Insect phylogenomics: results, problems and the impact of matrix composition

Harald O. Letsch 1,*, Karen Meusemann 2, Benjamin Wipfler 3, Kai Schütte 4, Rolf Beutel 3 and Bernhard Misof 2

1Department für Tropenökologie und Biodiversität der Tiere, Universität Wien, Rennweg 14, 1030 Wien, Austria
2Zoologisches Forschungsmuseum Alexander Koenig, Zentrum für Molekulare Biodiversitätsforschung, Adenauerallee 160, 53113 Bonn, Germany
3Institut für Spezielle Zoologie und Evolutionsbiologie mit Phyletischem Museum, Friedrich-Schiller-Universität Jena, Erbertstr. 1, 07743 Jena, Germany
4Biozentrum Grindel and Zoologisches Museum, Martin-Luther-King Platz 3, 20146 Hamburg, Germany

In this study, we investigated the relationships among insect orders with a main focus on Polyneoptera (lower Neoptera: roaches, mantids, earwigs, grasshoppers, etc.), and Paraneoptera (thrips, lice, bugs in the wide sense). The relationships between and within these groups of insects are difficult to resolve because only few informative molecular and morphological characters are available. Here, we provide the first phylogenomic expressed sequence tags data (‘EST’: short sub-sequences from a c(opy) DNA sequence encoding for proteins) for stick insects (Phasmatodea) and webspinners (Embioptera) to complete published EST data. As recent EST datasets are characterized by a heterogeneous distribution of available genes across taxa, we use different rationales to optimize the data matrix composition. Our results suggest a monophyletic origin of Polyneoptera and Eumetabola (Paraneoptera + Holometabola). However, we identified artefacts of tree reconstruction (human louse Pediculus humanus assigned to Odonata (damselﬂies and dragonﬂies) or Holometabola (insects with a complete metamorphosis); mayfly genus Baetis nested within Neoptera), which were most probably rooted in a data matrix composition bias due to the inclusion of sequence data of entire proteomes. Until entire proteomes are available for each species in phylogenomic analyses, this potential pitfall should be carefully considered.

Keywords: insect phylogeny; Polyneoptera; Embioptera; Phasmatodea; phylogenomics; expressed sequence tags

1. INTRODUCTION

Even though molecular studies on the insect phylogeny are legion [1–9], a consensus on the relationships between the insect orders is not yet in sight. The difficulties in reaching a robust phylogenetic tree are most probably due to an ‘ancient rapid radiation’ phenomenon [10–12]: most of the modern lower neopteran orders appeared in a geologically relatively short time span in the Early Mesozoic [13], followed by a long period of intra-ordinal diversiﬁcation. Thus, only traces of phylogenetic signal are left in the data. Molecular analyses based on ribosomal RNA (rRNA) genes, complete mitochondrial (mt) genomes, and few housekeeping protein-coding genes, are not only plagued by this loss of phylogenetic signal, but also hampered by lineage-speciﬁc substitution rates and base compositional biases potentially misleading tree inference. For example, genes of flies (Diptera) show highly accelerated substitution rates, whereas genes of roaches (Blattodea) evolve comparatively slowly [3,10,13].

Recently, investigators have tried to overcome the limitations of rRNA and mitochondrial genome sequences by using ‘expressed sequence tags’ (ESTs) [6,14] as molecular markers, which enormously increase the available data for tree inference. ESTs are short sequences of c(opy) DNA (100–800 bp long), gained by reverse transcription from messenger RNA (mRNA) that encompass transcripts of protein-coding regions (exons) being expressed in an organism at a particular time [15]. With next generation sequencing (NSG) techniques, up to 1 million ESTs can randomly be sequenced. However, even these extensive data supported partly implausible relationships in current phylogenetic studies. For example, several groups of bugs (Hemiptera), were placed as the sister group to all other Neoptera (insects capable of folding their wings over their abdomen), or the human louse Pediculus humanus as a close relative of crickets and grasshoppers (Orthoptera) [14]. Both results clearly contradict the widely accepted monophyletic origin of thrips, lice and bugs in the wide sense, (Paraneoptera). One problem with these EST data is certainly the lack of several phylogenetically problematic groups, such as stick insects (Phasmatodea), earwigs (Dermaptera) and web spinners (Embioptera). Sufﬁcient EST data are currently only available for insects of economic or medical interest, with a considerable bias on bugs (Hemiptera) and holometabolous insects, for example: sawﬂies, woodwasps, bees and their relatives (Hymenoptera), beetles (Coleoptera), butterﬂies and moths (Lepidoptera), and ﬂies (Diptera). In contrast, the so-called ‘lower Neoptera’ (Polyneoptera) are only represented by roaches and termites (Blattodea) and grasshoppers and their relatives (Orthoptera).

* Author for correspondence: (harald.letsch@univie.ac.at).

More comprehensive taxon sampling is obviously necessary. However, the generation of EST data produces character matrices with lots of missing data because of variable gene expression patterns in tissues of different specimens, the random sequencing of ESTs and the varying success of the orthology prediction of genes. In most extreme cases, super-matrices are composed of species represented by the entire gene set derived from a fully sequenced genome or represented only by sparse EST (transcriptome) data. Additionally, it can be expected that genes contribute differently to the overall signal (information content) of a super-matrix. In such a case, the influence of missing data on tree reconstructions is unclear [16–20].

In this study, we aimed to improve the signal-noise ratio in a super-matrix derived from insect genomic and EST data by optimizing data coverage and overall signal of the data matrix. Our goal was to investigate the potential and pitfalls of the phylogenomic data currently available for neopteran insects. Including new EST data for two polyneopteran orders, stick insects (Phasmatodea) and webspinners (Embioptera), we present the first EST-based phylogenomic investigation of Neoptera.

2. MATERIAL AND METHODS

(a) cDNA library construction from Embioptera sp. (Embioptera) and Phyllium philippinicum (Phasmatodea)

ESTs were sequenced with the 454 pyro-sequencing technique using GS-FLX Titanium systems. The whole process was outsourced to different companies (please refer to electronic supplementary material A for detailed protocols for specimen preservation, RNA extraction, cDNA library construction and subsequent processing of raw sequence data). For both species, EST sequences were assembled with the software Newbler (Roche, Indianapolis, IN, USA).

(b) Expressed sequence tag processing, sequence alignment and tree reconstruction

Additional EST sequences were extracted from GenBank (NCBI). We downloaded assembled EST contigs (February 2011) for all available poly- and paraneopteran orders and several representatives of holometabolous insects from http://www.deep-phylogeny.org. If a taxon was represented by EST data and a complete proteome (amino acid level), both sets were downloaded. For these data, sequence processing (cleaning and trimming sequences, softmasking, sequence assembly) had been conducted earlier by the Center for Integrative Bioinformatics Vienna (CIBIV) using a pipeline which had been developed in the priority program SPP 1174 ‘Deep Metazoan Phylogeny’ [6,14]. Prior to orthology prediction, we replaced all sequence ambiguities at the nucleotide level by ‘N’. The core orthologue set insecta_hmmer3–2 set (http://www.deep-phylogeny.org/hmmer3/download/datasets/hmmer3/) was used to assign genes to orthologous groups with the HaMStR approach v. 3.4 [21]. This set was built upon the inparanoid transitive closure (TC) approach [21] using proteomes from six ‘core’ or ‘primer’ taxa: Apis mellifera, Bombyx mori, Tribolium castaneum, Daphnia pulex, Isodes scapularis and Cappella sp. The core orthologue set includes 1579 sets of orthologous genes. We ran HaMStR with the option-representative and an e-value cut-off (-eval_limit = 1e-05) for the hidden Markov model (HMM) search. For the subsequent reciprocal blast of the candidate EST contigs, we chose the following species (option -relaxed): A. mellifera, B. mori, T. castaneum, D. pulex or I. scapularis. While summarizing the HaMStR output for alignment processing, we omitted all primer taxa.

Sequences of each orthologous gene were separately aligned with MAFFT v. 6.5 using the L-INS-i algorithm [22,23]. All gene alignments were individually checked for putative randomly similar sections using ALISCORE [24,25] in amino acid mode with the default sliding window size, the maximum number of pairwise comparisons and the option -e. For each alignment, only sequences comprising more than one half of the sequence information were included in the ALISCORE analyses. After discarding all randomly similar positions with ALICUT [26], all genes alignments were concatenated to a super-matrix using FASconCAT [27]. Unfortunately, we could not test for a potential bias in amino acid composition. Recent methods, which compare the observed amino acid frequency for each taxon with the expected average amino acid frequency, cannot be applied here, as the unevenly distributed genes in the present data matrices does not allow a reasonable estimation of the average amino acid frequency. Consequently, we cannot exclude a possible impact of convergence in amino acid composition on our tree reconstructions.

Maximum-likelihood tree reconstructions were conducted using RAxML v. 7.2.8 [28,29]. We applied rapid bootstrap and subsequent tree search in one step (-f a, 1000 bootstrap replicates). All ML analyses were calculated with the PROTCAT model, which optimizes individual per-site substitution rates and classifies those individual rates under 25 specified rate categories [30]. In a subsequent step to evaluate the final tree topology, the PROTGAMMA model with four discrete GAMMA rate categories (shape parameter estimated from the data) was used to estimate substitution rate heterogeneity across sites. Prior to the tree reconstruction, we used MODELGENERATOR v. 0.85 [31] to select the best fitting amino acid substitution matrix. The implemented Akaike information criterion (AIC), the Bayesian information criterion (BIC) and the hierarchical likelihood-ratio tests (hLRTs) suggested the use of the LG matrix [32], which incorporates the variability of evolutionary rates across sites in the matrix estimation + GAMMA. However, MODELGENERATOR v. 0.85 can currently handle datasets of only limited size. Thus, we applied model tests for the reduced datasets, and the unreduced dataset comprising EST data only. For the unreduced data matrices, including data from species for which a proteome is available, we applied an identical model setup. All analyses were computed on HPC Linux clusters, eight nodes with 12 cores each, at the Regionales Rechenzentrum Köln (RRZK) using Cologne High-Efficient Operating Platform for Science (CHEOPS). Trees were edited with TreeView v. 1.6.6. [33]. Figure 1 depicts a graphical workflow of the conducted analyses.

We refrained from inferring a phylogenetic tree and estimating branch lengths within a Bayesian framework as statistical problems are associated with this approach (e.g. unclear impact of priors on results, failure of Markov chains to reach stationarity, risk of becoming trapped in regions of parameter space [17,34–36]).

(c) Selection of optimal expressed sequence tag data matrices

To increase matrix saturation (i.e. the number of present genes in relation to the total number of entries, see [37]),
we first excluded all poly- and paraneopteran species represented by less than 100 orthologous genes (prior to the alignment process). Based on the remaining taxa, we compiled a first dataset including only EST data. In a second dataset, EST data were substituted by entire sequences data derived from proteomes, if available. Both resulting datasets comprise 40 taxa each, with one mayfly and one damselfly as outgroups (for a complete taxon sampling, see electronic supplementary material B, table S1).

The information content of the data matrices (i.e. the sum of tree-likeness values in relation to the total number of entries) was further optimized with MARE (MAtrix REduction) [14,37]. MARE employs a weighted geometry quartet mapping algorithm [38] extended to amino acid level to calculate the ‘tree-likeness’ (potential information content) [37] of each gene (partition). MARE successively discards genes and taxa (species) in a super-matrix based on their contribution to the total information content and matrix saturation of the data matrix (please refer to the manual of MARE [37] for a detailed description of the algorithm). With the two original datasets as a starting point, we selected two optimal data subsets with a reduced number of genes using MARE (table 1). The taxon weighting option in MARE was set to a value of 2.

The taxon weighting option in MARE was set to a value of 2.

Figure 1. Work flow representing the setups for the data matrix compilation and the tree reconstructions.

3. RESULTS AND DISCUSSION
(a) Matrix characteristics
The first unreduced dataset (including only EST data), represents 1561 orthologous genes and spans around 450 000 aligned amino acid positions. The second unreduced dataset (including sequence data derived from entire proteomes) represents all 1579 orthologous genes of the core orthologue set and spans about 790 000 aligned amino acid positions. Species for which the proteome is available ('proteome species') show a complete coverage of the orthologous gene set. Both unreduced matrices show a saturation of approximately 40 per cent and an information content of 0.12. The size of both reduced data matrix matrices is about 15 per cent of the original, unreduced data matrices, whereas the matrix saturation of both selected data subsets increases to approximately 70 per cent and the information content to approximately 0.37.

(b) Implications for the use of expressed sequence tag data reconstructing the insect phylogeny
In general, the super-matrices show congruent and robust phylogenetic trees (figures 2 and 3). However, our analyses also revealed some problems: the most interesting one is the extreme discrepancy between the incongruent placement of the human louse Pediculus humanus and the mayfly Baetis sp. on the one side, and the high topological congruence between all the remaining taxa in accordance with traditional phylogenetic concepts: (i) In the analysis of the unreduced data matrix based exclusively on EST data (figure 2a), the human louse groups with the damselfly Ischnura elegans and the mayfly Baetis sp. as sistergroup of the remaining insects. In the tree inferred from the reduced data matrix (figure 2b), the human louse is placed as sistergroup of bugs in a wide sense (Hemiptera) in a combined clade Paraneoptera, corroborating traditional systematic views. (ii) Both approaches including proteome data (figure 3a,b) yielded a very implausible clade human louse + Holometabola: all holometabolous insects have a very characteristic life cycle with a complete metamorphosis which is never the case in lice. The analysis of the reduced data matrix also yielded the mayfly as sistergroup to roaches and termites (figure 3b). This is similar to results obtained with entire mitochondrial genomes [39] and EST-based studies [14], thus challenging the monophyly of Neoptera, which is very strongly supported by morphological data, such as the ability to fold back the wings with specific modifications of the wing base, the presence of an arculus as attachment pad and a specifically modified ovipositor) [40–43]. Highly unusual placements of lice, either as a sistergroup to crickets and grasshoppers + mantids, roaches and termites [14], within holometabolous insects [7], or closely related to a clade, including the remaining Paraneoptera + Holometabola [1], have previously been retrieved in molecular studies. In contrast, the monophyly of Paraneoptera is well supported by morphological characters, for example, by modifications of their mouthparts (lacinia stylet-like, detached from stipes) [13,40] and other body regions (cerci absent, only one complex of abdominal ganglia, only four Malpighian tubules, only three tarsal segments or less [41,42]).

A possible explanation for the artificial grouping human louse + Holometabola might be the specific data matrix composition in both datasets, including data received from entire proteomes. Proteome data are available for the human louse and four holometabolan species, but only for one additional species in the polyneopteran and hemipteran clade. Thus, these 'proteome species' (see above) show an almost complete coverage of the orthologous gene set and comprise consequently nearly all genes in each dataset, resulting in a data overlap of almost 100 per cent. In contrast, the number of
common genes between these taxa and those for which no entire proteome and only transcriptome data (ESTs) are available is much smaller (see electronic supplement material C, figures S1 and S2). In this context, the datasets, including ‘proteome species’ bias the data composition of the super-matrix, with lots of missing data for those taxa with vastly sparse EST data. We hypothesize that this uneven data composition biases tree reconstruction.

Our results corroborate previous studies on missing data and gappy data matrices, which did not suggest a negative impact on tree reconstruction [19]: although the gene presence in all datasets is unevenly distributed (approx. between 150 and 1580 genes depending on the respective taxon and dataset), most inferred relationships are consistent in the resulting topologies and seem congruent with previous plausible hypotheses of the insect phylogeny [40, 41, 42]; see also [43]. The obviously wrong placement of the human louse must rather be seen as a problem of high data coverage, in terms of a biased gene (and thus, potential information) overlap between unrelated taxa. This interpretation is corroborated by the results of our analyses exclusively based on EST data, in which the problematic placement of the human louse and the Holometabola seems to be neutralized. In these data, the human louse either appears as sistergroup of bugs in a wide sense (Hemiptera), which is consistent with monophyletic Paraneoptera, or as sistergroup of the damselfly. This indicates that the data, in principle, contain phylogenetic signal for a correct placement of the human louse. As all tree inferences use identical estimates of model parameters, the discrepancies are likely rooted in matrix composition. The artificial exclusion of the human louse from Neoptera only appears in the unreduced data matrix. This is likely due to a problem of a small gene overlap between related taxa (human louse + bugs in a wide sense). In this context, the matrix reduction approach increased this overlap. However, the low bootstrap support for Paraneoptera still indicates a considerable impact of a misleading data matrix composition.

(c) Implications for the phylogeny of Neoptera

(i) The status of Polynota

The monophyletic origin of the lower neopteran orders: roaches, mantids, earwigs, grasshoppers, etc. (Polynota), is still disputed, and this issue is one of the last major problems in systematic entomology. From a morphological point of view, the status of Polynota depends on the studied character system. For example, monophyletic Polynota are supported by wing venation patterns [44], while head morphology does favour a paraphyletic situation [45, 46]. Analyses by Beutel & Gorb [41, 42, 47] covering all major polynotean lineages and using a broad spectrum of morphological characters also provide incongruent results. In molecular analyses, nuclear protein-coding genes and complete mitochondrial genomes support the monophyly [3, 7, 48, 49], while nuclear 18S and 28S rRNA and/or Histone 3, as well as previous EST analyses, do not [4, 8, 14]. Our present results support monophyletic Polynota, except for one approach showing an unlikely placement of the mayfly within Polynota (but also paraphyletic crickets and grasshoppers). We interpret this result as a problem of biased gene overlap. In contrast, the sensitivity of nuclear rRNA genes to even slight changes in the alignment parameter settings and tree reconstruction methods [5, 8, 50, 51] must be interpreted as a result of non-stationary processes in these genes [5, 10]. Previous studies also show that trees based on protein-coding genes are more robust to parameter settings [6]. Nevertheless, whole mitochondrial genomes have been shown inappropriate to target the relationships among insect orders, mostly due to base compositional biases and unequal rates of nucleotide substitution across groups [52]. Consequently, the use of EST data and nuclear protein-coding genes generally outperform approaches based on nuclear rRNA data and mitochondrial genomes in the investigations of inter-ordinal relationships of Polynota. Unfortunately, at present only rRNA and Histone 3 are available for all polynotean orders. As five polynotean groups (ice crawlers (Grylloblattodea), heel walkers (Mantophasmatodea), stone flies (Plecoptera), earwigs (Dermaptera) and mantids (Mantodea)), could not be included in our analyses, reliable statements about the monophyletic origin of Polynota are presently not possible. Nevertheless, our results tentatively support the monophyly of this lineage, as our sampling covers taxa widely separated in recently proposed lower neopteran phylogenies [1, 42, 45, 46, 48, 53].

(ii) The status of stick insects (Phasmatodea) and webspinners (Embioptera)

All results of the present analyses show a strongly supported clade stick insects + webspinners (i.e. Eukinolabia [4]). This relationship has first been proposed by Raehle [54] and is corroborated by recent studies based on nuclear rRNA, protein-coding genes [4, 7, 55, 56] and morphological features [45, 46, 57]. Our results do not support the traditional clade stick insects + crickets, katydids, grasshoppers and related groups (i.e. Orthopterida) [4, 58–60], and the morphological arguments for this relationship have recently been refuted [45, 57]. Although ongoing molecular and morphological studies corroborate the monophyly of

---

Table 1. Characteristics of the used super-matrices. sat., saturation; inf., information.

<table>
<thead>
<tr>
<th>No. taxa</th>
<th>No. genes</th>
<th>Matrix length (in ESTs)</th>
<th>Inf. content</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unreduced</td>
<td>40</td>
<td>1579</td>
<td>787,330</td>
<td>0.444</td>
</tr>
<tr>
<td>Reduced</td>
<td>40</td>
<td>240</td>
<td>53,998</td>
<td>0.693</td>
</tr>
<tr>
<td>Exclusive proteome data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. taxa</td>
<td>No. genes</td>
<td>Matrix length (in ESTs)</td>
<td>Inf. content</td>
<td>Ratio</td>
</tr>
<tr>
<td>----------</td>
<td>-----------</td>
<td>------------------------</td>
<td>--------------</td>
<td>-------</td>
</tr>
<tr>
<td>Unreduced</td>
<td>40</td>
<td>1561</td>
<td>445,742</td>
<td>0.393</td>
</tr>
<tr>
<td>Reduced</td>
<td>40</td>
<td>224</td>
<td>46,197</td>
<td>0.687</td>
</tr>
</tbody>
</table>

---

Eukinolabia, further studies should be carried out with an increased taxon sampling of stick insects and crickets, katydids, grasshoppers, etc., including the presumably ancestral phasmatodean taxa *Timema* and *Agathemera*, which are crucial for understanding the relationships and evolution of this order.

(iii) The relationships within Paraneoptera and Holometabola

Our datasets strongly support a clade comprising cicadas, cicadas, leaf hoppers, tree hoppers and spittlebugs (Cicadomorpha), and true bugs (Heteroptera). Previous molecular studies based on 18S rRNA sequences suggested a clade comprising plant hoppers (Fulgoromorpha) + true
bugs [61–67], whereas a recent multi-gene analysis supports monophyletic but only moderately supported Auchenorrhyncha (plant hoppers + cicadas, leafhoppers, etc.) [68]. In contrast to our study, which was based on a comprehensive dataset, molecular studies corroborating the fulgoromorph-heteropteran clade or Auchenorrhyncha were based on a much smaller gene sample. We thus consider our results here as comparatively robust. Nevertheless, our study comprises only representatives of one family of plant hoppers and leaf hoppers, respectively. The status of Auchenorryhncha and its relationships to true bugs is therefore still unclear.

The relationships within Holometabola are mostly congruent among different analyses. Wasps and allies (Hymenoptera) appear as the first split within Holometabola, which is congruent to comprehensive morphological

---

Figure 3. RAxML trees based on data matrices including taxa, for which entire proteomes are available: (a) unreduced dataset, (b) reduced dataset; PROTCAT model (final ML tree search under GAMMA) + LG matrix. Node support represented by bootstrap values (based on 1000 bootstrap replicates). Filled grey boxes indicate monophyly, blank boxes indicate paraphyly.
and molecular data [14,69,70]. The cat flea Ctenocephalides felis generally turned out as sistergroup of flies, but as the sistergroup of a clade flies + moths and butterflies in the topology based on the reduced dataset, excluding proteome data. In the light of previous molecular and morphological studies, the first result seems much more plausible as it is congruent with the generally accepted Antiophora/Mecoptera hypothesis, with moths and butterflies + caddisflies (Trichoptera) as sistergroup of a clade of flies (Diptera) + fleas (Siphonaptera) and scorpionflies and their relatives (Mecoptera) [71,72].

4. CONCLUSIONS

Our approach provides the first phylogenomic support for the monophyletic origin of several lineages of Polynoptera, and for a sistergroup relationship between webspinners and stick insects (i.e. Eukinolabia). Furthermore, our analyses suggest a sistergroup relationship between true bugs and cicadas, leaf hoppers, and their allies. Since our analyses did not include several polyneopteran orders and thrips, as well as only one family of plant hoppers and leaf hoppers, respectively, they must be seen as preliminary and as a first step towards a phylogenomic approach to the reconstruction of interordinal relationships. Placing the human louse revealed inherent problems with EST data in phylogenetics: our results clearly indicate an impact of matrix composition. The inclusion of taxa where entire proteomes had been available and thus covered the complete orthologue gene set led to a biased gene overlap, which resulted in an artificial placement of the human louse as closest relative of Holometabola. Until a maximum coverage of used orthologous gene sets or entire proteomes is available for each terminal in phylogenomic analyses, this potential pitfall should be carefully considered.

We would like to express our appreciation for many discussions and important comments coming from all members of the Molecular Systematics Unit at the ZFMK. Special thanks to Janus Borner, University of Hamburg, for the help with installing the HaMStR software and providing several ruby scripts that data during orthology assignment could be properly handled. We also thank Malte Petersen (zmb at the ZFMK), who wrote several Perl scripts to properly check EST data for submission. We further acknowledge Björn M. von Reumont (ZFMK) for advice regarding sequence submission and Christian Schulze, University of Vienna, for help with statistics. We acknowledge Viktor Achter and Volker Winkelmann for help with ML analyses on the Cologne High Efficiency Operating Platform for Sciences (CHEOPS, HPC cluster at the RRZK, University of Cologne, Cologne, Germany, http://rrzk.uni-koeln.de/cheops.html). We specially acknowledge Konrad Fiedler, University of Vienna, for financial support. We also want to thank Chris Simon and one anonymous reviewer, who provided insightful comments that helped to improve an earlier version of this article. B.M. and K.M. were supported by the DFG grant MI 649/6.

REFERENCES

22 Katoh, K., Kuma, K. & Toh, H. 2005 MAFFT v. 5: improvement in accuracy of multiple sequence


64. Sweet, M. H. 1996 *Comparative external morphology of the pregenital abdomen of the Hemiptera.* Lanham, MD: Entomological Society of America.


66. Sorensen, J. T., Campbell, B. C., Gill, R. J. & Setephen-Campbell, J. D. 1995 Non-monophyly of Auchenorrhyncha (homoptera), based upon 18s rDNA phylogeny—eco-evolutionary and Cladistic implications within pre-heteropterodea Hemiptera (sl) and a proposal for new monophyletic suborders. *Pan-pacific Entomol.* 71, 310.


