



ITS2 secondary structure and phylogeny of cyst-forming nematodes of the genus *Heterodera* (Tylenchida: Heteroderidae)

Hongwei Ma^a, Robin M. Overstreet^{a,*}, Sergei A. Subbotin^{b,c}

^aThe University of Southern Mississippi, 703 East Beach Drive, Ocean Springs, MS 39564, USA

^bPlant Pest Diagnostics Center, California Department of Food and Agriculture, 3294 Meadowview Road, Sacramento, CA 95832-1448, USA

^cCentre of Parasitology of A.N. Severtsov Institute of Ecology and Evolution of the Russian Academy of Sciences, Leninskii Prospect 33, Moscow 117071, Russia

Received 14 February 2007; accepted 18 September 2007

Abstract

The internal transcribed spacer 2 (ITS2) of the ribosomal RNA gene is a double-edged tool for eukaryote evolutionary comparison. In this paper, we re-evaluate the putative ITS2 secondary structures proposed for 29 species of cyst-forming nematodes in the genus *Heterodera* and present optimized variants. The models are organized in four main helices emerging from a central core. Helix III is always the longest and has two branches exhibited as a Y-shaped structure near the apex. In most species, helices III and IV have merged to form a segment of the central core. Two conservative motifs, GAUCR and UURC, were found in helices II and III, respectively. Using the MARNA program taking into consideration both the primary sequence and the secondary structure, we generated an optimal alignment for *Heterodera* sequences. The alignment was analyzed by Bayesian inference under a general-time-reversible model and the complex models, and by a maximum parsimony approach using original data and sequence data converted according to secondary-structure information. The consensus ITS2 secondary structure model for Heteroderidae can serve as a useful guide for reconstruction of optimal alignment and can be used as an additional source of data incorporating structural parameters of molecules for the study of nematode phylogeny. Application of the secondary ITS2 structure data allows a more resolved and realistic picture of relationships within Heteroderidae.

© 2008 Gesellschaft für Biologische Systematik. Published by Elsevier GmbH. All rights reserved.

Keywords: Cyst-forming nematode; *Heterodera*; Homology modeling; Internal transcribed spacer 2; Phylogeny; Secondary structure

Introduction

The internal transcribed spacer 2 (ITS2) of the ribosomal RNA gene is a broadly used standard marker for the phylogenetic study of many organisms at low taxonomic levels (Lott et al. 1998; Gómez-Zurita et al. 2000; Murrell et al. 2001; Hlinka et al. 2002; Chen et al.

2004). Because of the assumed conservation in secondary structure, it is also referred to as a double-edged tool for eukaryote evolutionary comparison (Mai and Coleman 1997; Morgan and Blair 1998; Joseph et al. 1999; Michot et al. 1999; Suh and Kim 2002; Alvarez and Wendel 2003; Coleman 2003; Goertzen et al. 2003; Young and Coleman 2004). Although the biological role of the ITS spacers is not well understood, the utilization of yeast models has definitely shown their importance for production of the mature rRNA. Among the

*Corresponding author. Tel.: +1 228 8724243; fax: +1 228 8724204.
E-mail address: Robin.Overstreet@usm.edu (R.M. Overstreet).

approximately 140,000 ITS2 sequences presently in GenBank, there are more than 25,000 exhibiting several common features of secondary structure (Schultz et al. 2005; Wolf et al. 2005). For many organisms, the ITS2 in premature rRNA is organized around a preserved central core of secondary structure from which four helices emerge (Coleman 2003). Knowledge of RNA secondary structure is becoming increasingly important in assisting phylogenetic analysis by three main approaches: (i) finding the best homology between nucleotides during alignment procedures and generation of optimal alignment, (ii) determining the source of additional data incorporating both structural and morphometric parameters of rRNA molecules, and (iii) selecting an increasingly appropriate model of rRNA evolution (Subbotin et al. 2007). A recently designed structural ITS2 database (Wolf et al. 2005) contains more than 25,000 pre-calculated structures for the currently known ITS2 sequences and provides new possibilities for incorporating structural information in phylogenetic studies (Schultz et al. 2005; Wolf et al. 2005).

The plant-parasitic cyst-forming nematodes of the genus *Heterodera* Schmidt, 1871 (Heteroderidae, Tylenchida) comprise more than 60 species parasitizing herbaceous and woody plants. Some species are considered as highly pathogenic and as important nematode pests to cereals, legumes, sugar beet, cabbage, carrot, and other crops (Baldwin and Mundo-Ocampo 1991). The species of *Heterodera* are classified in several morphologically defined groups: the *H. avenae*, *goettingiana*, *humuli*, *schachtii*, *cyperi*, and *sacchari* groups (Baldwin and Mundo-Ocampo 1991; Subbotin et al. 2001). Phylogenetic relationships among groups and species have been studied intensively using the ITS-rRNA sequences (Ferris et al. 1993; Sabo et al. 2001; Subbotin et al. 2001; Tanha Maafi et al. 2003; Madani et al. 2004). However, none of these studies has considered structural components of the ITS-rRNA molecule. Since the structural ITS2 database contains several *Heterodera* ITS2 secondary structures and templates, re-analysis of the phylogeny of cyst-forming nematodes of the genus *Heterodera* using secondary-structure information seems useful.

In this paper we re-evaluate the putative ITS2 secondary structures proposed for some cyst-forming nematodes and deposited in the structural ITS2 database (Schultz et al. 2005; Wolf et al. 2005), and we present optimized variants. Using the MARNAs program (Siebert and Backofen 2005), which created a multiple alignment based on primary sequences and secondary-structure information, we generated optimal alignment for cyst-forming nematode sequences. We analyzed the alignment using Bayesian inference both under the complex model with 16 states of nucleotide doublets for the stem region and under the standard

model for DNA substitution with nucleotide states for loops and bulges. We also tested the previously proposed coding method (Subbotin et al. 2007) for converting unpaired and paired bases of secondary structure into 28 symbol codes. Original and converted sequence data were analyzed by maximum parsimony. Finally, we discuss phylogeny of cyst-forming nematodes as inferred from these methods as well as problems with using secondary-structure information for phylogenetic reconstructions.

Material and methods

Secondary-structure prediction and sequence alignment

The ITS2 sequences of 29 cyst-forming nematode species (Subbotin et al. 2001, 2003; Tanha Maafi et al. 2003; Madani et al. 2004) and of *Cryphodera brinkmani* Karssen & Van Aelst, 1999 selected as an outgroup taxon (Subbotin et al. 2001) were obtained from GenBank. Secondary structure was predicted separately for each molecular helix and for each sequence. For some sequences, the 5.8S and 28S rRNA flanking regions as conservative fragments were added to improve folding. For structural modeling, we first reconstructed a consensus structure for a general model of each nematode group, then proceeded to optimization of an individual model for each taxon. Thermodynamic criteria and covariation analyses have been applied for reconstruction of secondary-structure models. This procedure included several steps (Subbotin et al. 2007): (i) the Mfold software version 3 (Zuker 2003; <http://www.bioinfo.rpi.edu/~zukerm/>) was applied to predict the secondary structure for whole ITS2 molecules and their helices using the energy minimization approach, (ii) for a general model, a consensus structure for each helix was designed considering all optimal and suboptimal foldings obtained from all sequences, (iii) a general model of the ITS2 was manually reconstructed assembling helix structures, (iv) a complete secondary structure for each sequence was manually reconstructed or improved 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 Photoshop. The secondary structures predicted by the above-mentioned approach were compared with those predicted by homology modeling for corresponding species from the ITS2 database (Wolf et al. 2005; <http://its2.bioapps.biozentrum.uni-wuerzburg.de>).

For reconstruction of consensus secondary structures for the *H. avenae* and *humuli* groups, 11 and 14 sequences were used, respectively. Accession numbers

of these sequences are listed in the caption to Fig. 1. The ITS2 consensus structure for Heteroderidae was generated from all sequences used in the present study. Sequence statistics and consensus sequence for nematode groups were created using RNAsat (<http://www.nemamex.ucr.edu/rna/stat>) (Subbotin et al. 2007). The ITS2 sequence in dot-bracket structural format was formatted for all species studied here, then deposited in

the online database NEMrRNA (<http://www.nemamex.ucr.edu/rna>) (Subbotin et al. 2007).

The ITS2 sequences with secondary-structure format were aligned using the MARNA web server (Siebert and Backofen 2005; <http://biwww2.informatik.uni-freiburg.de/Software/MARNA/index.html>) based on both the primary and secondary structures. As the default setting, base deletion was scored 2.0, base mismatch 1.0, arc

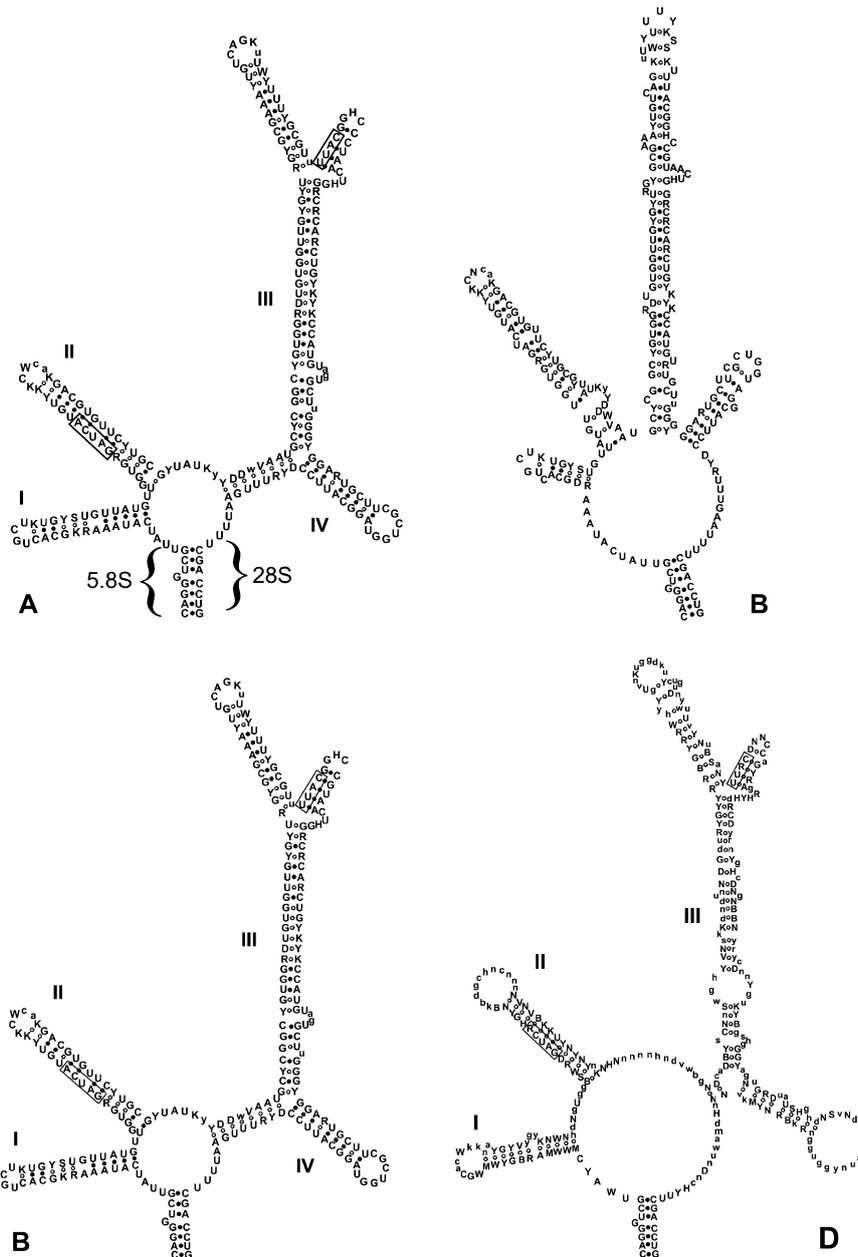


Fig. 1. Consensus putative secondary structure of the ITS2 region: (A) *Heterodera avenae* group (reconstructed with sequences AF274396, AF274403, AY148368, AY148380, AY148389, AY148394, AY148402, AY148407, AY347923, AY347925), predicted in present study; (B) *H. avenae* group, predicted by homology modeling approach (ITS2 database); (C) *H. humuli* group (AF274407–AF274410, AF393841, AF399384, AF498384, AF498385, AF498386, AY045755, AY347926, AY347927, DQ846902), predicted in present study; (D) Heteroderinae (all sequences as indicated in trees, see Fig. 3), predicted in present study. Lower-case letters indicate deletion/insertion events. Standard nucleotide ambiguity codes are used.

removing 2.0, arc breaking 1.5, and mismatch 1.8 with ensemble of shaped structures. The alignment is available from the senior author upon request.

Phylogenetic analysis

Sequence alignment was analyzed using maximum parsimony (MP) and Bayesian inference (BI). Using secondary-structure information, we converted unpaired and paired bases of each ITS2 sequence into 28 symbol codes as described by Subbotin et al. (2007) using the RNAsat program. Datasets from both the original and converted sequence alignment were analyzed by unweighted parsimony. The MP analysis was performed with PAUP* 4b4a (Swofford 2003). Heuristic search settings included 100 replicates of random taxon addition, tree bisection-reconnection branch swapping, multiple trees retained, no steepest descent, and accelerated transformation. Robustness of the clades was assessed using MP bootstrap analysis yielding a bootstrap percentage for each node estimated from 1000 replicates.

The BI analysis of the dataset was conducted using MrBayes 3.0 (Huelsenbeck and Ronquist 2001) under the general-time-reversible (GTR) model of nucleotide substitution and the complex model. The complex model 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). Bayesian analysis was initiated with random starting trees and was run with four chains for 1.0×10^6 generations. The Markov chains were sampled at intervals of 100 generations. The log-likelihood values of the sample points stabilized after approximately 10^3 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) were calculated for appropriate clades.

Results

Putative secondary structure of the ITS2 for cyst-forming nematodes

Using the energy minimization approach and comparative sequence analysis, we predicted the secondary structure of the ITS2 for 29 species of cyst-forming nematodes and one outgroup species (*Cryphodera brinkmani*). Consensus ITS2 secondary structures were reconstructed for the *H. avenae* (Fig. 1A) and *humuli*

(Fig. 1C) groups. The models were organized in four main helices emerging from the central core. The third helix was always the longest and had two branches exhibited as a Y-shaped structure near the apex. In most species, helices III and IV merged to form a segment (S) of the central core. The segment varied from 2 to 8 nucleotide pairs in most species; however, it was absent in *H. turcomanica*. For *H. cajani*, two alternative foldings could occur, one with and another without the segment (Fig. 2). The majority (50%) consensus secondary structure for cyst-forming nematodes and *Cryphodera* species used in this study is presented in Fig. 1D. There are two conservative motifs found, GAUCR and UURC in helices II and III, respectively (Figs. 1, 3). The GAUCR motif of helix II in our models shows the same U–U mismatch as helix II structures obtained from the ITS2 database (Wolf et al. 2005).

The nucleotide lengths of *Heterodera* ITS2 in this study ranged from 207 to 235 bp, with 236 bp for *Cryphodera*. The average base composition was as follows: A = 15.9%, U = 33.5%, C = 21.4%, and G = 29.2%; for loops: A = 14.6%, U = 45.4%, C = 18.8%, and G = 21.2%; and for stems: A = 21.0%, U = 25.8%, C = 18.8%, and G = 34.4%. Approximately 70–80% of these nucleotides were involved in stem formation in the proposed structures. In stem regions, average base pairs were as follows: CG/GC = 17.3/23.6%, GU/UG = 8.7/9.4%, and AU/UA = 17.3/10.7%. The number of pairing bases in helices I and III and the total number of pairing bases in the four helices were higher in our models than in the ITS2 database; the number of pairing bases in helix II was lower in our models than in the ITS2 database; concerning the pairing bases in helix IV, there was no significant difference between the two models (Table 1, Fig. 4). We also found one to four non-Watson base pairs (such as AA, AC, AG, UU, GG, and CU) in the stem connecting the central core to helices III and IV. The non-Watson base pairs were also found in helices I, II, III, and IV, but they were not common in the S region. Taxon-specific structural variations in the helices have been not observed among studied sequences.

Proposed putative structures for cyst-forming nematodes (Fig. 1B) were different from those reconstructed and deposited in the ITS2 database (Schultz et al. 2005; Wolf et al. 2005). Our consensus structures for the *H. humuli* and *avenae* groups involved 16 and 19% more nucleotides in the helices, respectively, than the corresponding structures proposed by Wolf et al. (2005). Only about 50% of those nucleotide pairs were identical in both structures. Significant differences were observed in helix III. The helix III generated by homology modeling (Wolf et al. 2005) does not contain a Y-shaped structure; in our study it constantly does if the energy minimization approach was applied.

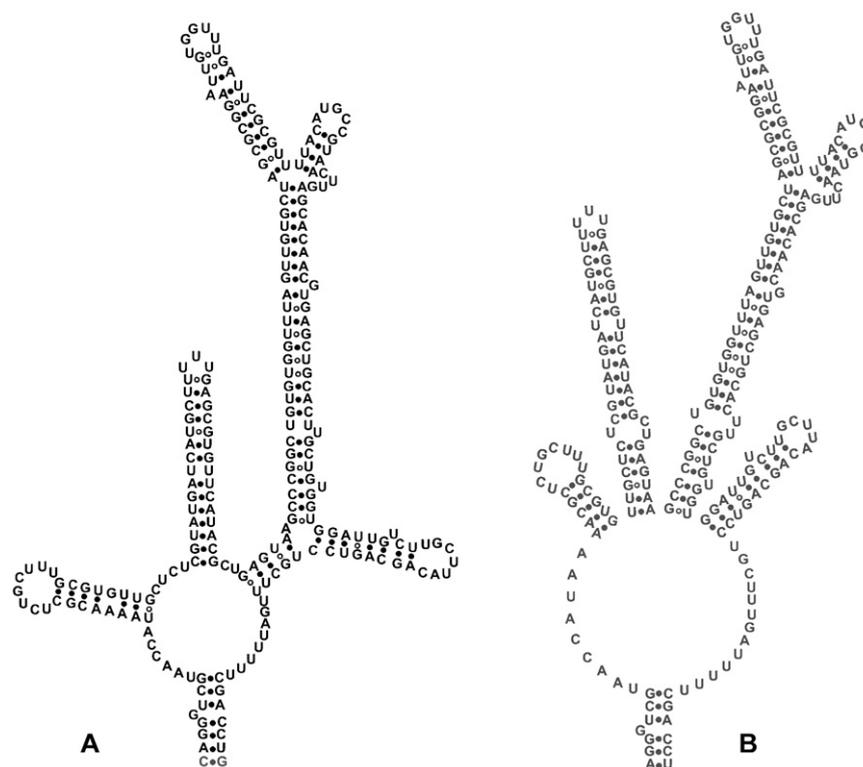


Fig. 2. Two alternative foldings of the ITS2 for *Heterodera cajani* predicted by thermodynamic criteria and covariation analysis.

Phylogenetic relationships as inferred from maximum parsimony and Bayesian analysis

Strict consensus maximum parsimonious trees obtained from original sequence data and converted data are presented in Fig. 5. Relationships among the main cyst-forming nematode groups were not well resolved. MP analysis of converted sequences revealed that the number of parsimony informative characters was 1.3 times lower and led to altered bootstrap support for some relationships: higher support for clustering the *H. avenae* and *sacchari* groups (53% vs. 79%) and lower support for clustering the *H. humuli* group with *Heterodera salixophila* (80% vs. 49%). Changed positions for *H. cyperi* and *H. cajani* were also revealed; however, their relationships with other nematodes were left unresolved.

The influence of a complex model of RNA evolution was also evident when considering paired nucleotides in stems and unpaired nucleotides in loops for tree topology with BI. The BI under the complex, structure-informed model produced a consensus tree with unresolved polytomy for several cyst-forming nematode groups (Fig. 6B), in contrast with the tree obtained under the GTR model (Fig. 6A). In both BI trees, the *H. avenae* and *humuli* groups were sister taxa. The type of model applied had minimal influence on the relationships within nematode groups.

Ten species of the *H. avenae* group (bifenestrated cyst with short vulval slit; parasites of Poaceae plants) formed a moderately or highly supported clade under different methods. Species with a wide vulval bridge, *H. latipons* and *H. hordecalis*, always occupied a basal position within this clade (Figs. 5, 6).

The *H. humuli* group (bifenestrated cyst, except for *H. fici* having ambifenestrated cyst, with long vulval slit; parasites of dicotyledonous plants) united six species in a moderately or highly supported clade. The BI and MP trees differed in the position of *H. turcomanica*. In BI trees, *H. fici* (ambifenestrated cyst) was always in a basal position within this clade, indicating that the bifenestrated cyst might be a derived characteristic in cyst-forming nematodes.

The relationships among four species of the *H. schachtii* group (ambifenestrated cyst with distinct bullae and underbridge; parasites of dicotyledonous plants) were not resolved in all trees. Sister relationship of these species with *H. cajani*, a subtropical and tropical representative of the *H. schachtii* group, was revealed by BI analysis under the complex model with low support.

Heterodera bifenestrata (bifenestrated cyst) clustered with high support in all trees with *H. cynodontis* (ambifenestrated cyst). Both species have a second stage juvenile with three incisures and are parasites of grasses. Two species (*H. goettingiana* and *H. scutellaria*) of the *H. goettingiana* group are in a basal position within *Heterodera* in BI trees (Fig. 6).

Table 1. Comparison of secondary structures in *Heterodera* in present study with that from ITS2 database

Characteristics	Min.	Max.	Mean	SD	CV	<i>n</i>
Length of sequence (bp)	207	235	215	6.5	3.0	29
G + C content (%)	52.3	71.4	63.1	4.9	7.7	29
No. of base pairs in helix I in present study ^{a,b}	4	11	9	1.6	17.1	29
No. of base pairs in helix I in ITS2 database ^{a,b}	5	7	6	0.5	7.7	29
No. of base pairs in helix II in present study ^{a,b}	8	16	13	1.3	9.8	29
No. of base pairs in helix II in ITS2 database ^{a,b}	15	19	17	1.1	6.6	29
No. of base pairs in helix III in present study ^{a,b}	25	42	39	3.0	7.6	29
No. of base pairs in helix III in ITS2 database ^{a,b}	33	40	35	1.9	5.5	29
No. of base pairs in helix IV in present study ^a	7	14	9	1.1	12.2	29
No. of base pairs in helix IV in ITS2 database ^a	7	14	9	1.2	13.8	29
No. of base pairs connecting central core and merged helices III and IV in present study ^{a,b}	0	8	4	2.4	59.4	29
No. of base pairs connecting central core and merged helices III and IV in ITS2 database ^{a,b}	0	0	0	0.0	0.0	29
Total base pairs in four helices in present study ^{a,b}	56	82	76	5.6	7.4	29
Total base pairs in four helices in ITS2 database ^{a,b}	62	74	67	3.0	4.5	29
Percentage of total paired bases in present study ^{a,b}	53.6	76.3	70.3	4.9	7.0	29
Percentage of total paired bases in ITS2 database ^{a,b}	57.0	68.6	62.5	2.8	4.5	29

CV = coefficient of variation; Max. = maximum; Mean = arithmetic mean; Min. = minimum; SD = standard deviation.

^aOnly canonical base pairs (GC, GU, AU) counted.

^bStatistically different between two models.

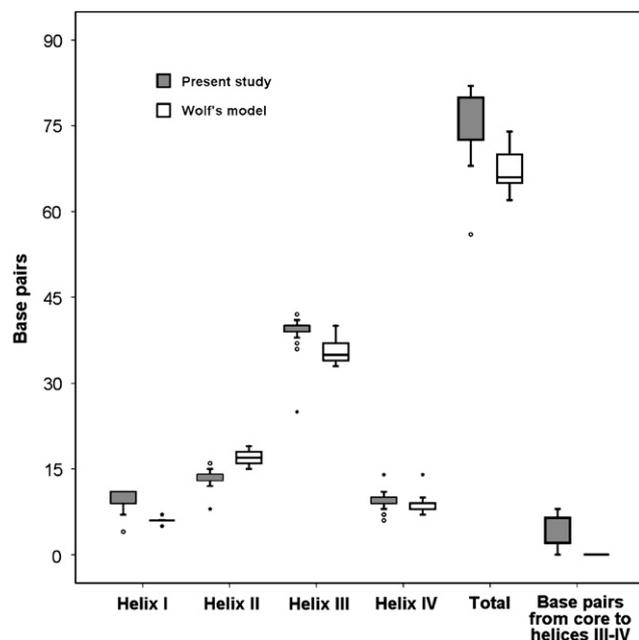


Fig. 4. Box plots of the helix-wise and total lengths of pairing base distribution of the predicted ITS2 structures in *Heterodera* predicted by the present study and the homology modeling approach (ITS2 database), respectively.

Discussion

Prediction of the ITS2 secondary structures for cyst-forming nematodes

The ITS2 sequence has been shown to vary greatly in length among different organisms: from absent or very

short in amitochondrial protozoa (Katiyar et al. 1995) to reaching 2212 bp in the fungus *Plasmopara halstedii* (e.g. Thines et al. 2005). In the ITS2, some microsatellite repeats such as (GA)_n, (GT)_n, (AG)_n, (CA)_n, (CCAT)_n, or (AGCA)_n have been reported from corals (Chen et al. 2004), nematodes (Conole et al. 2001), and beetles (Gómez-Zurita et al. 2000). Nucleotide repeats as long as 82–109 bp have been found in ticks (Barker 1998; Murrell et al. 2001). Although ITS2 sequences show high rates of size and sequence variation in the evolution of eukaryotes, the secondary structure of ITS2 exhibits several common features. The ITS2 is organized around a preserved central core from which four helices emerge. Four domain models of the ITS2 have been observed in fungi (Joseph et al. 1999), green algae (Mai and Coleman 1997), flowering plants (Hershkovitz and Zimmer 1996; Mai and Coleman 1997), monogeneans (Morgan and Blair 1998), *Drosophila* (Young and Coleman 2004), and mammals (Joseph et al. 1999; Michot et al. 1999). In nematodes, the ITS2 secondary structure was predicted as a four-helix structure for trichostongyloids (Chilton et al. 2001). However, the four-domain structure is not the only predicted model. Among the presently available 140,000 ITS2 sequences in GenBank, only about 17.9% show the typical four-domain structure (Wolf et al. 2005). Using the energy minimization approach and comparative sequence analysis, Subbotin et al. (2005) predicted the secondary structure of ITS2 for plant-parasitic ditylenchids, and found that it was displayed as a three-helix structure. The three-domain structure has been observed also in nematodes of the subfamily Anguininae and in several fungi-feeding *Ditylenchus* species (Subbotin,

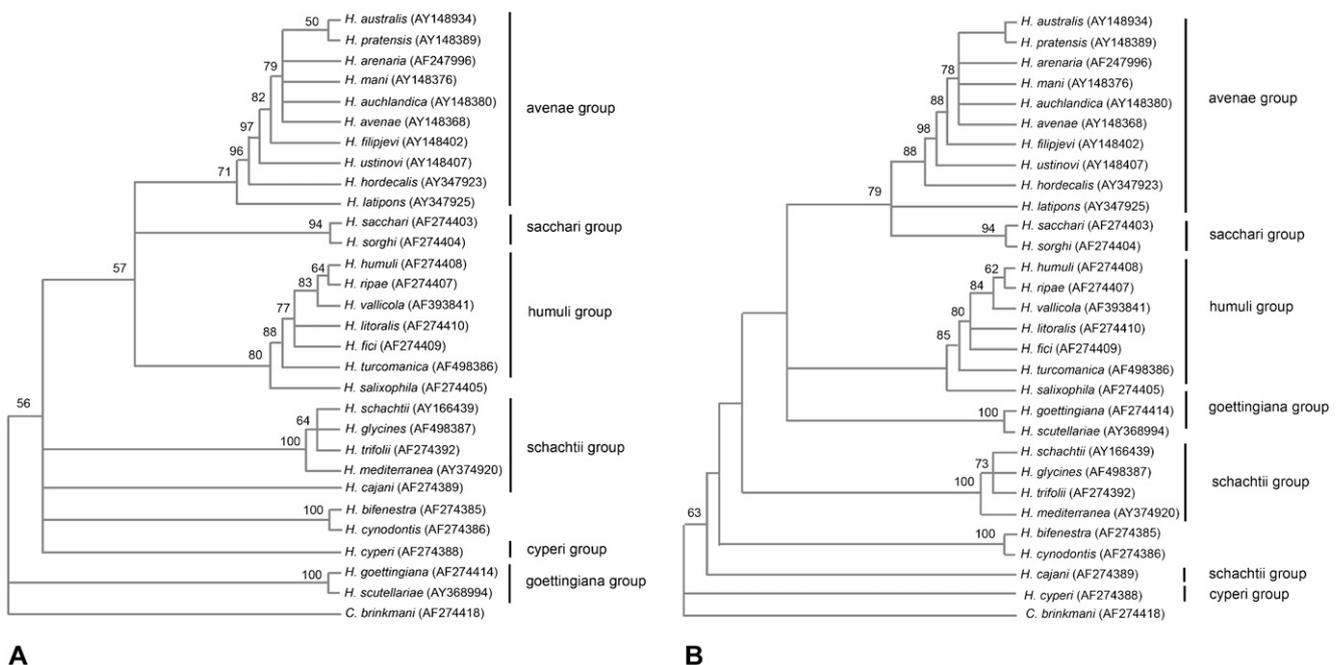


Fig. 5. Consensus maximum parsimony tree obtained from original (A) and converted (B) sequence alignments. Statistics tree A: number of trees = 20; tree length = 596; CI (excluding uninformative characters) = 0.5405; HI (excluding uninformative characters) = 0.4595; RI = 0.7123; RC = 0.4207; number of characters = 309; number of parsimony informative characters = 150. Statistics tree B: number of trees = 14; tree length = 560; CI (excluding uninformative characters) = 0.6323; HI (excluding uninformative characters) = 0.3677; RI = 0.7390; RC = 0.4896; number of characters = 224; number of parsimony-informative characters = 116. Bootstrap values greater than 50% are given for appropriate clades.

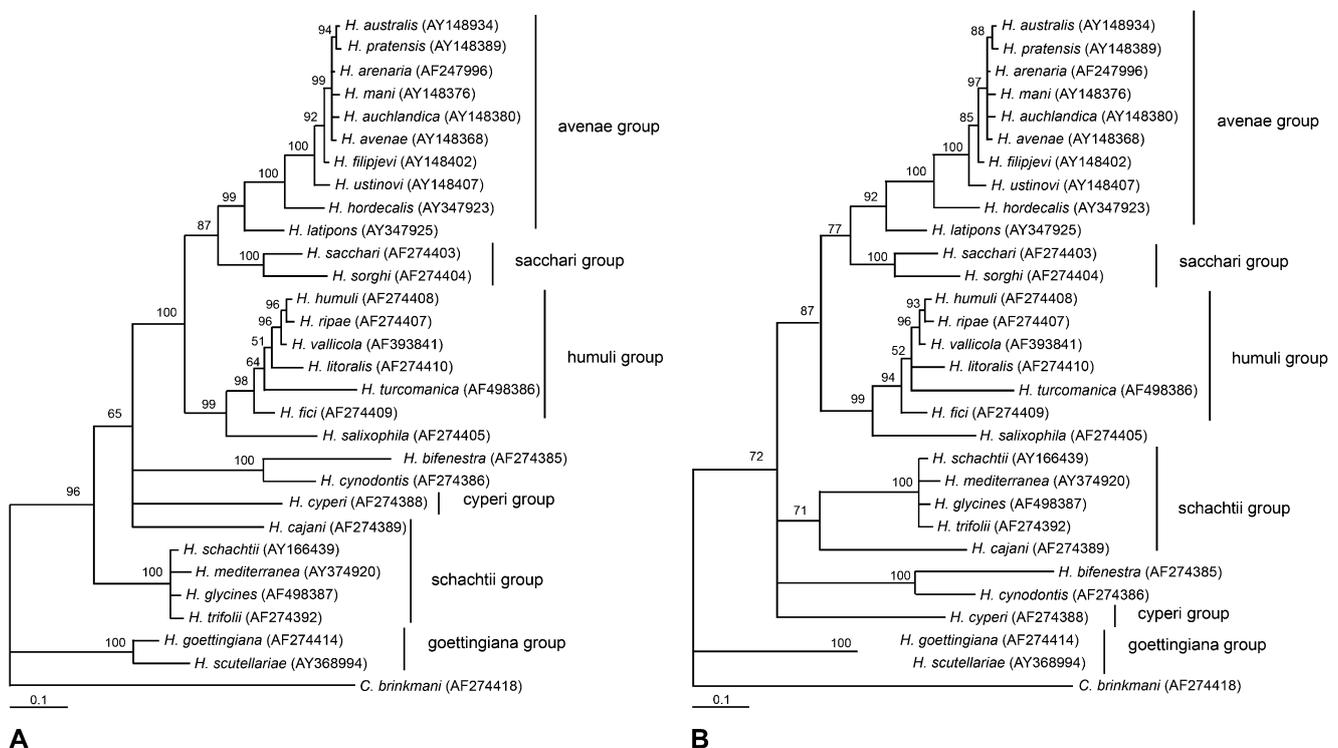


Fig. 6. The 50% majority rule consensus trees from Bayesian analysis generated with (A) the GTR+I+G model and (B) the complex model: 4 × 4 model for loops and 16-doublet model for stems of the secondary structure. Posterior probabilities greater than 50% are given for appropriate clades.

unpublished). The present senior author (Ma et al., unpublished), using the Mfold, Sfold (Ding et al. 2004) and RNAfold (Hofacker 2003) web servers, predicted the ITS2 structure for some ascaridoid nematodes of the genus *Contraecaecum* parasitizing fish and fish-eating birds, and has shown that the ITS2 sequence, at an average length of 300 bp, displays a five-helix structure. Thus, the ITS2 shows more structural variation in nematodes than in any other group of organisms studied so far.

Using the homology modeling approach, Wolf et al. (2005) predicted secondary structures of the ITS2-rRNA for 23 species (97 sequences) of *Heterodera*. Comparison of these structures deposited in the ITS2 database with the ones we obtained revealed some level of similarity in general structure. Both structural models display the same number of helices. However, helix lengths and the structures of helix III differed. In our model, helix I was longer and helix II shorter than in Wolf's model, and helix III formed a Y-shaped structure near the apex. A similar structure has been observed also in helix III of the ITS2 predicted for species of *Rotylenchus*, *Helicotylenchus*, *Hoplolaimus*, *Tylenchorhynchus*, and *Hirschmanniella* (Subbotin, unpublished). The conserved motif GAUCR–helix II is orientated in the same position as the U–U mismatch of helix II in the ITS2 database (Schultz et al. 2005; Wolf et al. 2005). This may mean that the GAUCR or U–U mismatch motif is important for structural stability or the function of ITS2. The most important finding was that the helices in Wolf's and our models exhibited only about 50% identity in base pairs. Interestingly, the stem connecting the central core and helices III–IV has 0–4 non-Watson base pairs, which may mean that this is not a stable but a rather sophisticated, dynamic motif. In certain life stages or environments, this motif may change its structure to be consistent with its functions. The secondary structure near the central core is the most important domain in homology modeling of the ITS2 database. To predict conserved homology modeling, one requires refinements such as closing free pairings, opening helix loops, cutting ends, and adding terminal sequences (Wolf et al. 2005).

Although selection of the correct model is a difficult task and is not possible without experimental confirmation, comparative analyses of such characteristics as energetic stability of the predicted structures, the numbers and distribution of compensatory and semi-compensatory base pair changes and indels, as well as consistency of structural variations along the sequence or alignment, could provide clues for preferences or justifications of the structural correctness of one model over another. In the present study, we propose improved models for the ITS2 for cyst-forming nematodes of the genus *Heterodera* using combinations of several approaches and optimization procedures. However, in

several cases such as the cyst nematodes of the *H. cyperi* group with highly divergent sequences, we were not able to reconstruct their ITS2 and include them in the present study. The problem of selecting the correct model for the ITS2 becomes more complicated than with 18S and 25/28S, if the dynamic conformational model as proposed by Côté et al. (2002) for the ITS2 is accepted. These authors distinguished a “ring model” and a “hairpin model”, two alternative secondary structures, for *Saccharomyces cerevisiae*. Initial formation of the ring structure may be essential for early events in processing complex assembly and may be followed by a transition to the hairpin structure that facilitates subsequent processing events (Côté et al. 2002). Alternative folding of the helices around the central core of the ITS2 structure have been observed also for mycorrhizal fungi of the genus *Cenococcum* (Shinohara et al. 1999). Our data for some cyst nematodes also suggest that ITS2 might be a dynamic structure, flexing between two alternative folding patterns, and the models proposed by Wolf et al. (2005) might contain structural elements of another alternative folding.

Phylogeny of cyst-forming nematodes under the ITS2 structural model

Several studies (e.g. Ferris et al. 1993; Subbotin et al. 2001) have shown that the ITS-rRNA gene in cyst-forming nematode taxa might be too variable in length to allow construction of a plausible alignment assuring the reliability of subsequent phylogenetic analysis. To overcome this problem, Subbotin et al. (2001) generated several alignments to analyze the phylogeny. Computer alignments with different gap-length and gap-opening penalties, an elision alignment combining a range of alignments into a single matrix, and culled alignment exclusion of the most variable fragments were used in that study. All those alignments were compared using different tree-building methods. They showed that the positions of some taxa could be influenced by alignment parameters and by the method of phylogenetic inference. In the present work, we applied a more objective approach with reconstruction of an optimal alignment using secondary-structure information for each studied ITS2 sequence. Secondary-structure information provided a guide to aligning molecule regions previously considered as ambiguous. Manual construction of a secondary-structure alignment is a complex and time-consuming procedure (Subbotin et al. 2005, 2007), whereas application of the MARNA program generates a reliable alignment and represents a powerful tool.

As our understanding of mechanisms of nucleotide substitution in rRNA structures expands, we recognize that ignoring such processes as specific substitution patterns with compensatory and semi-compensatory

changes can compromise attempts to estimate phylogeny. Therefore, structural model-based analyses of DNA sequence data have become increasingly important for phylogenetic inference. Recently, the complex model considering secondary structure 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. Our enhanced knowledge of secondary structure allows us to apply a more sophisticated model and, consequently, to generate a more realistic tree indicating relationships.

In our previous studies of several nematode groups, the differences in topologies obtained using traditional methods and the models considering secondary structure were conclusive (Subbotin et al. 2005, 2007). In the present work, however, the phylogenetic trees obtained under different models were not principally different but more resolved, and they are in good accordance with previously published results for heteroderids (Sabo et al. 2001; Subbotin et al. 2001; Tanha Maafi et al. 2003, Madani et al. 2004). The influence of the complex model on relationships corroborated the more resolved nematode groups, and patterns of relationships among species within the *H. humuli*, *avenae* and *schachtii* groups did not vary significantly among the different models. In the BI trees derived from both models, the cyst-forming nematodes are distributed in several distinct clusters: the *H. avenae*, *sacchari*, *humuli*, *schachtii*, and *goettingiana* groups. In both BI trees, close relations were evident between the *H. avenae* and *sacchari* groups and between the *H. humuli* group and *H. salixophila*. Notably, derived representatives of both the *H. humuli* and the *avenae* group have a similar structure of fenestration in the cyst stage and form so-called “bifenestral cysts”.

The BI analysis under the complex model indicates that relationships among several main groups, such as between the *H. schachtii* or the *cyperii* group and the *avenae* + *sacchari* + *humuli* clade, remain unresolved. It has been known that application of the secondary-structure model of rRNA to phylogenetic analyses leads to trees with less resolved relationships among clades and likely eliminates some artifactual support for misinterpreted relationships. The highly resolved topology in some tree parts suggests that a deep phylogenetic signal has been retained in the ITS2 sequences of extant species.

Although incorporating secondary-structure information allows improved estimates of phylogeny, relationships among several main heteroderid clades remain poorly resolved. This suggests that the ITS2 marker does not contain sufficient phylogenetic signal to resolve relationships at the higher taxonomic levels. Thus, the

main conclusions from our study are (i) our thermodynamic criteria and covariation analyses provide a more accurate method for determining the secondary structure of the ITS2 in cyst-forming nematodes than does the homology modeling method of Wolf et al. (2005); (ii) consensus ITS2 secondary-structure models developed here for Heteroderidae serve as a useful guide for reconstruction of optimal alignments and as a source of additional data for structural variation to study phylogeny; and (iii) application of secondary-structure information provides complex models of nucleotide substitution and gives a realistic picture of relationships within *Heterodera*.

Acknowledgments

We are grateful to Dr. M. Wolf, Department of Bioinformatics, University of Würzburg, Germany, for his helpful suggestions. The material is based upon work supported by the National Science Foundation under Grant nos. 0529684 and 0608603.

References

- Alvarez, I., Wendel, J.F., 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Mol. Phyl. Evol.* 29, 417–434.
- Baldwin, J.G., Mundo-Ocampo, M., 1991. Heteroderinae, cyst- and non-cyst-forming nematodes. In: Nickle, W.R. (Ed.), *Manual of Agricultural Nematology*. Marcel Dekker Inc., New York, pp. 275–362.
- Barker, S.C., 1998. Distinguishing species and populations of rhipicephaline ticks with ITS 2 ribosomal RNA. *J. Parasitol.* 84, 887–892.
- Chen, C.A., Chang, C.-C., Wei, N.V., Chen, C.-H., Lein, Y.-T., Lin, H.-E., Dai, C.-F., Wallace, C.C., 2004. Secondary structure and phylogenetic utility of the ribosomal internal transcribed spacer 2 (ITS2) in scleractinian corals. *Zool. Stud.* 43, 759–771.
- Chilton, N.B., Newton, L.A., Beveridge, I., Gasser, R.B., 2001. Evolutionary relationships of trichostrongyloid nematodes (Strongylida) inferred from ribosomal DNA sequence data. *Mol. Phyl. Evol.* 19, 367–386.
- Coleman, A.W., 2003. ITS2 is a double-edged tool for eukaryote evolutionary comparisons. *Trends Genet.* 19, 370–375.
- Conole, J.C., Chilton, N.B., Jarvis, T., Gasser, R.B., 2001. Mutation scanning analysis of microsatellite variability in the second internal transcribed spacer (precursor ribosomal RNA) for three species of *Metastrongylus* (Strongylida: Metastrongyloidea). *Parasitology* 122, 195–206.
- Côté, C.A., Greer, C.L., Peculis, B.A., 2002. Dynamic conformation model for the role of ITS2 in pre-rRNA processing in yeast. *RNA* 8, 786–797.
- De Rijk, P., Wuyts, J., De Wachter, R., 2003. RnaViz 2: an improved representation of RNA secondary structure. *Bioinformatics* 19, 299–300.

- Ding, Y., Chan, C.Y., Lawrence, C.E., 2004. Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Res.* 32, W135–W141.
- Ferris, V.R., Ferris, J.M., Faghihi, J., 1993. Variation in spacer ribosomal DNA in some cyst-forming species of plant parasitic nematodes. *Fund. Appl. Nematol.* 16, 177–184.
- Goertzen, L.R., Cannone, J.J., Gutell, R.R., Jansen, R.K., 2003. ITS secondary structure derived from comparative analysis: implications for sequence alignment and phylogeny of the Asteraceae. *Mol. Phyl. Evol.* 29, 216–234.
- Gómez-Zurita, J., Juan, C., Petitpierre, E., 2000. Sequence, secondary structure and phylogenetic analyses of the ribosomal internal transcribed spacer 2 (ITS2) in the *Timarcha* leaf beetles (Coleoptera: Chrysomelidae). *Insect Mol. Biol.* 9, 591–604.
- Han, K., Lee, Y., Kim, W., 2002. Pseudoviewer: automatic visualization of RNA pseudoknots. *Bioinformatics* 18 (Suppl. 1), S321–S328.
- Hershkovitz, M.A., Zimmer, E.A., 1996. Conservation patterns in angiosperm rDNA ITS2 sequences. *Nucleic Acids Res.* 24, 2857–2867.
- Hlinka, O., Murrell, A., Barker, S.C., 2002. Evolution of the secondary structure of the rRNA internal transcribed spacer 2 (ITS2) in hard ticks (Ixodidae, Arthropoda). *Heredity* 88, 275–279.
- Hofacker, I.L., 2003. Vienna RNA secondary structure server. *Nucleic Acids Res.* 31, 3429–3431.
- Huelsenbeck, J.P., Ronquist, F., 2001. MrBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754–755.
- Joseph, N., Krauskopf, E., Vera, M.I., Michot, B., 1999. Ribosomal internal transcribed spacer 2 (ITS2) exhibits a common core of secondary structure in vertebrates and yeast. *Nucleic Acids Res.* 27, 4533–4540.
- Katihar, S.K., Visvesvara, G.S., Edlind, T.D., 1995. Comparisons of ribosomal RNA sequences from amitochondrial protozoa: implications for processing, mRNA binding and paromomycin susceptibility. *Gene* 152, 27–33.
- Kjer, K.M., 2004. Aligned 18S and insect phylogeny. *Syst. Biol.* 53, 506–514.
- Lott, T.J., Burns, B.M., Zancope-Oliveira, R., Elie, C.M., Reiss, E., 1998. Sequence analysis of the internal transcribed spacer 2 (ITS2) from yeast species within the genus *Candida*. *Curr. Microbiol.* 36, 63–69.
- Madani, M., Vovlas, N., Castillo, P., Subbotin, S.A., Moens, M., 2004. Molecular characterization of cyst nematode species (*Heterodera* spp.) from the Mediterranean Basin using RFLPs and sequences of ITS-rDNA. *J. Phytopathol.* 152, 229–234.
- Mai, J.C., Coleman, A.W., 1997. The internal transcribed spacer 2 exhibits a common secondary structure in green algae and flowering plants. *J. Mol. Evol.* 44, 258–271.
- Michot, B., Joseph, N., Mazan, S., Bachelier, J.P., 1999. Evolutionary conserved structural features in the ITS2 of mammalian pre-rRNAs and potential interactions with the snoRNA U8 detected by comparative analysis of new mouse sequences. *Nucleic Acids Res.* 27, 2271–2282.
- Morgan, J.A.T., Blair, D., 1998. Trematode and monogenean rRNA ITS2 secondary structures support a four-domain model. *J. Mol. Evol.* 47, 406–419.
- Murrell, A., Campbell, N.J.H., Barker, S.C., 2001. Recurrent gains and losses of large (84–109 bp) repeats in the rDNA internal transcribed spacer 2 (ITS2) of rhipicephaline ticks. *Insect Mol. Biol.* 10, 587–596.
- Ronquist, F., Huelsenbeck, J.P., 2005. Bayesian analysis of molecular evolution using MrBayes. In: Nielsen, R. (Ed.), *Statistical Methods in Molecular Evolution*. Springer, New York, pp. 183–232.
- Sabo, A., Vovlas, N., Ferris, V.R., 2001. Phylogenetic relationships based on ribosomal DNA data for four species of cyst nematodes from Italy and one from Syria. *J. Nematol.* 33, 183–190.
- Schultz, J., Maisel, S., Gerlach, D., Mueller, T., Wolf, M., 2005. A common core of secondary structure of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota. *RNA* 11, 361–364.
- Shinohara, M.L., LoBuglio, K.F., Rogers, S.O., 1999. Comparison of ribosomal DNA ITS regions among geographic isolates of *Cenococcum geophilum*. *Curr. Genet.* 35, 527–535.
- Siebert, S., Backofen, R., 2005. MARNA: multiple alignment and consensus structure prediction of RNAs based on sequence structure comparisons. *Bioinformatics* 21, 3352–3359.
- Subbotin, S.A., Madani, M., Krali, E., Sturhan, D., Moens, M., 2005. Molecular diagnostics, taxonomy, and phylogeny of the stem nematode *Ditylenchus dipsaci* species complex based on the sequences of the internal transcribed spacer-rDNA. *Phytopathology* 95, 1308–1315.
- Subbotin, S.A., Sturhan, D., Rumpfenhorst, H.J., Moens, M., 2003. Molecular and morphological characterization of the *Heterodera avenae* species complex (Tylenchida: Heteroderidae). *Nematology* 5, 515–538.
- Subbotin, S.A., Sturhan, D., Vovlas, N., Castillo, P., Tambe, J.T., Moens, M., Baldwin, J.G., 2007. 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. *Mol. Phyl. Evol.* 43, 881–890.
- Subbotin, S.A., Vierstraete, A., De Ley, P., Rowe, J., Waeyenberge, L., Moens, M., Vanfleteren, J.R., 2001. Phylogenetic relationships within the cyst-forming nematodes (Nematoda, Heteroderidae) based on analysis of sequences from the ITS region of ribosomal DNA. *Mol. Phyl. Evol.* 21, 1–16.
- Suh, S.-J., Kim, J.-G., 2002. Secondary structure and phylogenetic implications of ITS2 in the genus *Tricholoma*. *J. Microbiol. Biotechnol.* 12, 130–136.
- Swofford, D.L., 2003. PAUP*: Phylogenetic Analysis Using Parsimony (* and Other Methods), Version 4.0b 10. Sinauer Associates, Sunderland, MA.
- Tanha Maafi, Z., Subbotin, S.A., Moens, M., 2003. Molecular identification of cyst-forming nematodes (Heteroderidae) from Iran and a phylogeny based on ITS-rDNA sequences. *Nematology* 5, 99–111.
- Telford, M.J., Wise, M.J., Gowri-Shankar, V., 2005. Consideration of RNA secondary structure significantly improves likelihood-based estimates of phylogeny: examples from the Bilateria. *Mol. Biol. Evol.* 22, 1129–1136.

- Thines, M., Komjáti, H., Spring, O., 2005. Exceptional length of ITS in *Plasmopara halsredii* is due to multiple repetitions in the ITS-2 region. *Eur. J. Plant Pathol.* 112, 395–398.
- Wolf, M., Achtziger, M., Schultz, J., Dandekar, T., Mueller, T., 2005. Homology modeling revealed more than 20,000 rRNA internal transcribed spacer 2 (ITS2) secondary structures. *RNA* 11, 1616–1623.
- Young, I., Coleman, A.W., 2004. The advantages of the ITS2 region of the nuclear rDNA cistron for analysis of phylogenetic relationships of insects: a *Drosophila* example. *Mol. Phyl. Evol.* 30, 236–242.
- Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31, 3406–3415.