Application of the secondary structure model of rRNA for phylogeny: D2–D3 expansion segments of the LSU gene of plant-parasitic nematodes from the family Hoplolaimidae Filipjev, 1934

Sergei A. Subbotin a,g,*; Dieter Sturhan b; Nicola Vovlas c; Pablo Castillo d; James Tanyi Tambe e; Maurice Moens f; James G. Baldwin a

a Department of Nematology, University of California, Riverside, CA 92521, USA
b c/o Institut für Nematologie und Wirbeltierkunde, BBA, 48161 Münster, Germany
c Istituto per la Protezione delle Piante, Sezione di Bari, C.N.R., 70126 Bari, Italy
d Instituto de Agronomía Sostenible, Consejo Superior de Investigaciones Científicas, Apdo 4084, 14080 Córdoba, Spain
e Via Gasparri 5, 62100 Macerata, Italy
f CLO-Department for Crop Protection, 9820 Merelbeke, Belgium
g Institute of Parasitology, Russian Academy of Sciences, Leninskii Prospect 33, 117071 Moscow, Russia

Received 19 May 2006; revised 25 September 2006; accepted 26 September 2006
Available online 7 October 2006

Abstract

Knowledge of rRNA structure is increasingly important to assist phylogenetic analysis through reconstructing optimal alignment, utilizing molecule features as an additional source of data and refining appropriate models of evolution of the molecule. We describe a procedure of optimization for alignment and a new coding method for nucleotide sequence data using secondary structure models of the D2 and D3 expansion fragments of the LSU-rRNA gene reconstructed for fifteen nematode species of the agriculturally important and diverse family Hoplolaimidae, order Tylenchida. Using secondary structure information we converted the original sequence data into twenty-eight symbol codes and submitted the transformed data to maximum parsimony analysis. We also applied the original sequence data set for Bayesian inference. This used the doublet model with sixteen states of nucleotide doublets for the stem region and the standard model of DNA substitution with four nucleotide states for loops and bulges. By this approach, we demonstrate that using structural information for phylogenetic analyses led to trees with lower resolved relationships between clades and likely eliminated some artefactual support for misinterpreted relationships, such as paraphyly of Helicotylenchus or Rotylenchus. This study as well as future phylogenetic analyses is herein supported by the development of an on-line database, NEMrRNA, for rRNA molecules in a structural format for nematodes. We also have developed a new computer program, RNAstat, for calculation of nucleotide statistics designed and proposed for phylogenetic studies.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Character coding; Bayesian inference; D2–D3 expansion segments; Structural database; Helicotylenchus; Maximum parsimony; Rotylenchus; rRNA; Secondary structure model; Variability map

1. Introduction

Phylogenetic studies using rRNA often ignore that this molecule is transcribed from genes performing complex secondary and tertiary structures mediated by base pairing between, sometimes distant, regions of the molecule. Ribosomal RNA forms helical structures where two regions of the sequences are complementary. Pairing within a helix involves canonical base pairs, that is, Watson–Crick base pairs and G–U wobble pairs (Wuyts et al., 2001). Knowledge of RNA secondary structure is becoming increasingly important in assisting phylogenetic analysis through three main approaches. First, secondary structure is used to
identify homologous positions for sets of nucleotides and sequence regions, which are otherwise difficult to determine using multiple alignment algorithms (Kjer, 1995). Second, secondary structure can be used as an additional source of data incorporating both structural and morphometric parameters of rRNA molecules (Aleshin et al., 1998; Billoud et al., 2000; He et al., 2005). Third, secondary structure is used for selecting increasingly appropriate models of evolution. It has been shown in many studies that the mode of evolution within helical regions of rRNA is via compensatory or semi-compensatory mutations, and thus, the evolution of bases in structurally related positions is highly dependent. Yet, most substitution models used in phylogenetic studies treat changes along nucleotide sequences as independent. The probability method of Bayesian phylogeny (Hudelot et al., 2003) and approaches considering empirical substitution rates for unpaired and paired bases matrices (Smith et al., 2004) have been recently applied to modeling sequence and structure of rRNA molecules. Several likelihood-based studies have already shown the superiority of considering base pair correlation in RNA stems over methods assuming independent evolution of nucleotides (Muse, 1995; Savill et al., 2001; Telford et al., 2005). Recently, this approach was applied to phylogenetic analyses of 18S rRNA of insects (Kjer, 2004) and Bilateria (Telford et al., 2005) using the Markov Chain Monte Carlo Bayesian search procedure toward resolving problems of incongruence among phylogenies yielded by other methods. Thus, knowledge of secondary structure allows applying a more sophisticated model, and consequently generating a picture of relationships argued to be more realistic. With increasing realization of the importance of secondary structure for accurate sequence alignment and phylogenetic analysis, the need for secondary structure models of rRNA for diverse taxonomic groups is becoming more pressing. Such models can best be developed and tested within a monophyletic group of tractable size and for which there are plausible hypotheses of relationships and outgroups including those based on classical morphological studies.

Nematodes, putatively with more than one million species, are of particular interest in molecular phylogenetics with special challenges including apparent rapid rates of evolution, and conserved or convergent morphology that confounds understanding evolution (Baldwin et al., 1999). In the face of these difficulties, SSU based phylogenies using limited representatives have identified major clades (Holsaterman et al., 2006), but considering the scope of the phylum, complex models such as considering secondary structure of rRNA are best developed and first applied within more discrete monophyletic groups within Nematoda. The family Hoplolaimidae (order Tylenchida), with more than 400 species, is particularly promising for developing such methods because as cosmopolitan plant parasites their taxonomy is better understood than most groups, and they are widely regarded as monophyletic. Furthermore, refining their phylogeny is particularly relevant to understanding the evolution of pathways for pathogenesis because they are demonstrably the outgroup for worldwide one of the most agriculturally important groups of pests, the cyst nematodes, Heteroderidae (Baldwin, 1992; Baldwin et al., 2004; Subbotin et al., 2006). By contrast to sedentary Heteroderinae, adults of Haplolaimidae are vermiiform (0.5–2 mm long) migratory ecto- or semi-endoparasites of roots. As pathogens on crops, many including Helicotylenchus multicinctus, Rotylenchus robustus, Rotylenchus uniformis, and Steleleonema brachyurus result in significant economic loss (Castillo and Vovlas, 2005; Siddiqi, 2000). On the basis of morphological diversity Haplolaimidae currently is comprised of three subfamilies (Aphasmatylenchinae, Hoploclaininae, and Rotylenchoidinae) and 11 genera (Siddiqi, 2000). Comparative morphology has been the basis to postulate evolutionary transformation of hoplolaimids (Geraert, 1990), but a phylogenetic analysis based on morphological characters or DNA markers has not been attempted.

The D2 and D3 expansion segments of the 28S rRNA are often sequenced in studies of nematode phylogenetics due to the availability of conserved primers amplifying DNA from many taxa, and the presence of phylogenetically informative sites. Thus, there has been a growing shift from 18S rRNA toward including the D2 and D3 segments for analyzing relationships including higher taxonomic levels; examples include studies among orders of the phylum Nematoda (Litvaitis et al., 2000), within the order Tylenchida (Subbotin et al., 2006), suborder Criconematina (Tylenchida) (Subbotin et al., 2005), and Cephalobina (Nadler et al., 2006) as well as within genera of several orders: Pratylenchus (Al-Banna et al., 1997; Duncan et al., 1999), Acrrobolodes (De Ley et al., 1999), Steirernema (Stock et al., 2001), Meloidogyne (Castillo et al., 2003; Tenente et al., 2004), Longidorus (Rubtsova et al., 2001; He et al., 2005). The D2-D3 expansion segments are also considered as promising loci for DNA barcoding of nematodes (De Ley et al., 2005).

Herein, we reconstruct and examine the secondary structure of the D2 and D3 expansion segments of LSU rRNA to improve an automatic alignment; also we use secondary structure information for construction of phylogenetic trees of hoplolaimids. We present a variability map of D2 and D3 expansion fragments superimposed on the putative secondary structure. We propose a new coding method for unpaired and paired bases of secondary structure similar to those used by Smith et al. (2004), but herein converting sequence alignment into a 28 symbol code. Original and converted sequence data are analyzed by maximum parsimony and Bayesian inference considering a complex model of evolution for Haplolaimidae with outgroup representatives from suborder Haplolaimina.

2. Materials and methods

2.1. Nematode populations, DNA extraction, PCR, and sequencing

Original data including 28 DNA sequences were collected from 22 nematode samples belonging to 11 nominal and four
unidentified species. Representatives of the Hoplolaimidae outgroup were selected based on Subbotin et al. (2006) data as *Radopholus* sp. (family Pratylenchidae) and *Meloidodera alni* (family Heteroderidae) (Table 1). Three to fifteen nematode specimens of each sample were transferred to an Eppendorf tube containing 16/afii9839l ddH2O, 2/afii9839l 10£PCR buffer and 2/afii9839l proteinase K (600/afii9839g/ml) (Promega, Benelux, The Netherlands) and crushed during 2 min with a microhomogeniser, Vibro Mixer (Zürich, Switzerland). The tubes were incubated at 65 °C (1 h) and then at 95 °C (15 min). Detailed protocols for PCR, cloning and automated sequencing were as described by Subbotin et al. (2005). The forward D2A (5/H11032CAAGTACCGTGAGGGAAAGTTG-3/H11032) and reverse D3B (5/H11032TCGGAAGGAACCAGCTACTA-3/H11032) primers were used for amplification and sequencing of the fragment of the 28S rRNA gene. The sequences reported here have been deposited in the GenBank database under the accession numbers given in Table 1. Voucher specimens are stored in nematode collections of Istituto per la Protezione delle Piante, Italy, Institut für Nematologie und Wirbeltierkunde, Germany and Instituto de Agricultura Sostenible, Spain.

### 2.2. Secondary structure prediction and sequence alignment

Secondary structure was predicted separately for each molecular segment and for each sequence. For structural modeling, we first reconstructed consensus structure for general model of hoplolaimids and then proceeded to optimization of an individual model for each taxon. This procedure included several steps: (i) each individual sequence was submitted to the Mfold software program Version 3 (http://www.bioinfo.rpi.edu/~zukerm/) (Zuker, 1989) to predict structure using the energy minimization approach; (ii) consensus structure for each helix for a general model was designed considering all optimal and suboptimal foldings obtained from all sequences; (iii) general model of D2 or D3 segments was manually reconstructed assembling helix structures; (iv) complete secondary structure for each sequence was manually reconstructed using a general model and comparative sequence analysis. Structures were visualized using RnaViz (De Rijk et al., 2003) and PseudoViewer 3 (Han et al., 2002) and drawn using vector graphs editor Adobe Illustrator® v10. Variability of sites for the

### Table 1

List of nematode species and populations used in the study

<table>
<thead>
<tr>
<th>Genus and species</th>
<th>Locality</th>
<th>GenBank Accession No.</th>
<th>Collector and/or identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subfamily Rotylenchoidinae Whitehead, 1958</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>H. pseudorobustus</em> (Steiner, 1914) Golden, 1956, type “A”</td>
<td>Germany, Münster</td>
<td>DQ328751</td>
<td>D. Sturhan</td>
</tr>
<tr>
<td><em>H. pseudorobustus</em>, type “B”</td>
<td>China, Beijing</td>
<td>DQ328747, DQ328749</td>
<td>D. Peng, D. Sturhan</td>
</tr>
<tr>
<td><em>H. pseudorobustus</em>, type “B”</td>
<td>California, Fresno, USA</td>
<td>DQ328754</td>
<td>S.A. Subbotin</td>
</tr>
<tr>
<td><em>H. vulgaris</em> Yuen, 1964</td>
<td>Italy, Ancona</td>
<td>DQ328759- DQ328761</td>
<td>J. Tanyi Tambe, S.A. Subbotin</td>
</tr>
<tr>
<td><em>H. digonicus</em> Perry, Darling &amp; Thorne, 1959</td>
<td>Italy</td>
<td>DQ328758</td>
<td>N. Vovlas</td>
</tr>
<tr>
<td><strong>Hoplolaimus</strong> Steiner, 1945</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>R. uniformis</em></td>
<td>Belgium, Poppel</td>
<td>DQ328739, DQ328740</td>
<td>J. Tanyi Tambe, D. Sturhan</td>
</tr>
<tr>
<td><em>R. uniformis</em></td>
<td>Belgium, Bruges</td>
<td>DQ328735, DQ328736</td>
<td>J. Tanyi Tambe, D. Sturhan</td>
</tr>
<tr>
<td><em>R. goodeyi</em> Loof &amp; Oostenbrink, 1958</td>
<td>The Netherlands, Elst</td>
<td>DQ328737</td>
<td>J. Tanyi Tambe, D. Sturhan</td>
</tr>
<tr>
<td><em>R. laurentinus</em> Scognamiglio &amp; Talame, 1973</td>
<td>Italy</td>
<td>DQ328757</td>
<td>N. Vovlas</td>
</tr>
<tr>
<td><em>R. eximius</em> Siddiqi, 1964</td>
<td>Spain, Cádiz</td>
<td>DQ328765</td>
<td>P. Castillo</td>
</tr>
<tr>
<td><em>R. eximius</em></td>
<td>Spain, Huelva</td>
<td>DQ328741</td>
<td>P. Castillo</td>
</tr>
<tr>
<td><em>R. eximius</em></td>
<td>Belgium, Ghent</td>
<td>DQ328743</td>
<td>N. Vovlas</td>
</tr>
<tr>
<td><em>R. eximius</em></td>
<td>Belgium</td>
<td>DQ328742</td>
<td>S.A. Subbotin</td>
</tr>
<tr>
<td><strong>Scutellonema</strong> Andrassy, 1938</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>S. brachyurus</em> (Steiner, 1938) Andrassy, 1958</td>
<td>Italy</td>
<td>DQ328753</td>
<td>N. Vovlas</td>
</tr>
<tr>
<td><strong>Hoplolaimus</strong> Daday, 1905</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>H. seinhorsti</em> (Luc, 1958) Shamsi, 1979</td>
<td>Italy</td>
<td>DQ328752</td>
<td>N. Vovlas</td>
</tr>
<tr>
<td><strong>Peltamigratus</strong> Sher, 1964</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>P. perscitus</em> Doucet, 1980</td>
<td>Spain, Cádiz</td>
<td>DQ328744</td>
<td>P. Castillo</td>
</tr>
<tr>
<td><strong>Outgroup taxa</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Radopholus</em> sp.</td>
<td>Viet Nam</td>
<td>DQ328712</td>
<td>Subbotin et al. (2006)</td>
</tr>
</tbody>
</table>

* For purposes of clarity and consistency, taxonomic categories referred to in this work are those of Siddiqi (2000).
LSU segments was calculated using maximum parsimony with PAUP* 4b4a (Swofford, 2003). Gaps were coded as the fifth characters. The number of changes were mapped on the consensus structural model.

The thirty-one sequences of ingroup and outgroup taxa were aligned using ClustalX1.64 (Chenna et al., 2003) with default parameters for gap opening and gap extension. One sequence for each species was included for alignment and submitted for phylogenetic analysis. Secondary structural sequence alignment was created using several steps: (i) helices were manually aligned using a graph editor by insertion of gaps in one or both strands; (ii) these gaps were also manually inserted into automatic alignment using sequence editor GenDoc 2.5 (Nicholas et al., 1997); (iii) each sequence was converted into dot-bracket format and correctness of insertion/deletion position for each sequence was verified by visually checking each secondary structure model with PseudoViewer 3. Statistics for nucleotide composition in stems and loops, nucleotide pairings, calculation of a consensus sequence, new sequence coding and converting individual sequences into dot-bracket RNA structure format were made using our new program, RNAstat (http://www.nemamex.ucr.edu/rna/stat). D2 and D3 sequence information on secondary structure, in dot-bracket structure, was formatted for all species studied here and was then deposited in the new on-line database NEMrRNA (http://www.nemamex.ucr.edu/rna).

2.3. Phylogenetic analysis of molecular data and sequence coding

Sequence data were analyzed using maximum parsimony (MP), and Bayesian inference (BI). Unweighted and weighted MP analyses were performed with PAUP* 4b4a. Heuristic search settings included: 10 replicates of random taxon addition, tree bisection-reconnection branch swapping, multiple trees retained, no sleepest descent, and accelerated transformation. Dataset from original sequence alignment was submitted for weighted MP analysis, in which transversion were weighted twice as much as transition. Datasets from both the original and converted sequence alignment were submitted by unweighted parsimony. Original sequences were converted to treat stem and loop regions uniformly. We converted unpaired and paired bases of secondary structure into a 28 symbol code, designating one symbol for each of the four unpaired bases and one symbol for each of the 16 bp combinations. We also assigned eight additional numerical symbols to unpaired bases appearing in stems; this approach is similar to that of Smith et al. (2004) in which a 20-letter symbol code was proposed. The code used herein is presented in Fig. 1 with an example showing conversion of the RNA sequence generated by our RNAstat program. Robustness of the clades was assessed using the MP bootstrap analysis yielding a bootstrap percentage (BP) for each node estimated from 1000 replicates.

Bayesian inference analysis (BI) of the dataset was conducted using MrBayes 3.0 (Huelsenbeck and Ronquist, 2001) with two approaches. For the first analysis, a general-time-reversible (GTR) model of nucleotide substitution was used; this included a proportion of invariable sites (I) and a gamma distribution (G) of among-site-rate heterogeneity with six rate categories. This model was estimated by Akaike Information Criterion with ModelTest as the best-fit to the present dataset (Posada and Crandall, 1998). The second analysis was conducted for the secondary structure alignment under the complex model. This included the doublet model with 16 states of nucleotide doublets for the stem region. It also included the standard model of DNA substitution with four nucleotide states for loops and bulges and a gamma distribution (G) of among-site-rate heterogeneity with six rate categories (Ronquist and Huelsenbeck, 2005). All Bayesian analyses were initiated with random starting trees and were run with four chains for $1.0 \times 10^6$ generations.

![RNA conversion code](http://www.nemamex.ucr.edu/rna/stat)

**Fig. 1.** Code for RNA secondary structure and an example of conversion of RNA sequence using RNAstat program.
The Markov chains were sampled at intervals of 100 generations. Two runs were performed for each analysis. The likelihood values of the sample points stabilized after approximately 10^6 generations. After discarding burn-in samples and evaluating convergence, the remaining samples were retained for further analysis. The topologies were used to generate a 50% majority rule consensus tree. Posterior probabilities (PP) are given on appropriate clades.

3. Results

3.1. Sequence variability and secondary structure of the D2 and D3 expansion segments

Length of the whole D2–D3 expansion segments, including junction sequences between these two segments, varied from 555 bp (S. brachyurus) to 566 bp (R. goodeyi and R. laurentinus). Percentage sequence divergences ranged from 0 to 17.8% among ingroup species, to 21.1% among all species. Significant intraspecific variation was detected for H. pseudorobustus. Sequences of the D2–D3 segment from China, USA and Italy were identical, whereas they differed by 14 substitutions (2.5%) from those of H. pseudorobustus from Germany; such sequence variation might indicate sibling species within this putative extant species. We tentatively designate these divergent populations of H. pseudorobustus as type ‘A’ and type ‘B’, respectively. Differences in sequences among populations and clones of R. uniformis were estimated as only in four nucleotides (0.7%). Three sequenced clones of H. vulgaris differed by five nucleotides (0.9%) and two clones of H. multicinctus differed by six nucleotides (1.1%). R. goodeyi and R. laurentinus showed identical sequences.

Predicted consensus secondary structures of D2 and D3 expansion fragments of the LSU rRNA with mapping nucleotide changes for hoplolaimids are shown in Fig. 2. A largely similar folding pattern was observed for all examined hoplolaimids and outgroup species; homology includes folding of the D2 domain into several helices C1–C1/e4 and the D3 includes folds D2–D6 as named by Wuys et al. (2001). For hoplolaimids the D2 and D3 average base composition was as follows: $A = 14.8\%$, $U = 21.9\%$, $C = 26.2\%$, and $G = 37.0\%$; in stems average base pairs were as follows: CG/GC = 29.0/22.7\%, GU/UG = 12.6/10.1\% and AU/UA = 7.3/8.4\%. Several non-canonical base pairs: CU/UC, GA/AG, AA and UA and unpaired nucleotides were also encountered in most variable sites of D2 segments (Figs. 2 and 3).

Variability of the nucleotide sites mapped in the shape of the model is shown in Fig. 2. Several variable areas can be distinguished in D2 and D3 segments; the helix C1/e1 of the D2 segment contained sites with the highest rate of substitution and insertion/deletion events. Most variable regions were on stems rather than on loops, except for a bulge between C1/e3 and C1/e4. Almost 50% of the base pairs in the stem regions, as revealed from the analysis of the consensus structure (Fig. 2), were constant with dominance of GC/GC in 62.5\%, and AU/UA and GU/UG in 17.3\% and 20.2\%, respectively. In most cases nucleotides of base pairs of both strands have approached identical or nearly identical rates of change as expected from the compensatory nature of mutations. There was a strong bias in mutation rates for one of the nucleotides in pairs. More than 70% of base pairs with a constant position for one strand and a variable position for the other occurred for constant G mainly paired with U, C or A, or as GY or GM pairs in the consensus structure (Fig. 2).

Use of the secondary structure information for each sequence allowed reconstruction of an optimal alignment with a relatively precise estimation of homology of nucleotide positions. Using graph editor we were able to align helices introducing insertions in both strands of the RNA molecule (Fig. 3).

3.2. Phylogenetic relationships within Hoplolaimidae

After converting original sequence data to a symbol code, the total number of characters decreased by 1.56 times, whereas the number of informative characters decreased by 1.43 times. A single tree obtained from our original and converted data sets using maximum parsimony had a similar topology, differing only by the position of Peliamigratus, which clustered with Hoplolaimus in a tree obtained from the converted data (Fig. 4). Comparison of these MP results revealed a reduction in the BP values for several clades in the parsimonious tree generated from the converted dataset, but not for Decay index. The influence of a complex model of RNA evolution, considering paired nucleotides in stems and unpaired in loops, for tree topology under BI was more evident than for those under MP. BI under the complex, structure-informed model resulted in a consensus tree with unsolved polytomy for several main hoplolaimid clades, in contrast to trees obtained under the GTR model (Fig. 5). Paraphyly of Helicotylenchus and Rotylenchus, observed in the majority rule consensus tree under the GTR model (Fig. 5A), was not evident under the complex model (Fig. 5B). Notably, under MP (Fig. 4B) using converted data and BI (Fig. 5B) with the complex model, Scutellonema clustered with the clade of H. pseudorobustus and H. multicinctus, suggesting paraphyly of Hoplolaiminae.

4. Discussion

4.1. Universal structure of D2 and D3 segments for nematodes

The secondary structure model for the taxa used in this study is a good fit for a universal model of the D2 and D3 fragments of LSU rRNA (Michot and Buchellerie, 1987; Wuys et al., 2001; Gillespie et al., 2005) for eukaryotic organisms. The structural model has the same helix numbers and similar conserved features, indicating a strong evolutionary functional constraint in the structural pattern
Fig. 2. Variability maps of D2 (A) and D3 (B) expansion fragments of the LSU rRNA superpose on the putative secondary structure with the consensus sequence for hoplolaimids. Variability sites were calculated using maximum parsimony analysis with gap coded as the fifth character. Variability sites are divided into five categories: 0, constant; 1, one change; 2, two changes; 3, three changes; 4, four or more changes. Lower case letters indicate deletion/insertion events. The standard ambiguity codes for nucleotides are used.
Fig. 3. Putative aligned structures for C1/e helices of the D2 segment for hoplolaimid and outgroup species. Nucleotides in which substitutions have occurred are indicated in gray ovals, as are insertions/deletions.

Fig. 4. A single maximum unweighted parsimony tree obtained from original (A) and converted (B) sequence alignments (Tree statistics: (A) Tree length = 569, parsimony informative characters = 148, CI excluding uninformative characters = 0.5066, HI excluding uninformative characters = 0.4934. (B) Tree length = 504, parsimony informative characters = 103, CI excluding uninformative characters = 0.6304, HI excluding uninformative characters = 0.3966). Bootstrap values more than 50% obtained for unweighted and weighted parsimony analyses are given for appropriate clades. Decay index less than six are given for unweighted parsimony tree in brackets.

of this part of the rRNA molecule. Only one structural component of our model differs from some previously described models for some other nematodes. The C1/4e helix in Caenorhabditis elegans (Michot and Bachellerie, 1987) and Labiostrongylus bipapillosus (Chilton et al., 2003) was reported as drastically truncated and composing 14–8 nucleotides, whereas in longidorids (He et al., 2005), ricketsi nematids (Subbotin et al., 2005) and hoplolaimids (present study) this helix is significantly longer and contains more than 100 nucleotides.

One of the main problems with utilizing secondary structure is rather low accuracy for its prediction from sequence data. Although several approaches for reconstruction of secondary structure have been proposed (Knudsen and Hein, 1999), the most efficient algorithm of folding is based on the energy minimization approach considering many
thermodynamic parameters. This algorithm does not consider functional constraints of the RNA molecule and often shows a low percentage of prediction accuracy for large molecules. Comparative sequence analyses with mapping of compensatory mutations across a large taxonomic sample can serve as an efficient guide to determine structural components for most ambiguous regions. However, this work requires careful manual estimation of each position which cannot be always objective. In the present study we applied such a stepwise approach for reconstruction of this part of the molecule combining an energy minimization approach for separate prediction of helix structures and comparative sequence analyses for reconstruction of consensus structure. Future experimental research and more comprehensive analyses will allow optimizing proposed models.

Several ways of measuring the variability site of rRNA molecules have been proposed and applied (Van de Peer et al., 1996). This study is the first to present the variability map for D2 and D3 segments of the LSU for nematodes as deduced by maximum parsimony analysis. It has an advantage over other methods, because it allows considering insertion/deletion events for calculation. Detailed information about the variability and conservation of the rRNA molecule is important for understanding fundamental questions of evolution of the main component of ribosomes as well as applications of molecular diagnostics and phylogeny. Specifically, this information is useful for recognizing taxon-specific sequence signatures and designing diagnostic probes. In the case of phylogenetic studies this information is important for construction of precise alignments and selection of the most sophisticated and appropriate evolution model.

4.2. rRNA structural database

The rapidly growing body of rRNA sequence data provides a resource from which biologists can infer evolutionary histories of organisms. Today many phylogenetic and comparative biology studies use rRNA sequence data obtained from public databases. Increasing recognition of the value of secondary structure models of RNA molecules for alignment and phylogenetic procedures raises the problem of broad access to such structural information for a wide range of taxa (Wuyts et al., 2004; Schultz et al., 2006). To overcome this problem we developed a specialized database NEMrRNA (http://www.nemamex.ucr.edu/rna) for storing rRNA structural information for nematodes. The present version of the database, which is maintained by our research group, contains information on secondary structures of the D2, D3 of the 28 rRNA and the ITS2 rRNA. Sequences in the database can be obtained via the query interface searching for species name, genus name, or GenBank accession number.

4.3. New method of sequence coding and its application for phylogenetic study

Although knowledge of the secondary structure allows applying more sophisticated models of nucleotide substitution by BI and theoretically gives a more realistic picture of relationships, it is nevertheless useful to compare such findings with simpler approaches including parsimony. A new coding method for nucleotide base pair developed here, is similar to that of Smith et al. (2004). However, it differs by a 28 symbol code (versus a 20 letter symbol code) including eight numerical symbols for unpaired bases appearing in each strand, and specifically in the most variable sites of molecule. This new method of coding secondary structure allows application of the maximum parsimony method for analysis of secondary structure and thus, comparison of trees derived from MP and BI analyses using similar data sets. As anticipated, the MP and BI trees obtained using converted data and a complex
model that considers paired nucleotides in stems and unpaired in loops, respectively, showed reduced statistical support compared to traditional models. This loss of resolution is predicted by consideration of a reduced number of independently evolving characters.

4.4. Phylogenetic reconstruction of hoplolaimid evolution

Application of the secondary structure model of rRNA for phylogenetic analyses of hoplolaimids led to trees with lower resolved relationships between clades and likely eliminated some artefactual support for misinterpreted relationships, such as paraphyly of *Helicotylenchus* or *Rotylenchus*. Paraphyly of these taxa derived from the models considering secondary structure are less evident and the results are more consistent with traditional morphologically-based views of an evolutionary framework including these genera.

Geraert (1990) pointed out that genera united in the subfamily Hoplolaiminae could have divergent phylogenetic origins. These nematodes share large scutellum-like phasmids (a pair of complex sensory organs), which in *Scutellonema* are located near the anus as in Rotylenchoidinae whereas in other genera scutella are more anteriorly positioned as pre- or postvulval. Geraert (1990) also speculated that the subfamily may reflect a morphological transformation series of small and pore-like phasmids as in Rotylenchoidinae and some *Scutellonema* species, to progressive enlargement and more anterior placement as expressed in *Peltamigratus* and *Aorolaimus*. Geraert further speculated that evolution of *Hoplolaimus* could have involved overall body enlargement from a *Peltamigratus*-like ancestor. This speculation is supported by trees obtained using secondary structure information; in the MP tree *Hoplolaimus* and *Peltamigratus* clustered together. The position of *Scutellonema* clustered with *Helicotylenchus* in these trees, suggests that the relatively posterior position of phasmids is more informative of phylogenetic relationships, than is the size of phasmids. This observation further suggests convergent origin of enlarged phasmids. Although it suggests possible polyphyly of the Hoplolaiminae within Hoplolaimidae, in some trees derived from a larger data set for Tylenchida the genera of Hoplolaiminae clustered together (Subbotin et al., 2006), and thus the matter remains unresolved pending further investigation.

Species groupings for *Helicotylenchus* and *Rotylenchus* raise additional important hypotheses for future testing. Surprisingly, molecular analysis revealed almost identical sequences for *R. goodeyi* and *R. laurentinus* suggesting that they are very closely related or co-specific. This grouping clearly contradicts morphological differences with respect to numbers of lip annuli, cuticular patterns, stylet length, position of the dorsal gland orifice and vulva position (Castillo and Vovlas, 2005). Although our data might indicate that some classical characters, such as longitudinal striations, are not phylogenetically informative in this context, these hypotheses must be further tested including molecular analysis of multiple geographically divergent populations of these species.

The present study strongly supports a unique shared lineage of *H. pseudorobustus* and *H. multicinctus* and this is strengthened by additional phylogenetic analyses including several populations of both species (Amiri, personal com.). Nevertheless these species are described as having significantly different morphologies with respect to tail shape, stylet knob morphology and morphometrics of the body and stylet. Such unexpected clustering of morphologically divergent species raises interesting questions of incongruence including relative rates of morphological and molecular evolution in hoplolaimids.

Although, incorporating secondary structure information allows improved estimates of phylogeny, relationships among several main hoplolaimid clades remain poorly resolved. Greater phylogenetic resolution will require sequencing additional taxa, rRNA fragments and genes. The present study of the D2 and D3 expansion regions of the 28S rRNA gene thus provides initial insight toward resolving phylogenetic relationships among hoplolaimids, while still pointing to key areas for further testing. To obtain a more robust perspective, it is necessary to analyze a larger number of species including some rare genera, for which the phylogenetic position remains controversial. Furthermore, more intensive sampling of hoplolaimids from different regions should be implemented, thereby would improve our knowledge about patterns of hoplolaimid relationships among species.

The present work uses hoplolaimids as a model to emphasize the importance of establishing the rRNA secondary structure model for phylogenetic studies in nematodes; the result is sequence alignment with greater confidence and selection of increasingly appropriate models of evolution. This approach can be easily generalized to additional groups of nematodes as the proposed structure provides guidelines for aligning new sequences from related taxa and for selecting evolution models for reconstruction of deeper-level nematode phylogenies.

Acknowledgment

The authors acknowledge support from the US National Science Foundation Grant DEB 0228692, USDA Grant 2005-00903, and Grant RFBR 04-04-48051.

References


