. 5:2082–2092.

Broadhead E., Richards A.M. 1982. The Psocoptera of East Africa—a taxonomic and ecological survey. Biol J Linn Soc. 17:137–216.

Broadhead E., Wapshere A.J. 1966. Mesopsocus Populations on Larch in England--The Distribution and Dynamics of Two Closely-Related Coexisting Species of Psocoptera Sharing the Same Food Resource. Ecol Monogr. 36:327–388.

Capella-Gutiérrez S., Silla-Martínez J.M., Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 25:1972–1973.

Clayton D.H., Bush S.E., Johnson K.P. 2015. Coevolution of Life on Hosts: Integrating Ecology and History. Chicago ; London: University of Chicago Press.

Cox C.J., Li B., Foster P.G., Embley T.M., Civáň P. 2014. Conflicting Phylogenies for Early Land Plants are Caused by Composition Biases among Synonymous Substitutions. Syst Biol. 63:272–279.

Duchêne D.A., Duchêne S., Ho S.Y.W. 2017. New Statistical Criteria Detect Phylogenetic Bias Caused by Compositional Heterogeneity. Mol Biol Evol. 34:1529–1534.

Durden L.A. 2019. Chapter 7 - Lice (Phthiraptera). In: Mullen G.R., Durden L.A., editors. Medical and Veterinary Entomology (Third Edition). Academic Press. p. 79–106.

Emeljanov A.F., Golub N.V., Kuznetsova V.G. 2001. (Evolutionary Transformation of Testes and Ovaries in Booklice, Birdlice, and Sucking Lice (Psocoptera, Phthiraptera: Mallophaga, Anoplura). 81:20.

Foster P.G., Jermiin L.S., Hickey D.A. 1997. Nucleotide composition bias affects amino acid content in proteins coded by animal mitochondria. J Mol Evol 44:282-288.

Galtier N., Gouy M. 1995. Inferring phylogenies from DNA sequences of unequal base compositions. PNAS. 92:11317–11321.

Gordon A., Hannon G.J. 2010. Fastx-toolkit. .

Houseley J., Tollervey D. 2009. The Many Pathways of RNA Degradation. Cell. 136:763–776.

Ishikawa S.A., Inagaki Y., Hashimoto T. 2012. RY-Coding and Non-Homogeneous Models Can Ameliorate the Maximum-Likelihood Inferences from Nucleotide Sequence Data with Parallel Compositional Heterogeneity. Evol Bioinform Online. 8:EBO.S9017.

Jarvis E.D., Mirarab S., Aberer A.J., Li B., Houde P., Li C., Ho S.Y.W., Faircloth B.C., Nabholz B., Howard J.T., Suh A., Weber C.C., Fonseca R.R. da, Li J., Zhang F., Li H., Zhou L., Narula N., Liu L., Ganapathy G., Boussau B., Bayzid M.S., Zavidovych V., Subramanian S., Gabaldón T., Capella-Gutiérrez S., Huerta-Cepas J., Rekepalli B., Munch K., Schierup M., Lindow B., Warren W.C., Ray D., Green R.E., Bruford M.W., Zhan X., Dixon A., Li S., Li N., Huang Y., Derryberry E.P., Bertelsen M.F., Sheldon F.H., Brumfield R.T., Mello C.V., Lovell P.V., Wirthlin M., Schneider M.P.C., Prosdocimi F., Samaniego J.A., Velazquez A.M.V., Alfaro-Núñez A., Campos P.F., Petersen B., Sicheritz-Ponten T., Pas A., Bailey T., Scofield P., Bunce M., Lambert D.M., Zhou Q., Perelman P., Driskell A.C., Shapiro B., Xiong Z., Zeng Y., Liu S., Li Z., Liu B., Wu K., Xiao J., Yinqi X., Zheng Q., Zhang Y., Yang H., Wang J., Smeds L., Rheindt F.E., Braun M., Fjeldsa J., Orlando L., Barker F.K., Jønsson K.A., Johnson W., Koepfli K.-P., O'Brien S., Haussler D., Ryder O.A., Rahbek C., Willerslev E., Graves G.R., Glenn T.C., McCormack J., Burt D., Ellegren H., Alström P., Edwards S.V., Stamatakis A., Mindell D.P., Cracraft J., Braun E.L., Warnow T., Jun W., Gilbert M.T.P., Zhang G. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science. 346:1320–1331.

Jermiin L.S., Ho S.Y.W., Ababneh F., Robinson J., Larkum A.W.D. 2004. The Biasing Effect of Compositional Heterogeneity on Phylogenetic Estimates May be Underestimated. Syst Biol. 53:638–643.

Johnson K.P. 2019. Putting the genome in insect phylogenomics. Curr Opin Insect Science. 36:111–117.

Johnson K.P., Cruickshank R.H., Adams R.J., Smith V.S., Page R.D.M., Clayton D.H. 2003. Dramatically elevated rate of mitochondrial substitution in lice (Insecta: Phthiraptera). Mol Phylogenet Evol. 26:231–242.

Johnson K.P., Dietrich C.H., Friedrich F., Beutel R.G., Wipfler B., Peters R.S., Allen J.M., Petersen M., Donath A., Walden K.K.O., Kozlov A.M., Podsiadlowski L., Mayer C., Meusemann K., Vasilikopoulos A., Waterhouse R.M., Cameron S.L., Weirauch C., Swanson D.R., Percy D.M., Hardy N.B., Terry I., Liu S., Zhou X., Misof B., Robertson H.M., Yoshizawa K. 2018a. Phylogenomics and the evolution of hemipteroid insects. PNAS. 115:12775–12780.

Johnson K.P., Nguyen N., Sweet A.D., Boyd B.M., Warnow T., Allen J.M. 2018b. Simultaneous radiation of bird and mammal lice following the K-Pg boundary. Biol Lett. 14:20180141.

Johnson K.P., Yoshizawa K., Smith V.S. 2004. Multiple origins of parasitism in lice. P R Soc B-Biol Sci. 271:1771–1776.

Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S., Cooper A., Markowitz S., Duran C., Thierer T., Ashton B., Meintjes P., Drummond A. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28:1647–1649.

Kozlov A.M., Aberer A.J., Stamatakis A. 2015. ExaML version 3: a tool for phylogenomic analyses on supercomputers. Bioinformatics. 31:2577–2579.

Lanfear R., Frandsen P.B., Wright A.M., Senfeld T., Calcott B. 2017. PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. Mol Biol Evol. 34:772–773.

Laumer C.E., Gruber-Vodicka H., Hadfield M.G., Pearse V.B., Riesgo A., Marioni J.C., Giribet G. 2018. Support for a clade of Placozoa and Cnidaria in genes with minimal compositional bias. eLife. 7:e36278.

Lienhard C., Smithers C.N. 2002. Psocoptera (Insecta): world catalogue and bibliography. Psocoptera (Insecta): world catalogue and bibliography.

Lyal C.H.C. 1985. Phylogeny and classification of the Psocodea, with particular reference to the lice (Psocodea: Phthiraptera). Syst Entomol. 10:145–165.

Mirarab S., Nguyen N., Guo S., Wang L.-S., Kim J., Warnow T. 2014a. PASTA: Ultra-Large Multiple Sequence Alignment for Nucleotide and Amino-Acid Sequences. J Comput Biol. 22:377–386.

Mirarab S., Reaz R., Bayzid M.S., Zimmermann T., Swenson M.S., Warnow T. 2014b. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics. 30:i541–i548.

Mirarab S., Warnow T. 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics. 31:i44–i52.

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

Mockford E.L. 1993. North American Psocoptera (Insecta). North American Psocoptera (Insecta).

Mockford E.L., Lienhard C., Yoshizawa K., Mockford E.L., Lienhard C., Yoshizawa K. 2013. Revised classification of "Psocoptera" from Cretaceous amber, a reassessment of published information. Insecta matsumurana. New series : journal of the Faculty of Agriculture Hokkaido University, series entomology. 69:1–26.

de Moya R.S., Allen J.M., Sweet A.D., Walden K.K.O., Palma R.L., Smith V.S., Cameron S.L., Valim M.P., Galloway T.D., Weckstein J.D., Johnson K.P. 2019. Extensive host-switching of avian feather lice following the Cretaceous-Paleogene mass extinction event. Comm Biol. 2:1–6.

de Moya R.S. 2020. Psocodea Phylogenomic dataset, v2, Dryad, Dataset, https://doi.org/10.5061/dryad.c59zw3r50

New T.R. 1970. The Relative Abundance of Some British Psocoptera on Different Species of Trees. J Anim Ecol. 39:521–540.

New T.R. 1987. Biology of the Psocoptera. Orient Insects. 21:1–109.

Nguyen L.-T., Schmidt H.A., von Haeseler A., Minh B.Q. 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol. 32:268–274.

Pattengale N.D., Alipour M., Bininda-Emonds O.R.P., Moret B.M.E., Stamatakis A. 2009. How Many Bootstrap Replicates Are Necessary? Lect N Bioinformat.:184–200.

Pease J.B., Brown J.W., Walker J.F., Hinchliff C.E., Smith S.A. 2018. Quartet Sampling distinguishes lack of support from conflicting support in the green plant tree of life. Am J Bot. 105:385–403.

Peters R.S., Krogmann L., Mayer C., Donath A., Gunkel S., Meusemann K., Kozlov A., Podsiadlowski L., Petersen M., Lanfear R., Diez P.A., Heraty J., Kjer K.M., Klopfstein S., Meier R., Polidori C., Schmitt T., Liu S., Zhou X., Wappler T., Rust J., Misof B., Niehuis O. 2017. Evolutionary History of the Hymenoptera. Curr Biol. 27:1013–1018.

Petersen M., Meusemann K., Donath A., Dowling D., Liu S., Peters R.S., Podsiadlowski L., Vasilikopoulos A., Zhou X., Misof B., Niehuis O. 2017. Orthograph: a versatile tool for mapping coding nucleotide sequences to clusters of orthologous genes. BMC Bioinformatics. 18.

Philippe H., Brinkmann H., Lavrov D.V., Littlewood D.T.J., Manuel M., Wörheide G., Baurain D. 2011. Resolving Difficult Phylogenetic Questions: Why More Sequences Are Not Enough. PLOS Biology. 9:e1000602.

Posada D., Crandall K.A. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics. 14:817–818.

Price R.D., Hellenthal R.A., Palma R.L., Johnson K.P., Clayton D.H. 2003. Chewing Lice: World Checklist and Biological Overview. Champaign, Ill.: Illinois Natural History Survey.

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

Rambaut A., Drummond A.J., Xie D., Baele G., Suchard M.A. 2018. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Syst Biol. 67:901–904.

Regier J.C., Shultz J.W., Zwick A., Hussey A., Ball B., Wetzer R., Martin J.W., Cunningham C.W. 2010. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. Nature. 463:1079–1083.

Romiguier J., Cameron S.A., Woodard S.H., Fischman B.J., Keller L., Praz C.J. 2016. Phylogenomics Controlling for Base Compositional Bias Reveals a Single Origin of Eusociality in Corbiculate Bees. Mol Biol Evol. 33:670–678.

Roure B., Philippe H. 2011. Site-specific time heterogeneity of the substitution process and its impact on phylogenetic inference. BMC Evol Biol. 11:17.

RS Team. 2015. RStudio: integrated development for R. .

Sayyari E., Mirarab S. 2016. Fast Coalescent-Based Computation of Local Branch Support from Quartet Frequencies. Mol Biol Evol. 33:1654–1668.

719 Scholtz C.H. 2016. The higher classification of southern African insects. Afr Entomol. 24:545–
720 555.

721 Sheffield N.C., Song H., Cameron S.L., Whiting M.F. 2009. Nonstationary Evolution and
722 Compositional Heterogeneity in Beetle Mitochondrial Phylogenomics. Syst Biol. 58:381–
723 394.

724 Simion P., Philippe H., Baurain D., Jager M., Richter D.J., Di Franco A., Roure B., Satoh N.,
725 Quéinnec É., Ereskovsky A., Lapébie P., Corre E., Delsuc F., King N., Wörheide G.,
726 Manuel M. 2017. A Large and Consistent Phylogenomic Dataset Supports Sponges as the
727 Sister Group to All Other Animals. Curr Biol. 27:958–967.

728 Simmons M.P. 2017. Relative benefits of amino-acid, codon, degeneracy, DNA, and purine-
729 pyrimidine character coding for phylogenetic analyses of exons - Simmons - 2017 -
730 Journal of Systematics and Evolution - Wiley Online Library. Available from
731 http://onlinelibrary.wiley.com/wol1/doi/10.1111/jse.12233/full.

732 Simon S., Blanke A., Meusemann K. 2018. Reanalyzing the Palaeoptera problem - The origin of
733 insect flight remains obscure. Arthropod Struct Dev. 47:328–338.

734 Simon S., Letsch H., Bank S., Buckley T., Donath A., Liu S., Machida R., Meusemann K., Misof
735 B., Podsiadlowski L., Zhuo X., Wipfler B., Bradler S. 2019. Old World and New World
736 Phasmatodea: Phylogenomics Resolve the Evolutionary History of Stick and Leaf
737 Insects. Front Ecol Evol. 7.

738 Simpson J.T., Wong K., Jackman S.D., Schein J.E., Jones S.J.M., Birol İ. 2009. ABySS: A
739 parallel assembler for short read sequence data. Genome Res. 19:1117–1123.

740 Skinner R.K., Dietrich C.H., Walden K.K.O., Gordon E., Sweet A.D., Podsiadlowski L.,
741 Petersen M., Simon C., Takiya D.M., Johnson K.P. 2020. Phylogenomics of
742 Auchenorrhyncha (Insecta: Hemiptera) using transcriptomes: examining controversial
743 relationships via degeneracy coding and interrogation of gene conflict. Syst Entomol.
744 45:85–113.

745 Slater G.S.C., Birney E. 2005. Automated generation of heuristics for biological sequence
746 comparison. BMC Bioinformatics. 6:31.

747 Song F., Li H., Liu G.-H., Wang W., James P., Colwell D.D., Tran A., Gong S., Cai W., Shao R.
748 2019. Mitochondrial Genome Fragmentation Unites the Parasitic Lice of Eutherian
749 Mammals. Syst Biol. 68:430–440.

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

752 Strimmer K., Haeseler A. von. 1997. Likelihood-mapping: A simple method to visualize
753 phylogenetic content of a sequence alignment. PNAS. 94:6815–6819.

Sukumaran J., Holder M.T. 2015. SumTrees: Phylogenetic Tree Summarization. 4.0.0. Available from https://github.com/jeetsukumaran/DendroPy.

Vaidya G., Lohman D.J., Meier R. 2011. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. Cladistics. 27:171–180.

Vasilikopoulos A., Balke M., Beutel R.G., Donath A., Podsiadlowski L., Pflug J.M., Waterhouse R.M., Meusemann K., Peters R.S., Escalona H.E., Mayer C., Liu S., Hendrich L., Alarie Y., Bilton D.T., Jia F., Zhou X., Maddison D.R., Niehuis O., Misof B. 2019. Phylogenomics of the superfamily Dytiscoidea (Coleoptera: Adephaga) with an evaluation of phylogenetic conflict and systematic error. Mol Phylogenet Evol. 135:270–285.

Wang R., Yao Y., Ren D., Shih C. 2019. Psocoptera – Barklice and Booklice. Rhythms of Insect Evolution. John Wiley & Sons, Ltd. p. 185–188.

Wappler T., Smith V.S., Dalgleish R.C. 2004. Scratching an ancient itch: an Eocene bird louse fossil. P R Soc B-Biol Sci. 271:S255–S258.

Wipfler B., Letsch H., Frandsen P.B., Kapli P., Mayer C., Bartel D., Buckley T.R., Donath A., Edgerly-Rooks J.S., Fujita M., Liu S., Machida R., Mashimo Y., Misof B., Niehuis O., Peters R.S., Petersen M., Podsiadlowski L., Schütte K., Shimizu S., Uchifune T., Wilbrandt J., Yan E., Zhou X., Simon S. 2019. Evolutionary history of Polyneoptera and its implications for our understanding of early winged insects. Proc Natl Acad Sci USA. 116:3024–3029.

Yang Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol Biol Evol. 24:1586–1591.

Yoshizawa K. 2002. Phylogeny and higher classification of suborder Psocomorpha (Insecta: Psocodea: 'Psocoptera'). Zool J Linn Soc. 136:371–400.

Yoshizawa K., Johnson K.P. 2010. How stable is the "Polyphyly of Lice" hypothesis (Insecta: Psocodea)?: A comparison of phylogenetic signal in multiple genes. Mol Phylogenet Evol. 55:939–951.

Yoshizawa K., Johnson K.P. 2013. Changes in base composition bias of nuclear and mitochondrial genes in lice (Insecta: Psocodea). Genetica. 141:491–499.

Yoshizawa K., Johnson K.P. 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., Lienhard C. 2010. In search of the sister group of the true lice : A systematic review of booklice and their relatives, with an updated checklist of Liposcelididae (Insecta: Psocodea). Arthropod Syst Phylo. 68:181–195.

Yoshizawa K., Lienhard C., Johnson K.P. 2006. Molecular systematics of the suborder Trogiomorpha (Insecta: Psocodea: 'Psocoptera'). Zool J Linn Soc. 146:287–299.

Yoshizawa K., Lienhard C., Yao I., Ferreira R.L. 2019. Cave insects with sex-reversed genitalia had their most recent common ancestor in West Gondwana (Psocodea: Prionoglarididae: Speleketorinae). Entomol Sci. 22:334–338.

Zwick A., Regier J.C., Zwickl D.J. 2012. Resolving Discrepancy between Nucleotides and Amino Acids in Deep-Level Arthropod Phylogenomics: Differentiating Serine Codons in 21-Amino-Acid Models. PLOS ONE. 7:e47450.

800 Figure Captions

801

802 Fig. 1: The result of phylogenomic analyses using degeneracy recoded nucleotide data. Clade
803 support is depicted as bootstrap support. The timescale provides an estimate of divergences
804 suggested by MCMCtree dating analyses using correlated rates. Taxonomic names marked with
805 an asterisk represent samples that are derived from transcriptomes. Names which lack the
806 asterisk represent samples derived from shotgun whole genome sequencing. The H. following
807 names indicate the taxon is classified within the Homilopsocidea.

808

809 Fig. 2: The result of likelihood-mapping derived from IQtree. Results show the percentage of
810 quartets derived from analyses supporting relationships among the Trogiomorpha, Psocomorpha,
811 Nanopsocetae, and Amphientometae.

812

813 Fig. 3: A cladogram of the result of quartet sampling based on the analysis of all nucleotide sites.
814 Clade support is depicted as: Quartet Concordance (QC)/Quartet Differential (QD)/Quartet
815 Informativeness (QI). Taxonomic names marked with an asterisk represent samples that are
816 derived from transcriptomes. Names which lack the asterisk represent samples derived from
817 shotgun whole genome sequencing. The H. following names indicate the taxon is classified
818 within the Homilopsocidea.

819

820 Fig. 4: A cladogram of the result of quartet sampling based on the analysis of recoded
821 nucleotides using degeneracy methods. Clade support is depicted as: Quartet Concordance
822 (QC)/Quartet Differential (QD)/Quartet Informativeness (QI). Taxonomic names marked with an

asterisk represent samples that are derived from transcriptomes. Names which lack the asterisk represent samples derived from shotgun whole genome sequencing. The H. following names indicate the taxon is classified within the Homilopsocidea.

Fig. 5: Box and whisker plot showing the distribution of GC content in third codon positions from the alignments analyzed.

Fig. 6: Box and whisker plot showing the distribution of GC content in first codon positions from the alignments analyzed.

Fig. 7: Box and whisker plot showing the distribution of GC content in second codon positions from the alignments analyzed.

Table Captions

Table 1: A summary of all species of Psocodea sampled and data analyzed. Transcriptome sequences are available from a previous study (Johnson et al. 2018a).

Table 2: A summary of the phylogenetic analyses completed, and respective data type analyzed.

Table 3: A comparison between historical ordinal taxonomic schemes for Psocoptera and Phthiraptera to the newly proposed classification scheme for a single order Psocodea.

Supplemental Figure Captions

Supplemental Fig. 1: Maximum likelihood results of the concatenated maximum likelihood analysis of second codon positions only. Taxa marked with stars represent transcriptome-based samples.

Supplemental Fig. 2: The ASTRAL coalescent result of 2,370 gene alignments of all nucleotide sites. Clade support is based upon local posterior probabilities. Taxa marked with stars represent transcriptome-based samples.

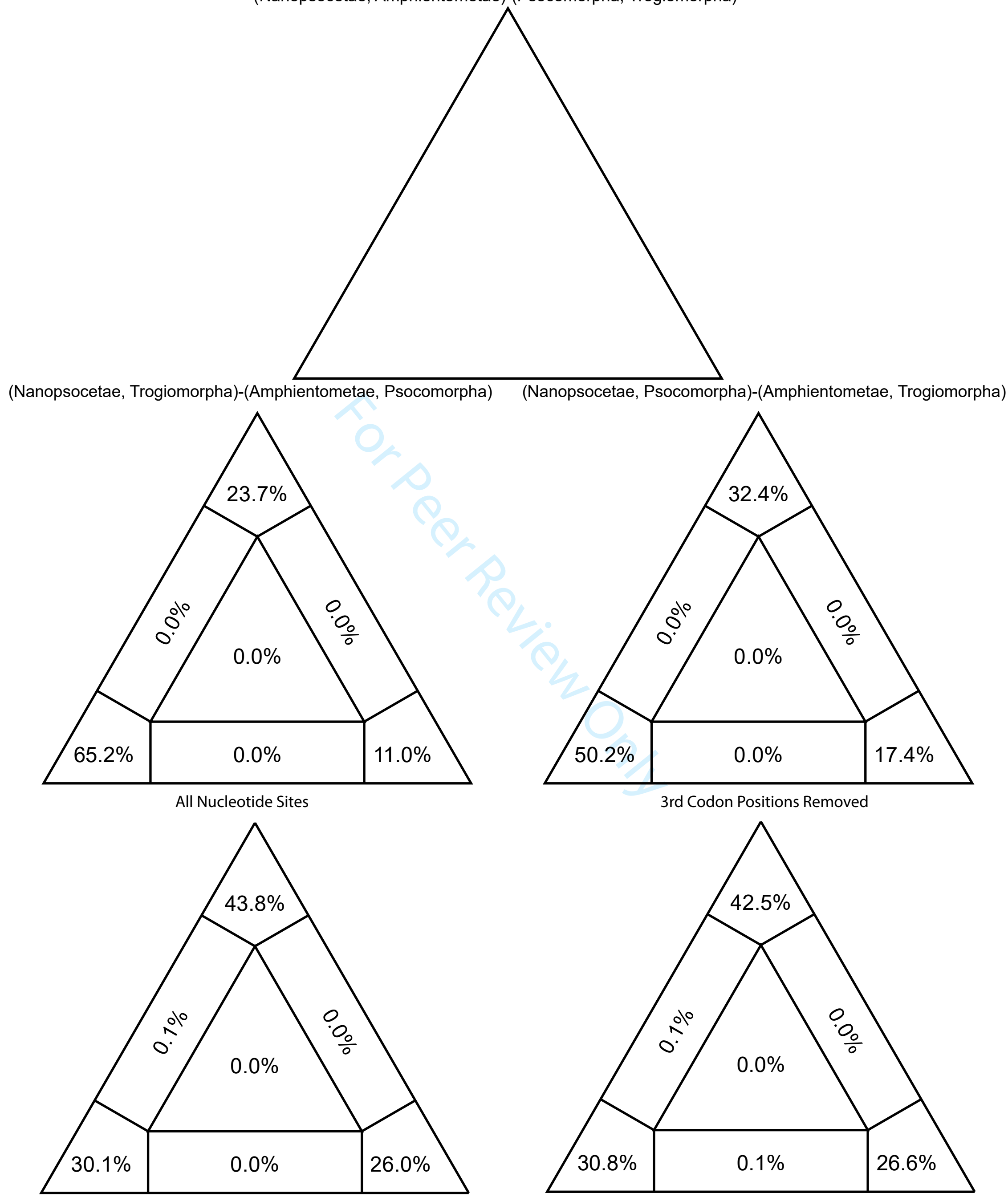
Supplemental Fig. 3: The ASTRAL coalescent result of 2,370 degeneracy recoded gene alignments. Clade support is based upon local posterior probabilities. Taxa marked with stars represent transcriptome-based samples.

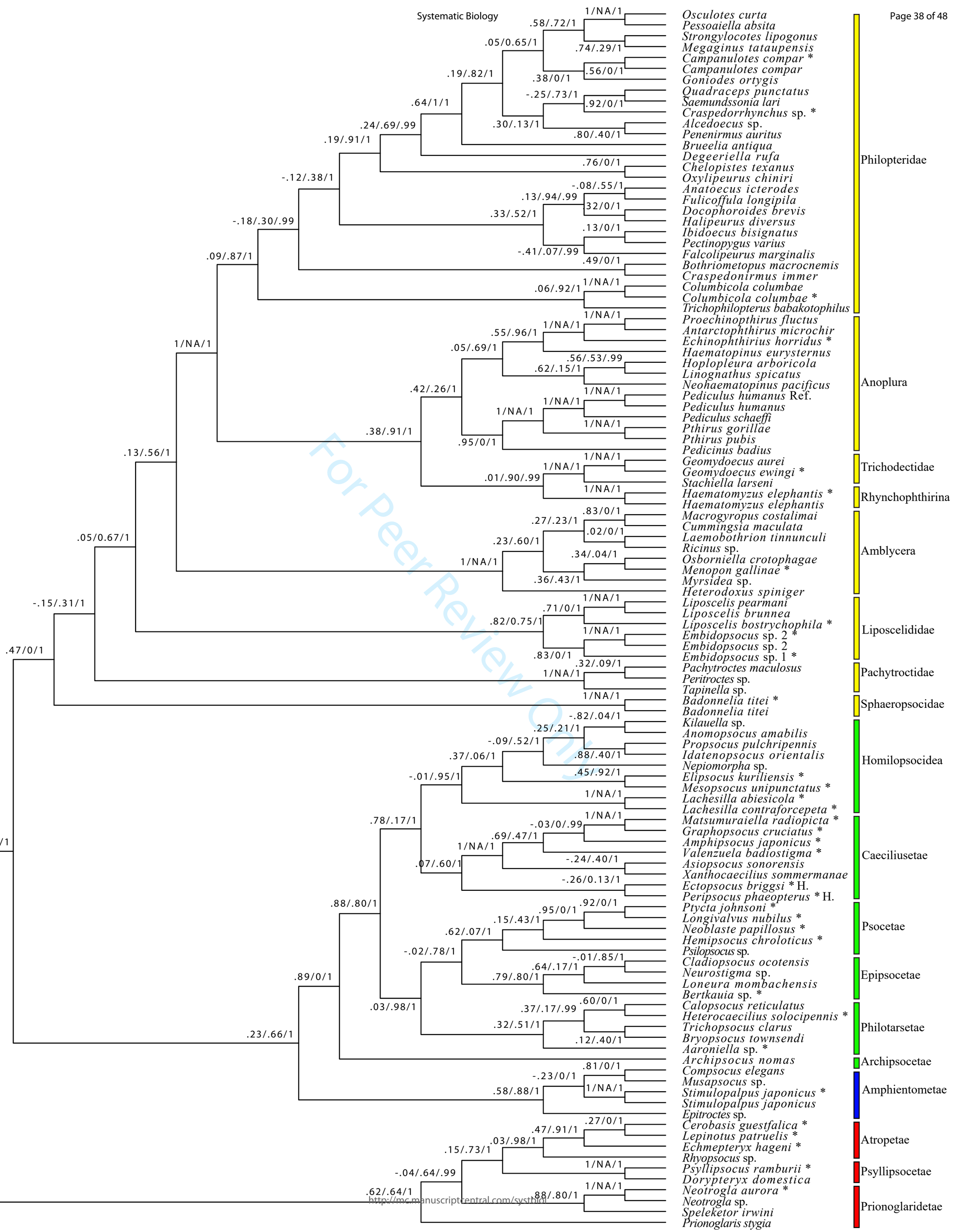
Osculotes curta
Pessoaiella absita
Megaginus tataupensis
Strongylocotes lipogonus
Campanulotes compar
Campanulotes compar *
Goniodes ortygis
Saemundssonia lari
Quadraceps punctatus
Penenirmus auritus
Alcedoecus sp.
Craspedorrhynchus sp. *
Brueelia antiqua
Degeeriella rufa
Chelopistes texanus
Oxylipeurus chiniri
Anatoecus icterodes
Fulicoffula longipila
Halipeurus diversus
Docophoroides brevis
Ibidoecus bisignatus
Pectinopygus varius
Falcolipeurus marginalis
Bothriometopus macrocnemis
Trichophilopterus babakotophilus
Craspedonirmus immer
Columbicola columbae
Columbicola columbae *
Proechinopthirus fluctus
Antarctophthirus microchir
Echinophthirius horridus *
Haematopinus eurysternus
Hoplopleura arboricola
Linognathus spicatus
Neohaematopinus pacificus
Pediculus humanus
Pediculus humanus Ref.
Pediculus schaeffi
Pthirus gorillae
Pthirus pubis
Pedicinus badius
Haematomyzus elephantis
Haematomyzus elephantis *
Geomydoecus ewingi *
Geomydoecus aurei
Stachiella larseni
Laemobothrion tinnunculi
Heterodoxus spiniger
Ricinus sp.
Macrogyropus costalimai
Cummingsia maculata
Myrsidea sp.
Osborniella crotophagae
Menopon gallinae *
Embidopsocus sp. 2
Embidopsocus sp. 2 *
Embidopsocus sp. 1 *
Liposcelis brunnea
Liposcelis pearmani
Liposcelis bostrychophila *
Tapinella sp.
Pachytroctes maculosus
Peritroctes sp.
Badonnelia titei
Badonnelia titei *
Musapsocus sp.
Compsocus elegans
Stimulopalpus japonicus
Stimulopalpus japonicus *
Epitroctes sp.
Kilauella sp.
Propsocus pulchripennis
Idatenopsocus orientalis
Anomopsocus amabilis
Nepiomorpha sp.
Mesopsocus unipunctatus *
Elipsocus kuriliensis *
Lachesilla contraforcepeta *
Lachesilla abiesicola *
Ectopsocus briggsi *
Matsumuraiella radiopicta *
Graphopsocus cruciatus *
Amphipsocus japonicus *
Valenzuela badiostigma *
Xanthocaecilius sommermanae
Asiopsocus sonorensis
Peripsocus phaeopterus * H.
Longivalvus nubilus *
Ptycta johnsoni *
Neoblaste papillosus *
Hemipsocus chroloticus *
Psilopsocus sp.
Cladiopsocus ocotensis
Neurostigma sp.
Loneura mombachensis
Bertkauia sp. *
Bryopsocus townsendi
Trichopsocus clarus
Calopsocus reticulatus
Heterocaecilius solocipennis *
Aaroniella sp. *
Archipsocus nomas
Cerobasis guestfalica *
Lepinotus patruelis *
Echmepteryx hageni *
Rhyopsocus sp.
Dorypteryx domestica
Psyllipsocus ramburii *
Neotrogla sp.
Neotrogla aurora *
Speleketor irwini
Prionoglaris stygia

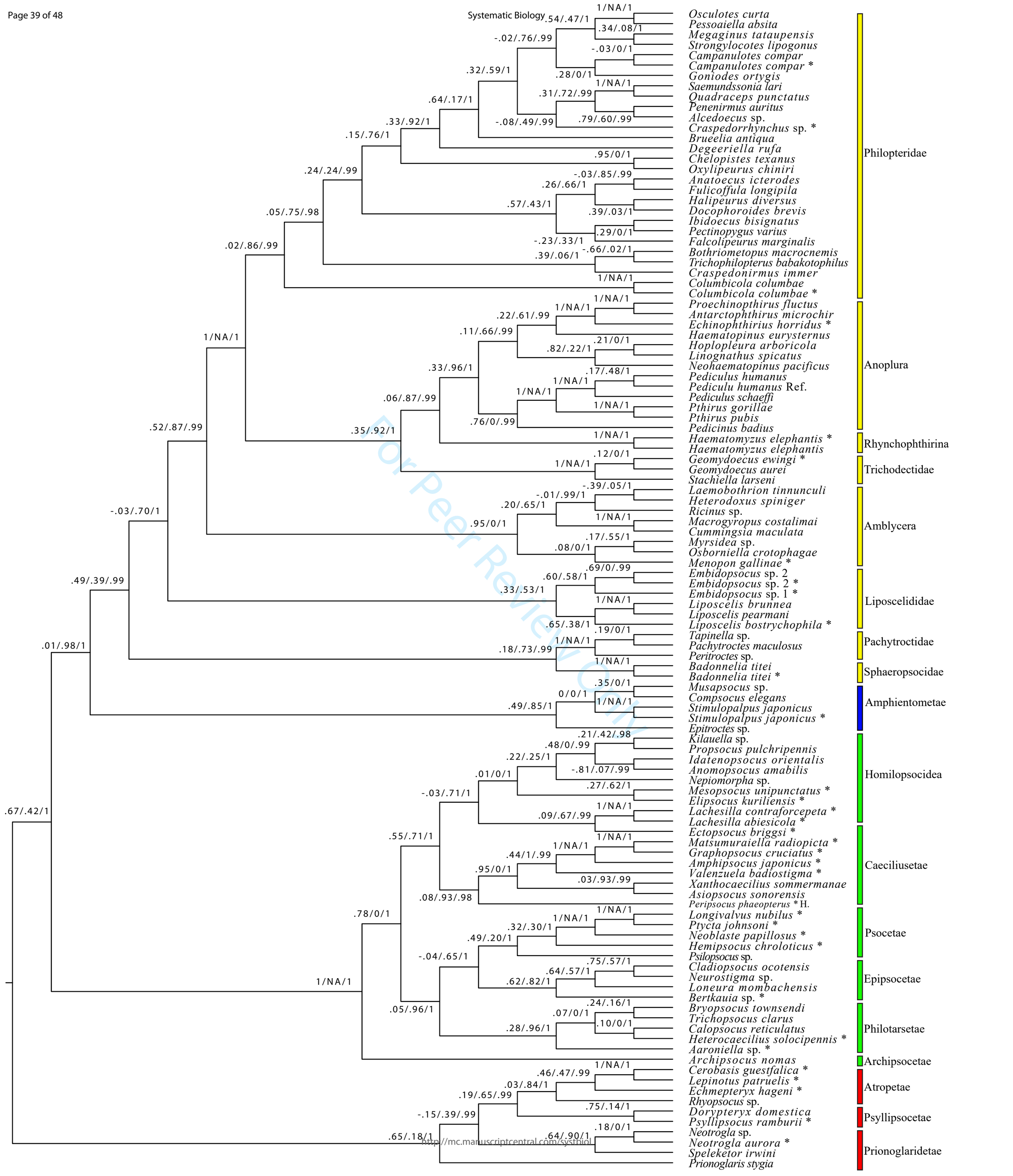
Philopteridae
Anoplura
Rhynchophthirina
Trichodectidae
Amblycera
Liposcelididae
Pachytroctidae
Sphaeropsocidae
Amphientometae
Troctomorpha
Homilopsocidea
Caeciliusetae
Psocetae
Epipsocetae
Philotarsetae
Archipsocetae
Psocomorpha
Atropetae
Psyllipsocetae
Prionoglaridetae
Trogiomorpha

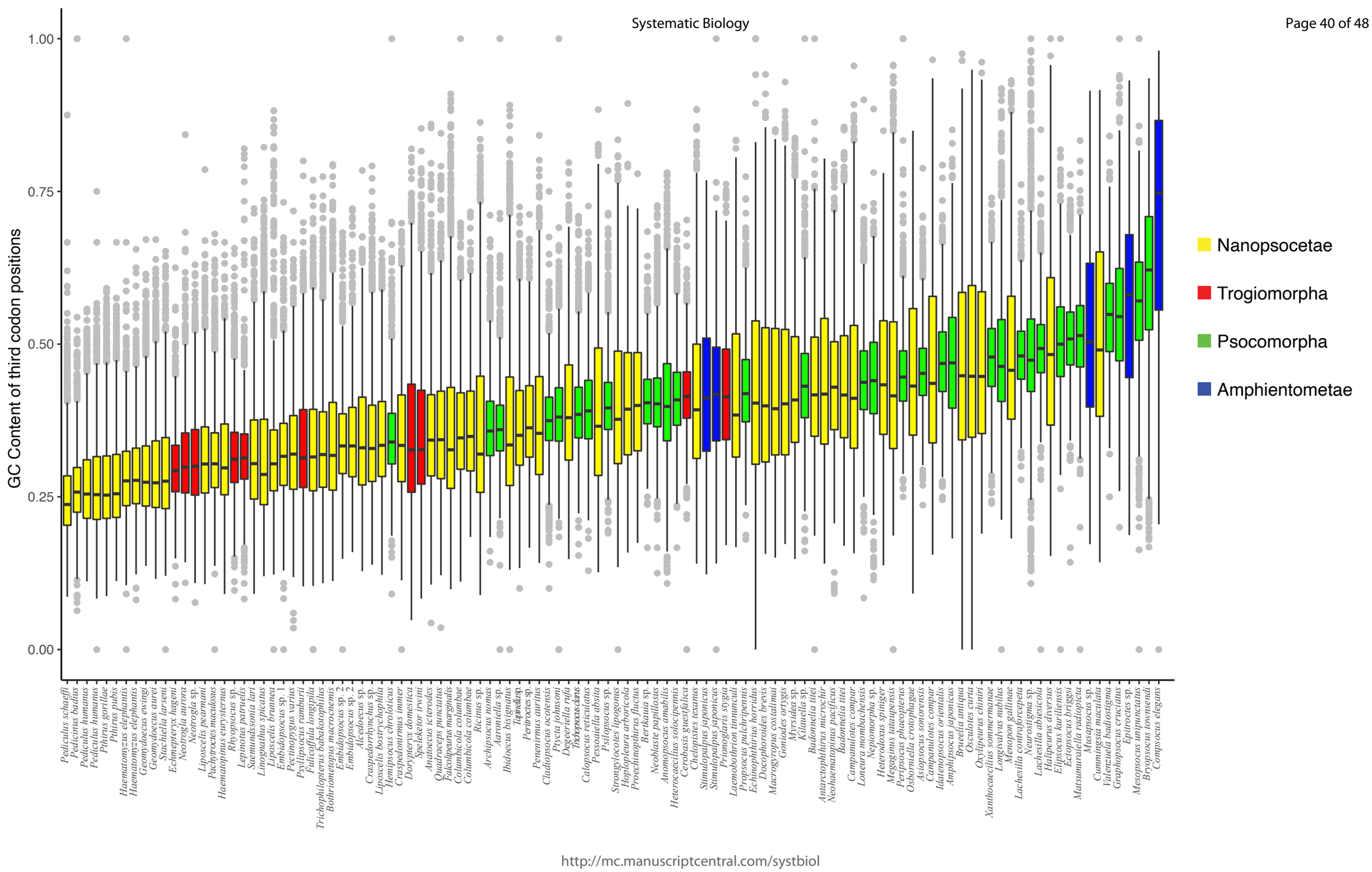
175 150 125 100 75 50 25 0 MYA

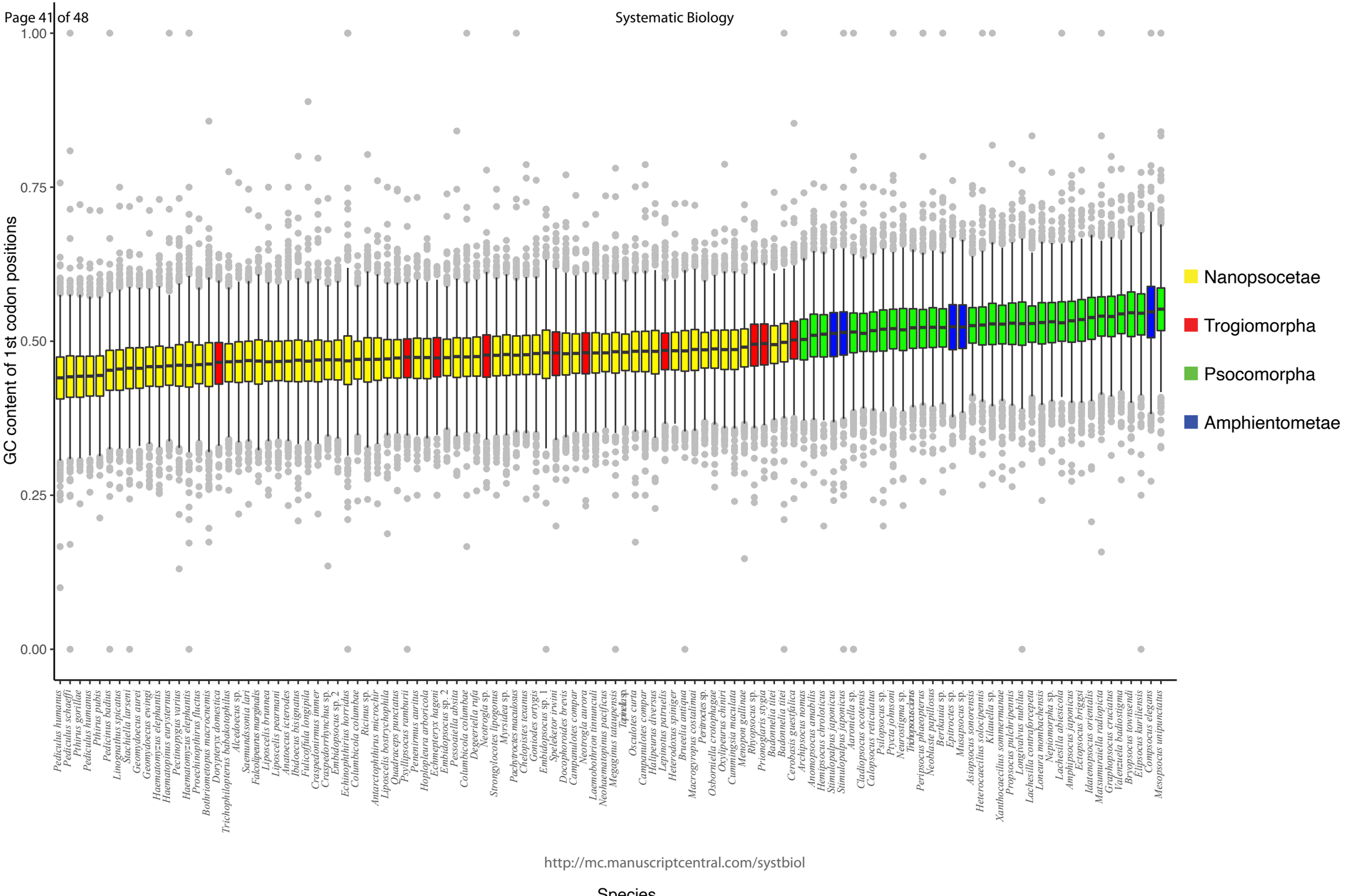
Mesozoic | Cenozoic
Jursassic | Cretaceous | Paleogene | Neogene | Quartenary

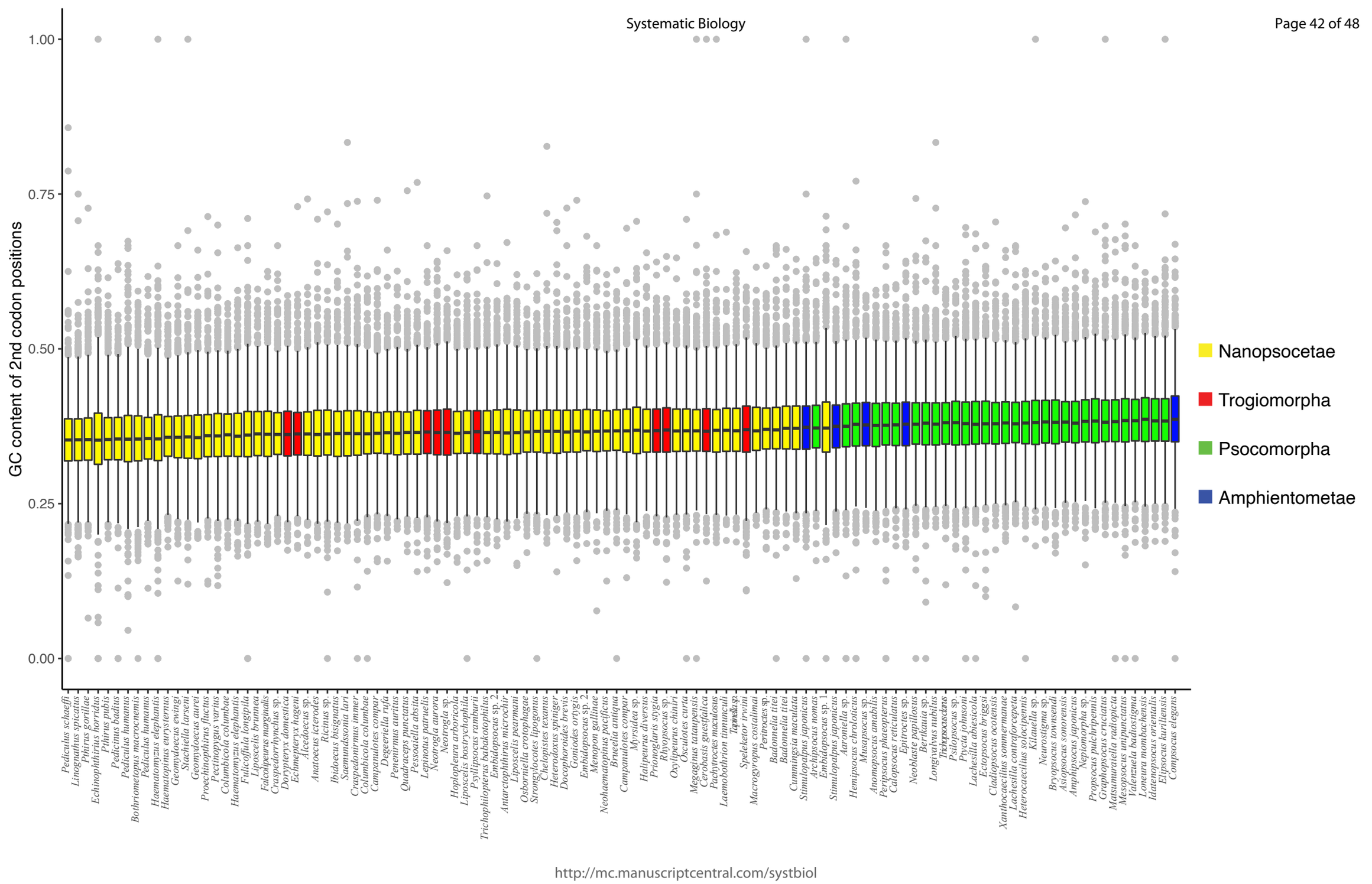







GC Content of third codon positions
1.00
0.75
0.50
0.25
0.00
Nanopsocetae
Trogiomorpha
Psocomorpha
Amphientometae
Pediculus schaeffi
Pediculus baditus
Pediculus humanus
Pediculus humanus
Pthirus gorillae
Pthirus pubis
Haematomyzus elephantis
Haematomyzus elephantis
Geomydoecus ewingi
Geomydoecus aurei
Stachiella larseni
Echmepteryx hageni
Neotrogla aurora
Neotrogla sp.
Liposcelis pearmani
Pachytroctes maculosus
Haematopinus eurysternus
Rhyposocus sp.
Lepinotus patruelis
Saemundssonia lari
Linognathus spicatus
Liposcelis brunnea
Embidopsocus sp. 1
Pectinopygus varius
Psyllipsocus ramburii
Fulicoffula longipila
Trichophilopterus babakotophilus
Bothriometopus macrocnemis
Embidopsocus sp. 2
Embidopsocus sp. 2
Aleodoecus sp.
Craspedorrhynchus sp.
Liposcelis bostrychophila
Hemipsocus chrolonicus
Craspedonirmus inmer
Dorypteryx domestica
Speleketor irwini
Anatoecus icterodes
Quadraceps punctatus
Falcolipeurus marginalis
Columbicola columbae
Columbicola columbae
Ricinus sp.
Archipsocus nomas
Auroniella sp.
Ibidoecus bisignatus
Peritroctes sp.
Penenirmus auritus
Cladiopsocus ocotensis
Prycta johnsoni
Degeeriella rufa
Calopsocus reticulatus
Pessoaiella absita
Psilopsocus sp.
Strongylocotes lipogonus
Hoplopleura arboricola
Proechinophthirus fluctus
Berkautia sp.
Neoblaste papillosus
Anomopsocus amabilis
Heterocaecilius solocipennis
Cerobasis guestfalica
Chelopistes texanus
Stimulopalpus japonicus
Stimulopalpus japonicus
Prionoglaris stygia
Laemobothrion tinnunculi
Propsocus pulchripennis
Echinophthirus horridus
Docophoroides brevis
Macrogyropus costalimai
Goniodes ortygis
Myrsidea sp.
Kilauella sp.
Badonnelia titei
Antarctophthirus microchir
Neohaematopinus pacificus
Badonnelia titei
Campanulotes compar
Loneura mombachensis
Nepiomorpha sp.
Heterodoxus spiniger
Megaginus tataupensis
Peripsocus phaeopterus
Osbornella crotophagae
Asiopsocus sonorensis
Campanulotes compar
Idiotenopsocus orientalis
Amphipsocus japonicus
Brueelia antiqua
Osculotes curta
Oxylipeurus chiniri
Xanthocaecilius sommermanae
Longivalvus nubilus
Menopon gallinae
Lachesilla contraforcepeta
Neurostigma sp.
Lachesilla abiesicola
Halipeurus diversus
Elipsocus kuriliensis
Ectopsocus briggsi
Matsumuraiella radiopicta
Musapsocus sp.
Cummingsia maculata
Valenzuela badiostigma
Graphopsocus cruciatus
Epitroctes sp.
Mesopsocus unipunctatus
Bryopsocus townsendi
Compsocus elegans
http://mc.manuscriptcentral.com/systbiol
Species


1.00
0.75
0.50
0.25
0.00
GC content of 1st codon positions
Nanopsocetae
Trogiomorpha
Psocomorpha
Amphientometae
Pediculus humanus
Pediculus schaeffi
Pthirus gorillae
Pediculus humanus
Phthirus pubis
Pediculus baduis
Linognathus spicatus
Stachiella larseni
Geomydoecus aurei
Geomydoecus ewingi
Haematomyzus elephantis
Haematopinus eurysternus
Pectinopygus varius
Haematomyzus elephantis
Proechinophthirus fluctus
Bothriometopus macrocnemis
Dorypteryx domestica
Trichophilopterus babakotophilus
Alcedoecus sp.
Saemundssonia lari
Falcolipeurus marginalis
Liposcelis brunnea
Liposcelis pearmani
Anatoecus icterodes
Ibidoecus bisignatus
Fulicoffula longipila
Craspedonirmus inner
Craspedorrhynchus sp.
Embidopsocus sp. 2
Echinophthirius horridus
Columbicola columbae
Ricinus sp.
Antarctophthirus microchir
Liposcelis bostrychophila
Quadraceps punctatus
Psylipsocus ramburii
Penenirmus auritus
Hoplopleura arboricola
Echmepteryx hageni
Embidopsocus sp. 2
Pessoatella absita
Columbicola columbae
Degeeriella rufa
Neotrogla sp.
Strongylocotes lipogonus
Myrsidea sp.
Pachytroctes maculosus
Chelopistes texanus
Goniodes ortygis
Embidopsocus sp. 1
Speleketor irwini
Docophoroides brevis
Campanulotes compar
Neotrogla aurora
Laemobothrion tinnunculi
Neohaematopinus pacificus
Megagnius tataupensis
Tapinella sp.
Osculotes curta
Campanulotes compar
Halipeurus diversus
Lepinotus patruelis
Heterodoxus spiniger
Brueelia antiqua
Macrogyropus costalimai
Peritroctes sp.
Osborniella crotophagae
Oxylipeurus chiniri
Cummingsia maculata
Menopon gallinae
Rhyopsocus sp.
Prionoglaris stygia
Badonnelia titei
Badonnelia titei
Cerobasis guestfalica
Archipsocus nomas
Anomopsocus amabilis
Hemipsocus chrolaticus
Stimulopalpus japonicus
Stimulopalpus japonicus
Aaroniella sp.
Cladiopsocus ocotensis
Calopsocus reticulatus
Psilopsocus sp.
Prycta johnsoni
Neurostigma sp.
Trichopsocus sp.
Peripsocus phaeopterus
Neoblaste papillosus
Berthauia sp.
Epitroctes sp.
Musapsocus sp.
Asiopsocus sonorensis
Heterocaecilius solocipennis
Kilauella sp.
Xanthocaecilius sommermanae
Propsocus pulchripennis
Longivalvus nubilus
Lachesilla contraforcepeta
Loneura mombachensis
Nepiomorpha sp.
Lachesilla abiesicola
Amphipsocus japonicus
Ectopsocus briggsi
Idatenopsocus orientalis
Matsumuraiella radiopicta
Graphopsocus cruciatus
Valenzuela badiostigma
Bryopsocus townsendi
Elipsocus kurilensis
Compsocus elegans
Mesopsocus unipunctatus
Species


GC content of 2nd codon positions
1.00
0.75
0.50
0.25
0.00
Nanopsocetae
Trogiomorpha
Psocomorpha
Amphientometae
Pediculus schaeffi
Linognathus spicatus
Pthirus gorillae
Echinophthirius horridus
Pthirus pubis
Pediculus baduis
Pediculus humanus
Bothriometopus macrocnemis
Pediculus humanus
Haematomyzus elephantis
Haematopinus eurysternus
Geomydoecus ewingi
Stachiella larseni
Geomydoecus aurei
Proechinophthirus fluctus
Pectinopygus varius
Columbicola columbae
Haematomyzus elephantis
Fulicoffula longipila
Liposcelis brunnea
Falcolipeurus marginalis
Craspedorrhynchus sp.
Dorypteryx domestica
Echmepteryx hageni
Alcedoecus sp.
Anatoecus icterodes
Ricinus sp.
Ibidoecus bisignatus
Saemundssonia lari
Craspedonirmus immer
Columbicola columbae
Campanulotes compar
Degeeriella rufa
Penenirmus auritus
Quadraceps punctatus
Pessoatella absita
Lepinotus patruelis
Neotrogla aurora
Neotrogla sp.
Hoplopleura arboricola
Liposcelis bostrychophila
Psyllipsocus ramburii
Trichophilopterus babakothipilus
Embidopsocus sp. 2
Antarctophthirus microchir
Liposcelis pearmani
Osbornella crotophagae
Strongylocotes lipogonus
Chelopistes texanus
Heterodoxus spiniger
Docophoroides brevis
Goniodes orygis
Embidopsocus sp. 2
Menopon gallinae
Neohaematopinus pacificus
Brueelia antiqua
Campanulotes compar
Myrsidea sp.
Halipeurus diversus
Prionoglaris stygia
Rhyopsocus sp.
Oxylipeurus chiriri
Osculotes curta
Megaginus tataupensis
Cerobasis guestfalica
Pachytroctes maculosus
Laemobothrion tinnunculi
Tapinella sp.
Spelekelor irwini
Macrogyropus costalimai
Pentrocles sp.
Badonnelia titei
Badonnelia titei
Cummingsia maculata
Stimulopalpus japonicus
Archipsocus nomas
Embidopsocus sp. 1
Stimulopalpus japonicus
Aaroniella sp.
Hemipsocus chrolonicus
Musapsocus sp.
Anomopsocus amabilis
Peripsocus phaeopterus
Calopsocus reticulatus
Epitroctes sp.
Neoblaste papillosus
Berkania sp.
Longivalvus nubilus
Trichopsocus sp.
Psilopsocus sp.
Pnyeta johnsoni
Lachesilla abiesicola
Ectopsocus briggsi
Cladiopsocus ocotensis
Xanthocaecilius sommermanae
Lachesilla contraforcepeta
Heterocaecilius solocipennis
Kilauella sp.
Neurostigma sp.
Bryopsocus townsendi
Asiopsocus sonorensis
Amphipsocus japonicus
Nepiomorpha sp.
Propsocus pulchripennis
Graphopsocus cruciatus
Matsumuraiella radiopicta
Mesopsocus unipunctatus
Valenzuela badiostigma
Loneura mombachensis
Idatenopsocus orientalis
Elipsocus kurilensis
Compsocus elegans
Species

Psocodea Taxnomic Sampling Scheme

Suborder or Order	Family	Taxon	Data Type	Total BP	# Genes	SRA Accession
Anoplura	Haematopinidae	*Haematopinus eurysternus*	WGS	2549976	2357	SRR5308123
Anoplura	Echinophthiriidae	*Proechinopthirus fluctus*	WGS	2253525	2335	SRR5308138
Anoplura	Echinophthiriidae	*Antarctophthirus microchir*	WGS	2440791	2349	SRR5088465
Anoplura	Echinophthiriidae	*Echinopthirus horridus*	RNA seq	785123	1599	SRR2051484
Anoplura	Linognathidae	*Linognathus spicatus*	WGS	2567331	2332	SRR5308129
Anoplura	Polyplacidae	*Neohaematopinus pacificus*	WGS	2787132	2364	SRR5088469
Anoplura	Hoplopleuridae	*Hoplopleura arboricola*	WGS	2336922	2359	SRR5088468
Anoplura	Pedicinidae	*Pedicinus badius*	WGS	2344773	2356	SRR5308136
Anoplura	Pthiridae	*Pthirus gorillae*	WGS	2789559	2368	SRR5088474
Anoplura	Pthiridae	*Pthirus pubis*	WGS	2754231	2365	SRR5088475
Anoplura	Pediculidae	*Pediculus schaeffi*	WGS	2562033	2364	SRR1182279
Anoplura	Pediculidae	*Pediculus humanus*	WGS	2277984	2354	SRR5088472
Anoplura	Pediculidae	*Pediculus humanus*	Reference	2945068	2370	PRJNA19807
Rhynchophthirina	Haematomyzidae	*Haematomyzus elephantis*	WGS	2439042	2358	SRR5308122
Rhynchophthirina	Haematomyzidae	*Haematomyzus elephantis*	RNA seq	1194114	1867	SRR2051491
Ischnocera	Trichodectidae	*Stachiella larseni*	WGS	2524617	2353	SRR5308143
Ischnocera	Trichodectidae	*Geomydoecus aurei*	WGS	2507592	2355	SRR5308121
Ischnocera	Trichodectidae	*Geomydoecus ewingi*	RNA seq	2596821	2310	SRR1821919
Ischnocera	Philopteridae	*Trichophilopterus babakotophilus*	WGS	2497134	2359	SRR5308144
Ischnocera	Philopteridae	*Bothriometopus macrocnemis*	WGS	1835208	2259	SRR5088466
Ischnocera	Philopteridae	*Craspedonirmus immer*	WGS	2193294	2348	SRR5308116
Ischnocera	Philopteridae	*Columbicola columbae*	WGS	2225532	2356	SRR5308115
Ischnocera	Philopteridae	*Columbicola columbae*	RNA seq	1968492	2218	SRR1821984
Ischnocera	Philopteridae	*Docophoroides brevis*	WGS	2313018	2355	SRR5308117
Ischnocera	Philopteridae	*Halipeurus diversus*	WGS	2281860	2354	SRR5308124
Ischnocera	Philopteridae	*Fulicoffula longipila*	WGS	2129253	2345	SRR5308119
Ischnocera	Philopteridae	*Anatoecus icterodes*	WGS	2208450	2338	SRR5308111
Ischnocera	Philopteridae	*Falcolipeurus marginalis*	WGS	2280087	2350	SRR5308118
Ischnocera	Philopteridae	*Ibidoecus bisignatus*	WGS	2341512	2353	SRR5308126
Ischnocera	Philopteridae	*Pectinopygus varius*	WGS	2034798	2315	SRR5308135
Ischnocera	Philopteridae	*Chelopistes texanus*	WGS	2589447	2363	SRR5308114
Ischnocera	Philopteridae	*Oxylipeurus chiniri*	WGS	2668434	2368	SRR5308134
Ischnocera	Philopteridae	*Degeeriella rufa*	WGS	2485305	2353	SRR5088467
Ischnocera	Philopteridae	*Brueelia antiqua*	WGS	2700864	2357	SRR5308112
Ischnocera	Philopteridae	*Penenirmus auritus*	WGS	2497635	2361	SRR5308137
Ischnocera	Philopteridae	*Alcedoecus* sp.	WGS	2417226	2359	SRR5308110
Ischnocera	Philopteridae	*Quadraceps punctatus*	WGS	2566107	2362	SRR5308139
Ischnocera	Philopteridae	*Saemundssonia lari*	WGS	2501502	2353	SRR5308141
Ischnocera	Philopteridae	*Goniodes ortygis*	WGS	2498694	2356	SRR5308120
Ischnocera	Philopteridae	*Campanulotes compar*	WGS	2336796	2355	SRR5308113
Ischnocera	Philopteridae	*Campanulotes compar*	RNA seq	2134521	2251	SRR1821983
Ischnocera	Philopteridae	*Strongylocotes lipogonus*	WGS	2739207	2363	SRR5308142
Ischnocera	Philopteridae	*Megaginus tataupensis*	WGS	2582559	2364	SRR5308131
Ischnocera	Philopteridae	*Pessoaiella absita*	WGS	2403510	2358	SRR5308145
Ischnocera	Philopteridae	*Osculotes curta*	WGS	2480868	2363	SRR5308133
Ischnocera	Philopteridae	*Craspedorrhynchus* sp.	RNA seq	2560936	2301	SRR1821912
Amblycera	Ricinidae	*Ricinus* sp.	WGS	2187996	2310	SRR5308140
Amblycera	Menoponidae	*Osborniella crotophagae*	WGS	2399703	2342	SRR5088470
Amblycera	Menoponidae	*Myrsidea* sp.	WGS	1997502	2284	SRR5308132
Amblycera	Menoponidae	*Menopon gallinae*	RNA seq	2405619	2263	SRR921619
Amblycera	Boopiidae	*Heterodoxus spiniger*	WGS	2451921	2340	SRR5308125
Amblycera	Trimenoponidae	*Cummingsia maculata*	WGS	2314959	2335	SRR5308146
Amblycera	Gyropidae	*Macrogyropus costalimai*	WGS	2432211	2347	SRR5308130
Amblycera	Laemobothriidae	*Laemobothrion tinnunculi*	WGS	2425458	2344	SRR5308127
Troctomorpha	Liposcelididae	*Liposcelis brunnea*	WGS	2112042	2329	SRR5308128
Troctomorpha	Liposcelididae	*Liposcelis pearmani*	WGS	2199234	2346	SRR5308268
Troctomorpha	Liposcelididae	*Liposcelis bostrychophila*	RNA seq	2125494	2201	SRR921613
Troctomorpha	Liposcelididae	*Embidopsocus* sp. 2	WGS	2330187	2353	SRR5308269
Troctomorpha	Liposcelididae	*Embidopsocus* sp. 2	RNA seq	2557295	2318	SRR5134727
Troctomorpha	Liposcelididae	*Embidopsocus* sp.	RNA seq	963150	1709	SRR2051486
Troctomorpha	Pachytroctidae	*Pachytroctes maculosus*	WGS	2317935	2313	SRR5308279
Troctomorpha	Pachytroctidae	*Tapinella* sp.	WGS	2370138	2336	SRR5308286
Troctomorpha	Pachytroctidae	*Peritroctes* sp.	WGS	2442492	2340	SRR5308280
Troctomorpha	Sphaeropsocidae	*Badonnelia titei*	WGS	2394783	2343	SRR5308262
Troctomorpha	Sphaeropsocidae	*Badonnelia titei*	RNA seq	2183178	2186	SRR2051472
Troctomorpha	Amphientomidae	*Stimulopalpus japonicus*	WGS	2270973	2300	SRR5088476
Troctomorpha	Amphientomidae	*Stimulopalpus japonicus*	RNA seq	1828280	2061	SRR2051511
Troctomorpha	Musapsocidae	*Musapsocus* sp.	WGS	2363169	2348	SRR5308275
Troctomorpha	Compsocidae	*Compsocus elegans*	WGS	2367267	2322	SRR5308266
Troctomorpha	Electrentomidae	*Epitroctes* sp.	WGS	2400144	2330	SRR5308270
Psocomorpha	Elipsocidae	*Kilauella* sp.	WGS	2329824	2313	SRR5308272
Psocomorpha	Elipsocidae	*Nepiomorpha* sp.	WGS	2570349	2309	SRR5308276
Psocomorpha	Elipsocidae	*Propsocus pulchripennis*	WGS	2325882	2309	SRR5308281
Psocomorpha	Elipsocidae	*Elipsocus kuriliensis*	RNA seq	1843671	2158	SRR2051485
Psocomorpha	Pseudocaeciliidae	*Calopsocus reticulatus*	WGS	2301513	2288	SRR5308264
Psocomorpha	Pseudocaeciliidae	*Bryopsocus townsendi*	WGS	2188992	2268	SRR5308263
Psocomorpha	Pseudocaeciliidae	*Heterocaecilius solocipennis*	RNA seq	1906914	2126	SRR2051493
Psocomorpha	Trichopsocidae	*Trichopsocus clarus*	WGS	2332443	2298	SRR5308287
Psocomorpha	Cladiopsocidae	*Cladiopsocus ocotensis*	WGS	2312040	2290	SRR5308265
Psocomorpha	Ptiloneuridae	*Loneura mombachensis*	WGS	2059170	2240	SRR5308274
Psocomorpha	Archipsocidae	*Archipsocus nomas*	WGS	2373123	2304	SRR5308260
Psocomorpha	Epipsocidae	*Neurostigma* sp.	WGS	2158530	2267	SRR5308277
Psocomorpha	Epipsocidae	*Bertkauia* sp.	RNA seq	1686357	1939	SRR2051473
Psocomorpha	Asiopsocidae	*Asiopsocus sonorensis*	WGS	2368782	2298	SRR5308261
Psocomorpha	Caeciliusidae	*Xanthocaecilius sommermanae*	WGS	2256093	2284	SRR5308288
Psocomorpha	Caeciliusidae	*Valenzuela badiostigma*	RNA seq	1980198	2179	SRR2051514
Psocomorpha	Lachesillidae	*Anomopsocus amabilis*	WGS	2308974	2294	SRR5308259
Psocomorpha	Lachesillidae	*Lachesilla contraforcepeta*	RNA seq	2164547	2279	SRR1821927
Psocomorpha	Lachesillidae	*Lachesilla abiesicola*	RNA seq	1658454	2106	SRR2051497
Psocomorpha	Mesopsocidae	*Idatenopsocus orientalis*	WGS	2362071	2302	SRR5308271
Psocomorpha	Mesopsocidae	*Mesopsocus unipunctatus*	RNA seq	1546809	1987	SRR2051502
Psocomorpha	Dasydemellidae	*Matsumuraiella radiopicta*	RNA seq	2026917	2212	SRR2051500
Psocomorpha	Psilopsocidae	*Psilopsocus* sp.	WGS	2111205	2239	SRR5308283
Psocomorpha	Stenopsocidae	*Graphopsocus cruciatus*	RNA seq	1820073	2095	SRR2051490
Psocomorpha	Amphipsocidae	*Amphipsocus japonicus*	RNA seq	2010662	2177	SRR2051466
Psocomorpha	Peripsocidae	*Peripsocus phaeopterus*	RNA seq	1665303	2038	SRR2051507
Psocomorpha	Ectopsocidae	*Ectopsocus briggsi*	RNA seq	1870215	2125	SRR645929
Psocomorpha	Psocidae	*Longivalvus nubilus*	RNA seq	1299474	1716	SRR2051498
Psocomorpha	Psocidae	*Ptycta johnsoni*	RNA seq	1602536	2006	SRR1821962
Psocomorpha	Psocidae	*Neoblaste papillosus*	RNA seq	1716579	2052	SRR2051505
Psocomorpha	Hemipsocidae	*Hemipsocus chloroticus*	RNA seq	1638022	2025	SRR2051492
Psocomorpha	Philotarsidae	*Aaroniella* sp.	RNA seq	1578200	1891	SRR2051465
Trogiomorpha	Prionoglaridae	*Speleketor irwini*	WGS	1898916	2200	SRR5308285
Trogiomorpha	Prionoglaridae	*Prionoglaris stygia*	WGS	2282502	2290	SRR5308282
Trogiomorpha	Prionoglaridae	*Neotrogla* sp.	WGS	1977267	2244	SRR5308278
Trogiomorpha	Prionoglaridae	*Neotrogla aurora*	RNA seq	1725501	2005	SRR5134732
Trogiomorpha	Psyllipsocidae	*Dorypteryx domestica*	WGS	2330406	2308	SRR5308267
Trogiomorpha	Psyllipsocidae	*Psyllipsocus ramburii*	RNA seq	2394803	2277	SRR5134716
Trogiomorpha	Psoquillidae	*Rhyopsocus* sp.	WGS	2012127	2213	SRR5308284
Trogiomorpha	Trogiidae	*Lepinotus patruelis*	RNA seq	2391270	2286	SRR5134710
Trogiomorpha	Trogiidae	*Cerobasis guestfalica*	RNA seq	1947314	2186	SRR2051476
Trogiomorpha	Lepidopsocidae	*Echmepteryx hageni*	RNA seq	1546566	1859	SRR1821982

Summary of Analyses Completed				
Data Type	Maximum Likelihood	Astral	Quartet Sampling	Likelihood Mapping
All Sites	Partitioned and Concatenated	Yes	Yes	Yes
Third Positions Removed	Concatenated Only	No	Yes	Yes
Second Positions Only	Concatenated Only	No	Yes	Yes
Degeneracy Recoded	Partitioned and Concatenated	Yes	Yes	Yes

Traditional Classficiation:			Newly Proposed Classification:			
Order	**Suborder**	**Infraorder**	**Order**	**Suborder**	**Infraorder**	**Parvorder**
Psocoptera:			Psocodea:			
	Trogiomorpha:			Trogiomorpha:		
		Prionoglaridetae			Prionoglaridetae	
		Psyllipsocetae			Psyllipsocetae	
		Atropetae			Atropetae	
	Psocomorpa:			Psocomorpa:		
		Archipsocetae			Archipsocetae	
		Philotarsetae			Philotarsetae	
		Epipsocetae			Epipsocetae	
		Psocetae			Psocetae	
		Caeciliusetae			Caeciliusetae	
		Homilopsocidea			Homilopsocidea	
	Troctomorpha:			Troctomorpha:		
		Amphientometae			Amphientometae	
		Nanopsocetae			Sphaeropsocetae	
Phthiraptera:					Pachytroctetae	
	Amblycera				Liposcelidetae	
	Anoplura				Phthiraptera:	
	Rhynchophthirina					Amblycera
	Ischnocera					Anoplura
						Rhynchophthirina
						Trichodectera
						Ischnocera