Available online at www.sciencedirect.com
SCIENCE DIRECT®
Fungal Genetics and Biology 41 (2004) 473–481
FUNGAL GENETICS AND BIOLOGY
www.elsevier.com/locate/yfgbi

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*.

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 (Vossbrinck 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.

* Corresponding author. Fax: +886-2-2736-4329.
E-mail address: wangch@ccms.ntu.edu.tw (C.-H. Wang).
[1] These two authors equally contributed to this work.

1087-1845/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/j.fgb.2003.12.005

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 (2×10^7 spores in 0.25 ml TE buffer) was mixed with an equal volume of zirconia/silica beads (0.1 mm diameter) in a 10×75 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.*

Table 1
Primers used for amplification and sequencing of *N. bombycis* rRNA

Primer	Sequence	Amplicon size (bp)
Large subunit rRNA (LSU)		2108
LS228F	5′-GGA GGA AAA GAA ACT AAC-3′	
ILSUR	5′-ACC TGT CTC ACG ACG GTC TAA AC-3′	
5′ end of LSU		
LSUF	5′-ACT CTC CTC TTT GCC TCA ATC A-3′	
HG4R	5′-CGC CGA ATT AAA CTG AGT TG-3′	
Internal transcribed spacer (ITS)		501
ILSUF	5′-TGG GTT TAG ACC GTC GTG AG-3′	
S33R	5′-ATA GCG TCT ACG TCA GGC AG-3′	
Small subunit rRNA (SSU)		1232
18f	5′-CAC CAG GTT GAT TCT GCC-3′	
1537r	5′-TTA TGA TCC TGC TAA TGG TTC-3′	
Intergenic spacer (IGS) and 5S rRNA		852
HG4F	5′-GCG GCT TAA TTT GAC TCA AC-3′	
5SR	5′-TAC AGC ACC CAA CGT TCC CAA G-3′	
LS228R	5′-CCT CCT TTT CTT TGA TTG-3′	
Nosema bombycis putative pseudogene		
KAI01N	5′-GTA GTA GAG ACC CAA ATA TC-3′	
KAI02N	5′-ACT GTT CAG ATA TGG TCC TTA TCG-3′	
(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.

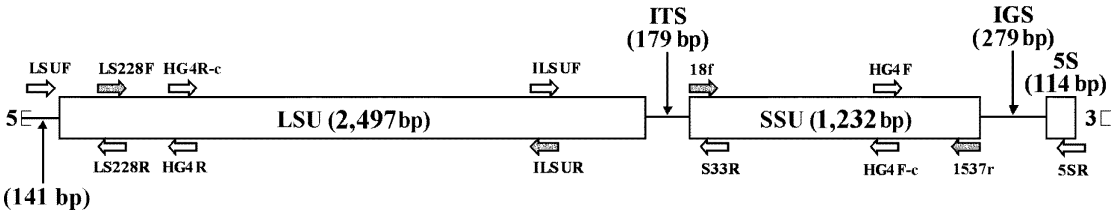

Fig. 1. Schematic diagram of *N. bombycis* rRNA gene. Mature rRNA gene domains are boxed. Details of the primers are given in Table 1. The gray-shaded primer sets were used to amplify the SSUrRNA coding region and the main part of the LSUrRNA coding region. HG4R-c and HG4F-c are the complementary sequences of HG4R and HG4F, respectively.

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 $MgCl_2$, 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/5SR with Platinum *Pfx* DNA polymerase (Invitrogen). The amplicon was then eluted and used as a template to amplify fragments with the primer sets HG4R-c/HG4F-c and ILSUF/HG4F-c for partial LSUrRNA-ITS-partial SSUrRNA; ILSUF/1537R for partial LSUrRNA-ITS-SSUrRNA; and ILSUF/5SR 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/5SR

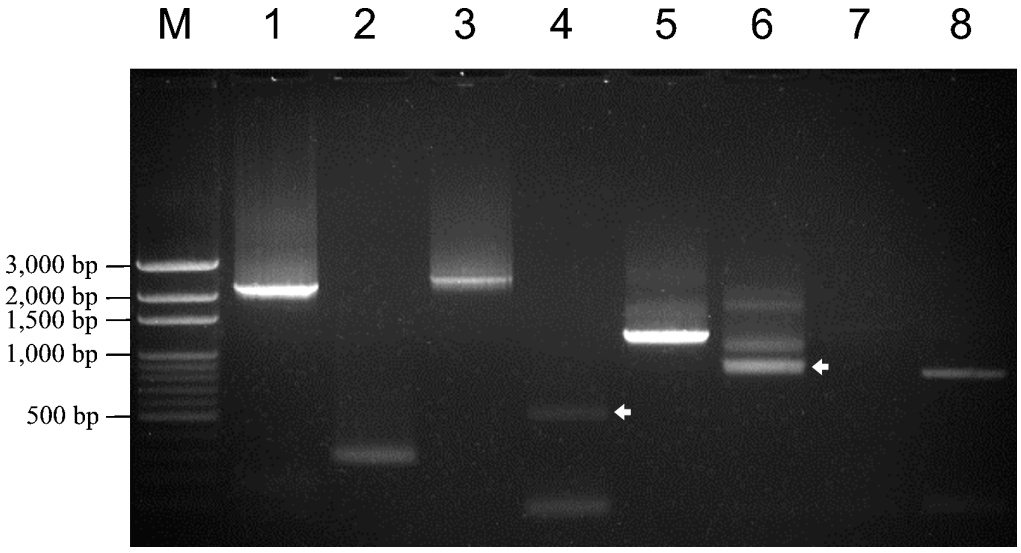

Fig. 2. Agarose gel electrophoresis of PCR products. Lane 1, main part of LSUrRNA (primer set LS228F/ILSUR; 2108 bp amplicon). Lane 2, 5′ region of LSUrRNA (primer set LSUF/LS228R). Lane 3, 5′ region and main part of LSUrRNA (primer set LSUF/ILSUR). Lane 4, 3′ end of LSUrRNA- ITS-5′end of SSUrRNA (arrow) (primer set ILSUF/S33R). Lane 5, SSUrRNA (primer set 18f /1537r). Lane 6, 3′ end of SSUrRNA-IGS-5S (arrow) (primer set, HG4F/5SR). Lane 7, the result of amplification with HG4F/LS228R. Lane 8, (DNA quality control) the result of amplification with the *N. bombycis* specific primer set KAI01N/KAI02N (Tsai et al., 2003).

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

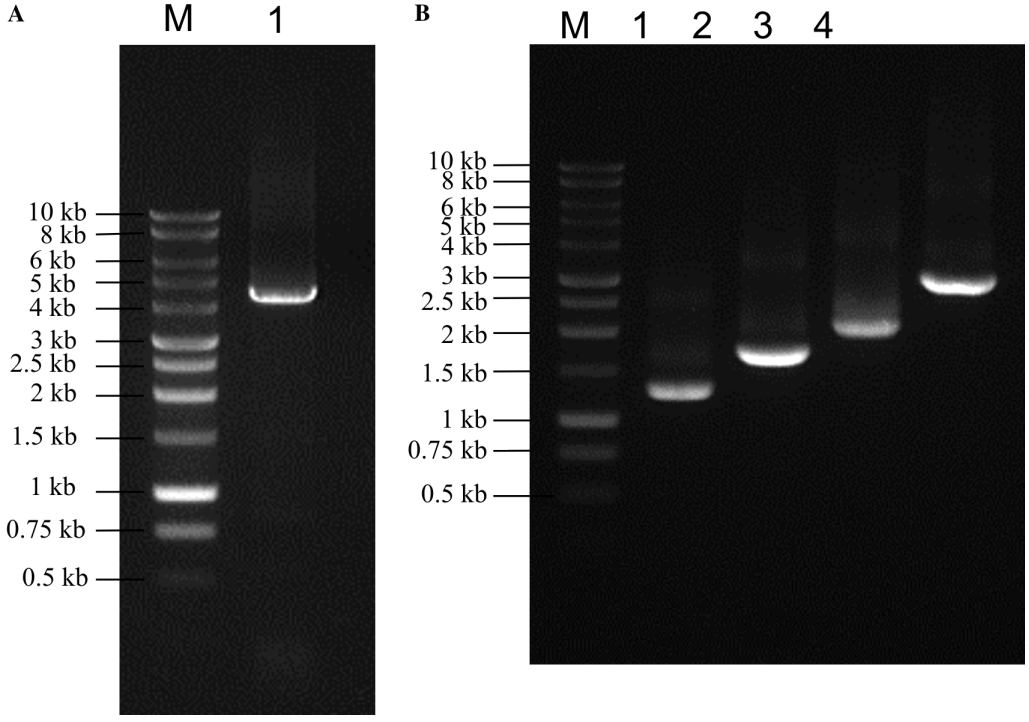

Fig. 3. Confirmation of the organization of the *N. bombycis* rRNA gene. (A) Amplification with the primer set LSUF/5SR yielded an amplicon of the predicted size (4401 bp). M, 1 kb DNA ladder (Promega). (B) Amplification of four internal rRNA fragments. Lane 1, 3′ end of LSUrRNA- ITS-main part of SSUrRNA (1261 bp; primer set ILSUF/HG4F-c). Lane 2, 3′ end of LSUrRNA-ITS-SSUrRNA (1700 bp; primer set ILSUF/1537R). Lane 3, 3′ end of LSUrRNA-ITS-SSUrRNA-IGS-5S (2093 bp; primer set ILSUF/5SR). Lane 4, main part of LSUrRNA-ITS-main part of SSUrRNA (3031 bp; primer set HG4R-c/HG4F-c). M, 1 kb DNA ladder (Promega).

5′ 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

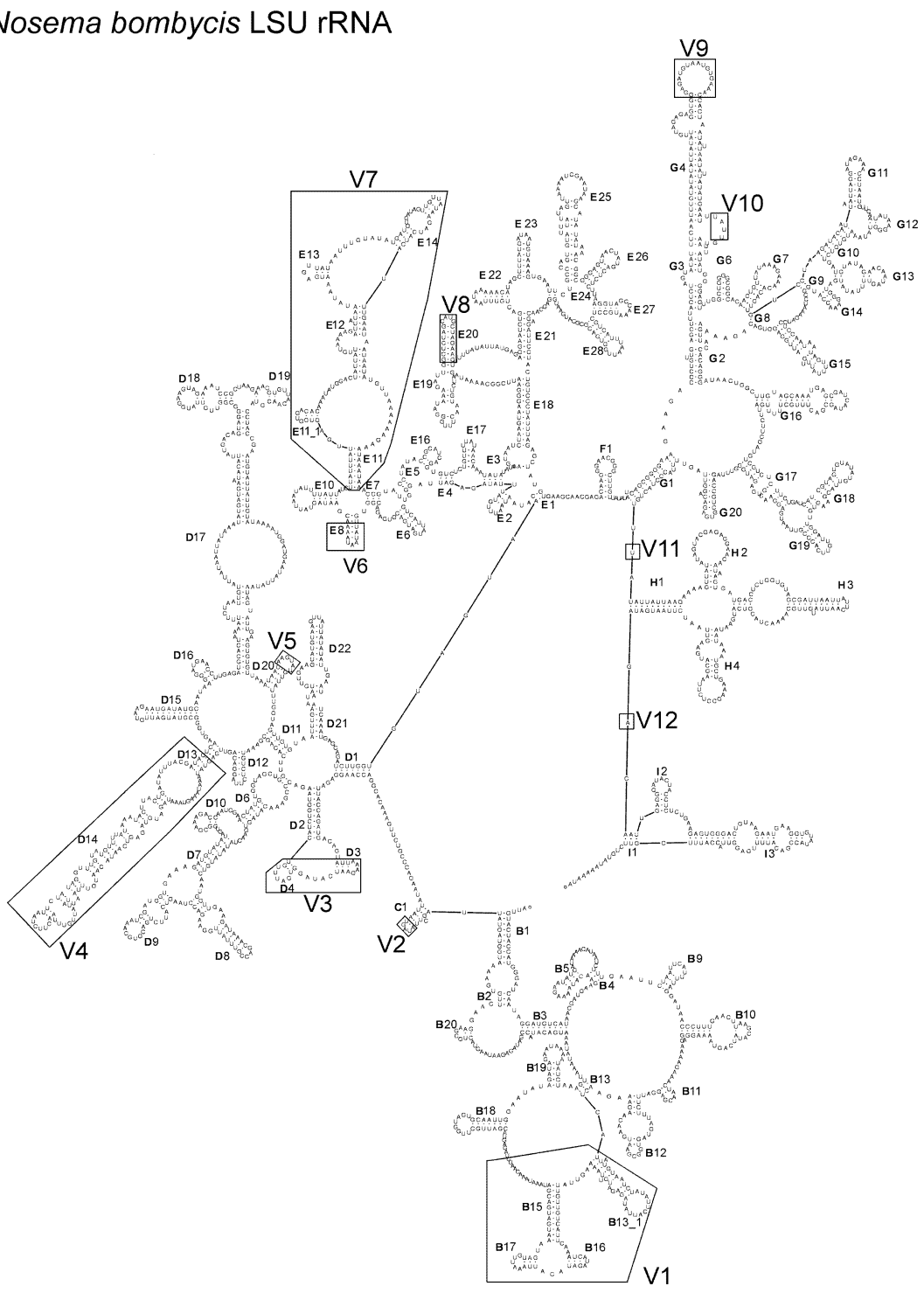

Fig. 4. A model of the secondary structure of *N. bombycis* large subunit (LSU) rRNA. Helix numbering is according to De Rijk et al. (1998a). The boxed regions indicate parts of the structure where hypervariable areas are found in the eukaryotic rRNAs (De Rijk et al., 1998b; Wuyts et al., 2001).

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* LSUrRNA 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 LSUrRNA genes include a sequence that corresponds to 5.8S (Gatehouse and Malone, 1998; Vossbrinck and Woese, 1986), the sequence of nucleotides in the *N. bombycis* LSUrRNA gene located at 1–160 from the 5′ end corresponds to the known fungal 5.8S rRNA sequences. Compared to *Cystofilobasidium bisporidii* (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 LSUrRNA 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 LSUrRNA 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* LSUrRNA diverges most markedly from the LSUrRNAs of *N. apis*, Microsporidium 57864, *H. anguillarum*, and *E. cuniculi* in the V4 area. The V3 areas of *N. bombycis* and *N. apis* LSUrRNAs are similar in conformation. By comparison, the LSUrRNA of *E. cuniculi* s has its own specific conformation in V3 (lack of the D3 helix), while the LSUrRNA 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 also 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 CTCTCCTCTT TGCCTCAATC AATCAATTTC ATCAAATCAA ACATCACCCA TCAACCCCAT

-81 CATTGACCAG AACTTCCAGA AAATAAAGAC GTGAAAGAAG AAGTAATAGT ATTCTTTCAT AATTATAAAA CATTTTGATA

|→ LSU rRNA

1 ATTGTACTAC CATGGGATCA ATAGGATGTC ATAACGATGA AGAACATAAA AGAATATGAT AAAACATAAT CTTTGAATTC

LS228F

81 TAATTCATTT TAGGATAACC CTTTGAACTT AAGCATATCA GTAAAAGGAG GAAAAGAAAC TAACAAGGAT TTCTTTAGTA

161 GTGGCGAATG AACAAGAAAG AACTCATTAT GTAATCTATA TTCATTATAG AGATGTTAAA AGTTATTGTT GTCATTCAAA

241 TCATAGATAC ATTAAATTGT AGTAAATGAT GACGATAAAT AAACAATGTA GAGTACGATT GCTTGGTAGT GCAATTGGAA

321 TATAGATAGA ATAAGATATC TAAGTTAAAT ATAAATGACA TACCGATAGA GAATAAGTAC TGCGAAGGAA CTTGTGAAAA

401 TGGTAGTATT AGCTTATAGT AATTATATAA GACCCGTCTT GAAACACGGA CCAAGGAGAT TACCAGATGA CAGAATTAAA

HG4R

481 GAATCATAGG TGATTTGTCA TCTGGTAAGA CCCGAAACAT AGTGAACTAT ATATGTCAAT GGTTGAAGAT AAACGACCGT

561 TTATTGGAAG ACCTAAGTCA TTCTGACGTG CAAATCGATG GTGAAAGATG TATATAGTGG CGAAAGACCA ATCGAACTAT

641 GTGGTAGCTG GTTCACAGCG AAATGTCTCT TGGGACAGTT GAGTATAAAT AAAGTAAATA GATGTAGAGT CAATACAATG

721 TTTAATTATA TTGTTTAACT TCTAATTCTA TAGGTTTGTA TGTTTTATAA ATACTTACTG ATATTTACGA TACTCAAGTG

801 GGCGTATGAT TGTAAGAATG ATATGCAATA AGGGATGAAC CTTGAGATGC ACTAAAATTT CTAATTGTAT TATAATTATA

881 AATATGATGA ATACATTATG ACAGTAGGGC GGTTGTTATG GAAGTAGAAA TCCGCTAAGA AACGTGTTAC AACGTACCTA

961 CCGAATGTAT TATTGTATAA AATGGATGAA TATTATAATA GTATTGAAGT GTGTTAAATA ACAAGTATGT TATTTGGTAG

1041 TTGTGTAATT TGAATATGTT GAAGTATGTA AGTTATTATA TATTGAATAA TCAAATGAGC AGATCTTGGT GTAGTAACAT

1121 AATAATTTGT ATTTATAGAG ATTAGGGTTG TCATTATGAT GACGTGAGAC GGTGTTATAA TAAAACGAAT AGTATTAATA

1201 TTTTATTATC TTATTTATTG AGTGGTGACA CAAATATGGA CTATTATTGT AAAAGATTAT AATATAAATG AGTTATAATT

1281 TGTATATGAT GTGATAGTTG TTATAAGATC ATCATTTGAA TAATAATATG TTAAAAAAGA AATAAATAAA CCGTACCTAT

1361 ACCGCATCAG GTGTCTCTGT TTTATAACAA ATATATATAA AGTAATCTAA GTAAGGGAAT TCGGCAAAAT AGATCTGTAA

1441 CTTTGGGATA AAGATTGGCT CTAGCATGCT AGAACTTTTT ATATTATGAG AGGAATCTGA CTGTTTAATA AAAACATAGC

1521 TTTATGATAA TGTAAAGTGA ATTCTGCCCA GTGTATTTAT TAGTTAAAAT CGAATAAGCA TATATAAACG GCGGGAGTAA

1601 CTATGACTCT CTTAAGGTAG CCAAATGCCT CGTCATTTAA TTGGTGACGC GCATGAATGG AGCAACGAGA TTCCTACTGT

1681 CCCTATTTAG AGCTATGTGA AGCAACGAGA CAAGGGAACG GGCTTGTTAT ATATCAGCGG GGAAAGAAGA CCCTGTTGAG

1761 CTTTACTCTA GTATATTTCA ATTTGTATAT ATTATATTGT AGAGAGGTGG GAGATGTAAT GTGAAACCAC TAATATTAAT

1841 ATTATATGAA TTTATTGTAA TTATATGGGG AGTTTGGCTG GGGCGGCACA ACTGTTATAA AGAAACACAG TTGTCCTAAG

1921 ATACAATTAA TTAGGATAGA AACCTAATGT TCAATATAAG GGTATAAATG TGTCTGATGT ATAGAACAAT AGTTTTATAT

2001 GATGGGAAAC CATGGCCTAA AGATCCTCCA ATAATAGTTT TATTTGATTG GAGGTGACAG
AAAAGTTACC ACAGGGATAA

2081 CTGGCTTGTA GCAAATGAGC GATCATAGCG ACTTTGCTTT TTGATTCTTC GATGTCGTCT
CTTCTGATCA TCGTAGTGTA

├————ILSUF———┤

2161 TATGTTACGA AGTGTTGGAT TGTTCACCCG TTAATGAGGA ACGTGAGATG GGTTTAGACC
GTCGTGAGAC AGGTTAGTTT

├————————ILSUR————————┤

2241 TACCCTACTG TTTATATTAT TAAGAAAAGT TATATAGTAC GAGAGGAACA ATAAGTGATG
ACCTCTGGTG TAGCGATTAA

2321 TTATTCAATT ATGTTGCAAA ACTACGTCAT GAATAATAAA TGCTGAAAGC CTCTTAAGCA
TGAAGTTAAT CTTAATGATA

2401 GACAATTGAG GGATACTACC TCTCTGAAGA GTGGGACTGT AAGAATGAAG GTGTATACCG
ACATTTTGAG GTTACCATTT

LSU rRNA ←|→ ITS region

2481 CGTTCTGTAT AAAAATAcct tcacgtggta cagaatgtat tgattatgtt tttatcaatc taaaagatct aacttgtatt

2561 tttttatatt catattgtgt attgttgttt taattttgat tattatattt tatcattctt tttattatta ttttctagtt

ITS region ←□→ SSU rRNA

├————18f————┤

2641 gtgttatcgt attttcattt ataaataagt tgtaaaCACC AGGTTGATTC TGCCTGACGT AGACGCTATG CTCTAAGATT

├————S33R————┤

2721 AACCCATGCA TGTTTATTGA ATATAAAGAA AAGACGAACA GCTCAGTAAC TCTTATTTGA
TTTGATGTAT TAGGATTATA

2801 ACTATGTTAA ATTATAGGTA ACAATAATAC AATAAGAATA AGATCTATCA GTTAGTTGTT
AAGGTAATGG CTTAACAAGA

2881 CTATGACGGA TAACGGTATT ACTTTGTAAT ATTCCGGAGA AGGAGCCTGA GAGATTGCTA
CTAAGTCTAA GGATTGCAGC

2961 AGGGGCGAAA CTTGACCTAT GATATTATAT TGAGGCAGTT ATGAGTAGTA TTTTATAATT
ATTGTAGTAT TGTAAGTACA

3041 TATTACAAGA TAAATCGGAG GGCAAATCGA GTGCCAGCAG CCGCGGTAAT ACTTGTTCCG
ATAGTGTGTA TGATGATTGA

3121 TGCAGTTAAA AAGTCTGTAG TTTATTTATA ATAAGCATTG TAAGGTATAC AGTATGGTTA
GGAGAGAGAT GAAATGTGAT

3201 AACCCTAACT GGATGAACAG AAGCGAAAGC TGTATACTTA AATGTATTAT TAGAACAAGG
ACGTAAGCTA GAGGATCGAA

3281 GATGATTAGA TACCATTGTA GTTCTAGCAG TAAACTATGT TGAATCATAG ATATATTTTG
ATATATATTT ATGTAGAGAA

3361 ATTAAGATTA TATTGACTCT GGGGATAGTA TGATCGCAAG ATTGAAAATT AAAGAAATTG
ACGGAAGAAT ACCACAAGGA

HG4F

3441 GTGGATTGTG CGGCTTAATT TGACTCAACG CGGGGTAATT TACCAGGTAT AACATGGTAT
AATATTTTAT CATGATAGTG

3521 GTGCATGGCC GTTTCCAATG GATGCTGTGA AGTAATGATT AATTTCAACA AGATGTGAGA
CCCTCATTTA GACAGATGTA

3601 GTGATACATA TGAAGGAGAG GATTAAAACA GGTCCGTTAT GCCCTAAGAT AATCTGGGTT
GCACGCGCAA TACAATAATA

3681 TTTGATATTA TAAGGGATAA TATAATGTAA GATATATTTG AACATGGAAT TGCTAGTAAA
TTTTATTTAA TAAGTAGAAT

3761 TGAATGAGTC CCTGTTCTTT GTACACACCG CCCGTCGCTA TCTAAGATGG TATTATCTAT
GAACAAATTT ATAAAGTGAA

SSU rRNA ← | → IGS region

3841 TAGATAGTAC TAGATCTGAT ATAAGTCGTA ACATGGTTGC TGTTGGAGAA CCATTAGCAG
GATCATAAat aaaaaagatg

1537r

3921 ggattattat aatgaaagtg tatgctgaca ttataataat cttgtctatg ttctaaatac aaaatacctt cacgaggttc

4001 agaatgggtg atataaggag gttatattgg ctaaatttat gtatgtgaga aaatgatgtt gtgatgttaa aaggcgtgtg

4081 tagttatgtg tagagaattt tattcaagta taatgtatgt ggtttctgct ttgctttata ctttttatgt tcaggtagca

IGS region ←□→ 5S rRNA

4161 tacatataac aacacacttg gttcatcAGA TACGGTCATA TCTACTGAAA ACCACTGGAA CCCACCAGAA
CTCCGAAGTT

5S rRNA ←□

4241 AAACCAGATG AGCTTAATCA GTACTAAGAA GGGAGACCAC TTGGGAACGT TGGGTGCTGT A

5SR

References

Baker, M.D., Vossbrinck, C.R., Maddox, J.V., Undeen, A.H., 1994. Phylogenetic relationships among *Vairimorpha* and *Nosema* species (Microspora) based on ribosomal RNA sequence data. J. Invertebr. Pathol. 64, 100–106.

Baker, M.D., Vossbrinck, C.R., Didier, E.S., Maddox, J.V., Shadduck, J.A., 1995. Small subunit ribosomal DNA phylogeny of various microsporidia with emphasis on AIDS related forms. J. Eukaryot. Microbiol. 42 (5), 564–570.

Becnel, J.J., Andreadis, T.G., 1999. Microsporidia in insect. In: Wittner, M., Weiss, L.M. (Eds.), The Microsporidia and Microsporidiosis. ASM Press, Washington, DC, pp. 447–501.

De Rijk, P., De Wachter, R., 1993. DCSE, an interactive tool for sequence alignment and secondary structure research. Comput. Appl. Biosci. 9, 735–740.

De Rijk, P., De Wachter, R., 1997. RnaViz, a program for the visualization of RNA secondary structure. Nucleic Acids Res. 25, 4679–4684.

De Rijk, P., Gatehouse, H.S., De Wachter, R., 1998b. The secondary structure of *Nosema apis* large subunit ribosomal RNA. Biochim. Biophys. Acta 1442, 326–328.

De Rijk, P., Wuyts, J., De Wachter, R., 2003. RnaViz 2: an improved representation of RNA secondary structure. Bioinformatics 19 (2), 299–300.

De Rijk, P., Caers, A., Van de Peer, Y., De Wachter, R., 1998a. Database on the structure of large ribosomal subunit RNA. Nucleic Acids Res. 26, 183–186.

Galtier, N., Gouy, M., 1995. Inferring phylogenies from DNA sequences of unequal base compositions. Proc. Natl. Acad. Sci. USA 92, 11317–11321.

Gatehouse, H.S., Malone, L.A., 1998. The ribosomal RNA gene region of *Nosema apis* (Microspora): DNA sequence for small and large subunit rRNA genes and evidence of a large tandem repeat unit size. J. Invertebr. Pathol. 71, 97–105.

Germot, A., Philippe, H., LeGuyader, H., 1997. Evidence for loss of mitochondria from a mitochondrial-type HSP 70 kDa in *Nosema locustae*. Mol. Biochem. Parasitol. 87, 159–168.

Gresoviac, S.J., Khattra, J.S., Nadler, S.A., Kent, M.L., Devlin, R.H., Vivares, C.P., Fuente, E. DeLa, Hedrick, R.P., 2000. Comparison of small subunit ribosomal RNA gene and internal transcribed spacer sequences among isolates of the intranuclear microsporidian *Nucleospora salmonis*. J. Eukaryot. Microbiol. 47, 379–387.

Hatakeyama, Y., Kawakami, Y., Iwano, H., Inoue, T., Ishihara, R., 1997. Analyses and taxonomic inferences of small subunit ribosomal RNA sequences of five microsporidia pathogenic to the silkworm, *Bombyx mori*. J. Seric. Sci. Jpn. 66, 242–252.

Hirt, R.P., Healy, D., Vossbrinck, C.R., Canning, E.U., Embley, T.M., 1997. Identification of a mitochondrial HSP 70 orthologue in *Vairimorpha necatrix*: molecular evidence that microsporidia once contained mitochondria. Curr. Biol. 7, 995–998.

Huang, H.W., Lo, C.F., Tseng, C.C., Peng, S.E., Chou, C.M., Kou, C.H., 1998. The small subunit ribosomal RNA gene sequence of *Pleistophora anguillarum* and the use of PCR primers of diagnostic detection of the parasite. J. Eukaryot. Microbiol. 45, 556–560.

Kawakami, Y., Inoue, T., Ito, K., Kitamizu, K., Hanawa, C., Ando, T., Iwano, H., Ishihara, R., 1994. Identification of a chromosome harboring the small subunit ribosomal RNA gene of Nosema bombycis. J. Invertebr. Pathol. 64, 147–148.

Kawakami, Y., Inoue, T., Kikuchi, M., Takayanagi, M., Sunairi, M., Ando, T., Ishihara, R., 1992. Primary and secondary structures of 5S ribosomal RNA of *Nosema bombycis* (Nosematidae, Microsporidia). J. Seric. Sci. Jpn. 61, 321–329.

Keeling, P.J., 2003. Congruent evidence from α-tubulin and β-tubulin gene phylogenies for a zygomycete origin of microsporidia. Fungal Genet. Biol. 38, 298–309.

Keeling, P.J., Doolittle, W.F., 1996. α-Tubulin from early-diverging eukaryotic lineages and the evolution of the tubulin family. Mol. Biol. Evol. 13, 1297–1305.

Keeling, P.J., Fast, N.M., 2002. Microsporidia: biology and evolution of highly reduced intracellular parasites. Annu. Rev. Microbiol. 56, 93–116.

Keeling, P.J., Luker, M.A., Palmer, J.D., 2000. Evidence from β-tubulin phylogeny that microsporidia evolved from within the fungi. Mol. Biol. Evol. 17, 23–31.

Peyretaillade, E., Biderre, C., Peyret, P., Duffieux, F., Metenier, G., Gouy, M., Michot, B., Vivares, C.P., 1998. Microsporidian *Encephalitozoon cuniculi*, a unicellular eukaryote with an unusual chromosomal dispersion of ribosomal genes and a LSUrRNA reduced to the universal core. Nucleic Acids Res. 26, 3513–3520.

Sprague, V., Becnel, J.J., Hazard, E.I., 1992. Taxonomy of phylum Microspora. Crit. Rev. Microbiol. 18, 285–395.

Steinhaus, E.A., 1949. Principles of Insect Pathology. McGraw-Hill, New York.

Tanada, Y., Kaya, H.K., 1993. Insect Pathology. Academic Press, New York.

Tsai, S.J., Kou, G.H., Lo, C.F., Wang, C.H., 2002. Complete sequence and structure of ribosomal RNA gene of *Heterosporis anguillarum*. Dis. Aquat. Org. 49, 199–206.

Tsai, S.J., Lo, C.F., Soichi, Y., Wang, C.H., 2003. The characterization of microsporidian isolates (Nosematidae: Nosema) from five important lepidopteran pest in Taiwan. J. Invertebr. Pathol. 83, 51–59.

Undeen, A.H., Cockburn, A.F., 1989. The extraction of DNA from microsporidia spores. J. Invertebr. Pathol. 54, 132–133.

Van de Peer, Y., Ali, A.B., Meyer, A., 2000. Microsporidia: accumulating molecular evidence that a group of amitochondriate and suspectedly primitive eukaryotes are just curious fungi. Gene 246, 1–8.

Van de Peer, Y., Caers, A., De Rijk, P., De Wachter, R., 1998. Database on the structure of small ribosomal subunit RNA. Nucleic Acids Res. 26, 179–182.

Vossbrinck, C.R., Woese, C.R., 1986. Eukaryotic ribosomes that lack a 5.8S RNA. Nature 320, 287–288.

Vossbrinck, C.R., Baker, M.D., Didier, E.S., Debrunner-Vossbrinck, B.A., Shadduck, J.A., 1993. Ribosomal DNA sequences of Encephalitozoon hellem and Encephalitozoon cuniculi: species identification and phylogenic construction. J. Eukaryot. Microbiol. 40, 354–362.

Vossbrinck, C.R., Maddox, J.V., Friedman, S., Debrunner-Vossbrinck, B.A., Woese, C.R., 1987. Ribosomal RNA sequence suggests microsporidia are extremely ancient eukaryotes. Nature 326, 411–414.

Weiss, L.M., Vossbrinck, C.R., 1999. Molecular biology, molecular phylogeny, and molecular diagnostic approaches to the microsporidia. In: Wittner, M., Weiss, L.M. (Eds.), The Microsporidia and Microsporidiosis. ASM Press, Washington, DC, pp. 129–171.

Williams, B.A.P., Hirt, R.P., Lucocq, J.M., Embley, T.M., 2002. A mitochondrial remnant in the microsporidian *Trachipleistophora homins*. Nature 418, 865–869.

Wittner, M., 1999. Historic perspective on the microsporida: expanding horizons. In: Wittner, M., Weiss, L.M. (Eds.), The Microsporidia and Microsporidiosis. ASM Press, Washington, DC, pp. 1–6.

Wuyts, J., De Rijk, P., Van de Peer, Y., Winkelmans, T., De Wachter, R., 2001. The European large subunit ribosomal RNA database. Nucleic Acids Res. 29, 175–177.