A preliminary phylogeny of the Pentatomomorpha (Hemiptera: Heteroptera) based on nuclear 18S rDNA and mitochondrial DNA sequences

Hong-Mei Li a,b, Ri-Qiang Deng a, Jin-Wen Wang a, Zhen-Yao Chen a, Feng-Long Jia a, Xun-Zhang Wang a,*

a State Key Laboratory for Biocontrol, School of Life Sciences, Sun Yat-sen University, Guangzhou, People’s Republic of China
b Zhongkai University of Agriculture and Technology, Guangzhou, People’s Republic of China

Received 15 October 2004; revised 30 June 2005; accepted 1 July 2005
Available online 30 August 2005

Abstract

Pentatomomorpha is the second suborder in size only to Cimicomorpha in Heteroptera. However, the phylogenetic relationships among members of the suborder are not well established. Sequences from partial nuclear ribosomal 18S gene and mitochondrial COX1 gene were analyzed separately and in combination to generate a preliminary molecular phylogeny of Pentatomomorpha based on 40 species representing 17 putative families. Analyses of the combined sequence data provided a better-resolved and more robust hypothesis of Pentatomomorpha phylogeny than did separate analyses of the individual genes. The phylogenies were mostly congruent with morphological studies. Results strongly supported the monophyly of the infraorder Pentatomomorpha, and the placement of Aradoidea as sister to Trichophora. The monophyletic Trichophora was grouped into two major lineages, one being the superfamily Pentatomoidea, and the other comprising Lygaeoidea, Coreoidea, and Pyrrhocoroidea. The analysis of the ML and ME trees of combined dataset supported the monophyletic Pentatomoidea. In all analysis the Pyrrhocoroidea was polyphyletic; the monophyletic Lygaeoidea was supported only in the analysis of ME tree, and Coreoidea was polyphyletic except in the MP tree of combined dataset. Our phylogenetic results suggested that the COX1 segment alone might not be an optimal molecular marker for the phylogeny of Pentatomomorpha.

© 2005 Elsevier Inc. All rights reserved.

Keywords: Molecular phylogeny; Pentatomomorpha; 18S rDNA; COX1

1. Introduction

The Pentatomomorpha is the second large suborder and one of the most important groups in Heteroptera, numbering from 12,500 to 15,000 species worldwide (Dolling, 1991; Schaefer, 1990; Schuh and Slater, 1995). All of the taxa in this infraorder are terranean, mostly plant-feeding. They are economically important in agriculture, as some of them are pests. However, the phylogenetic relationships among members of the group are not well established.

Leston et al. (1954) established the infraorder Pentatomomorpha based on pretarsal structure, male and female genitalia, wing venation, egg morphology, and salivary gland structures. Approximately two-thirds of the families previously placed in the terrestrial bug group Geocorisaen make up the Pentatomomorpha, a grouping that coincides with Tullgren’s (1918) trichophoran assemblage, less the Aradidae and Termitaphididae.
(Henry, 1997). Within the Heteroptera, the Pentatomomorpha are the most distal position among the seven infraorders, with the Cimicomorpha hypothesized as their sister group (Schuh, 1979; Wheeler et al., 1993).

Within the Pentatomomorpha, the relationship among the superfamilies is uncertain. There have been recognized 4 (Schaef er, 1964), 5 (Stys, 1961, 1967; Schaefer, 1990; Schuh and Slater, 1995), 6 (Carver et al., 1991; Henry, 1997), or 7 (Schuh, 1986; Henry and Froeschner, 1988; Li and Zheng, 1994) superfamilies. Only the Aradoidea and Pentatomoinidea are consistently recognized as monophyletic within these schemes. No consensus exists for the relative position of the Piesmatidae and Idiostoloidea. The other superfamilies Coreoidea, Lygaeoidea, and Pyrrhocoroidea have been defined with some difference. The position of the family Piesmatidae has been controversial since Drake and Davis (1958) placed it in Pentatomomorpha. Stys (1961, 1967) grouped the Coreoidea, Lygaeoidea, and Pyrrhocoroidea into 1 superfamily, the Coreoidea; and recognized 5 superfamilies Aradoidea, Idiostoloidea, Pentatomoidea, Coreoidea, and Piesmatoidea. Schuh (1986), Henry and Froeschner (1988) accepted 7 superfamilies as they have been presented in the literature (Aradoidea, Pentatomoinidea, Coreoidea, Pyrrhocoroidea, Idiostoloidea, Lygaeoidea, and Piesmatoidea). Li and Zheng (1994) favored the 7 groups, and emphasized that Piesmatidae is actually a specialized group that should be separated as a superfamily, Piesmatoidea. Whereas, Southwood (1956) separated the Largidae and Pyrrhocorididae from the Lygaeoidea as a superfamily Pyrrhocoroidea, let the Piesmatidae still in the Lygaeoidea. Also, Schaefer (1993); Schuh and Slater (1995) favored 5 groups, including the Aradoidea, Pentatomoinidea, Coreoidea, Pyrrhocoroidea, but put the Idiostoloidea and Piesmatoidea into the Lygaeoidea. More recently, Henry (1997) brought together all the greatly scattered morphological character information found in the literature and to analyze it cladistically, recognized 6 superfamilies (Aradoidea, Pentatomoinidea, Coreoidea, Pyrrhocoroidea, Idiostoloidea, Lygaeoidea), put the Piesmatoidea into the Lygaeoidea. All these arguments show that there is little agreement in the literature as to the phylogenetic relationships among the superfamilies in Pentatomomorpha. It is clear that morphology alone may be unable to clarify these controversies, combining molecular data will provide additional needed information.

Thus far, attempts to incorporate molecular data into phylogenetic analyses of the Pentatomomorpha have not been made. The nuclear small subunit ribosomal RNA (18S rDNA) has been very useful to infer phylogenetic affiliations of ancient evolutionary lineages (Hillis and Dixon, 1991). It provides readily obtainable nucleotide sequences and ease of PCR primer design (Kjer, 1995; Woese, 1987). Recent phylogenetic studies using partial or complete sequences of the 18S rDNA have successfully examined relationships among the major lineages of Hemiptera (Campbell et al., 1995; Sorensen et al., 1995; Wheeler et al., 1993). Mitochondrial DNA (=mtDNA) is widely used for comparisons among closely related species (Martin and Palumbi, 1993). Meanwhile, mitochondrial sequences are also used for analyzing relationship among distantly related organisms (Kocher et al., 1989). The mitochondrial cytochrome c oxidase subunit I gene (=COXI) is conserved among Metazoa (Jacobs et al., 1988) and have been used to examine phylogenetic relationships within the Arhynchobdellida order (Borda and Siddall, 2004) and in subclass Pteriomorphia (Matsumoto, 2003), although it is often used among families and lower lineages (Damgaard and Sperling, 2001; Pollock et al., 1998).

In this work, we choose 18S rDNA and COXI as proper molecular markers (1) to study the molecular phylogenetic relationships among the higher taxa of Pentatomomorpha, (2) to provide molecular data to assist with interpreting Pentatomomorpha evolution and reassessing higher classification, and (3) to select and assess a proper molecular marker to reconstruct the phylogeny of Pentatomomorpha.

2. Materials and methods

2.1. Taxa

Forty taxa representing 5 superfamilies and 17 putative families were analyzed to study the relationship of high rank level of Pentatomomorpha. The specimens used, their current taxonomy are shown in Table 1.

We collected 22 species in Guangdong Province during the year of 2003 and sequenced their 18S rDNA and COXI. The nucleotide sequence data have been deposited in GenBank and their Accession numbers were listed in Table 1. Although great efforts were made to include multiple representatives of larger families and at least one representative of all families, sampling was limited by the availability of specimens.

Other sequences were collected directly from GenBank including two other infraorder taxa (Enesaya brevipennis and Campyloneura virgula) that were used as outgroups. The outgroup taxa belong to Cimicomorpha and the infraorder was regarded as sister group of Pentatomomorpha (Wheeler et al., 1993). Accession numbers of these sequences were also listed in Table 1.

2.2. Laboratory procedure

Total genomic DNA was extracted from frozen, ethanol-preserved (95% ethanol) and a few pinned-dried specimens, usually obtained from thoracic muscles. Heads, abdomens and legs were stored in 70% alcohol as voucher specimens and deposited in the State Key Lab...
Table 1

<table>
<thead>
<tr>
<th>Current family Classification</th>
<th>Taxa</th>
<th>Accession number 18S rDNA</th>
<th>COX1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aradoidea</td>
<td>Mezira granulata</td>
<td>AY252221*</td>
<td>AY252962*</td>
</tr>
<tr>
<td>Pentatomomidea</td>
<td>Pocilocoris latus (Dallas)</td>
<td>AY627311</td>
<td>AY627330</td>
</tr>
<tr>
<td>Scutellerida</td>
<td>Cantao ocellatus (Thunberg)</td>
<td>AY627316</td>
<td>AY627335</td>
</tr>
<tr>
<td>Tesseratomidae</td>
<td>Tesseractoma papillosa (Drury)</td>
<td>AY627312</td>
<td>AY627331</td>
</tr>
<tr>
<td>Plataspidida</td>
<td>Megacopta cribraria (Fabricius)</td>
<td>AY627313</td>
<td>AY627332</td>
</tr>
<tr>
<td>Pentatomidae</td>
<td>Aponisia montana (Distant)</td>
<td>AY627314</td>
<td>AY627333</td>
</tr>
<tr>
<td>Acanthosomatidae</td>
<td>Staurocruta compuncta</td>
<td>AY252269*</td>
<td>AY253001*</td>
</tr>
<tr>
<td>Lygaeoidea</td>
<td>Neobium armiger (Fabricius)</td>
<td>AY627321</td>
<td>AY627339</td>
</tr>
<tr>
<td>Lygaeida</td>
<td>Spilostethus hosipes (Fabricius)</td>
<td>AY627319</td>
<td>AY627338</td>
</tr>
<tr>
<td>Rhyparochromidae</td>
<td>Pseudopachyactylus guttatus (Dallas)</td>
<td>AY627327</td>
<td>AY627346</td>
</tr>
<tr>
<td>Berytida</td>
<td>Neolepidotes muticus</td>
<td>AY252414*</td>
<td>AY253130*</td>
</tr>
<tr>
<td>Piesmatidae</td>
<td>Mecaela sp. WC-2003a</td>
<td>AY252164*</td>
<td>AY253096*</td>
</tr>
<tr>
<td>Pyrrhocoroidea</td>
<td>Largus sp. WC-2003</td>
<td>AY252227*</td>
<td>AY252967*</td>
</tr>
<tr>
<td>Reduviidae</td>
<td>Enessa brevispennis</td>
<td>AY252321*</td>
<td>AY253047*</td>
</tr>
<tr>
<td>Miridae</td>
<td>Campyloneura virgula</td>
<td>AY252317*</td>
<td>AY253045*</td>
</tr>
</tbody>
</table>


a Direct submission to NCBI GenBank Database by Wheeler and Schub.

b The 18S rDNA sequence of Serinthea sp. WC-2003a has been combined with the COX1 sequence of Serinthea sp. WC-2003g.

c The 18S rDNA sequence of Mecaela sp. WC-2003a has been combined with the COX1 sequence of Mecaela sp. WC-2003g.

d The 18S rDNA sequence of Mecaela sp. WC-2003b has been combined with the COX1 sequence of Mecaela sp. WC-2003f.

e The 18S rDNA sequence of Dysdercus sp. WC-2003a has been combined with the COX1 sequence of Dysdercus sp. WC-2003g.

For Biocontrol, Sun Yat-sen University. The method of DNA extraction was referred to Wen and He (2003). The protocol was as follows: each specimen was softly ground in 100μl of buffer (10mM Tris, 1mM EDTA, and 0.1M NaCl, pH 8.0). After adding Proteinase K (200μg/ml), the homogenate was incubated at 56°C for...
at 1–2 h and 95°C for 45 s and centrifuged for 5 min. The supernatant was collected as the template of PCR. Some specimens, especially a few pinned-dried specimens was extracted DNA with EaZY Nucleic Acid Isolation Kit (Omega Bio-tek).

PCR were conducted in 25 µl volume containing 2–6 µl of DNA, 0.5 U Taq polymerase (Takara Biotechnology (Dalian) CO., Ltd.), 1.5 µl 10 µM primers, 2.0 µl 25 mM dNTPs (Takara), and 2.5 µl 10× buffer. Amplification conditions were 1 cycle, 95°C (5 min); 35 cycles, 95°C (30 s), 50°C (40 s), 72°C (2 min); 1 cycle, 72°C (8 min). In most cases, target products from PCR, which contained one single, strong 1.8-kb band from 1% agarose gel electrophoresis with ethidium bromide staining under UV light were purified by Agarose Gel DNA Purification Kit (Takara) and sequenced directly. Some sequences were obtained by cloning the PCR products in pMD18-T Vector. DNA sequencing was performed in a PE/ABI 377 automated sequencer using the ABI PRISM BigDye Terminator Cycle Sequencing v 2.0 Ready Reaction Kit (PE Biosystems) with same primers used for PCR amplification.

The primers are listed in Table 2. Primer F and R of 18S rDNA were derived from Loxdale and Lushai (1998). The target segment of COX1 was delimited by the primers C1-J-2183 to TL2-N-3014 (reproduced from Simon et al. 1994). Because the segments were too long to be amplified and sequenced in one step, two internal primers were used for amplification with the end-primers, that is C1-N-2609 (Damgaard and Sperling, 2001) to work with C1-J-2183, and a new primer, C1-J-2530 to work with TL2-N-3014. The sequences were obtained from Loxdale and Lushai (1998). The target segment of COX1 was delimited by the primers C1-J-2183 to TL2-N-3014 (reproduced from Simon et al. 1994). Because the segments were too long to be amplified and sequenced in one step, two internal primers were used for amplification with the end-primers, that is C1-N-2609 (Damgaard and Sperling, 2001) to work with C1-J-2183, and a new primer, C1-J-2530 to work with TL2-N-3014.

2.3. Sequence alignment and phylogenetic analyses

Sequence reliability was checked by reading chromatograms in the CHROMAS V2.3 (Technelysium Pty Ltd., 2004). The sequences were aligned using default parameters of CLUSTAL X (Thompson et al., 1997) and the peloridiid 18S RNA model of Van de Peer et al. (2000) and the peloridiid 18S RNA model of (Ouvrard et al., 2000). Some highly variable regions, particularly the E10, 23, 41, which were not able to be aligned unambiguously across all taxa even considering the secondary structure were excluded from further analyses. The aligned data are available from author or they may be downloaded from the following web site: http://life.zsu.edu.cn/insect_tax/

Phylogenetic analyses were used maximum parsimony (MP), maximum likelihood (ML) and distance-based methods (minimum evolution, ME). Parsimony analyses of various datasets were performed using PAUP* 4.0b10 (Swofford, 2002) without considering gaps. As the number of taxa and the size of the data matrix often precluded more thorough searches, Heuristic search were performed with 100 random-taxon-addition replicates, TBR branch swapping, and no maxtrees restrictions. Clade stability was evaluated using two different parameters: bootstrap (Felsenstein, 1985) and branch support (Bremer, 1994). Bootstrap values were generated in PAUP* from 1000 replicates, each with ten random-addition heuristic searches. Branch support values (a.k.a. Decay indices) were estimated with program Autodecay V5.0 (Eriksson, 2001).

ML analyses were performed on both the individual and the combined data partitions. To identify the most appropriate substitution model for the ML analysis of the combined data partition, we used the computer program MrModeltest 2.0 (Nylander, 2004), which applies hierarchical likelihood ratio tests (LRTs) and the Akaike Information Criterion (AIC) to compare likelihood scores for nested and nonnested sets of substitution models for a given user tree.

Then we used the distance method to obtain a minimum-evolution (ME) tree using PAUP*.

Epochs of 100,000 iterations of simulated annealing were performed with TBR branch swapping and no maxtrees restrictions. Clade stability was evaluated using two different parameters: bootstrap (Felsenstein, 1985) and branch support (Bremer, 1994). Bootstrap values were generated in PAUP* from 1000 replicates, each with ten random-addition heuristic searches. Branch support values (a.k.a. Decay indices) were estimated with program Autodecay V5.0 (Eriksson, 2001).

Table 2
Primer for Amplification and Sequencing

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer name</th>
<th>Direction</th>
<th>Primer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>18S rDNA</td>
<td>F</td>
<td>Forward</td>
<td>5’-TCCCCTGTGTGTCTTTCAGAT-3’</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>Reverse</td>
<td>5’-TGTATGTTGCAAGGGTGTCGTA-3’</td>
</tr>
<tr>
<td></td>
<td>2F</td>
<td>Forward</td>
<td>5’-GGGAGAGTATTCCCAAATTACAGC-3’</td>
</tr>
<tr>
<td></td>
<td>2R</td>
<td>Reverse</td>
<td>5’-GTGTTATATCGCTAATCTCGT-3’</td>
</tr>
<tr>
<td></td>
<td>3F</td>
<td>Forward</td>
<td>5’-GGTTAGAATTCTTGGATCGTGC-3’</td>
</tr>
<tr>
<td></td>
<td>3R</td>
<td>Reverse</td>
<td>5’-ACATACTGCGGAAATGCTTTCG-3’</td>
</tr>
<tr>
<td></td>
<td>4R</td>
<td>Reverse</td>
<td>5’-GAATTAGGTGGCCGT-3’</td>
</tr>
<tr>
<td>COX1</td>
<td>C1-J-2183</td>
<td>Forward</td>
<td>5’-CACAATTTATTTTATTTTATTTG-3’</td>
</tr>
<tr>
<td></td>
<td>TL2-N-3014</td>
<td>Reverse</td>
<td>5’-CCTCAATCTCATTGCATATATAT-3’</td>
</tr>
<tr>
<td></td>
<td>C1-J-2530</td>
<td>Forward</td>
<td>5’-GGGATGTTATTCAGCCTCAGTCT-3’</td>
</tr>
<tr>
<td></td>
<td>C1-N-2609</td>
<td>Reverse</td>
<td>5’-AATTTGTGTTAGTTGACAGG-3’</td>
</tr>
</tbody>
</table>

Table 2
Primers for Amplification and Sequencing
To determine whether significant incongruence exists between the two molecular datasets, incongruence length difference (ILD) tests (Farris et al., 1995) were conducted by excluding all invariant sites and using the partition homogeneity test in PAUP*4.0b10 with 100 iterations.

3. Results

3.1. Analyses of data characteristics

The length of aligned 18S rDNA data comprised 888 positions. The numbers and percentages of variable sites and potentially parsimony informative characters found are presented in Table 3.

The aligned region of COXI used in analyses comprised 364bp (see Table 3), of which 156(42.9%) were phylogenetically informative including 31 in first (19.9%), 10 in second (6.4%), and 115 (73.7%) in third codon positions. The ratio of transitions to transversions was 1.039 for all sites (1.401 in first + second codon positions and 1.350 in third codon positions). The predominance of transitions has been documented widely for insect mtDNA (Simon et al., 1994; Damgaard and Sperling, 2001). The nucleotide sequences were A/T rich (see Table 3), that is a common characteristic for insect mitochondrial DNA sequences. Nucleotide frequencies for A ranged from 29.6% to 36.8%, for T from 25.7% to 40.5%, for C from 11.8% to 22.2%, and for G from 15.1% to 18.4%.

3.2. Saturation analyses

To test the level of substitution saturation in the COXI data, whole codons, the first, second, and third codon positions separately were analyzed using scatter plot graphics (presented in Fig. 1), comparing the uncorrected p-distances with the distances calculated by the Kimura two-parameter (K2P + Γ) model (Kimura, 1980). First and second positions showed no saturation. Whole codons presented a medium level of saturation. Third positions were fully saturated.

Table 3

<table>
<thead>
<tr>
<th>Data partition</th>
<th>18S</th>
<th>COXI</th>
<th>18S + COXI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data partition length (bp)</td>
<td>888</td>
<td>364</td>
<td>1252</td>
</tr>
<tr>
<td>No. variable sites/% data partition</td>
<td>258/29.1</td>
<td>192/52.7</td>
<td>558/44.6</td>
</tr>
<tr>
<td>No. parsimony informative Site/% data partition</td>
<td>132/14.9</td>
<td>156/42.9</td>
<td>300/23.96</td>
</tr>
<tr>
<td>%A</td>
<td>24.6</td>
<td>33.9</td>
<td>27.6</td>
</tr>
<tr>
<td>%T</td>
<td>24.5</td>
<td>34.7</td>
<td>27.8</td>
</tr>
<tr>
<td>%C</td>
<td>22.5</td>
<td>14.9</td>
<td>20.0</td>
</tr>
<tr>
<td>%G</td>
<td>28.4</td>
<td>16.5</td>
<td>24.6</td>
</tr>
<tr>
<td>Sequence divergence</td>
<td>0.2–11.4%</td>
<td>0.3–28.8%</td>
<td>0.3–28.8%</td>
</tr>
<tr>
<td>Tree length</td>
<td>629</td>
<td>762</td>
<td>986</td>
</tr>
<tr>
<td>Consistency index</td>
<td>0.391</td>
<td>0.278a</td>
<td>0.493</td>
</tr>
<tr>
<td>Retention index</td>
<td>0.606</td>
<td>0.381a</td>
<td>0.539</td>
</tr>
</tbody>
</table>

* The value was obtained with the condon positions weighting of nt1, nt2, and nt3 ( = 2:2:1).

![Fig. 1. Saturation analysis of the COXI nucleotide substitution. X axis: K2P+Γ distance; Y axis: uncorrected p-distance. (A) Scatter plot graphic for whole codons; (B–D) Same for first, second, and third positions in alignment.](image-url)
in Pentatomoidea and the lowest divergence was 17.6% in Coreoidea. The result suggested that mutation rate of COXI gene was about four times quicker than that of 18S rDNA. We analyzed the sequence divergence (K2P + Γ) between superfamilies based on 18S rDNA data and 18S rDNA data, which are presented at Table 4. From the sequence variation results, it was shown that the superfamly Aradoidea was distant from other superfamilies (= Trichophora) whether based on COXI data or 18S rDNA data. This was accordant with our analyses based on combined dataset and morphological studies.

3.4. Phylogenetic analyses

3.4.1. 18S rDNA partition

Trees reconstructed using 18S rDNA data with MP, ME and ML methods were mainly congruent. Unweighted parsimony analyses yielded 9 equally parsimonious trees of 629 steps (CI = 0.591, RI = 0.606); the strict consensus tree was presented in Fig. 2. The monophyly of infraorder Pentatomomorpha was strongly supported (98% bootstrap and 8 branch support value). The superfamly Aradoidea was sister group to the remainder of the Pentatomomorpha (= Trichophora), which was at a basal position of the tree topology. The monophyletic Trichophora was subgrouped into two major lineages. Superfamilies affiliations recovered were largely concordant with morphology-based groupings. One of the two major lineages recovered by 18S rDNA was polyphylectic superfamly Pentatomoidea, which was monophyletic in ML and NJ trees (results not shown); the family Pyrrhocoridae (superfamly Pyrrhocoroidea) was arising from it. The second major lineage recovered comprised Lygaeoida, Coreoidea, and Pyrrhocoroidea. The three superfamilies were all polyphyletic (Fig. 2). In Lygaeoida, four families Piesmatidae, Berytidae, Lygaeidae and Rhyparochromidae were analyzed, grouping into two clades. One clade was Lygaeidae and Rhyparochromidae, which recovered as sister group with 53% bootstrap probability. The other was Piesmatidae and Berytidae, which were also sister group with 68% bootstrap probability. However, the Piesmatidae and Berytidae were distant to the clade of Lygaeidae and Rhyparochromidae.

3.4.2. COXI partition

Trees reconstructed using COXI partition with MP, ME, and ML methods were not mainly congruent. Considering the saturation of substitutions (see Fig. 1), different weighted parsimony analyses were performed. However, these tree topologies were not congruent and most groups were unresloved (trees not shown). Therefore, we utilized ML methods to evaluate the phylogeny among superfamilies for estimating the rate heterogeneity in the data sets. Based on the result from the program MrModeltest, the ML analysis was performed using a GTR + I + G model with the following parameters: base frequencies = freq A: 0.4510, freq C: 0.1169, freq G: 0.1078, freq T: 0.3243; I = 0.3366; Γ = 0.3672. The ML tree topology (Fig. 3) showed that the monophyly of infraorder Pentatomomorpha was not recovered; the superfamly Aradoidea was not sister group to Trichophora but arising from within it. It was obvious that the phylogenetic relationship was not concordant with morphology-based groupings.

3.4.3. Combined data partition

The result from the partition homogeneity test gave \( P = 0.97 \), which indicated that there is insufficient evidence for rejecting the hypothesis of congruence. We take this result as sufficient basis for combination of the data partitions for phylogenetic analyses and consider the results of the combined analyses the best hypotheses of Pentatomomorpha phylogeny based on molecular evidence (Farris, 1983).

Trees reconstructed based on combined dataset with MP, ML, and ME methods were mainly congruent. First, we used MP method with combined dataset to reconstruct the phylogeny of Pentatomomorpha. Considering the saturation at third condon positions in COXI partition, different downweighted parsimony analyses of COXI partition were performed. Analysis of combined molecular data yielded 2 most-parsimonious trees with 0.05 weight- ing at third condon positions in COXI partition (986 steps, CI = 0.493, RI = 0.539). The strict consensus tree (Fig. 4) was similar to the consensus tree obtained by the independent analysis of 18S rDNA (Fig. 2). However within Trichophora, the paraphyly of Pentatomoidea was recovered, and relationships within the large clade including Lygaeoida, Coreoidea, and Pyrrhocoroidea were more resolved with Coreoidea paraphyletic.

Table 4

<table>
<thead>
<tr>
<th>Superfamily</th>
<th>Aradoidea</th>
<th>Lygaeoida</th>
<th>Coreoidea</th>
<th>Pentatomoidea</th>
<th>Pyrrhocoroidea</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aradoidea</td>
<td>—</td>
<td>0.238</td>
<td>—</td>
<td>0.240</td>
<td>0.225</td>
</tr>
<tr>
<td>Lygaeoida</td>
<td>0.103</td>
<td>—</td>
<td>0.173</td>
<td>0.182</td>
<td>0.176</td>
</tr>
<tr>
<td>Coreoidea</td>
<td>0.082</td>
<td>0.063</td>
<td>—</td>
<td>0.183</td>
<td>0.158</td>
</tr>
<tr>
<td>Pentatomoidea</td>
<td>0.090</td>
<td>0.073</td>
<td>0.032</td>
<td>—</td>
<td>0.178</td>
</tr>
<tr>
<td>Pyrrhocoroidea</td>
<td>0.095</td>
<td>0.082</td>
<td>0.042</td>
<td>0.050</td>
<td>—</td>
</tr>
</tbody>
</table>

*Note.* The upper triangle represent sequence divergence (K2P + Γ) between superfamilies based on COXI data. The lower triangle represent sequence divergence (K2P + Γ) between superfamilies based on 18S rDNA data.
Then based on the results of the MrModeltest and AIC evaluations, the ML analysis of combined molecular data was performed using the general time-reversible substitution model (GTR) with a gamma correction for among-site rate variation and a correction for significant invariable sites, with the following values: Base frequencies: \( \text{freq A: 0.2636, freq C: 0.2166, freq G: 0.2377, freq T: 0.2821} \); \( \text{I: 0.411, ID: 0.4213} \). This analysis yielded the topology shown in Fig. 5 (\( \text{log likelihood = 10344.45} \)), which was largely concordant with combined analysis MP tree (Fig. 5). As in MP tree, topology of the ML tree also strongly supported the monophyletic Pentatomomorpha with Aradoidea being sister to Trichophora. Further, the superfamily Pentatomoida was recovered as a monophyletic group with 65% bootstrap value.

Additionally, we utilized ME method with the K2P model (Kimura, 1980) to reconstruct the phylogeny of Pentatomomorpha. The topology of the tree (Fig. 6) was similar to that of the ML tree. However, the Lygaeoida
was recovered as monophyletic with relatively weak support (<50% bootstrap value).

4. Discussion

The COX1 appears to be excellent for phylogenetic comparisons of not too distantly related species of insects (Damgaard, 2000). However, at deep levels of divergence, multiple substitutions in third positions might contribute to problems in resolving the correct phylogeny. Thus, in deep-rooted phylogenies there is the potential for considerable homoplasy in third positions, especially due to saturation for A-T transversions. This appears to be the case of our present COX1 data set, which obviously showed a high degree of saturation in third positions (Fig. 1).
Meanwhile, the phylogenetic relationships inferred from analyses of COXI sequence data (Fig. 3) were not congruent with those obtained from analyses of 18S rDNA and combined molecular data (Figs. 2 and 4–6). The phylogenetic information contained in COXI data seemed to be inadequate to independently resolve relationships at or above superfamily level (Fig. 3), although combining these data with 18S rDNA sequences increased support for most clades recovered by 18S rDNA alone. Therefore, it was suggested that the COXI segment might not be an optimal molecular marker for the phylogeny of Pentatomomorpha.

Analyses of the combined data partition provided more resolved and robust estimates of Pentatomomorpha phylogeny overall than the analyses of either gene separately. The phylogenies (Figs. 2 and 4–6) were most congruent with morphological studies. The recovered monophyletic Pentatomomorpha was congruent with morphological studies (Schuh, 1979; Wheeler et al., 1993). The placement of the Aradoidea within...
Pentatomomorpha was strongly supported, which was the sister group to Trichophora. In morphological study, this is supported by the presence of pulvilli in some members (Tullgren, 1918), tubular salivary glands (Southwood, 1955), egg architecture, and trichophoran-type spermathecae (Pendergrast, 1957). Another well-supported clade in our analyses was Trichophora. It was grouped into two major lineages, one being the superfamily Pentatomoidea, and the other comprising Lygaeoidea, Coreoidea, and Pyrrhocoroidea. This was entirely identical with morphological research of Schaefer (1993). He grouped Trichophora into two lines: the Pentatomoidae, and the Lygaeoidea + Coreoidea + Pyrrho coroidea, with distinguished differences in trichobothrial patterns (2 pairs on abdominal sterna 3–7 in the former, 3:3:3:2 pairs on these sterna in the latter). Therefore, our molecular analyses supported that the morphological characteristic abdominal trichobothria was synapomorphy of Trichophora. And this was significant to determining polarities of morphological characteristics in phylogenetic researches of Pentatomomorpha.

Fig. 5. Maximum-likelihood phylogenetic reconstruction of Pentatomomorpha based on 18S rDNA and COX1 gene sequences. The topology was reconstructed under the GTR + I + G model of nucleotide substitution, −log likelihood = 10544.45. Numbers represent the percentage of 100 bootstrap replicates. Current superfamily and family taxa following Henry (1997) are indicated on the right.
As to the superfamily Pentatomoidae lineage, the monophyly was supported by the analyses of the ML and ME trees of combined dataset, being consistent with the morphological studies. However, the relationships among families were confused with different trees having different topologies, which is consistent with the appalling confusion status of this superfamily in morphological analysis (Stys and Kerzhner, 1975). The 18S rDNA did not provide sufficient information to separate families inside this superfamily, suggesting requirement of more sequence data in phylogeny study.

Within the second major Trichophora lineage, the Pyrrhocoroidea was polyphyletic in all analyses; the monophyletic Lygaeoidea was supported only in the
analysis of ME tree of combined dataset, and Coreoidea was polyphyletic except in the MP tree of combined dataset.

There is no doubt that Piesmatidae is a family of Pentatomomorpha (Drake and Davis, 1958), however its dataset was polyphyletic except in the MP tree of combined analysis of ME tree of combined dataset, and Coreoidea.

324

H.-M. Li et al. / Molecular Phylogenetics and Evolution 37 (2005) 313–326

ME trees (Fig. 6). So our limited taxon sample failed to two clades formed a monophyletic Lygaeoidea in the clade of Lygaeidae and Rhyparochromidae, whereas the tion, and the MP and ML trees of combined dataset, the MP and ML trees of combined dataset, the clade of Piesmatidae and Berytidae were distant to the clade of Lygaeidae and Rhyparochromidae, whereas the two clades formed a monophyletic Lygaeoidea in the ME trees (Fig. 6). So our limited taxon sample failed to reject an alternative hypothesis suggested by the morphological analysis of Henry (1997) that Piesmatidae is a family of Lygaeoidea. This will need more taxa to be sampled.

With regard to the phylogeny of Coreoidea, there are many arguments in morphological research. Henry (1997) and Li (1996) recovered the monophyletic Coreoidea using cladistic analyses with morphological data, whereas our analyses failed to recover the Coreoidea as monophyletic. The conflict may partly account for the low bootstrap value and branch support of superfamily Coreoidea in our analyses. This indicated that such areas of a cladogram tend to be most sensitive to taxon sampling bias. With regard to the family Corei- dae, which is a large family and currently comprising about 250 genus worldwide, traditional classification is primarily based on the concepts of “synthetic systematics” (Hsiao et al., 1977; Schuh and Slater, 1995). Lacking of morphological synapomorphies suggests this family polyphyletic. Although without strong support, the tree topologies of our results also rendered this family polyphyletic. The molecular data and morphological data both indicate that taxonomy of this family should be revised subsequently.

The effects of taxon sampling on phylogenetic reconstruction have been explored in a number of studies (Graybeal, 1998; Lecointre et al., 1993; Poe, 1998); a conclusion common to such studies is that the addition or deletion of taxa often affects the resulting topology. We acknowledge that the addition of taxa in our work, especially within missing or underrepresented families, could alter phylogenetic hypotheses of the Pentatomomorpha. However, the results of our molecular analyses for the first time contribute to our understanding of the molecular phylogeny of Pentatomomorpha. Evidence from other genes is clearly necessary to unequivocally resolve Pentatomomorpha relationships. Ultimately, combined analyses of molecular and morphological data sets should strengthen current estimates of phylogenetic relationships of Pentatomomorpha.

Acknowledgments

We gratefully thank Dr. Pang Hong for providing the specimens and help (Sun Yat-sen University, Institute of Entomology). We are indebted to Jakob Damgaard (Department of Evolutionary Biology Biological Institute University of Copenhagen) for advice in our research and providing literature. The authors thank Bruce C. Campbell and Li Xin-Zheng for providing literature and the anonymous reviewers for their thorough and helpful reviews of the manuscript. This work was funded by the National Natural Science Foundation of China under Grant 30170039 and the Natural Science Foundation of Guangdong Province under Grant 011134 and 20023002.

References


Graybeal, A., 1998. Is it better to add taxa or characters to a di
Felsenstein, J., 1985. Con
Farris, J.S., 1983. The logical basis of phylogenetic analysis. In: Plat-
Kocher, T.D., Thomas, W.K., Meyer, A., Edwards, S.V., Paabo, S., Vil-
Li, X.Z., 1996. Cladistic analysis and higher classification of the Coro
Kimura, M., 1980. A simple method for estimating evolutionary rate of
Henry, T.J., Froeschner, R.C., 1988. Catalog of the Heteroptera, or true
Kumar, M., Tamura, K., Jakobsen, I.B., Nei, M., 2001. MEGA2: Molecular Evolutionary Genetics Analysis Software. Arizona State University, Tempe, AZ, USA.
Svensen, J.T., Campbell, B.C., Gill, R.J., Steffen-Campbell, J.D., 1995. Non-monomophy of Anoectochyphina ("Homoptera"), based upon 18S rDNA phylogeny: Eco-evolutionary and cladistic implications within pre-Heteropterodea Hemiptera (s.l.) and a proposal for new monophyletic suborders. Pan-Pac. Entomol. 73, 31–60.