Phylogentic utility of secondary structure of ribosomal ITS2 and Cytochrome oxidase subunit-I (COI) in sarcoptes isolates from different hosts

A Valappil, S Angadi

Citation

Abstract
A comparative study of common core secondary structure in the ribosomal internal transcriber 2 (ITS2) and cyotchrome oxidase subunit I (COI) of 11 and 5 isolates respectively were selected from different hosts was carried out. Sarcoptes cause scabies of different nature in different hosts. Multiple sequence alignment and secondary structural analysis of ITS2 and COI was performed to elucidate the phylogengetic relationships. This study reveals a phylogenetic relationship among the different isolates of sarcoptes form different hosts. Several common features of secondary structures are shared among these isolates, with some of them support by compensatory changes, suggesting the significant role by ITS2 and COI as an RNA domain during ribosome biogenesis. The study also shows ITS2 in combination with COI sequence analysis gives a better understanding of intra generic variation in sarcoptes.

INTRODUCTION
Internal transcribed spacer 2 (ITS2) refers to a piece of non-functional RNA situated between structural ribosomal RNA (rRNA) on a common precursor transcripts. DNA sequence of the two internal transcribed species (ITS1 and ITS2) of the rRNA transcription unit have proven useful in resolving phylogenetic relationships for closely related taxa due to their relatively rapid evolution rate (Baldwin, 1992; Schlotterer et al., 1994; Mai and Coleman, 1997; Weekers et al., 2001; Oliverio et al., 2002). Zahler et al. (1999) used ITS2 to investigate the relationships within the genus sarcoptes and in psoroptes but less conclusive results. Navajas et al. (1998) used ITS2 in combination with COI (Cytochrome oxidase I) to investigate intra specific variation in Tetranchus urticae and found species -wide homologyneity of ITS2 sequence, but extensive COI polymorphism. To resolve deeper phylogeny ITS2 in combination with COI is better (Robert, 2002). Cytochrome oxidase is important in terminal oxidation in respiratory chin in aerobes. Murrell et al. (2000) successfully used 12S in combination with COI to resolve relationships among Rhipicephaline genera. COI has similar range of uses to ITS2 but appears to evolve slightly faster. Many workers in the systematic used COI to phylogeny of many genera (Anderson and Trueman, 2000; Toda, 2000; Salmone et al., 2002; Soller et al., 2001).

The transcripts folding structure of the ITS2 provide some signals that guide the ribosomal coding region when they are processed into small, 5.8S and large ribosomal RNA (Van der Sande et al., 1992; Van Nues et al., 1995). The potential to predict the folding structure has enhanced the role of ITS in phylogenetic studies, since it is important to guide reliable sequence (Michot et al., 1999). The secondary structure can be predicted by many methods like electron microscopy (Gonzales et al., 1990), chemical and structure probing (Yeh and Lee,1992; Van Nues et al., 1995 ) and computer software program (e.g. mfold and sfold) which utilize minimum free energy values (Zuker and Steigler, 1981). By using mfold software a secondary structure for the ITS2 with 4 domains (I~IV) has been proposed for green algae, flowering plants, fruit flies, parasitic flat worms, gastropods and mouse (Schlotte, 1994; Mai and Coleman, 1997; Michot et al., 1999, Coleman and Vacquier, 2002; Gottschling and Plotner, 2004). A highly conserved sequence is situated around a central loop and at the apex of a long stem in the 3´- half (Joseph et al., 1999). Due to high rate of sequence variation of transcribed spacers this may exhibit dramatic size variation and extensive sequence divergence even among moderately distant species (Michot et al., 1983; Furlong and Made, 1983). Nevertheless, the presence of phylogenetically conserved secondary structure elements in the 5´ externally transcribed spacer was recently
Phylogentic utility of secondary structure of ribosomal ITS2 and Cytochrome oxidase subunit-I (COI) in sarcoptes isolates from different hosts

revealed by the comparative analysis of a limited set of vertebrate sequence (Michot and Bachellerie, 1991). The coding region which has been most widely used in mites is the ITS2. Molecular phylogenetic study in mite Tetranchus (Navajas et al., 1992) agreed closely with morphology but sequence two sympatric species of Eotetrancthus from different hosts show substantial genetic divergence.

The over all literature cited shows that ITS2 and COI are widely used to resolve the phylogeny of aracina and many other invertebrates. Utility of ITS2 in phylogenetic study is limited with regards to sarcoptes. ITS2 in combination with COI is not don so far in sarcoptes isolates. Thus the present investigation was concentrated to reveal the phylogeny of sarcoptes isolates from different hosts using ITS2 and COI regions.

MATERIALS AND METHOD

ITS2 sequence of 11 sarcoptes isolates from different hosts (human, dog, silver fox, red fox, raccoon, dromedary, lynx, goat, swine, wombat and rabbit) and COI sequence of five available isolates (human, wombat, canine (dog), rabbit and goat) that are deposited in GenBank were investigated. The accession numbers of ITS2 and COI isolates are:-

<table>
<thead>
<tr>
<th>Isolates</th>
<th>ITS2</th>
<th>COI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>AB369382</td>
<td>AF 46241</td>
</tr>
<tr>
<td>Canine (Dog)</td>
<td>AF 129154</td>
<td>AY 46241</td>
</tr>
<tr>
<td>Rabbit</td>
<td>DQ 991138</td>
<td>AY 493396</td>
</tr>
<tr>
<td>Wombat</td>
<td>AF 129156</td>
<td>AY 493397</td>
</tr>
<tr>
<td>Goat</td>
<td>DQ 99137</td>
<td>AY 493748</td>
</tr>
<tr>
<td>Dromedary</td>
<td>AF 129158</td>
<td>Not available</td>
</tr>
<tr>
<td>Racoon</td>
<td>AF 129157</td>
<td>Not available</td>
</tr>
<tr>
<td>Red fox</td>
<td>AF 129159</td>
<td>Not available</td>
</tr>
<tr>
<td>Silver fox</td>
<td>AF 129155</td>
<td>Not available</td>
</tr>
<tr>
<td>Lynx</td>
<td>AF129160</td>
<td>Not available</td>
</tr>
<tr>
<td>Swine</td>
<td>DQ 991139</td>
<td>Not available</td>
</tr>
</tbody>
</table>

SEQUENCE ALIGNMENT

Multiple sequence alignment were performed by using CLC free workbench version 4.5.1 (http://www.clebio.com) with gap open penalty 15 and gap extension penalty 07. This program align nucleotides using a progress alignment algorithm (Feng and Doolittle, 1987).

SECONDARY STRUCTURE PREDICTION

The RNA secondary structure for ITS2 and COI were predicted by using RNADRAW online software (Christoffersen et al., 1994). RNADRAW predict RNA structure by identifying suboptimal structure using the free energy optimization methodology at a default temperature 37° C. In the current study ITS2, 5.8S and COI sequence were used separately for RNA structure prediction. The algorithm used in RNADRAW was ported from RNAFOLD program included in the Vienna RNA package. (Hofacker et al., 1994). The dynamic programming algorithm used in RNADRAW was based on the work of Zuker and Stiegler (1981) and uses energy parameters taken from Frier et al. (1986) and Jaeger et al. (1989).

RNA FOLD

The Sriobo program in Sfold (Statistical and Rational Design of Nucleic Acids) was used to predict the probable target accessibility sites (loop) for trans-cleaving ribozymes ITS2 (Ding et al., 2004). The prediction of accessibility is based on a statistical sample of the Boltzman ensemble for secondary structures. Here, we assessed the likelihood of unpaired sites for potential ribozyme target. Each mRNA exists as population of different structures. Hence, stochastic approach to the evolution of accessible sites was found appropriate (Christoffersen et al., 1994). The probability profiling approach by Ding and Lawrence (2001) reveals target sites that are commonly accessible for a large number of statistically representative structures in the target RNA. This novel approach bypasses the longstanding difficulty in accessibility evaluation due to limited representation of probable structures and high statistical confidence in predictions. The probability profile for individual bases (W=1) is produced for the region that includes a triple and two flanking sequences of 15 bases each in every site of the selected cleavage triplet (e.g. GUC).

PHYLOGENETIC ANALYSIS

The phylogenetic service of CLC free workbench was used for phylogenetic tree construction (CLC free workbench 4.5.1 http://www.clebio.com). Unweighted Pair Group Method using Arithmetic average (UPGMA) (Michener and Sokal, 1957; Sneath and Sokal, 1973). Clustering algorithm was used, for interpreting phylogenies bootstrap values are used.

RESULTS

SEQUENCE ANALYSIS OF ITS2 AND COI REGIONS

The length of ITS2 of 11 selected sarcoptes isolates ranged in size 296bp and 361bp. Eleven dispersed but unambiguously conserved sequence segments encompassing about a third of the ITS2 length have been identified. They
were interspersed with variable regions and gaps where size variations accumulate. The characteristics of sequences for each isolates are shown in the table 1. The length variations were observed with maximum length being 361bp in goat isolate and minimum 276bp in human isolate. The G+C content for the 2 regions of rDNA (5.8S and ITS2) of all the isolates ranged from 32% to 40%. For ITS2 regions the sequence identities ranges, with maximum 99% similarity between human isolate and wombat isolates; 83% similarity between rabbit isolate and goat isolate and rabbit isolate and swine isolate; 77% similarity between wombat isolate and silver fox isolates and wombat isolate and dromedary isolate; 49% between silver fox isolates and raccoon isolate and dromedary and raccoon; 35% between raccoon isolate and red fox isolate; 23% between red fox isolates and lynx isolate and minimum 9% similarity between dog isolate and lynx isolate. Alignment of ITS2 region is shows simple tandem repeats were present at various locations along the ITS2. The sequence similarity is more towards the 5’ end and with dispersed conservedness in the middle than towards the 3’ end.

The characteristics of COI sequence of each isolates were shown in the table 2. In COI regions the sequence identities ranges, with maximum 98% similarity between human and wombat isolate; 94% similarity between rabbit and human isolates, and between canine and human isolates; where as minimum being the 34% between rabbit and canine isolates. Alignment of COI is shown in the figure 2. Simple tandem repeats were present at various locations along the COI sequences. The sequence similarity is more at the 5’ end with dispersed conservedness in the middle towards the 3’.

**Figure 2**

Figure 1: Phylogenetic tree of sarcoptes isolates ITS2 sequences

**Figure 3**

Figure 2: Phylogenetic tree of sarcoptes isolates COI sequences

**SECONDARY STRUCTURE IN ITS2 AND COI REGIONS**

Secondary structure of ITS2 region was given in the Table 2. The secondary structures of the isolates from different hosts were classified into 5 groups based on the analysis of conserved stem and loops (Fig 3). Class I includes isolates of dog, red fox and class II includes isolates of dromedary and rabbit, class III includes isolates of wombat and human, class IV includes isolates of raccoon and swine and class V includes isolates of goat. Three common motifs having sequence AAAA and GCUUU respectively were conserved in all classes (Figure 5). Apart from the common conserved motifs shared among the species that are categorized into different classes, variable regions also do exists. The observed similarities at the secondary structural level are further reflected at energy level and of ITS2 and regions of various isolates.
Phylogenetic utility of secondary structure of ribosomal ITS2 and Cytochrome oxidase subunit-I (COI) in sarcoptes isolates from different hosts

**Figure 4**
Table 1: Secondary structural Features of ITS2 and 5.8S sequence of sarcoptes isolates from different hosts

<table>
<thead>
<tr>
<th>Isolates</th>
<th>ITS2</th>
<th>5.8S+ ITS2</th>
<th>Energy (Kcal)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Length (kbase)</td>
<td>GC+G%</td>
<td>GU (Numbers)</td>
</tr>
<tr>
<td>AFI291154 (Rat)</td>
<td>320</td>
<td>32</td>
<td>33</td>
</tr>
<tr>
<td>AFI291155 (Rat)</td>
<td>305</td>
<td>34</td>
<td>33</td>
</tr>
<tr>
<td>AFI291156 (Wombat)</td>
<td>305</td>
<td>35</td>
<td>36</td>
</tr>
<tr>
<td>AFI291157 (Dromedary)</td>
<td>305</td>
<td>36</td>
<td>35</td>
</tr>
<tr>
<td>AFI291158 (Dromedary)</td>
<td>305</td>
<td>35</td>
<td>34</td>
</tr>
<tr>
<td>AFI291159 (Dromedary)</td>
<td>305</td>
<td>35</td>
<td>34</td>
</tr>
</tbody>
</table>

The secondary structure features of COI were is given in the table 2. The secondary structures of the isolates were classified into four groups (Fig 4). Class I includes human isolate, class II includes rabbit and isolates, class III includes goat isolate and class IV includes wombat isolate. The common UUU AND AUAU respectively were conserved in all classes (Figure 6).

**Figure 5**
Table 2: Secondary structural Features of COI sequence of sarcoptes isolates from different hosts

<table>
<thead>
<tr>
<th>Isolates</th>
<th>COI</th>
<th>Length (kbase)</th>
<th>GC+G%</th>
<th>AU (Numbers)</th>
<th>GU (Numbers)</th>
<th>Number of stems</th>
<th>Energy (Kcal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFI462951 (Human)</td>
<td>747</td>
<td>747</td>
<td>174</td>
<td>40</td>
<td>23</td>
<td>-182</td>
<td></td>
</tr>
<tr>
<td>AY493396 (Wombat)</td>
<td>747</td>
<td>747</td>
<td>174</td>
<td>40</td>
<td>23</td>
<td>-182</td>
<td></td>
</tr>
<tr>
<td>AY493397 (Rabbit)</td>
<td>747</td>
<td>747</td>
<td>174</td>
<td>40</td>
<td>23</td>
<td>-182</td>
<td></td>
</tr>
<tr>
<td>AY493399 (Goat)</td>
<td>747</td>
<td>747</td>
<td>174</td>
<td>40</td>
<td>23</td>
<td>-182</td>
<td></td>
</tr>
</tbody>
</table>

DISCUSSION
Sarcoptes isolates are serious pathogen for scabies of different nature that affect vertebrate host like human, rabbit, bat, dog, swine, fox etc. in each host it cause scabies with characteristic nature. In the present investigation, the ITS2 and COI sequence reflected the trend observed in phylogeny. The more distantly related the less was the convergence at the ITS2 level (Figure-1 and 2). However accumulated substitutions in ITS2 leading to length variation also had a profound effect on the conserved ness among the structure. On the contrast the length is constant in the sequence of COI in all the isolates selected, indicating selective pressure cause less effect on the COI sequence of the isolates. The length of ITS2 is similar in all the isolates except rabbit, swine and wombat. This may be due to insertion affected by many factors including genetic drifts, rate of unequal cross
Phylogenetic utility of secondary structure of ribosomal ITS2 and Cytochrome oxidase subunit-I (COI) in sarcoptes isolates from different hosts

over, relative number of and size of repeats, gene conversion, immigration and the number of loci. (Levinson and Gutman., 1987). The length variation in human isolates may be due to deletion, less tandem repeats and genetic drifts. In all other isolates there is high level of sequence conservation even in the tandem repeats. The conservation is more predominant in the case of COI sequence of all isolates selected including human. This conservation was further reflected at the secondary structure and energy level. The predicted features of ITS2 and COI using RNADRAW are given in the table 1 and 2 respectively. ITS2 RNA structure from goat and swine isolate have the highest negative free energy (-117.78 K.cal.), followed by isolates from rabbit (-103.74 K.cal.) and isolates from raccoon (-101.64 K.cal.), dromedary (-101.92 K.cal), red fox (-101.96 K.cal.), silver fox (-101.64 K.cal.). Visual comparison shows that this is related to the trend in the cladogram given in the figure 1 and 2. The COI RNA structure from all the selected isolates except human shows identical negative free energy. This exception of human correlates the feature of the sequence of ITS2 and COI. The relative uniform negative free energy in COI comparing to ITS2 negative free energy proves that COI sequence analysis is better than ITS2 in establishing the phylogeny of lower taxa. This convergence at secondary structural level among a few isolate from different hosts may be due to the evolutionary pressure on ITS2 and COI to maintain the RNA secondary structure involved in post-transcriptional processing of rRNA (Shinohara et al., 1999). Secondary structure prediction for ITS2 and COI regions shows that these domains base pair to be form a core region central to several stem features implying that conserved ness is more important for the proper rRNA folding pattern (Wesson et al., 1992). The conserved ness is more obvious in the case COI than ITS2 indicating COI region is a better marker for phylogenetic study than ITS2 or in combination of ITS2 and COI provide a better understanding the deep phylogeny of the lower taxa (Navajas and Fenton, 2000). Studies in Tetranychus urticae revealed very low level of variation at the ITS2 locus in this species. In Casava green mite (Mononychellus progressivus) Navajas et al. (1994) found a similar pattern of variation with intraspecific diversity being lows for both ITS2 and COI, but lower for ITS2 alone. In the present investigation all the isolates studied shows sequence variation in both ITS2 and COI regions. The sequence variation is more in ITS2 than COI. The result suggests that the difference and conserved ness observed between ITS2 and COI of different isolates are not 'neutral' and are not simple accumulated random nucleotide changes, but bare a significant functional trend. The study of Wesson et al. (1992) in mosquito genera found that intra spacer variable regions appear to co-evolve and that ITS2 and COI variation is constrained to some extent by its secondary structure. Study by Van der Sande et al. (1992) in yeast have demonstrated that ITS2 is essential for the correct and efficient processing and maturation of certain ribosomal unit and finally for the efficient functionary of the rDNA cluster.

CONCLUSION

The selected 11 sarcoptes isolates from different hosts occur world wide and they cause scabies of various extensions in their hosts. In the present investigation, the ITS2 and COI sequences reflected the trend observed in the phylogeny. The isolates from more distantly related hosts the less was the convergence at the ITS2 and COI level. This study reveals, particularly with COI, although sequence variation was found in ITS2 sequence than in COI sequence, this did not correlate well with morphology and host suggesting that all the sequences studied belong with a single polymorphic species. The study also point out that

Figure 8

Figure 5: Secondary structures fold of ITS2 common to all sarcoptes isolate

Figure 9

Figure 6: Secondary structures fold of COI common to all sarcoptes isolate
COI sequence is more reliable, than ITS2, for sequence analysis and secondary structure prediction to resolve the phylogeny of sarcoptes.

References


r-8. Furlong JC and Maden BE.1983. Pattern of major divergence between the internal transcribed spacers of ribosomal DNA in Xenopus borealis and Xenopus levis, and of minimal divergence within ribosomal coding regions. EMBO J 1: 443-448.


Phylogenetic utility of secondary structure of ribosomal ITS2 and Cytochrome oxidase subunit-I (COI) in sarcoptes isolates from different hosts


Phylogenetic utility of secondary structure of ribosomal ITS2 and Cytochrome oxidase subunit-I (COI) in sarcoptes isolates from different hosts

Author Information

Ashokan K. Valappil, Ph.D.
Department of biological science, P.V.P.Colleg, Kavathe Mahankal

S.M. Angadi, M.Phil
Department of Zoology, Kasturbha Walchnd College