Evolution and phylogenetic information content of the ribosomal DNA repeat unit in the Blattodea (Insecta)

Dmitry Mukha a, *, Brian M. Wiegmann b, Coby Schal b

* Vavilov Institute of General Genetics, Russian Academy of Sciences, Gubkin St. 3, 117809, Moscow, Russia
b Department of Entomology, Box 7613, North Carolina State University, Raleigh, NC 27695, USA

Received 5 December 2000; received in revised form 25 October 2001; accepted 26 October 2001

Abstract

The organization, structure, and nucleotide variability of the ribosomal repeat unit was compared among families, genera, and species of cockroaches (Insecta; Blattodea). Sequence comparisons and molecular phylogenetic analyses were used to describe rDNA repeat unit variation at differing taxonomic levels. A ~1200 bp fragment of the 28S rDNA sequence was assessed for its potential utility in reconstructing higher-level phylogenetic relationships in cockroaches. Parsimony and maximum likelihood analyses of these data strongly support the expected pattern of relationships among cockroach groups. The examined 5’ end of the 28S rDNA is shown to be an informative marker for larger studies of cockroach phylogeny. Comparative analysis of the nucleotide sequences of the rDNA internal transcribed spacers (ITS1 and ITS2) among closely related species of Blattella and Periplaneta reveals that ITS sequences can vary widely in primary sequence, length, and folding pattern. Secondary structure estimates for the ITS region of Blattella species indicate that variation in this spacer region can also influence the folding pattern of the 5.8S subunit. These results support the idea that ITS sequences play an important role in the stability and function of the rRNA cluster. © 2002 Published by Elsevier Science Ltd.

Keywords: Phylogenetic analysis; 28S ribosomal DNA; Ribosomal internal transcribed spacer 1 and 2; Blattodea

1. Introduction

Ribosomal DNA (rDNA) has long been considered a useful marker for comparative evolutionary and phylogenetic studies (Hillis and Davis, 1986; Mindell and Honeycutt, 1990; Wesson et al., 1993; Schlotterer et al., 1994; Vogler and DeSalle, 1994; Tang et al., 1996). This utility results largely from the differential conservation and variability of different genes and regions that make up the rDNA repeat unit (Hillis and Dixon, 1991; Hamby and Zimmer, 1992). The basic organization of rDNA is conserved throughout eukaryotes. Eukaryotic ribosomal RNA genes are arranged in tandemly repeated clusters, with each cluster containing the genes for 18S-, 5.8S-, and 28S-like ribosomal RNAs. The genes are separated by several spacers, namely, the NTS (nontranscribed spacer), and ITS1 and ITS2 (internal transcribed spacers). The NTS separates neighboring repeat units; ITS1 is located between the 18S- and 5.8S-like coding regions; ITS2 lies between the 5.8S- and 28S-like genes (Fig. 1A) (Gerbi, 1985). Ribosomal repeats are usually localized to one or a few chromosomes and form part of the nucleolar organizers (NO). Because the coding regions and spacers differ widely in their rate of evolution, they can reveal phylogenetic relationships ranging from the level of major phyla of living organisms to the population level (Hillis and Dixon, 1991; Wesson et al., 1992; Kuperus and Chapco, 1994; Honda et al., 1998; Muccio et al., 2000; Wiegmann et al., 2000).

Additional information about relationships and rDNA evolution can be obtained by examining the secondary structure of the ribosomal RNA. In the rDNA coding regions (18S-, 5.8S-, and 28S-like), sequence conservation is reflected in the high similarity of secondary structures in a variety of distantly related organisms (Gerbi et al., 1982; Rauê et al., 1988; Wesson et al., 1992). Sequence tracts that are highly variable between
species may still retain certain, apparently functionally important, components of secondary structure. For example, estimated secondary structures of the ITS1 sequences are nearly identical in human, chimpanzee, and gorilla (Gonzales et al., 1990), and those from mouse and rat also show high similarity (Michot et al., 1983). Integrity of the rDNA secondary structure is maintained through compensatory nucleotide changes to preserve base pairing in stems, and through indels which change stem length but do not alter the overall folding pattern. Structural analysis of the ITS in yeast has shown that the spacer regions also play a role in the maturation of precursor rRNA molecules (Musters et al., 1990; Van der Sande et al., 1992; Schulenburg et al., 1999).

Despite the importance of rDNAs in phylogenetic and molecular evolutionary studies of insects, very little is known about the rDNA of cockroaches. Because of their pest status and the ease with which they are cultured in the laboratory, cockroaches have served as model organisms for studies of insect physiology, molecular genetics,
and chemical ecology, but relatively few of these studies have been explicitly comparative or evolutionary in their focus. Our studies of the nuclear ribosomal genes in cockroaches are aimed at understanding the dynamics of sequence and structural variation in these genes and assessing the utility of that variation for comparative evolutionary studies. In this paper, we examine the organization, structure, and nucleotide variability of the ribosomal repeat unit of cockroaches (Insecta: Blattodea). Our comparison focuses on the ITS1 and ITS2 spacer regions, as well as on portions of the rDNA subunit coding regions that immediately flank them. These regions are compared across a broad range of taxonomic divergences within cockroaches.

Divergences between extant genera of cockroaches may be as old as 75–100 my before present (Labandeira, 1994; Nalepa and Bandi, 1999), but some genus and species-level divergences could be much more recent (Nalepa and Bandi, 1999). Phylogenetic relationships among cockroach groups are the subject of current debate in the insect systematics literature (McKittrick, 1964; Grandcolas, 1994, 1996, 1999; Kambhampati, 1995, 1996; Klass, 1997, 1998; Nalepa and Bandi, 1999). Despite renewed interest in cockroach phylogenetics and a wealth of new data, major differences remain among proposed phylogenetic arrangements for cockroach families (Fig. 2). For divergences as old as those hypothesized for cockroach families, it is likely that the more slowly evolving regions of the nuclear ribosomal DNA (18S rDNA, Lo et al., 2000; 28S rDNA, this study) and conserved nuclear protein encoding genes could be important sources of new evidence on cockroach relationships.

Our sequence comparisons and molecular phylogenetic analyses are used to describe rDNA repeat unit variation at differing taxonomic levels. First, we evaluate the phylogenetic utility of a ~1200 bp fragment of the 28S rDNA for reconstructing higher-level cockroach relationships. Our findings show that the 28S rDNA is highly informative for higher-level cockroach phylogeny. Second, comparative analysis of the nucleotide sequences and the secondary structures of the ITS1 and ITS2 among closely related Blattella and Periplaneta species reveals major structural and sequence-level constraints.

2. Materials and methods

2.1. Cockroach taxa sampled

rDNA sequences were obtained from 11 cockroach species from three families: Blattellidae — Blattella germanica, B. vaga, B. lituricollis, B. asahinai, Parcoblatta latta; Blattidae — Periplaneta americana, P. fuliginosa, P. brunea; Blaberidae — Diploptera punctata, Blaberus atropus, B. giganteus. Specimens were obtained from laboratory cultures (CS lab, NCSU), or from colleagues, and frozen at −80 °C to preserve nucleic acids.

2.2. Laboratory methods

Total genomic nucleic acids were extracted using a standard DNA extraction buffer (Tris–HCl, proteinase K, SDS) according to the protocol described in Mukha et al. (1995).

The primers used for amplification and sequencing are shown in Fig. 1B. The priming sites for DAMS-18 and DAMS-28 are highly conserved in eukaryotes; these primers are “universal” and may be used for amplification of the corresponding rDNA fragments across a broad taxonomic range of organisms (Mukha and Sidorenko, 1995, 1996; Mukha et al., 2000). In addition, we developed seven additional priming sites, specific for cockroaches, within the rDNA repeat unit (coc1–coc7, Fig. 1B).

PCR amplification of overlapping fragments within analyzed rDNA fragments was carried out using Taq DNA Polymerase (Promega) and the PTC-100 Thermal Cycler (MJ Research Inc.). Each reaction contained 0.1 mg DNA template, 1.5 mM MgCl and 1 mM each dNTP. The PCR regimen was as follows: initial template denaturation at 95 °C for 5 min, followed by 30 cycles of 94 °C for 1 min, 55 °C for 2 min, and 72 °C for 3 min, and a final 7 min elongation cycle at 72 °C.

2.3. DNA sequencing

Sequences were obtained by dye terminator cycle sequencing using the ABI Taq FS enzyme (PE Applied Biosystems, Foster City CA), gel fractionated, and base-called on the ABI PRISM™ 377 DNA sequencer (PE Applied Biosystems, Foster City, CA) of the North Carolina State University DNA Sequencing Facility. Opposite strands were confirmed for all templates. ABI trace files were edited and contigs assembled using the program GAP4 in the STADEN software package (Staden, 1996) on a SUN Ultra1 workstation.

2.4. Nucleotide alignment and phylogenetic analysis

Alignments were constructed using the multiple sequence alignment algorithm in the program Clustal-W 1.7 (Thompson et al., 1994). The program FASTA (Pearson, 1990) was used to investigate pairwise sequence differences for ITS1 and ITS2 in closely related cockroach species. For the phylogenetic analysis of the 28S rDNA fragment, we analyzed two nucleotide alignment data sets. One included the entire optimal alignment generated in Clustal-W 1.7. For the second, we excluded positions within the 28S gene for which primary sequence homology was in doubt, thus treating...
Fig. 2. Alternative phylogenetic hypotheses for the higher-level relationships of cockroaches. (a) 12S rRNA (Kambhampati, 1996); (b) 16S and 12S rRNA (Kambhampati, 1995); (c) multiple gene sequences (Lo et al., 2000); (d) mt COII (Maekawa and Matsumoto, 2000); (e) comparative morphology (Grandcolas, 1996); (f) comparative morphology (Klass, 1997).

as missing data positions for which alternative ad hoc placement of indels could affect the phylogenetic outcome. The latter method uses only those positions whose homology is certain to infer a tree. All of the alignments generated in the current study can be obtained from the following website: http://www2.ncsu.edu/unity/users/b/bwiegman/public_html/align.html and from the EMBL nucleotide database. Gaps in unambiguously alignable regions were also treated as missing data. Phylogenetic trees were constructed using parsimony and maximum likelihood methods in PAUP*4b2a (Swofford, 1999). Parsimony searches were conducted using the branch-and-bound option in PAUP* (Swofford et al., 1996).

Heuristic likelihood searches were conducted under the following model parameters: HKY85 (Hasegawa et al., 1985) model of nucleotide substitution with empirical base frequencies and transition/transversion ratio set to 2.0. Bootstrap support values (Felsenstein, 1985) were obtained from 1000 replicate re-sampled data sets for parsimony analysis and 500 replicates for maximum likelihood.

2.5. rDNA secondary structure estimation

We estimated secondary structures for the rRNA fragment containing 5.8S–ITS1 sequences. Folding patterns were compared in two closely related species: B. germanica and B. lituricollis. Zuker’s dynamic programming algorithm (Zuker and Stiegler, 1981) was used to calculate secondary structures. Estimated structures were visualized using RNAdraw software (http://iubio.bio.indiana.edu/IUBio-Software+Data/molbio/ibmpc/rnadraw-readme.html).

3. Results and discussion

3.1. Phylogenetic utility of the 28S rDNA in Blattodea

Alignment of the 28S rDNA sequence data for the 11 cockroach species used in this study resulted in 883 sites included in the phylogenetic data set. Of these, 267 were variable and 156 were parsimony informative (see http://www2.ncsu.edu/unity/users/b/bwiegman/public_html/align.html). Analysis of the entire alignment generated in Clustal-W 1.7 with all positions included gave identical phylogenetic results to those reported below.

The length of the fragment was nearly identical for all of the sequenced taxa (Table 1). Uncorrected pairwise sequence divergence ranged from 0.1 to 8% for Blattella species, 11 to 19% between genera and families, with...
### Table 1
Length variation and nucleotide composition of sequenced cockroach rDNA regions

<table>
<thead>
<tr>
<th>Species</th>
<th>rDNA Region</th>
<th>GenBank Acc number</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Length (nt)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ITS1</td>
<td>ITS2</td>
</tr>
<tr>
<td>Blattella asahinai</td>
<td>667</td>
<td>443</td>
</tr>
<tr>
<td>Blattella germanica</td>
<td>661</td>
<td>445</td>
</tr>
<tr>
<td>Blattella lituricollis</td>
<td>841</td>
<td>493</td>
</tr>
<tr>
<td>Blattella vaga</td>
<td>291</td>
<td>442</td>
</tr>
<tr>
<td>Parabolatta latta</td>
<td>393</td>
<td>711</td>
</tr>
<tr>
<td>Blaberus atropus</td>
<td>342</td>
<td>282</td>
</tr>
<tr>
<td>Blaberus giganteus</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Periplaneta Americana</td>
<td>391</td>
<td>267</td>
</tr>
<tr>
<td>Periplaneta brunea</td>
<td>357</td>
<td>226</td>
</tr>
<tr>
<td>Periplaneta fuliginosa</td>
<td>343</td>
<td>264</td>
</tr>
<tr>
<td>Diploptera punctata</td>
<td>379</td>
<td>247</td>
</tr>
</tbody>
</table>

*28S* lengths represent sequenced and aligned lengths of the 5' fragment sampled in the current study only.

the most divergent comparisons between *Diploptera* and *Blattella*. Average base frequencies for this fragment were A=16.10, C=34.07, G=29.29, and T=20.06%, reflecting differential base substitution rates across deep divergence times. A Chi square test for base composition showed no significant deviation from these proportions among taxa (χ²=19.26, df=30, p=0.93). To assess whether the sequenced 28S rDNA fragment is saturated for the divergences among our taxa, total number of transitional and transversional changes were plotted against HKY model-corrected pairwise distances (Fig. 3). For both classes of substitution, numbers steadily accumulated as corrected pairwise divergence increased, indicating that saturation had not been reached (Fig. 3).

Parsimony analysis of the 28S rDNA data set yielded a single most parsimonious tree (Fig. 4; length=395; CI=0.85; RI=0.86). The resulting tree topology was well-labeled as corrected pairwise divergence increased, indicating that saturation had not been reached (Fig. 3).

Fig. 3. Summed transitions and transversions in the 883 bp 28S rDNA fragment as a function of HKY-corrected sequence divergence for each pairwise comparison of taxa.

![Fig. 3. Summed transitions and transversions in the 883 bp 28S rDNA fragment as a function of HKY-corrected sequence divergence for each pairwise comparison of taxa.](image-url)

Fig. 4. Single most parsimonious tree and identical maximum likelihood topology inferred from the 883 bp 28S rDNA fragment (length=604; CI=0.92; RI=0.76; −lnL=3250.52). Node support values above the line are bootstrap percentages based on 1000 replicate parsimony searches/500 replicate maximum likelihood searches. Branch lengths are proportional to the number of parsimony-based assigned character changes under ACCTRAN optimization in PAUP* 4b2a.
supported, with high bootstrap values (above 89%) supporting most of the major internal nodes. Maximum likelihood analysis of the same data yielded an identical tree with similar support values (−lnL=3250.52; Fig. 4). The relationships implied by this analysis are consistent with several published hypotheses of cockroach phylogeny (Kambhampati 1995; Grandcolas 1996; Fig. 2). Our tree supports the monophyly of the included Blattella species and a close relationship of these to Parcoblatta (Fig. 4).

Support for previous findings based on mitochondrial RNA genes (Kambhampati, 1995) and morphology (Grandcolas, 1996, 1999), which placed Blaberidae in or near the Blattellidae, depends on identifying the root position for our tree. Placement of blaberid species (Diploptera punctata, Blaberus atropus, B. giganteus) within our unrooted network is consistent with previous molecular and morphological results, but cannot rule out the alternative placement as sister group to Blattidae implied by a root node position for the blattids between the Blattellidae and all remaining taxa. Sequences from additional taxa, notably Cryptocercidae and Polyphagidae, as well as from close relatives to cockroaches, Isoptera (termites), would provide useful information about the phylogenetic relationships of these primitive taxa.

Nevertheless, our small taxon sample was only intended to gauge the utility of 28S rDNA as a potential source of phylogenetic information. Our finding of the same well-supported topology using both parsimony and maximum likelihood methods with strong support for nodes, as well as the observed intermediate levels of variation (11–19% pairwise sequence difference between families) among the taxa, demonstrates that this gene contains sufficient nucleotide variation to be highly informative of cockroach phylogeny. This finding supports similar conclusions for the nuclear rDNA genes in other orthopteroid insects (Flook et al., 1999), and insect order- and family-level analyses (Whiting et al., 1997; Wiegmann et al., 2000). The 28S rDNA would be an excellent candidate gene for higher-level phylogenetic studies in Blattodea.

### 3.2. ITS interspecies variation

The results of multiple nucleotide alignments among Blattella species (B. germanica, B. asahinai, B. lituricolis, B. vaga) and Periplaneta species (P. americana, P. fuliginosa, P. brunea) are shown in Figs. 5 (ITS1 sequences) and 6 (ITS2 sequences). Multiple insertions and deletions, as well as numerous point substitutions, are revealed. The most extensive changes were found within ITS sequences of the Blattella species. The ITS1 sequences of B. germanica–B. lituricolis and B. germanica–B. vaga were found to be 67.2 and 30.0% identical, respectively. Significant length differences were also found between B. germanica, B. lituricolis and B. vaga ITS1 sequences — 661 b, 841 b and 291 b, respecti-
Fig. 5. Comparison of the nucleotide sequences in ITS1; similar nucleotides are indicated by asterisks. (A) Blattella species; (B) Periplaneta species.

1. B. germanica; 2. B. asahinai; 3. B. lituricollis; 4. B. vaga; 5. P. americana; 6. P. fuliginosa; 7. P. brunea. The nucleotides that affect rRNA folding are shown in bold.
Fig. 6. Comparison of the nucleotide sequences in ITS2; similar nucleotides are indicated by asterisks. (A) Blattella species; (B) Periplaneta species.


In addition to the observed variation, the ITS2 sequences also contained specific conserved motifs (underlined sequences in Fig. 6A and B, respectively). Elements of these motifs were shared between Blattella and Periplaneta, as indicated by double underlining in Fig. 6A and B. Additionally, a "super motif" — gccgaccctcagccagg — was shared by all of the Blattodea ITS2 sequences we have examined to date, and was also found in the published ITS2 sequences of more evolutionarily distant insect species, such as Nebria castanea (Coleoptera), Trichogramma dendrolimi, Leptopilina victoriae, Ganaspis xanthopoda, Melittobia digitata (Hymenoptera), (GENBANK Accession numbers: AF173883, AF227949, AF015902, AF015892, MDU02950, respectively). It is commonly suggested that the high variability of rDNA spacer region sequences results from relaxed selection on the primary sequence, essentially that these regions are more "neutral" (sensu Kimura, 1983). However, evidence is accumulating that suggests that regions of sequence conservation in ITS sequences correspond to functionally important regions. The above described super motif is thus a likely candidate in the further identification of ITS structures that might play an important role in the function and organization of the insect rDNA.

3.3. ITS and secondary structure evolution

To infer whether ITS structural changes revealed here exert an effect on possible rDNA function, we compared secondary structures of the 5.8S rRNA compartment. This molecule is excised during processing and should fold as a separate unit, without the involvement of extended sequences of transcribed spacers (Gerbi, 1985). Folding of the 5.8S–ITS1 fragments of B. germanica rRNA is represented in Fig. 7A. The 5.8S rRNA did, in fact, fold as a separate compartment in this species. There was some minor variation in the observed size of the compartment among species; however, the inferred folding of the 5.8S when the ITS1 sequence was included corresponded to the fold inferred when 5.8S was analyzed without the ITS1 sequence (Fig. 7B). Folding of the B. lituricollis 5.8S–ITS1 sequence is represented in Fig. 7C. In B. lituricollis, sequences corresponding to the 5.8S rRNA formed extended contacts with the ITS1 region, with an inferred folding pattern entirely different from that observed in B. germanica. In a hypothetical sequence from the 5.8S–ITS1 region of B. lituricollis, but with the sequences shown in bold in Fig. 5A removed (i.e., excluding a large insertion and a dinucleotide insertion present in comparisons of B. germanica and B. lituricollis ITS1 sequences), the basic B. germanica-like 5.8S rRNA fold was restored. Without these two sequence data for the 11 cockroach taxa used in this study is available at http://www2.ncsu.edu/unity/users/b/bwiegm/public_html/align.html.
insertions the 5.8S sequence now formed a discrete secondary structure domain separate from the ITS1 (Fig. 7D). These results suggest to us that the differences observed between ITS1 of B. lituricollis and its closest relative species are not “neutral” and are not simple accumulated random nucleotide changes, but bear a significant functional load. At least two multinucleotide insertions within ITS1 of B. lituricollis critically influence the folding of the rRNA precursor. Structural analysis of the ITS in yeast has shown that spacers play a role in the maturation of precursor rRNA molecules (Musters et al., 1990; Schulenburg et al., 1999; Van der Sande et al., 1992).

Critical changes in the rRNA folding pattern brought about by sequence evolution in the ITS spacer regions may thus have an important influence on the kinetics of precursor rRNA formation, and ultimately on the efficient functioning of the rDNA cluster.

Acknowledgements

We thank the North Carolina State University DNA Sequencing Facility for help with automated DNA sequencing. This study was supported in part by the Blanton J. Whitmire Endowment and the W. M. Keck Center for Behavioral Biology at NCSU, and by the Russian State Program “Frontiers in Genetics”. We thank two anonymous reviewers for valuable suggestions.

References


