The novel organization and complete sequence of the ribosomal RNA gene of *Nosema bombycis*

Wei-Fone Huang, a,1 Shu-Jen Tsai, a,1 Chu-Fang Lo, b Yamane Soichi, c and Chung-Hsiung Wang a,*

a Department of Entomology, National Taiwan University, Taipei 106, Taiwan, ROC
b Department of Zoology, National Taiwan University, Taipei 106, Taiwan, ROC
c Biological Laboratory, Faculty of Education, Ibaraki University, Mito 310-8512, Japan

Received 10 September 2003; accepted 15 December 2003

Abstract

We present here for the first time the complete DNA sequence data (4301 bp) of the ribosomal RNA (rRNA) gene of the microsporidian type species, *Nosema bombycis*. Sequences for the large subunit gene (LSUrRNA: 2497 bp, GenBank Accession No. AY211393), the internal transcribed spacer (ITS: 179 bp, GenBank Accession No. AY211394), the small subunit gene (SSUrRNA: 1232 bp), intergenic spacer (IGS: 279 bp), and 5S region (114 bp) are also given, and the secondary structure of the large subunit is discussed. The organization of the *N. bombycis* rRNA gene is LSUrRNA-ITS-SSUrRNA-IGS-5S. This novel arrangement, in which the LSU is 5’ of the SSU, is the reverse of the organizational sequence (i.e., SSU-ITS-LSU) found in all previously reported microsporidian rRNAs, including *Nosema apis*. This unique character in the type species may have taxonomic implications for the members of the genus *Nosema.*

© 2003 Elsevier Inc. All rights reserved.

Index descriptors: *Nosema*; rRNA organization; microsporidia

1. Introduction

Microsporidia are tiny eukaryotic organisms (1–10 μm) that infest all major animal groups, more than 1200 species from 143 genera of animals are reported (Wittner, 1999). These obligate, intracellular parasites are well adapted in pathogenicity, transmission, ecology, and resistance to the defense mechanisms of their hosts. Insects in nearly all taxonomic orders are susceptible to this pathogen, but over half of the susceptible insect hosts occur in two orders, Lepidoptera and Diptera. Most of the entomopathogenic microsporidia occur in the genus *Nosema*, more than 150 described species found in 12 orders of insects (Becnel and Andreadis, 1999). *Nosema bombycis*, which is the type species of this genus (Sprague et al., 1992), has caused heavy losses in sericulture in Europe, as well as in Asia and America, especially in the middle of 19th century (Steinhaus, 1949).

Since microsporidia lack mitochondria, for a long time they were considered to be extremely ancient eukaryotes (Vosbrinck et al., 1987). However, recent molecular data and phylogenetic analysis suggest that mitochondrial endosymbiosis occurred before the emergence of microsporidia (Germot et al., 1997; Hirt et al., 1997; Williams et al., 2002). The small genomic size (2.9–19.5 Mb) of these organisms indicates that may have developed strategies of packing genetic information tightly into the genome or they may have lost genetic information for a metabolic pathway and depend on host cell sources for these compounds (Weiss and Vossbrinck, 1999). Evidence from protein coding genes, especially α- and β-tubulins (Keeling, 2003; Keeling and Doolittle, 1996; Keeling and Fast, 2002; Keeling et al., 2000), and phylogenetic analysis of microsporidia based on LSUrRNA sequences (Van de Peer et al., 2000) now suggest that in fact microsporidia share a common origin with fungi.
The sequences of microsporidian rRNAs are prokaryote-like and shorter than the known sequences of eukaryotic or prokaryotic rRNA (Galtier and Gouy, 1995; Vossbrinck et al., 1987). No distinct 5.8S rRNA gene has been found (Gatehouse and Malone, 1998; Tsai et al., 2002; Vossbrinck and Woese, 1986). The small subunit rRNAs of microsporidia are highly conserved, but in contrast to the many microsporidian SSUrRNA sequences in GenBank, only 4 complete LSUrRNA gene sequences have been published. These are for *Nosema apis* (Gatehouse and Malone, 1998), *Microsporidium 57864* (GenBank Accession No. U90885), *Heterosporis anguillarum* (Tsai et al., 2002), and *Encephalitozoon cuniculi* (Peyretaillade et al., 1998). There are also several partial sequences for microsporidian LSUrRNAs in GenBank, including a portion (approximately 350 nucleotides) of the LSUrRNA from *Vairimorpha* and *Nosema* species (Baker et al., 1994). Baker et al. (1995) noted that *N. apis* bears a closer resemblance (in terms of its SSUrRNA sequence) to some *Vairimorpha* species than it does to some other *Nosema* species. For *N. bombycis*, the full SSUrRNA sequence (1232 bp) but only a partial LSUrRNA sequence (292 bp) have been published (GenBank Accession Nos. D85503 and L28962) (Baker et al., 1994, 1995). Microsporidian rRNAs are hard to sequence completely, because it is difficult to design suitable primer sets when the microsporidan LSUrRNA sequences are highly diverse and, as we show here for *N. bombycis*, the rRNA gene has a novel organization.

### 2. Materials and methods

#### 2.1. Spore purification and nucleic acid preparation

Microsporidian spores of *N. bombycis* were a gift from Dr. R. Sugimoto of the MAFF GENE Bank of the National Institute of Agrobiological Science, Japan. The purification of spores was carried out as described previously (Huang et al., 1998; Tsai et al., 2002). Then, for DNA extraction, a suspension of purified spores (2 x 10^7 spores in 0.25 ml TE buffer) was mixed with an equal volume of zirconia/silica beads (0.1 mm diameter) in a 10 x 75 mm glass tube and shaken at maximum speed on a vortex mixer for 1 min (Undeen and Cockburn, 1989). The DNA was extracted using the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany) according to manufacturer’s instructions. The DNA was eluted in TE buffer and stored at -20°C. The DNA concentration was measured by a GeneQuant II RNA/DNA Calculator (Pharmacia, Uppsala, Sweden).

#### 2.2. Amplification and sequencing strategy of rRNA genes

The primer sets used for rRNA gene amplification and the expected sizes of the amplicons are shown in Table 1 and Fig. 1. Primer sets intended to amplify the ITS and adjoining 5' or 3' end of the LSUrRNA were also designed based on the published sequence for *N. bombycis*.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Amplicon size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Large subunit rRNA (LSU)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS228F</td>
<td>5'-GGA GGA AAA GAA ACT AAT-3'</td>
<td>2108</td>
</tr>
<tr>
<td>ILSUR</td>
<td>5'-ACC TGT CTC ACG ACG GTC TAA AC-3'</td>
<td></td>
</tr>
<tr>
<td><strong>5' end of LSU</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSUF</td>
<td>5'-ACT CTC CTC TTT GCC TCA ATC-3'</td>
<td></td>
</tr>
<tr>
<td>HG4R</td>
<td>5'-CGC CGA ATT AAA CTG AGT-3'</td>
<td></td>
</tr>
<tr>
<td><strong>Internal transcribed spacer (ITS)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ILSUF</td>
<td>5'-TGG GTT TAG ACC GTG AG-3'</td>
<td>501</td>
</tr>
<tr>
<td>S33R</td>
<td>5'-ATA GCG TCT ACG TCA GGC AG-3'</td>
<td></td>
</tr>
<tr>
<td><strong>Small subunit rRNA (SSU)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18f</td>
<td>5'-CAC CAG GGT GAT TCT GCC-3'</td>
<td>1232</td>
</tr>
<tr>
<td>1537r</td>
<td>5'-ATA TGG TCC TGC TAA TGG TCC-3'</td>
<td></td>
</tr>
<tr>
<td><strong>Intergenic spacer (IGS) and 5S rRNA</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HG4F</td>
<td>5'-GCC GCT TAA TTT GAC TCA AC-3'</td>
<td>852</td>
</tr>
<tr>
<td>5SR</td>
<td>5'-TAC AGC ACCCAA CTG TCC CAAG-3'</td>
<td></td>
</tr>
<tr>
<td>LS228R</td>
<td>5'-CCT CTT TTT TTT ATC TGA TAC-3'</td>
<td></td>
</tr>
</tbody>
</table>

*Nosema bombycis* putative pseudogene

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Amplicon size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>KAI01N</td>
<td>5'-GTA GGA GACCAA ATA ATC-3'</td>
<td></td>
</tr>
<tr>
<td>KAI02N</td>
<td>5'-ACT GGT GAG TCA AAT-3'</td>
<td></td>
</tr>
</tbody>
</table>

(modified from KAI01 and KAI02 by removing the restriction enzyme site)

---

Primers LS228F, ILSUR, 18f, and 1537r are from Vossbrinck et al. (1993). HG4F and HG4R are from Gatehouse and Malone (1998). ILSUF (Tsai et al., 2002) is the complementary sequence to ILSUR. KAI01N and KAI02N (Tsai et al., 2003) are modified from Kawakami et al. (1994). 5SR was designed based on the conserved region of 5S. HG4R-c is the complementary sequence to HG4R, and S33R is the reverse sequence of the SSUrRNA located between nucleotides 14 and 33.
apis (Gatehouse and Malone, 1998), but these primer sets failed; data not shown.]

For amplification, the genomic DNA (80 ng) of N. bombycis was mixed in a 100 μl PCR reaction mixture containing 10 mM Tris–HCl, pH 9.0, 50 mM KCl, 1.5 mM MgCl₂, 100 mM of each dNTP, 100 pmol of each primer (Table 1), and 2.5 U Taq DNA polymerase (Promega). The amplification was performed in an AG-9600 Thermal Station (Biotronics) for 40 cycles, each with the following profile: 94°C for 30 s, 50°C for 30 s, and 72°C for 2 min. A 10 μl aliquot from each reaction was run on a 1.0% agarose gel to visualize the PCR products. The gel was photographed using the Eagle-Eye II photo-documentation system (Stratagene). The PCR products were eluted by an E.Z.N.A. Gel Extraction Kit (Omega Bio-tek). The eluted DNAs were then sequenced directly on an automated DNA Sequencer (DNA Sequencer 377, Applied Biosystems).

2.3. Confirmation of the rRNA gene organization of N. bombycis

The total length of N. bombycis rRNA was amplified by the primer set LSUF/SSR with Platinum Pfx DNA polymerase (Invitrogen). The amplicon was then eluted and used as a template to amplify fragments with the primer sets HG4R-HELG4F-c and LSUF/HG4F-c for partial LSUrRNA-ITS-partial SSUrRNA; ILSUF/1537R for partial LSUrRNA-ITS-SSUrRNA; and ILSUF/SSR for partial LSUrRNA-ITS-SSUrRNA-IGS-5S. The amplification protocol was as described above.

2.4. Secondary structure construction

The secondary structures of N. bombycis LSUrRNA were constructed by a combined manual and automatic method in which the N. bombycis LSUrRNA sequence was aligned to the rRNA database to generate DCSE alignment files (De Rijk and De Wachter, 1993). The helices in the LSUrRNA secondary structure elements were then located and labeled based on database on the LSUrRNA secondary structure (De Rijk et al., 1998a), while the hypervariable areas (V1-12) were numbered in accordance with N. apis and with all known eukaryotic LSUrRNAs (De Rijk et al., 1998b; Wuyts et al., 2001). The final drawings were rendered by the RnaViz program (De Rijk and De Wachter, 1997; De Rijk et al., 2003). The secondary structures of N. bombycis SSUrRNA are already known and can be found in the European small subunit ribosomal RNA database (Van de Peer et al., 1998).

3. Results and discussion

3.1. The complete sequence and organization of N. bombycis rRNA

The complete DNA sequence of the N. bombycis rRNA gene contained 4301 bp (see Appendix A) and was submitted to the GenBank with Accession No. AY259631. The organization of the gene is shown in Fig. 1. Starting from the 5’ end, the N. bombycis rRNA gene consists of the large subunit gene (LSUrRNA: 2497 bp; submitted to GenBank with Accession No. AY211393), the internal transcribed spacer (ITS: 179 bp; GenBank with Accession No. AY211394), the small subunit gene (SSUrRNA: 1232 bp), the intergenic spacer (IGS: 279 bp), and the 5S region (114 bp). This organizational sequence, in which, the LSU gene is 5’ of the SSU, is unique among microsporidia and is the reverse of the organizational sequence for all previously reported microsporidian rRNAs (Gatehouse and Malone, 1998; Peyretaillade et al., 1998; Tsai et al., 2002).

PCR results with various combinations of the primers listed in Table 1 are shown in Figs. 2 and 3. The amplicon yielded by the primer set ILSUF/S33R (Fig. 2, lane 4) specifically confirms the novel organization of this gene (i.e., LSU-ITS-SSU). Conversely, the primer set HG4F/LS228R would usually amplify the ITS region between the SSU and LSUrRNA, but the absence of any amplicon with these primers (Fig. 2, lane 7) provides further evidence that the LSU is 5’ of the SSU (and also suggests that the N. bombycis rRNA gene is not a multiple gene). Further confirmation that the N. bombycis rRNA gene is organized as shown in Fig. 1 was provided by amplifying the whole rRNA gene with the LSUF/SSR
primer set (Fig. 3A). The resultant amplicon was of the predicted length (4442 bp), and when four internal fragments were amplified (Fig. 3B) and sequenced, results were completely consistent with the DNA sequence listed in Appendix A.

3.2. LSUrRNA gene sequence

The main part of the LSUrRNA gene was amplified with the LS228F/ILSUR primer set, which produced an amplicon of 2108 bp (Fig. 2, lane 1). For sequencing the
S' end of LSUrRNA, the primer set LSUF/LS228R was used (Fig. 2, lane 2). The putative start and terminal regions were determined by comparison to the N. apis LSUrRNA sequence (Gatehouse and Malone, 1998) and by the secondary structure construction of N. bombycis LSUrRNA gene (Fig. 4). The LSUrRNA gene contains 2497 bp and its base composition is 31.9% G + C, which is the lowest G + C content of all known microsporidian LSUrRNA genes. The LSUrRNA sequence identities between N. bombycis and N. apis (GenBank accession no. U97150; Gatehouse and Malone, 1998), Microsporidium 57864 (GenBank

**Nosema bombycis LSU rRNA**

Fig. 4. A model of the secondary structure of N. bombycis large subunit (LSU) rRNA. Helix numbering is according to De Rijk et al. (1998a). The boxed regions indicate parts of the structure where hypervariable areas are found in the eukaryotic rRNAs (De Rijk et al., 1998b; Wuyts et al., 2001).
Accession No. U90885), *H. anguillarum* (GenBank Accession No. AF402839; Tsai et al., 2002), and *E. cuniculi* (GenBank Accession No. AJ005581; Peyretaillade et al., 1998) are 71, 69, 46, and 53%, respectively. A previously reported partial sequence of the *N. bombycis* LSU rRNA gene (GenBank Accession No. L28962; Baker et al., 1994) had 100% identity with the sequence from 132 to 423 bp from the 5′ end.

Like all microsporidia, the internal transcribed spacer (ITS) region of *N. bombycis* lacks the 5.8S rRNA gene. However, as in *Vairimorpha necatrix* and *N. apis*, in which the 5′ ends of the LSU rRNA genes include a sequence that corresponds to 5′S (Gatehouse and Malone, 1998; Vossbrinck and Woese, 1986), the sequence of nucleotides in the *N. bombycis* LSU rRNA gene located at 1–160 from the 5′ end corresponds to the known fungal 5.8S rRNA sequences. Compared to *Cystofilobasidium hisporidi* (GenBank Accession No. M94511), *Lactarius acerrimus* (GenBank Accession No. AJ278139), *Thanatephorus cucumeris* (GenBank Accession No. AB019008), *Trichoderma reesei* (GenBank Accession No. L27800), and *Tuber cf. rapedorum* (GenBank Accession No. AJ278140), the homologies are 34, 44, 44, 44, and 42%, respectively, by Clustal X and GeneDoc.

The secondary structure of the LSU rRNA of *N. bombycis* (Fig. 4) is basically similar to that of *N. apis* and *H. anguillarum* (De Rijk et al., 1998b; Tsai et al., 2002; Van de Peer et al., 2000). Based on the secondary structures of the eukaryotic LSU rRNA of *Xenopus laevis* (De Rijk et al., 1998a) and the eukaryotic database (Wuyts et al., 2001), eight helical groups (B–I) can be distinguished clockwise from a core area. Eleven hypervariable areas (V1–5; V7–12) are also shown in Fig. 4. Nine helices (B6, B7, B8, B14, B21, D5, E9, E15, and G5) are missing. Six areas of the hypervariable areas are also almost entirely missing (V2, V3, V8, V10, V11, and V12), and two areas are extremely reduced (V5 and V9). V6 is almost absent. The secondary structure of *N. bombycis* LSU rRNA diverges most markedly from the LSU rRNAs of *N. apis*, Microsporidium 57864, *H. anguillarum*, and *E. cuniculi* in the V4 area. The V3 areas of *N. bombycis* and *N. apis* LSU rRNAs are similar in accordance. By comparison, the LSU rRNA of *E. cuniculi* has its own specific conformation in V3 (lack of the D3 helix), while the LSU rRNA of *H. anguillarum* has specific conformations in five other areas (V5, V6, V7, V9, and V10).

### 3.3. ITS sequence

In contrast to all known microsporidian rRNA genes, the ITS region of *N. bombycis* is 3′ of the LSU and 5′ of the SSUrRNA. It consists of 179 bp located between nucleotides 2498–2576 from the 5′ end of rRNA gene (Fig. 1), and its G + C content is 19.6%. The *N. bombycis* ITS region has no homology to the known microsporidian ITS sequences (Gresoviac et al., 2000).

### 3.4. SSUrRNA gene sequence

The SSUrRNA gene contains 1232 bp and is located between nucleotides 2677–3908 relative to the 5′ end of the rRNA gene (Fig. 1). The G + C content of the SSUrRNA gene is 34.2%. The complete DNA sequence of the SSUrRNA gene of *N. bombycis* has 99% homology to the *N. bombycis* SSUrRNA sequence already held in GenBank (GenBank Accession No. D85503; different nucleotides at 3497 and 3874), to another microsporidian isolate from Japan (GenBank Accession No. D85504; Hatakeyama et al., 1997), and to five *Nosema* isolates from Taiwan (Tsai et al., 2003).

### 3.5. IGS

The IGS region consisted of a 279 bp sequence located between the SSU and 5S rRNA genes (nucleotides 3909–4187). Its G + C content is 30%. Homology with other known microsporidian ITS or IGS sequences is low; comparisons using standard nucleotide–nucleotide BLAST [blastn], Nucleotide Blast, and NCBI found only 20 matching nucleotides.

### 3.6. 5S rRNA gene

The 5S rRNA of *N. bombycis* consists of 114 bp (including the putative end), and is located between nucleotides 4188 and 4301 from the 5′ end of the rRNA gene. Its G + C content is 47.3%. The sequence has a high homology, 91 and 92%, respectively, with the 5S rRNAs of two *N. bombycis* isolates from Japan (GenBank Accession Nos. D14631 and AB097401; Kawakami et al., 1992), but a homology of only 77% with the 5S region of *Microsporidium* 57864 (GenBank Accession No. U90885).

The members of the genus *Nosema* are often considered the most important and widely distributed group of insect microsporidia (150 described species found in 12 orders of insects) (Becnel and Andreadis, 1999; Tanada and Kaya, 1993). As *N. bombycis* is the type species of this genus, its characteristics—not only as they relate to life cycle, development, and morphology, but also in terms of their biochemical and molecular characteristics—are critical. The unique organization of the *N. bombycis* rRNA therefore has implications for the taxonomy of the *Nosema* group.

### Acknowledgments

This paper was supported by the National Science Council, ROC (Grant No. NSC 92-2313-B-002-030).
We would like to thank Dr. R. Sugimoto of the MAFF GENE Bank of the National Institute of Agrobiological Science, Japan for donating spores of the type species, *N. bombycis*. We would also like to thank Dr. J. Wuyts, University of Antwerp, Belgium, for his assistance with the secondary structure.

Appendix A. The complete sequence of *N. bombycis* rRNA

```
LSUF

-141 A CTCCTCTCTTT TTGGCTCAATC ATCAAATTTG ATCAAATCAA ACATCACCAC TCAACCCCAT
-81 CATTGACCAG AACCTCCAGA AAATAAAGAC GTGAAAAGAG AGATTAATGT ATTTTTCCAT

| ← LSU rRNA |
1 ATTTGACTAC CATGGGATCA ATAGGAGATC ATACACGTGA AGAACATATA AGAATATGAT
AAACATAATC CTGGTATCC

LS228F

81 TAAATCATTATTAGATAACCTTTGGAACATTCTAACA GAAAAGAGAG AGAATACCTATG

161 GTGCGCATATG AACGAGAAAG AAACCTCATTATG TTAACAATCTAT ACTTTATAGG AAGTTAAAG
GTGATGTTTCTGATCA

241 TCTAGATAC ATAAATTGTT AGTAAATGAT GAGCATAAAT AAACAAATGTAA GAGTACAGGT
GCTGTTAGTG ACTTTTTCAA

321 TATAGATAGA ATAGATATTCT TAAAGTTAAT ATACATGAA TACCCAAGAC TCAAGTGAAA

401 TGGTGAATTAGCCTTTATG ATATATTATA GACCCTGCTT GAACACCGGA CCAAGGAGAT
TACCAAGTAG CAAGATTACA

HG4R

481 GAATCATAAGG GTTTTTGTCA TCTGTGAAGA CCCCCACAT AGTGAACATAT ATATGCAAT
GCTGGAAGATG AAAAAACGT

561 TTTATAGGGAC ATCCTAGCTA TTGTAGCAAGC CCAAAGACAT ATCAGTTCA

641 GCTGGAGCTG GTGCAACGGG AATATGTCTT TGGGACAGTT GAGTAAATAT AGGAAAATAG
GATGAAATG CAATTTACTG

721 TTAATTTAATA GTGTTTACTT TTAATTTACTA TAGTATTTCTA TGTGTTTAAAT ATACTTTACGT ATATTTACGA
TACCAAGTG

801 GGCTGATATG TTAGAAGATG ATATGCAATA AGGGATAGAC CTTGAAGATGC ACTAAAAATT
CTAATGTTAT TTTATATTAA

881 AATATGATG ATACATATTAG ACAGTAGGGG CTTTTTTATG GAAGTAAAGA TCCCCAAAAG
AACCGTGTAC AAACCTACCA

961 CCAGTTATG TTTGTTAATA AATATGGAAT TTATTATATA GTATGAAAT GTTAAACAAT
ACAGTATTTGT TTTTGGAT

1041 TTTTATCAAT TGAATATGGT GAAGAGCTAT GAATTTATA TATTTTAAAT TAAATGAAA TCAATAGAC
GAGCTCATGT GAGTTAATAC

1121 AATATTTGTT ATTTATGAG ATTAGGTTTG ATCTTTATGAT GACGTGAGAT GGTGGTAAAT
TAAACTAGAT ATATTTAATA

1201 TTATATATCT TTTTATTTTG AGTTTTTGCA CAAAATATGAA CTATTATTGT AAAAAATTT
AAATATATG ATTTTTAATA

1281 GTCTATAGAT GTTTATCTTG TTTTAAAGAT ATACTTTGAA TAAATATAG CAAAATAGGA
AAATATACAA GCTATTTTGT

1361 ACCCGCATCG GTGCTCTCTG TTTTAAACAA ATATATATTT AGTTATACAA GATCTAGCAAT
TGCGCAAAAT GACATCAAA

1441 CTGTTGGATG AAAGTTGCGG ATCGAGTCTG AGA CTTTTTTG ATATTATGAG AGGAATCTGA
CCTTTAAAAA AAACACACAT

1521 TATAGTAAAT ATATAGTAG ATTTCTGCAA TGTTATTAT TAGTAAAAAT GAATAAGAAG
CTTATACAAAG CCAAGGGTAT

1601 CTATGACTG ATCAAGCTTG CCAATATGCG TCACTTTTTA TTTTGGTGAAGC CCAATTAG
GGAAACAGAA TTCATTGAG

1681 CCCTATTAGT ACAGTGTAGA AGCAACAGGA CAAGGAGAAG GGGCTTTATAT ATACACGCGG
GAAAAGAAGA CCGTCTGGAG

1761 CTTTACCTA GTATTTCCTA ATTTGTTAT ATTTATTTGT AGAGAGAGGAG GAGATTAAT
AGTTACACCA TGTCTAAAG

1841 ATATATGAA TTATGTGAAT TTATATGGAAG AGTTTGCGTG GGGCGCCACA ACTTTTATAC
AGAATAGAAG ATGTTTATAT

1921 ATACAGATTT ATAGGATAGA AAACCTATGCT TCAATATAAG GGTGAAAAAT GTCTGTGATG
ATAGAAACCAG ATGTTTATAT
```
2001. GATCGGAAAC CATGGCCTAA AGATCCCCTA ATAATAGTTT TATTTGATTG GAGGTGACAG AAAAAAGGACACCGGATAA

2081. CTGCGTGGTA GCAATAAGGC GATCATAGCG ACTTGGCTTT TTGATTCTTC GATGTCGCTG CTTCGCTGCA TCTGATTG

2161. TATGTTACGA AGTGGTTGGA TGTCACCCCG TTAATGAGGA AGCTGAGATG GGTTTAGACC GTGCGGAGAC AGGGATTG

2241. TACCTCTACTG TTATATTTAT AAAGAAAAGT TATATAATGC GAGAGGAACA ATAAAGTATG AGCTGATTGT TACGGATT

2321. TTTATCAATT ATGTGAAAA ACTACGTCAT GAATAATAAA TGCTGAAAAC CTCTTAAAGCA TGAAGTTAAT CTTAACAA

2401. GACAATTTGAG GGTACTACCT TCTCTGAAAG ATGGGACGCTT TAAAGATGAAG GTGTATACCG ACAATTTTAC GTTACCAAT

LSU rRNA ←→ ITS region

2481. CGTCTCGTGT AAAATAATG Acct tcaagtggta cgaagattat ttgattatgc tttacatac taaaagatct aacgtagtt

2561. ttctttatt ccatagcgtg tattggagt tattattttc tattattatt tttctatgt

ITS region ←→ SSU rRNA

2641. ggcttacgct atttacattt ttaataatg cacc aggttgatttc tgtcctgacgt agagcccatag tctcaagat

2721. AACCCATCGA TGTTATTTTG ATATAAAGAA AAAAGGACAA GCTCAGTAAC CTCTATTTTG TTATATATGC TACGGATT

2801. ACTATAGTTA ATATAGTTAA ACAATAACG AATAAAAGATA AGATCTCATT GTTATGGTTT AAGGAAAATG CTTACCAAT

2881. CTATGACCGA TAACGTATTATT CTATTGTATAT ATCCCGGAGA AGGAGCTCTGA GAGATGTCA TCAAGTCCTA GGGATGCTT

2961. AGGGCGAAA CTACGGCTAT ATGATTATTT ATGAGCGATT ATGAGGTATT TTTTATATT ATTTATCG TATGAGTAA

3041. TATACGAGA TAAACTTGGA ATGGCAATCTA ATGGACGCTG ATGGGAGCAC CAGGGATATG ATCTGGTAC

3121. TGGAGTTTAA AAGTGCTGTT TTATTTATTT TAAAGTTAAT AATAGGTGCTT TGAAGATCAG ATGATGTTTA

3201. AACCCTAATCC GAGTGAAACG AAGCGAAACG TGATACATTT ATATATTTAT TGAACAAATG AGCTAACGCTA GAGGATCAA

3281. GATGATTAGA TACCTACCTA GTTCTACGAG TAAACTATTT TGATACATG ATATATTTTG ATATATTTAT TTAAGAGAGA

3361. ATATAGATTA ATTGACGCTT CGGGATAGTA TGATTCGCAAG ATGGAATAAT AAAAGAATTG ACGGAGGAAAT ACCAACCAAG

HG4F

3441. GGGTATAGG GGGTCTTTTT TGGACTAACG GGGGATTAAT TACCAGCTT ATACATGGAT AATATTTAT TCAATGAGT

3521. GATGATGGGCT GTTCTCAATAG TATGCTGTTA AGTAATGATT AATTTAACA AGATGTGAGA CCCCCTATTA GACAGATAC

3601. GTGATACATA TGAAGGAGAG GATTAAACCA GGTCGGTTAT GCTCCTAGAT AAATCTGGTT GACCGCAGA TACCATATA

3681. TTTGATATTA TAAAGGATATA TATATTTGAA GATATATTGT AACATGGGAAT TGCTAGATAA TTTATATTTA TAAAGAGAG

3761. TGAAATGAGTC CGTGGTCTTTT GTACACACCG CCCGTGGCTA TCTAAAGATG TATTATCTAT GAAATTTAT TTAAAGGAG

SSU rRNA ←→ IGS region

3841. TAGATAGTAC TAGATCTGTA ATAGACGTGTA ACATGGTTGC TGTTGGAGAA CCATTAGCA GATCATAT Aaataaagagt

1537r

3921. gctgatttat aataaagagt tagttcagca ttaatattct cttgtctttg tattaatattt ctaaatttt ccaagaggtc

4001. aagatggtgt atataagaggt gttaatattt cttaaatttt tgaagttg tggatgtaa aataaggtggt

4081. tagttcagtt tagattgtttc tagattgtttc gttcctttt attttatttt tattttttt tcaagaggtc

IGS region ←→ 5S rRNA

4161. tacataatca aacaccattt gttcattacGA TACGTTACTA TCTACGAAA ACCACTGGAA CCCACCCGAA CTCCCCGAT

5S rRNA ←→

4241. AAACCCAGATG AGCTTTAATCA GTATACAGAA GGAGACCACT TGGGGACCTG TGGTGCTGT AA

5SR
References


