African Journal of Biotechnology Vol. 11(31), pp. 7819-7835, 17 April, 2012
Available online at http://www.academicjournals.org/AJB
DOI: 10.5897/AJB11.1229
ISSN 1684–5315 © 2012 Academic Journals

Full Length Research Paper

The characterization of cytoplasmic ribosomal protein genes in microsporidian *Nosema bombycis* Genome

Handeng Liu[1,2], Guoqing Pan[1]*, Tian Li[1], Wei Huang[1] and Zeyang Zhou[1,3]

[1]Institute of Sericulture and Systems Biology, Southwest University, Chongqing 400716, P.R. China.
[2]Experimental Teaching Center, Chongqing Medical University, Chongqing 400016, P.R. China.
[3]Laboratory of Animal Biology, Chongqing Normal University, Chongqing 400047, P.R. China.

Accepted 31 August, 2011

Microsporidia are obligate intracellular, eukaryotic parasites of medical and commercial importance, which can infect almost all animals including humans. However, their ribosomes are not of the 80S type as other eukaryotes, but like the prokaryotic 70S ribosome. In order to get the global composition of ribosomal protein genes of *Nosema bombycis*, the pathogen of Pébrine, and their comparative genomics' characteristics, a genome-wide survey in *N. bombycis* genome was performed. From the results, we identified 130 CDSs corresponding to 73 ribosomal protein genes. Among them, three ribosomal protein genes (*RPL19*, *RPS4* and *RPS18*) with short introns (23 or 24 bps) were verified by *N. bombycis* ESTs, and they have the same structure among microsporidia. The novel arrangements of 'AAATTT-like signal – CCC/GGG-like motif – transcription start site' are present in the upstream sequences of ribosomal protein genes, and several regulatory elements that may have synergy with introns of ribosomal protein genes for its high transcriptional frequency were detected too. 76.7% ribosomal protein genes of *N. bombycis* were located in syntenic blocks, indicating that their gene order was conserved among microsporidian species. And phylogenic trees show its ancient eukaryotic position too. The characterization of the total ribosomal proteins contributes a first step to ribosomal proteins' transcription regulation, evolution of microsporidia.

Key words: Microsporidian, ribosomal protein, *Nosema bombycis*, transcription regulation, evolution.

INTRODUCTION

Ribosome is a ribonucleoprotein complex that is composed of rRNAs and proteins, and is responsible for the synthesis of polypeptide chains in all living cells. The 70S ribosome belongs to prokaryote, while eukaryotes harbor 80S ribosome. Atomic resolution crystal structures of prokaryotic ribosome suggest that the two fundamental activities of ribosome (peptidyl transfer at the P site and decoding activity at the A site) are carried out by the RNA (Ban et al., 2000; Nissen et al., 2000; Ogle et al., 2001). The structure and function of both prokaryotic and eukaryotic ribosome have been investigated. The cytosolic ribosome is composed of a large number of ribosomal proteins (RPs) and distinct rRNAs (three distinct rRNAs in prokaryotes and four in eukaryotes). Prokaryotic ribosome and ribosomal proteins have provided knowledge of ribosome structure and composition. Three dimensional structures of small and large ribosomal subunits of thermophilic eubacteria have been earlier described at 5.5- and 2.5-Å resolution from crystallographic data (Ban et al., 1999, 2000; Clemons et al., 1999), respectively. In total, 55 ribosomal proteins have been identified in *Escherichia coli* and their amino acid sequences have been determined (Wittmann, 1982; Wittmann-Liebold et al., 1990). The ordered assembly process of eubacterial ribosome was also documented too (Nomura et al., 1984; Culver et al., 1999). As well known, ribosomes of an archaebacterial ancestor gave rise to the cytosolic ribosomes of eukaryotes (Matheson et al., 1990; Wittmann-Liebold et al., 1990; Wool et al., 1995). However, the ribosomal proteins of plastids and mitochondria indicate evolutionary similarity to those of

*Corresponding author. E-mail: gqpan@swu.edu.cn. Tel: +86-23-68251088. Fax: +86-23-68251128.

eubacteria and include organelle-specific proteins (Graack and Wittmann-Liebold, 1998; Koc et al., 2000; Yamaguchi and Subramanian, 2000; Yamaguchi et al., 2000). In eukaryotes, the ribosomal protein component of *Rattus norvegicus* was determined by direct protein sequencing followed by gene cloning and a presumed complete set of 79 ribosomal proteins was compiled (Wool et al., 1995). In addition, genes corresponding to 78 ribosomal proteins of *Saccharomyces* cerevisiae were also identified through genome sequencing efforts (Goffeau et al., 1996; Planta and Mager, 1998). The evolutionary conservation of ribosomal proteins is not surprising given the constraints of rRNA-protein interactions, coordinated ribosome assembly, and ribosome function. Actually, phylogenetic relationships between animal, fungi, and plant kingdoms have been obtained by comparison of orthologous ribosomal proteins (Veuthey and Bittar, 1998).

Also, the distribution of ribosomal protein genes of both prokaryotes and eukaryotes has been examined. In eubacteria, lots of the ribosomal protein genes are clustered in a few operons, which allows for coordinated regulation (Nomura et al., 1984). For humans, 75 ribosomal protein genes have been located and they are distributed on all chromosomes, with a bias toward chromosome 19 (Kenmochi et al., 1998b). Synthesis of ribosomal proteins in eukaryotes requires coordination of now unlinked genes. It is conspicuous that the regulation of ribosomal protein gene expression appears to occur at the transcriptional level in *S. cerevisiae* (Planta and Mager, 1998) and predominantly at the translational level in animals (Meyuhas, 2000; Meyuhas and Hornstein, 2000).

For the ribosome of microsporidia, which is same as the other two amitochondriate groups of unicellular eukaryotes, diplomonads and parabasalids; they exhibit certain distinct features when compared with ribosomes of 'typical' eukaryotes such as *S. cerevisiae* or *R. norvegicus*. These features are that the 70S sedimentation coefficient of microsporidian (*N. bombycis*) (Ishihara and Hayashi, 1968) and trichomonad (Champney et al., 1992) ribosomes have shorter small and large subunit ribosomal RNAs (Vossbrinck et al., 1987; Sogin et al., 1989; Chakrabarti et al., 1992; Healey et al., 1990; Philippe and Germot, 2000; Peyretaillade et al., 1998) than those found in 'typical' eukaryotes, and the presence of fewer proteins (40 to 56) in trichomonad ribosomes (Champney et al., 1992), which is absent in microsporidia of the internal transcribed spacer 2 present in the LSU rRNA region of all other eukaryotes (Weiss and Vossbrinck, 1999). Moreover, a covalent link joins the 5.8S region with the LSU as seen in prokaryotes. Previously, these features were regarded as one part of evidence that made them 'prokaryotic' or 'primitive eukaryotic', although, more recent work indicates that none of these provide a compelling argument for such status (Roger, 1999). In addition, recent phylogenetic studies have consistently supported a placement of these organisms as a basal lineage of the fungi (Franzen et al., 2006; Germot et al., 1997; Hirt et al., 1997; Peyretaillade et al., 1998). These parasites are characterized by small genomes ranging from only 2.3 Mb to 23 Mb (Peyretaillade et al., 1998; Belkorchia et al., 2008), a trait reflected in the short length of most putative proteins compared to eukaryote orthologues and compact gene organization. To date, over 1300 species of Microsporidia (in 160 genera) have been formally described in the literature, based on their cellular structure, life cycle, and host specificity (Corradi and Keeling, 2009). Unfortunately, the characterization of ribosome does not encompass these features for all three groups, and thus, it remains to be established whether the protein components really separate these groups from the 'typical' eukaryotes.

Till date, there are a few reports about the ribosomal proteins of microsporidia. Using a genome-wide investigation of synteny between a broader range of fungi and three microsporidia, the only conserved gene pair was *RPL21* and *RPS9* (Lee et al., 2008), which is intriguingly conserved across all fungi, but not in species outside the fungal kingdom, suggesting an ancient functional role for this gene pair within the fungi. After a small spliceosomal-type intron had been discovered in a ribosomal protein gene of Encephalitozoon cuniculi (Biderre et al., 1998), 11 ribosomal protein genes which harbor a short intron sequence have been found (Katinka et al., 2001). In *Nosema ceranae* genome, six ribosomal protein genes with predicted short introns have been identified, only five of them are orthologs to ribosomal protein genes of *E.* cuniculi (Cornman et al., 2009). However, no spliceosomal intron has been found in ribosomal protein genes in the Enterocytozoon *bieneusi* genome data (Akiyoshi et al., 2009). And *RPL5* gene, which contains an intron in *E. cuniculi*, does not contain an intron in *Edhazardia aedis* (Gill et al., 2008).

As the first discovered microsporidia, *N. bombycis* has been studied since the middle of the nineteenth century. *N. bombycis* is an obligate intracellular parasite of the domesticated silkworm, and as the causative agent of pebrine disease, is still prevalent in sericulture. Through pulse-field gel electrophoresis (PFGE), the number of chromosomes in *N. bombycis* was estimated at 18, ranging from approximately 380 to 1500 kbp, with a total size of approximately 15,330 kb (Kawakami et al., 1994). Recently, the genome project of *N. bombycis* had been performed in our laboratory. To date, a whole genome shotgun database with 7.8-fold coverage of the whole genome and a cDNA library with 11,155 expressed sequence tags were constructed successfully. For ribosomal RNA, we performed a genome-wide analysis and find that the ribosomal RNA of *N. bombycis* is multiple and distributed on all chromosomes (Liu et al., 2008), but the characteristics of ribosomal proteins are still unknown.

In this study, a genome-wide analysis was performed to the cytoplasmic ribosomal protein genes (*RPGs*) of *N.*

bombycis and the total ribosomal protein genes have been identified. At the same time, the characteristics and genome distribution of all *RPGs* have been displayed and there are three ribosomal protein genes containing short intron sequences verified by ESTs. Also, the core promoter and regulatory elements corresponding to ribosomal protein genes are predicted in the upstream sequence. In order to get more information among different microsporidian, syntenic maps and phylogenic trees are constructed for all ribosomal protein genes.

MATERIALS AND METHODS

Based on the *N.* bombycis genome, in total, we perform a genome-wide survey to search the genome annotation table using the key words “ribosomal protein”. Also, using the annotation ribosomal protein gene sequences in other organisms (including *E*. cuniculi, Antonospora locustae, E. bieneusi, N. ceranae and S. cerevisiae) as query, a BLAST searching was conducted on N. bombycis genome in order to find more ribosomal protein gene orthologues. In order to display the distribution of ribosomal proteins on N. bombycis genome, pictures of ribosomal protein genes located on genome have been plotted using a script of Perl programmer (drawContigStructure). Some ribosomal protein genes which clustered on one scaffold have been exhibited too. In other microsporidian, such as E. cuniculi, N. ceranae and A. locustae, ribosomal protein genes which harbor intron have been documented before (Katinka et al., 2001; Cornman et al., 2009; Peyretaillade et al., 2009). However, we haven’t found intron in another microsporidian E. bieneusi. So, using the same method, we searched introns in ribosomal protein genes of N. bombycis.

The core promoter regions of ribosomal protein genes were predicted by a promoter-finding program, which was designed on the basis of time-delay neural network, Neural Network Promoter Prediction (http://www.fruitfly.org/seq_tools/promoter.html). The parameter of minimum promoter score of each gene was set at 0.8 (Reese, 2001). After predicting the promoter of the ribosomal protein genes and aligning their sequences by ClustalX software, we treated the core promoters using Weblogo (http://weblogo.berkeley.edu/logo.cgi). The position of the nucleotide sequences was calculated based on the distance from the initiation site of the predicted transcription start site (TSS, +1). The regions 500 bps upstream from the transcription initiation site were used for analyzing the potential binding sites for the transcription regulatory motifs by NSITE program (http://www.softberry.com/ berry.phtml).

As a class of very conservative gene, ribosomal protein genes have some conserved characteristics on gene order among the five sequenced microsporidia. To compare the locations of different microsporidian RPGs, syntenic maps of RPGs containing regions among E. cuniculi, A. locustae, N. ceranae, E. bieneusi and N. bombycis were plotted with a Perl script based on the genome data of N. bombycis, E. cuniculi (http://www.ncbi.nlm.nih.gov), A. locustae (http: //jbpc.mbl.edu/Nosema), N. ceranae (http://www.ncbi.nlm.nih.gov) and E. bieneusi (http://www.ncbi.nlm.nih.gov). Othologous genes were identified using BLASTP program (Altschul et al., 1997). We retrieved almost all ribosomal protein-related sequences by searching NCBI database (http:// www.ncbi.nlm.nih.gov/) using BLASTP (Altschul et al., 1990). Multiple sequence alignments were initially made using the program ClustalX version 1.83. These alignments were then reconciled and further adjusted by eye to minimize insertion/deletion events. A completed ribosomal protein sequence region of each RPGs was used in the subsequent phylogenetic analyses, which included the RPGs signature motif. Phylogenetic trees were reconstructed using the neighbor-joining method (Saitou and Nei, 1987) implemented in MEGA 4.0 program (Tamura et al., 2007). Bootstrap support was evaluated based on 1000 replicates.

RESULTS

Characteristics of ribosomal protein genes in N. bombycis

As other organisms, N. bombycis also has lots of ribosomal proteins which combine with rRNA to form a complete ribosome. In N. bombycis genome, we identified a total of 130 ribosomal protein (RP) gene hits (including both complete and incomplete sequences) corresponding to 73 putative cytoplasmic ribosomal protein types, which include 30 putative small-subunit proteins encoded by 51 RP gene hits and 43 putative large-subunit proteins encoded by 79 RP gene hits. There are 72 ribosomal protein genes in E. cuniculi genome and the identity of all ribosomal proteins between N. bombycis and E. cuniculi is around 33 to 77% at amino acid level, and RPL10, RPL23, RPS23 and RPS26 are 70, 77, 75 and 72%, respectively. This indicates that these housekeeping genes are conserved in microsporidian species.

Furthermore, there were only six multiple copy genes in E. cuniculi. However, more than half of RPGs are multiple in N. bombycis in spite of containing some incomplete sequences and it is mainly the length difference among different copies of RPGs. Among all RPGs, we only found 29 ribosomal protein genes which have EST proofs (about 368 EST hits) in NbBmEST database. Domain of each ribosomal protein sequence was predicted on http://smart.embl-heidelberg.de/. Molecular weight and theoretical pI also were predicted by the BioXM software. All basic characteristics of each ribosomal protein have been displayed in Table 1.

Ribosomal protein genes’ distribution on N. bombycis genome

In N. bombycis genome, 130 ribosomal protein genes were distributed on 92 superscaffolds. Among them, there are 26 superscaffolds harbor at least two RP genes. Among them, there are four superscaffolds (nbo2, nbo6, nbo10 and nbo13), which contain more than four RP genes on each superscaffold (Figure 1). In addition, we found some RP genes are multiple in N. bombycis genome. Four pairs of RP genes were found not only multiple, but also in a tandem order distributed on different superscaffolds (Figure 2). RPL8-A gene and RPS17 gene were located on superscaffold nbo2, nbo18 and nbo1552 (Figure 2a); RPL10 gene and RPS19 gene were located on superscaffold nbo12 and nbo419 (Figure 2b); RPL13 gene and RPS16 gene were located on superscaffold nbo16 and nbo386 (Figure 2c); RPL17 and ubiquibitin/L40 were located on superscaffold nbo67 and nbo98 (Figure 2d). And the transcriptive directions of tandem RP genes are conserved among its corresponded superscaffolds,

Table 1. Identification and characteristics of ribosomal protein genes of *N. bombycis*.

Annotation of ribosomal protein	Location on genome	GenBank Accession No.	Number of amino acids (Aa)	Domain (SMART predicted)	intron	GC content (%)	Copy number in genome	Molecular weight (kDa)	Theoretical pI	EST proof
RPSA	NBO_554 g 0006	HQ291394	260	Ribosomal_S2	NF	34.36	1	29.39	5.22	N
	NBO_64 g 0055	HQ291395	99(NC)	NO	NF	32.33	1	11.50	4.52	N
RPS2	NBO_81 g 0008	HQ291367	116(NC)	Ribosomal_S5	NF	38.46	3	12.67	5.84	N
	NBO_1209 g 0001	HQ291366	197(NC)			39.06		21.37	9.15	N
	NBO_1208 g i001	HQ291365	240			37.81		26.46	9.38	N
RPS3	NBO_41 g 0036	HQ291390	214	Ribosomal_S3	NF	32.71	1	23.60	9.52	N
RPS3aE	NBO_76 g 0005	HQ291405	235	Ribosomal_S3Ae	NF	33.47	1	26.41	9.96	N
RPS4	NBO_13 g 0033	HQ291370	172	Ribosomal_S4e	YES	32.76	1	19.82	5.11	Y
RPS5	NBO_70 g 0002	HQ291404	205	Ribosomal_S7	NF	34.63	1	23.10	10.01	Y
RPS6	NBO_53 g 0020	HQ291392	218	Ribosomal_S6e	NF	35.01	2	24.85	9.8	Y
	NBO_53 g 0021	HQ291393	218			35.01				
RPS7	NBO_20 g 0021	HQ291369	74	NO	NF	32	2	8.28	6.71	Y
	NBO_1210 g 0002	HQ291368	163			30.28		18.85	6.33	Y
RPS8-B	NBO_10 g 0072	HQ291357	170	Ribosomal_S8e	NF	35.87	3	19.00	10.27	Y
	NBO_996 g 0002	HQ291359				35.67		19.00	10.27	Y
	NBO_1167 g 0001	HQ291358				35.28		19.03	10.27	Y
RPS9	NBO_72 g 0004	HQ291399	187	Ribosomal_S4	NF	35.64	2	21.56	9.99	N
	NBO_594 g i001	HQ291398	50(NC)			34.00		6.06	10.28	N
RPS11	NBO_64 g 0024	HQ291403	122	Ribosomal_S17	NF	36.59	1	13.96	10.26	N
RPS12	NBO_29 g 0033	HQ291382	149	Ribosomal_L7Ae	NF	29.56	1	17.03	8.35	N
RPS13	NBO_13 g 0045	HQ291372	1145	Ribosomal_S13; Ribosomal_S15	NF	31.27	2	132.25	5.32	Y
	NBO_445 g i001	HQ291373	149			34.44				
RPS14	NBO_54 g 0017	HQ291362	135	Ribosomal_S11	NF	41.18	2	14.48	10.49	Y
	NBO_1061 g 0002	HQ291361	111			42.26		11.84	10.94	Y
RPS15	NBO_56 g 0004	HQ291396	148	Ribosomal_S19	NF	33.11	2	16.84	9.59	N
	NBO_59 g i001	HQ291397				33.11		16.87	9.62	N
RPS15a	NBO_6 g 0022	HQ291400	127	Ribosomal_S8	NF	34.38	1	14.55	9.59	Y
RPS16	NBO_16 g 0058	HQ291376	146	Ribosomal_S9	NF	41.72	2	16.67	9.65	Y
	NBO_386 g 0016	HQ291377				41.5		16.65	9.60	Y
RPS17	NBO_2 g 0034	HQ291379	120	Ribosomal_S17e	NF	37.19	2	14.06	9.52	Y
	NBO_18 g 0032	HQ291378				36.64		14.06	9.52	Y
RPS18	NBO_66 g 0017	HQ291375	115	Ribosomal_S13	YES	37.93	2	13.16	10.72	Y
	NBO_1315 g 0002	HQ291374				37.36		13.16	10.72	Y

Table 1. Continues.

RPS19	NBO_419 g 0009	HQ291364	101	Ribosomal_S19	NF	24.42	2	11.90	9.59	N
	NBO_12 g 0023	HQ291363				23.76		11.86	9.59	N
RPS20	NBO_1009 g 0002	HQ291360	119	Ribosomal_S10	NF	34.17	1	13.75	5.38	Y
RPS21e	NBO_32 g 0001	HQ291384	68	Ribosomal_S21e	NF	33.82	2	7.50	6.14	Y
	NBO_86 g 0004	HQ291385				34.3		7.50	6.14	Y
RPS23	NBO_3 g 0051	HQ291383	389	Ribosomal_S12	NF	30.76	1	45.25	9.73	N
RPS24-A	NBO_6 g 0054	HQ291402	131	Ribosomal_S24e	NF	37.12	1	15.13	10.16	N
RPS25	NBO_13 g 0044	HQ291371	67	NO	NF	34.8	1	7.86	9.80	Y
RPS26-3	NBO_37 g 0009	HQ291387	103	Ribosomal_S26e	NF	33.97	3	11.90	10.17	N
	NBO_44 g 0012	HQ291388				33.65				
	NBO_641 g 0001	HQ291389				34.62				
RPS28e	NBO_41 g 0041	HQ291391	67	Ribosomal_S28e	NF	35.29	1	7.56	9.25	N
RPS30	NBO_6 g 0036	HQ291401	60	Ribosomal_S30	NF	39.89	1	7.10	11.31	N
RPS31	NBO_367 g 0007	HQ291386	138	Ubiquitin; Ribosomal_S27	NF	32.13	3	15.82	9.13	Y
	NBO_366					32.13		15.81	8.83	Y
	NBO_198					31.4		15.80	9.13	Y
RP S1 + IF2	NBO_304 g 0003	HQ291381	286	S1; EIF_2_alpha	NF	30.66	2	32.33	7.97	N
	NBO_20 g 0032	HQ291380				30.31		32.36	8.26	N
RPL3	NBO_11 g 0008	HQ291421	384	Ribosomal_L3	NF	36.36	2	43.31	10.06	N
	NBO_85 g 0009	HQ291422	664			33.79		76.43	9.04	N
RPL4	NBO_13 g 0061	HQ291439	334	Ribosomal_L4	NF	35.92	1	37.84	9.74	Y
	NBO_1546 g 0001	HQ291440	192(NC)	Ribosomal_L4	NF	36.79	1	21.44	10.14	Y
RPL5	NBO_73 g 0002	HQ291479	317	NO	NF	36.38	1	34.68	5.10	N
RPL5-B	NBO_66 g 0043	HQ291472	284	Ribosomal_L18p	NF	32.51	2	33.17	9.75	N
	NBO_941 g 0003	HQ291473				32.63		33.14	9.76	N
RPL6	NBO_3 g 0006	HQ291457	801	Ribosomal_L6e	NF	35.5	1	91.81	8.93	N
RPL7	NBO_919 g 0001	HQ291484	232	Ribosomal_L30	NF	33.62	1	27.25	9.54	N
RPL8	NBO_28 g 0076	HQ291455	237	Ribosomal_L2	NF	38.38	2	25.81	9.91	N
	NBO_30 g 0007	HQ291456				38.52		25.84	9.91	N
RPL8-A	NBO_2 g 0031	HQ291452	203	Ribosomal_L7Ae	NF	35.46	2	23.03	9.71	N
	NBO_18 g 0035	HQ291451				36.27		23.00	9.71	N
RPL9	NBO_914 g i001	HQ291483	56	NO	NF	37.06	1	6.41	8.66	N
RPL9-B	NBO_10 g 0064	HQ291406	186	Ribosomal_L6	NF	34.76	1	21.04	9.68	N
RPL10	NBO_419 g 0008	HQ291426	116	Ribosomal_L16	NF	37.04	3	13.27	10.13	Y
	NBO_12 g 0022	HQ291425	116	Ribosomal_L16		36.75		13.27	10.13	Y
	NBO_9 g 0005	HQ291427	520	NO		33.91		57.45	8.10	Y

Table 1. Continues.

RPL10A	NBO_53 g 0009	HQ291431	69(NC)	Ribosomal_L1	NF	31.9	4	8.02	5.22	N
	NBO_53 g 0010	HQ291432	123(NC)			31.99		14.42	9.86	N
	NBO_770 g 0001	HQ291433	89(NC)			31.85		10.21	5.05	N
	NBO_1221 g 0001	HQ291430	217			31.8		25.25	9.28	N
RPL11	NBO_2 g 0013	HQ291435	174	Ribosomal_L5	NF	35.43	2	19.51	9.94	N
	NBO_1288 g 0002	HQ291434				35.43		19.51	9.94	N
RPL12	NBO_32 g 0010	HQ291429	164	RL11	NF	36.36	2	17.73	9.56	N
	NBO_1214 g 0001	HQ291428				36.16		17.72	9.56	N
RPL13	NBO_16 g 0057	HQ291447	162	Ribosomal_L13e	NF	36.4	2	18.99	10.67	N
	NBO_386 g 0017	HQ291448				36.2		19.00	10.67	N
RPL13a	NBO_6 g 0082	HQ291412	197	Ribosomal_L13	NF	32.49	2	22.88	9.74	Y
	NBO_1020 g 0003	HQ291411	147(NC)			30.18		17.53	9.79	Y
RPL15	NBO_4 g 0015	HQ291459	204	Ribosomal_L15e	NF	38.37	1	24.07	10.77	N
RPL17	NBO_67 g 0006	HQ291476	171	Ribosomal_L22	NF	31.78	2	19.62	9.77	N
	NBO_98 g 0001	HQ291477				31.4		19.62	9.77	N
RPL18	NBO_508 g 0011	HQ291468	194	Ribosomal_L18e	NF	34.87	1	21.99	9.84	N
RPL18a	NBO_199 g 0002	HQ291453	180	Ribosomal_L18ae	NF	32.78	1	21.08	9.78	N
RPL19	NBO_66 g 0010	HQ291471	154	Ribosomal_L19e	YES	37.63	1	18.02	10.65	Y
RPL21	NBO_72 g 0006	HQ291470	198	Ribosomal_L21e	NF	35.18	2	23.39	10.07	N
	NBO_594 g 0004	HQ291469	160			35.61		18.68	10.33	N
RPL22	NBO_19 g 0024	HQ291420	109	Ribosomal_L22e	NF	33.03	3	12.72	9.49	N
	NBO_1303 g 0001	HQ291418	109			32.42		12.64	9.35	N
	NBO_1305 g 0001	HQ291419	113			30.41		13.11	9.46	N
RPL23	NBO_13 g 0023	HQ291436	146	Ribosomal_L14	NF	42.18	3	16.09	10.06	N
	NBO_1436 g 0002	HQ291437				42.4		16.03	10.13	N
	NBO_1440 g 0001	HQ291438				42.4		16.03	10.13	N
RPL23a2	NBO_69 g 0008	HQ291478	151	Ribosomal_L23	NF	29.61	1	17.28	10.24	Y
RPL24	NBO_18 g 0010	HQ291449	92	Ribosomal_L24e	NF	36.56	4	10.28	11.09	N
	NBO_34 g 0042	HQ291450	92			36.56		10.28	11.09	N
	NBO_998 g i001	HQ291417	66(NC)			35		7.35	11.39	N
	NBO_1033 g i001	HQ291416	64(NC)			37.44		7.91	10.31	N

Table 1. Continues.

RPL24P	NBO_417 g 0005	HQ291462	139	KOW	NF	33.57	2	16.06	9.68	N
	NBO_418 g 0002	HQ291463				33.1		16.06	9.75	N
RPL27	NBO_64 g 0036	HQ291461	125	Ribosomal_L27e	NF	35.45	2	14.48	9.93	N
	NBO_401 g 0009	HQ291460				34.92				
RPL27a	NBO_10 g 0092	HQ291407	107	L15	NF	35.49	2	12.40	9.70	N
	NBO_1278 g 0002	HQ291408	147			38.51		16.76	10.38	N
RPL30	NBO_15 g 0009	HQ291441	108	Ribosomal_L7Ae	NF	35.17	2	11.87	9.66	Y
	NBO_55 g 0011	HQ291442				35.17		11.90	9.66	Y
RPL31	NBO_73 g 0024	HQ291480	111	Ribosomal_L31e	NF	30.65	1	12.80	10.25	Y
RPL32	NBO_6 g 0084	HQ291414	139	Ribosomal_L32e	NF	36.43	3	16.02	10.32	N
	NBO_1021 g 0002	HQ291413	139			35.95		16.02	10.32	N
	NBO_1022 g i001	HQ291415	80(NC)			37.45		8.82	9.78	N
RPL34	NBO_15 g 0024	HQ291443	106	Ribosomal_L34e	NF	31.78	2	12.44	10.56	Y
	NBO_54 g 0010	HQ291444	193			31.27		22.59	10.49	Y
RPL35	NBO_16 g 0038	HQ291424	122	Ribosomal_L29	NF	33.88	2	14.60	9.88	Y
	NBO_1116 g 0001	HQ291423				34.15		14.60	9.88	Y
RPL35A	NBO_2 g 0091	HQ291410	73(NC)	Ribosomal_L35Ae	NF	35.14	2	8.12	10.20	Y
	NBO_10 g 0112	HQ291409	113			34.21		12.57	10.48	Y
RPL36e	NBO_792 g 0001	HQ291481	95	Ribosomal_L36e	NF	31.25	1	11.12	10.71	N
RPL37a	NBO_462 g 0004	HQ291466	86	Ribosomal_L37ae	NF	37.98	2	9.45	10.53	N
	NBO_6 g i002	HQ291467	92			37.63		10.16	10.53	N
RPL39e	NBO_443 g 0003	HQ291464	52	Ribosomal_L39	NF	37.11	2	6.39	11.60	N
	NBO_733 g 0001	HQ291465				36.48				
UBIQUITIN/L40 ribosomal protein FUSION	NBO_67 g 0005	HQ291474	131	UBQ; Ribosomal_L40e	NF	29.8	2	14.94	9.87	Y
	NBO_98 g 0003	HQ291475				29.29		15.00	9.87	Y
RPL44	NBO_80 g 0021	HQ291482	103	Ribosomal_L44	NF	35.58	1	11.80	10.45	Y
RPP0	NBO_514 g 0008	HQ291446	263	Ribosomal_L10	NF	29.67	2	29.90	9.01	N
	NBO_1574 g 0002	HQ291445	139(NC)			29.52		15.68	4.83	N
RPP1	NBO_375 g 0004	HQ291458	106	Ribosomal_60s	NF	38.63	1	11.43	4.55	Y
RPP2	NBO_27 g i004	HQ291454	136	NO	NF	36.5	1	14.95	4.21	Y

NC, The ribosomal protein sequence is not complete; NO, domain not found by SMART predicted; NF, intron not found in this sequence; YES, there is one intron in this sequence; Y, found EST proof; N, EST proof not found.

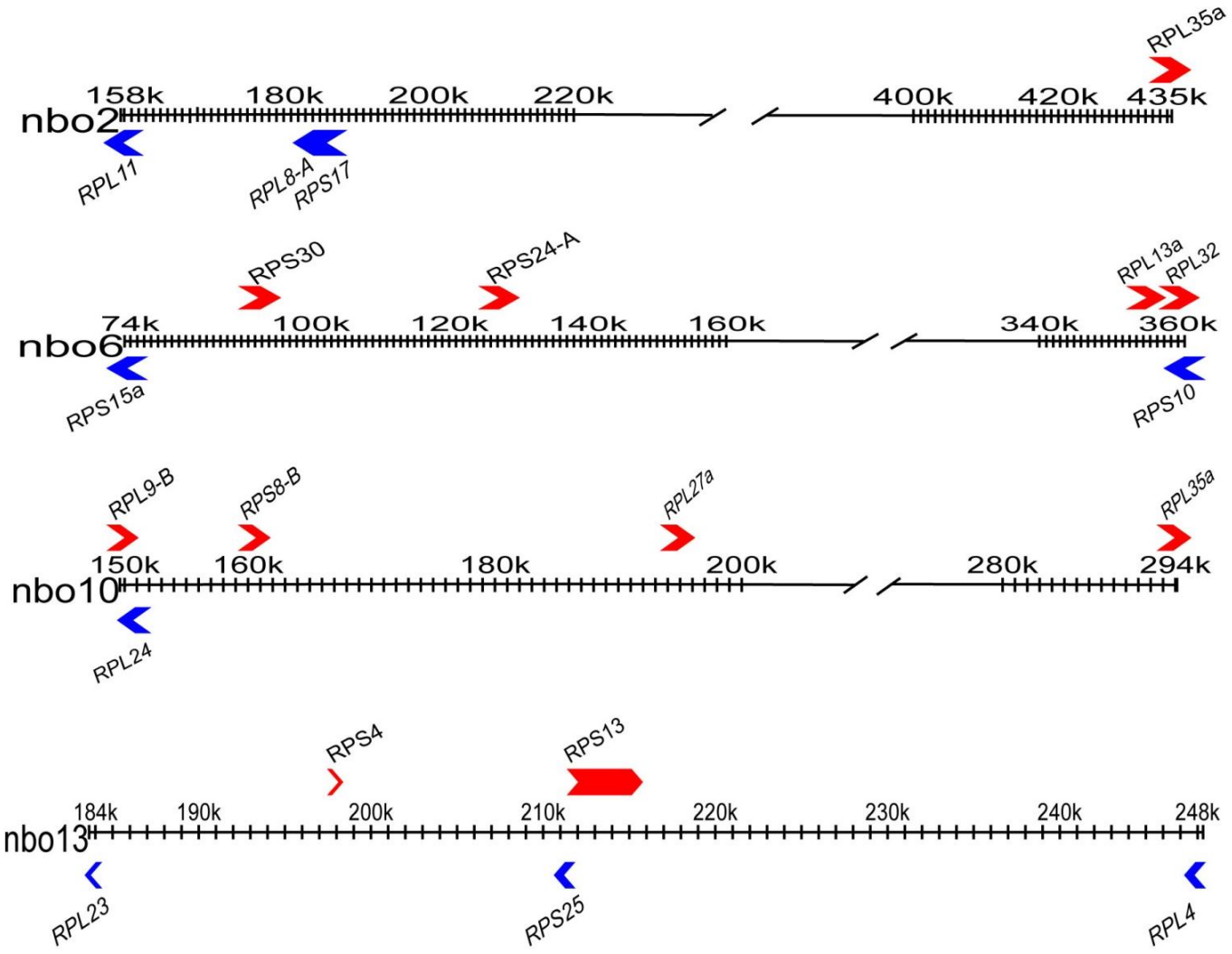

Figure 1. Distribution of *RP* genes on four superscaffolds. The superscaffold name is labeled at the left, and the number of the scale labeled on the top of each superscaffold. *RP* genes are signed with the arrow (red represents forward direction, and blue represents backward direction).

respectively.

Introns located in three ribosomal protein genes of N. bombycis

In N. bombycis genome, three ribosomal protein genes have been found harboring predicted short introns, which have the same sequence structure to the documented result on *E*. cuniculi, N. ceranae and Octosporea bayeri (Corradi et al., 2009). While Katinka et al. (2001) inferred eleven ribosomal protein genes with introns in E. cuniculi, Cornman et al. (2009) documented that five of the N. ceranae orthologues to E. cuniculi also contained an intron and the sixth N. ceranae gene containing an intron encodes the S4 ribosomal protein, which lacks an intron in E. cuniculi. Except for the L19 and S4 ribosomal proteins of *N. bombycis*, another ribosomal protein, *S18*, contains a short intron (24 bp). These introns harbor consensus spliceosomal boundaries (5′-GT...AG-3′) with a 5′ region identical to the consensus of higher eukaryotes (GTAAGT). In addition, the intronic sequences also show fairly robust conservation within and among species (Figure 3), indicating selection for efficient recognition by the spliceosomal machinery. Recently, one intron has been detected in *A. locustae* ribosomal protein genes (Peyretaillade et al., 2009). However, no intron has been detected in ribosomal protein genes in *E. bieneusi* (Akiyoshi et al., 2009) to date. Fortunately, using the NbBmEST database in our lab, EST proofs of these three *RP* genes have been found, but only the EST corresponding to *S18* gene is complete. Figure 4 shows the aligned result which the intron sequence has been spliced in the EST sequence.

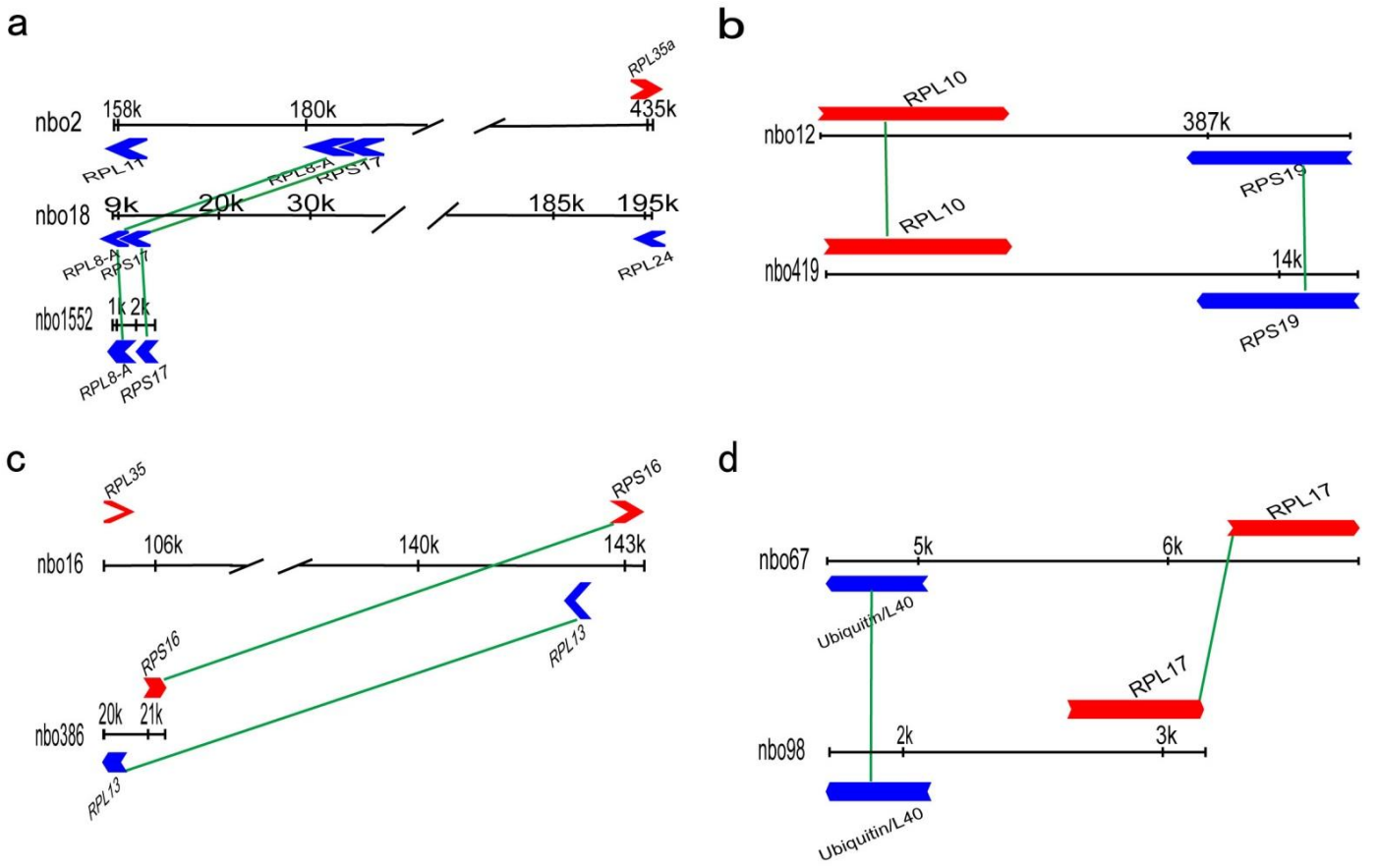

Figure 2. Tandem ribosomal protein genes in *N. bombycis* genome. a, b, c and d stand for the four groups of tandem *RP* genes in *N. bombycis,* respectively. The superscaffold number is labeled on the left of each superscaffold, and the red (forward) and blue (backward) arrows represent the *RP* genes located on the superscaffold. The tandem *RP* genes are connected by green lines between its corresponding genes.

Promoter and regulatory sequences prediction of ribosomal protein genes

Further analysis of the ribosomal protein genes showed that each encoding protein of these genes is located on one coding sequence, and core promoter sequences of partial (about 29.2%) of them are structurally characteristic of TATAA (Figure 5). Moreover, like other microsporidian such as *E. cuniculi*, *A. locustae*, *E. bieneusi*, *N. ceranae* and *A. algerae* (Peyretaillade et al., 2009), the genes coding for ribosomal proteins of *N. bombycis* were also identified with the AAATTT-like signal; and the CCC motif (three *RPGs* - *RPL4*, *RPL10A*, *RPS21*, harbour GGG-like motif) was identified between the AAATTT sequence and the translational transcription start (TSS) codon. In addition, the regions 500 bps upstream of all genes were searched for its potential regulation motifs with NSITE programs (Figure 6), and it was found that all the ribosomal protein genes have at least one SP1 and one GATA-1 factor binding site. SP1, a ubiquitous transcription factor binding GC-boxes in the regulating promoter elements, is required for the constitutive and inducible expression of a variety of genes, such as in cell cycle or mammalian development. GATA-1, which is expressed predominantly in hematopoietic cells, regulates differentiation and gene expression in T-lymphocytes, erythroids and megakaryocytes. *RPL22* have one transcription factor, NF-κB, which is involved in the regulation of many genes encoding mediators of the immune, acute phase and inflammatory responses. Genes of *RPS6*, *RPS7* and *RPS14* have one Oct-1 motif each. Each member of the Oct family is a factor involved in the first step of differentiation during embryogenesis and its functions were to keep an undifferentiated cell stable. The other three regulatory elements, Rap1, GAGA and HNF-3 motifs were also found in these ribosomal protein genes.

Syntenic analysis of ribosomal protein genes among five microsporidian

To compare the locations of different microsporidian *RPGs*, syntenic maps of *RPGs* containing regions among

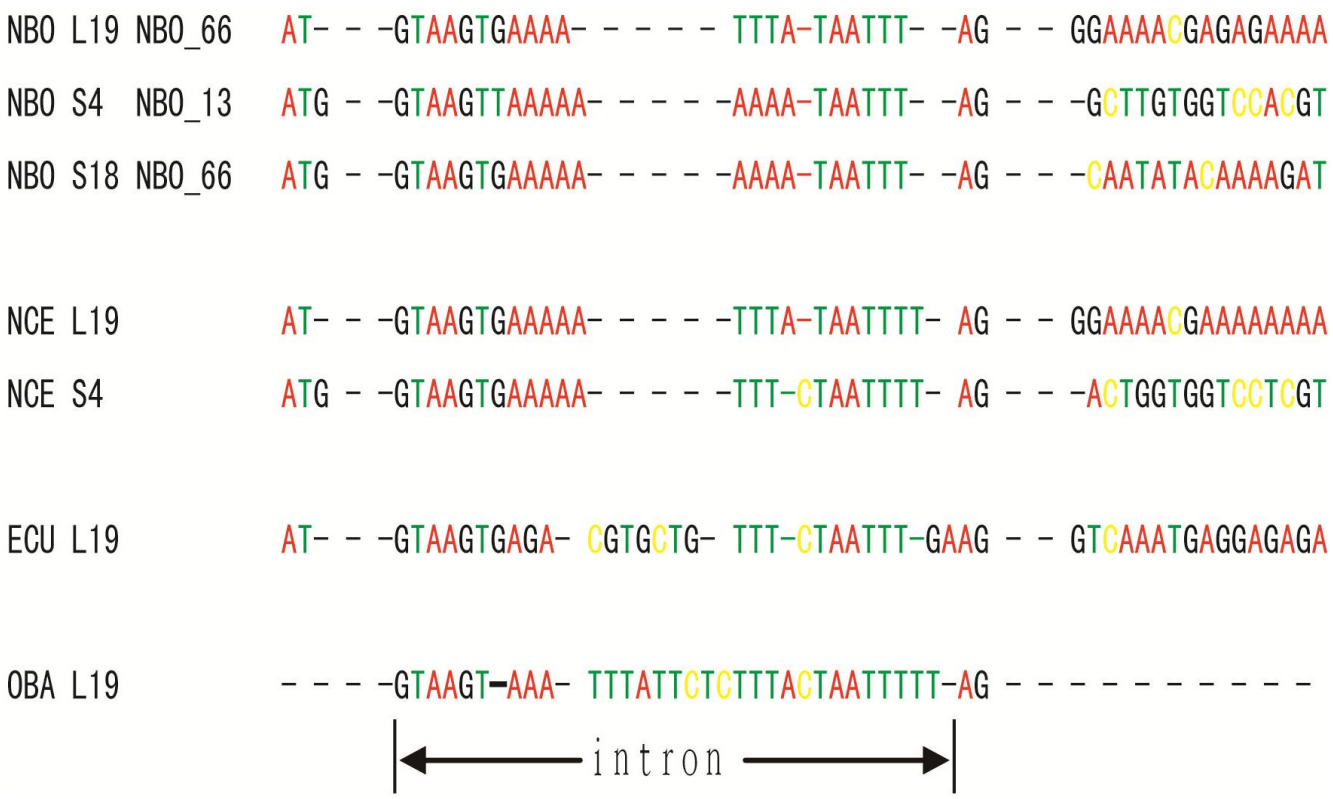

Figure 3. Alignment of the 5' region of three ribosomal protein genes in *N. bombycis* that contained predicted introns, together with the three orthologues of these genes in *N. ceranae*, one orthologue in *E. cuniculi* and one orthologue in *O. bayeri*. Alignment begins with the start codon, which is interrupted by an intron in the ribosomal protein L19. Introns were manually aligned to illustrate regions of sequence conservation. Nbo = *N. bombycis*, Nce = *N. ceranae*, Ecu = *E. cuniculi*, Oba = *O. bayeri*.

E. cuniculi, A. locustae, N. ceranae, E. bieneusi and N. bombycis were plotted with a Perl script based on the genome data of N. bombycis, E. cuniculi (http: //www.ncbi.nlm.nih.gov), A. locustae (http: //jbpc.mbl.edu/ Nosema), N. ceranae (http: //www.ncbi.nlm.nih.gov) and E. bieneusi (http: //www.ncbi.nlm.nih.gov). Othologous genes were identified using BLASTP program (Altschul et al., 1997). A syntenic map of each ribosomal protein gene has been drawn based on the five microsporidian genome sequences. All syntenic blocks flanking RP genes were observed among N. bombycis, E. cuniculi, A. locustae, N. ceranae and E. bieneusi, for example, RPL3 located on syntenic block among the five microsporidians (Figure 7). In total, about 56 ribosomal protein genes have been detected belonging to syntenic blocks among more than three microsporidian species, which is about 76.7% to the RPGs total number 73.

Phylogenetic trees of ribosomal protein genes of N. bombycis

The phylogenetic trees of all ribosomal protein genes were constructed from multiple alignments of *RP* genes (for example *RPL3* showed on Figure 8). In all trees, microsporidia lies on almost the same location as that of tree constructed by ribosomal RNA. According to this phylogenetic tree, it also indicated that microsporidia is a primitive eukaryote.

DISCUSSION

For the first time, we got the information of whole ribosomal protein genes of N. bombycis. Also, this is the first species that all ribosomal protein genes have been analyzed in microsporidia. N. bombycis has 130 ribosomal protein gene hits corresponding to 73 putative ribosomal protein genes, which is one more than ribosomal protein gene in E. cuniculi, and is close to the ribosomal proteins number of typical eukaryotes. In N. bombycis, more RPGs are multiple than that in E. cuniculi. This is consistent to the redundant characteristic of N. bombycis genome. In addition, the 130 ribosomal protein gene sequences of N. bombycis scatter on 92 super-scaffolds randomly. Comparing with other microsporidian, we found the phenomenon that gene orders of some RP genes are conserved among microsporidia (data not shown). And the RP genes which include in conserved gene orders are multiple in *N. bombycis* genome (such

Figure 4. Alignment of *S18* gene and its NbBmEST sequence. The start codon and stop codon have been labeled with green color. And spliceosomal boundaries (5′-GT...AG-3′) have been labeled with red color. The intron has been indicated on the top of its corresponding sequence.

as RPL8-A and RPS17, RPS16 and RPL13, et al.). Inspection of genes in the vicinity of ribosomal protein genes suggests maybe extensive duplication of large chromosome fragments has happened before in the N. bombycis genome.

Among all the ribosomal protein genes, three of them

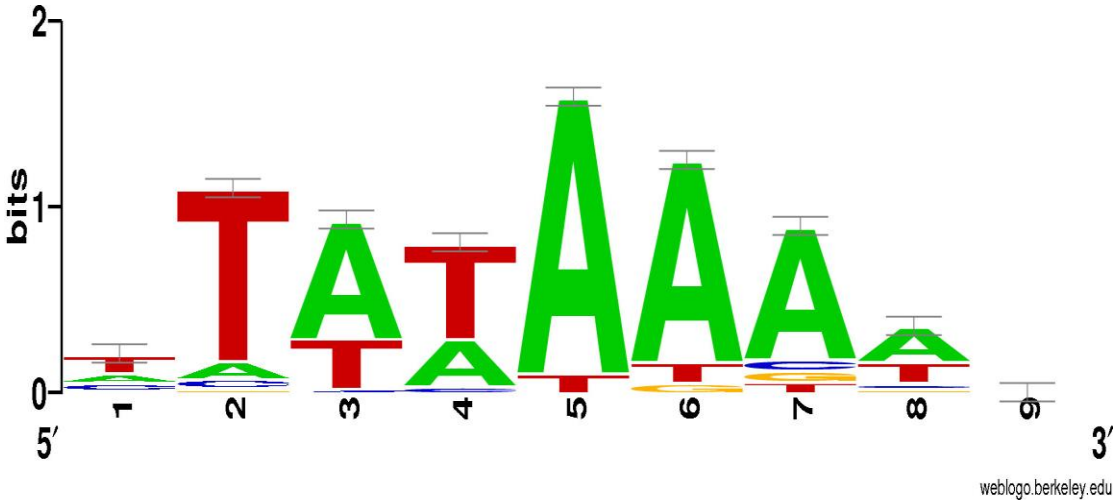

Figure 5. TATA-box of the core promoters treated on Weblogo (http://weblogo.berkeley.edu/logo.cgi). The bits show the frequency of the base of the sites.

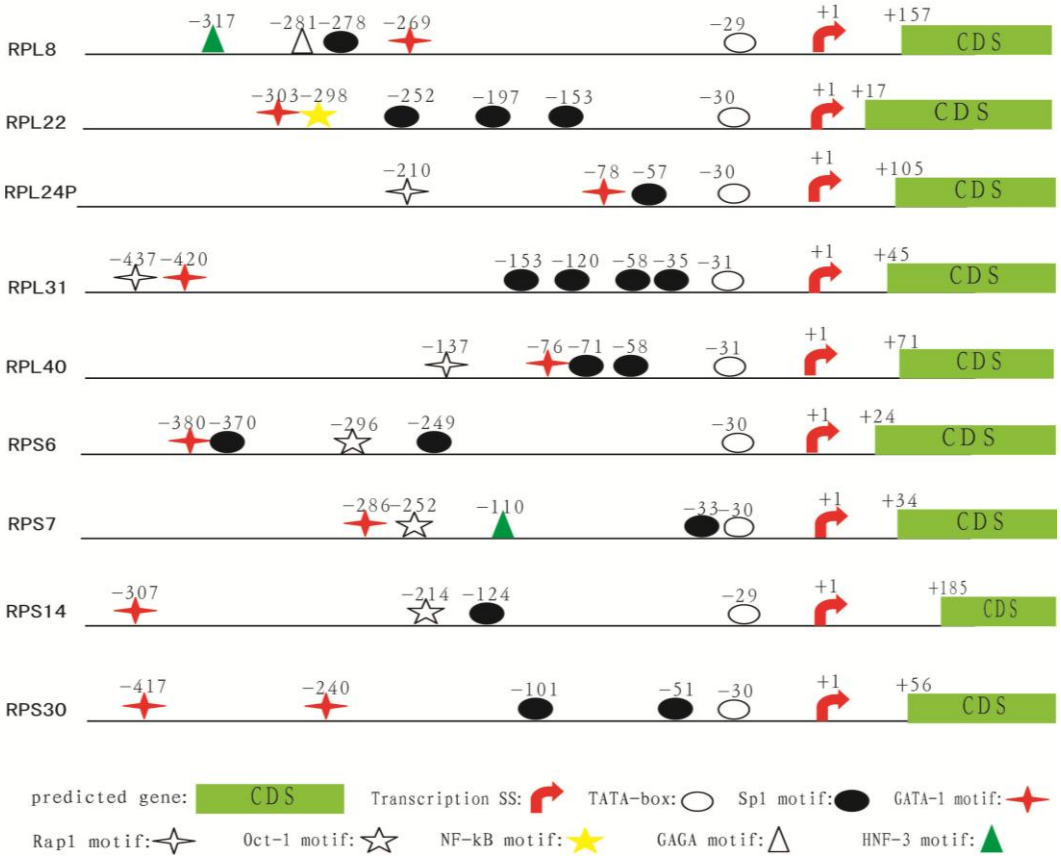

Figure 6. Some putative regulation elements of ribosomal protein genes in *N. bombycis*. The NSITE programs were used to identify the regulatory elements in the upstream regions from the genes available in *N. bombycis* database. The upstream (500 bps) regions were drawn by scale. Arc arrows indicate the transcription start sites. The position of the nucleotide sequences is calculated based on the distance from the predicted transcription initiation site. The open read frame is marked with a green box, and the numbers mean the distances to the transcription initiation sites. The binding sites of the SP1 factor and the GATA-1 factor are shown with black circles and red crux stars, respectively. The blank crux stars stand for Rap1 motifs and the blank pentangular stars refer to Oct-1 motifs. The yellow pentangular star and the blank and green triangles mean NF-κB, GAGA and HNF-3 motifs, respectively.

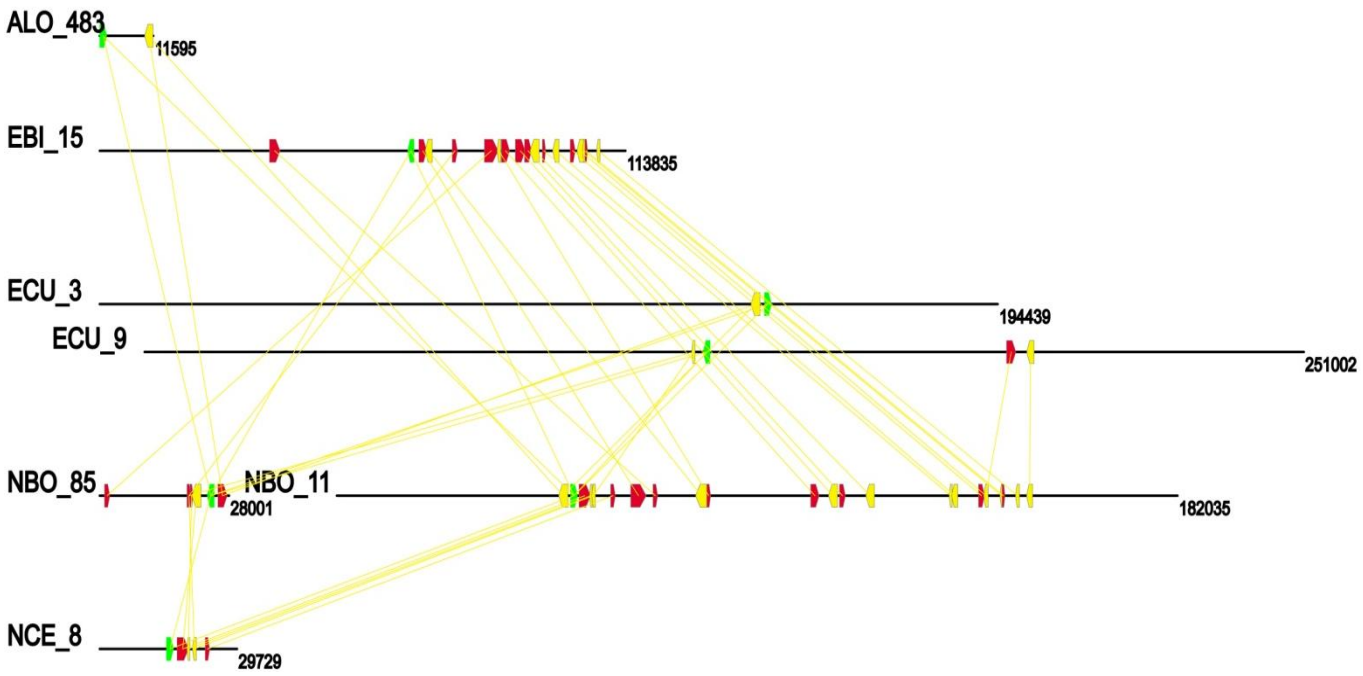

Figure 7. Syntenic maps of *RPL3* flanking regions on chromosomes/superscaffolds in five microsporidians. Syntenic maps of *RPL3* containing regions were plotted among *N. bombycis* (superscaffold NBO_11 and NBO_85), *A. locustae* (contig ALO_483), *E. bieneusi* (contig EBI_15), *E. cuniculi* (chromosome ECU_3 and ECU_9) and *N. ceranae* (contig NCE_8). Orthologulous genes were detected with BLSATP (Altschul et al., 1997) and linked with transparent yellow lines. The green (*RPL3*), red and yellow blocks represent syntenic genes.

contained short introns, which have the same sequence structure to the documented result on E. cuniculi, N. ceranae (Cornman et al., 2009) and O. bayeri (Corradi et al., 2009). And the intronic structures showed fairly robust conservation within and among species (Figure 3), which indicate election for efficient recognition by the spliceosomal machinery. Recently, one intron has been found in the ribosomal protein gene S6 of A. locustae (Peyretaillade et al., 2009). However, no intron has been detected in ribosomal protein genes of E. bieneusi to date. We have not found any rules of the distribution of intron in ribosomal protein genes till now. In S. cerevisiae, however, more than 70% of the RPGs contain introns in contrast to only 5% intron-containing genes in its total genes. And recently, the law of combinational regulation between upstream regions and introns has been characterized and the synergy between them may result in the high transcriptional frequency of yeast RPGs (Hu et al., 2010). As a type of species related to fungi, Microsporidia have the same intron structure (5'-end of RPGs) to S. cerevisiae. In order to satisfy the quantity requirement of protein synthesis in period of microsporidian spore proliferation, the synergy between upstream regions and introns may play important role. These structures of genes in N. bombycis and S. cerevisiae are not same to that of Rattus norvegicus, which has more introns in each ribosomal protein gene. As one type of parasite, maybe this event is either a new obtained or a degraded thing in the ribosomal protein genes evolution.

In order to get more information of the upstream sequences of ribosomal protein genes, we cut the 500 bps sequences for analysis. A structural characteristic of TATA-like motif has been found in the ribosomal protein genes. In N. ceranae, TATA-like promoters are important components of its gene regulation (Cornman et al., 2009), which is same to that of yeast genes (Basehoar et al., 2004). But there are still some ribosomal protein genes of N. bombycis of which the classic promoter structure cannot be found, maybe because the length 500 bps was not sufficiently long, or those ribosomal protein genes do not have this structure at all. In addition, the majority of ribosomal protein genes presented the CCC-like motif immediately upstream from the transcription start site, except for three ribosomal protein genes (RPL4, RPL10A, RPS21), which contained GGG-like motif. At the same time, an AAATTT-like signal was identified upstream from the CCC-like motif. These characteristics are same to other microsporidian species, such as E. cuniculi, A. locustae, E. bieneusi, N. ceranae and A. algerae for ribosomal protein genes (Peyretaillade et al., 2009). Maybe these characteristics are widespread among microsporidia. Like S. cerevisiae, we have not found the characteristic "terminal oligopyrimidine tract" in the 5' end of RPGs in N. bombycis. Certainly, lots of regulatory motifs have been detected in the upstream sequences, which can facilitate the transcription of ribosomal protein

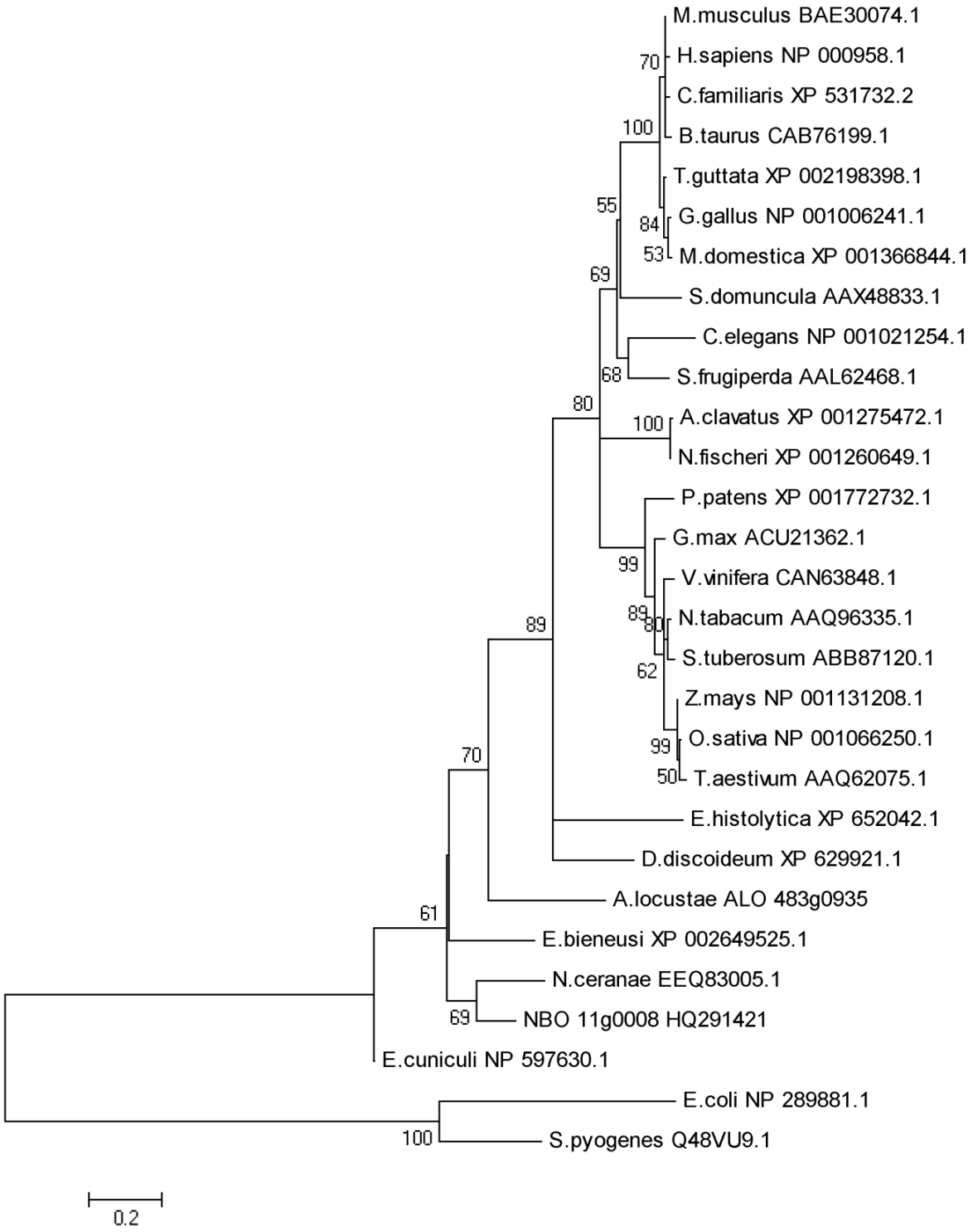

Figure 8. Phylogenetic tree of ribosomal protein *L3* gene. The neighbor-joining trees were generated with the MEGA4.0 software and sequences aligned without gaps. The bootstrap values given at the nodes were from 1,000 replicates.

genes.

From the syntenic maps of RP-containing regions among the five microsporidian, E. cuniculi, A. locustae, N. ceranae, E. bieneusi and N. bombycis, we can see more ribosomal protein genes belonging to syntenic blocks.

That is; to the house-keeping genes, they follow a

conserved gene order evolution mechanism. Recently, a novel approach has been applied, which is based on the conservation of gene order (Lee et al., 2008). It has been known for some time that while microsporidian gene sequences are evolving very quickly, the order of genes within the genome is highly conserved (Corradi et al., 2007; Slamovits et al., 2004). This observation was recently extended by the demonstration that microsporidian genomes share higher frequency of gene order conservation with zygomycetes than they do with any other group of fungi for which genome data are available (Dyer, 2008; Lee et al., 2008). From the syntenic blocks, similar genomic structures are shown among these five microsporidian, in spite of some inner transversion occurring in its syntenic regions. In addition, among these five microsporidian, if there is one or more ancestral organization of RPGs, the genome structures of other microsporidian which have gene transversion must have been greatly rearranged. This may be caused by homologous recombination, segmental duplication, whole genome duplication or transposition. Causally, the TEs harbored in N. bombycis, which had been documented (Xu et al., 2006) before and discovered in other microsporidian lately (Williams et al., 2008; Gill et al., 2008), to some extent may change these variations. Perhaps, the analysis of syntenic blocks of ribosomal protein gene flanking regions can give more information of microsporidian genome evolution.

The trees of ribosomal protein genes showed almost the same result to that of ribosomal RNA. According to this result of phylogenetic tree (Figure 8), it also indicated that microsporidia may be a primitive eukaryote. This however, is the long-branch artifacts due to increased evolutionary rates of their ribosomal RNAs (Palmer and Delwiche, 1996; Embley and Hirt, 1998; Keeling and McFadden, 1998). As we all know, the ribosome of plant is smaller than that of mammals (Cammarano et al., 1972; Verschoor et al., 1996). Despite an overall similarity in ribosomal architecture, the length of highly variable loop regions of the 23S-like rRNA is difference. The size ranges from approximately 3,300 bp (25 to 26S) in plants to approximately 4,700 bp (28S) in mammals (Schnare et al., 1996). According to this information, we compared the rRNA length and structure of N. bombycis to those of other eukaryotes and bacterium, and found that the rRNA length of N. bombycis and some nuclear acids are different. However, this is mainly because the rRNA sequence is a highly degenerated sequence (Van de Peer et al., 2000). To ribosomal proteins, the same condition may occur. This still need to be clarified in the future.

The number of ribosomal protein genes of N. bombycis is slightly smaller than that of typical eukaryote, but the phylogenetic location of these genes show that it belongs to primitive eukaryote. Combination with the length of rDNA sequence of microsporidia, maybe the 70S ribosome of Microsporidia, resulted from the shorter sequence length of rDNA and the smaller number of ribosomal protein genes. We may give a further hypothesis that the sedimentation coefficient (70S or 80S) of ribosome not only depend on phylogenetic position of species, but also depend on its environment. In short, the identification of the total ribosomal protein genes and the determination of their sequence organization and distribution constitute a first step to determine their biological role, modeling of ribosome structure and function, and genome evolution. Nucleotide sequences reported in this study have been submitted to the GenBank™, EMBL and DDBJ databases under the accession numbers: HQ291357-HQ291484.

ACKNOWLEDGEMENTS

We are grateful to all the authors for their free-charged software cited and used in this article. The authors appreciate the assistance of Dr. Hongjuan Cui in improving the manuscript. This work is supported by the National Natural Science Foundation of China under Grant No. 30930067, the Program of Introducing Talents of Discipline to Universities (No. B07045) and the Research Fund of the Doctoral Program of Southwest University under Grant No. swu109039.

REFERENCES

Akiyoshi DE, Morrison HG, Lei S, Feng X, Zhang Q, Corradi N, Mayanja H, Tumwine JK, Keeling PJ, Weiss LM, Tzipori S (2009). Genomic survey of the non-cultivatable opportunistic human pathogen, *Enterocytozoon bieneusi*. PLoS Pathog. 5(1): e1000261.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. J. Mol. Biol. 215(3): 403-410.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. 25: 3389-3402.

Ban N, Nissen P, Hansen J, Capel M, Moore PB, Steitz TA (1999). Placement of protein and RNA structures into a 5 Å-resolution map of the 50S ribosomal subunit. Nature, 400: 841–846.

Ban N, Nissen P, Hansen J, Moore PB, Steitz TA (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. Science, 289: 905-920.

Basehoar AD, Zanton SJ, Pugh BF (2004). Identification and distinct regulation of yeast TATA box-containing genes. Cell, 116: 699–709.

Belkorchia A, Biderre C, Militon C, Polonais V, Wincker P, Jubin C, Delbac F, Peyretaillade E, Peyret P (2008). In vitro propagation of the microsporidian pathogen *Brachiola algerae* and studies of its chromosome and ribosomal DNA organization in the context of the complete genome sequencing project. Parasitol. Int. 57(1): 62-71.

Biderre C, Metenier G, Vivares CP (1998). A small spliceosomal-type intron occurs in a ribosomal protein gene of the microsporidian *Encephalitozoon cuniculi*. Mol. Biochem. Parasitol. 94(2): 283-286.

Cammarano P, Pons S, Romeo A, Galdieri M, Gualerzi C (1972). Characterization of unfolded and compact ribosomal subunits from plants and their relationship to those of lower and higher animals: evidence for physiochemical heterogeneity among eukaryotic ribosomes. Biochim. Biophys. Acta. 28: 571-596.

Chakrabarti D, Dame JB, Gutell RR, Yowell CA (1992). Characterization of the rDNA unit and sequence analysis of the small subunit rRNA and 5.8S rRNA genes from *Tritrichomonas foetus*. Mol. Biochem. Parasitol. 52: 75-84.

Champney WS, Chittum HS, Samuels R (1992). Ribosomes from

trichomonad protozoa have prokaryotic characteristics. Int. J. Biochem. 24: 1125-1133.
Clemons WM, May Jr. JL, Wimberly BT, McCutcheon JP, Capel MS, Ramakrishnan V (1999). Structure of a bacterial 30S ribosomal subunit at 5.5 Å resolution. Nature, 400: 833–840.
Cornman RS, Chen YP, Schatz MC, Street C, Zhao Y, Desany B, Egholm M, Hutchison S, Pettis JS, Ian Lipkin W, Evans JD (2009). Genomic analyses of the microsporidian *Nosema ceranae*, an emergent pathogen of honey bees. PLoS Pathog. 5(6): e1000466.
Corradi N, Akiyoshi DE, Morrison HG, Feng X, Weiss LM, Tzipori S, Keeling PJ (2007). Patterns of genome evolution among the microsporidian parasites *Encephalitozoon cuniculi*, *Antonospora locustae* and *Enterocytozoon bieneusi*. PLoS ONE, 2: e1277.
Corradi N, Haag KL, Pombert JF, Ebert D, Keeling PJ (2009). Draft genome sequence of the Daphnia pathogen *Octosporea bayeri*: insights into the gene content of a large microsporidian genome and a model for host parasite interactions. Geno. Biol. 10: R106.
Corradi N, Keeling PJ (2009). Microsopridia: a journey through radical taxonomical revisions. Fungal Bio. Rev. 23: 1-8.
Culver GM, Cate JH, Yusupova GZ, Yusupov MM, Noller HF (1999). Identification of an RNA-protein bridge spanning the ribosomal subunit interface. Science. 285: 2133-2136.
Dyer PS (2008). Evolutionary biology: microsporidia sex a missing link to fungi. Curr. Biol. 18: 1012-1014.
Embley TM, Hirt RP (1998). Early branching eukaryotes. Curr. Opin. Genet. Dev. 8: 624-629.
Franzen C, Nassonova ES, Scholmerich J, Issi IV (2006). Transfer of the members of the genus Brachiola (microsporidia) to the genus Anncaliia based on ultrastructural and molecular data. J. Eukaryot. Microbiol. 53(1): 26-35.
Germot A, Philippe H, Le Guyader H (1997). Evidence for loss of mitochondria in Microsporidia from a mitochondrial-type HSP70 in *Nosema locustae*. Mol. Biochem. Parasit. 87(2): 159-168.
Gill EE, Becnel JJ, Fast NM (2008). ESTs from the microsporidian *Edhazardia aedis*. BMC Genomics. 9: 296.
Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG (1996). Life with 6000 genes. Science. 274: 546–567.
Graack HR, Wittmann-Liebold B (1998). Mitochondrial ribosomal proteins (MRPs) of yeast. Biochem. J. 329: 433-448.
Healey A, Mitchell R, Upcroft JA, Boreham PFL, Upcroft P (1990). Complete nucleotide sequence of the ribosomal RNA tandem repeat unit from *Giardia intestinalis*. Nucleic Acids Res. 18: 4006.
Hirt RP, Healy B, Vossbrinck CR, Canning EU, Embley TM (1997). A mitochondrial Hsp70 orthologue in Vairimorpha necatrix: molecular evidence that microsporidia once contained mitochondria. Curr. Biol. 7(12): 995-998.
Hu J, Li H, Zhang J (2010). Analysis of transcriptional synergy between upstream regions and introns in ribosomal protein genes of yeast. Comput. Biol. Chem. 34: 106-114.
Ishihara R, Hayashi Y (1968). Some properties of ribosomes from the sporoplasm of *Nosema bombycis*. J. Invertebr. Pathol. 11: 377-385.
Katinka MD, Duprat S, Cornillot E, Méténier G, Thomarat F, Prensier G, Barbe V, Peyretaillade E, Brottier P, Wincker P, Delbac F, El Alaoui H, Peyret P, Saurin W, Gouy M, Weissenbach J, Vivarès CP (2001). Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. Nature. 414(6862): 450-453.
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 ribosome RNA gene of *Nosema bombycis*. J. Invertebr. Pathol. 64: 147-148.
Keeling PJ, McFadden GI (1998). Origins of microsporidia. Trends Microbiol. 19: 19-23.
Kenmochi N, Ashworth LK, Lennon G, Higa S, Tanaka T (1998 b). High-resolution mapping of ribosomal protein genes to human chromosome 19. DNA Res. 5: 229-233.
Koc EC, Burkhart W, Blackburn K, Moseley A, Koc H, Spremulli LL (2000). A proteomics approach to the identification of mammalian mitochondrial small subunit ribosomal proteins. J. Biol. Chem. 275: 32585-32591.
Lee SC, Corradi N, Byrnes EJ 3rd, Torres-Martinez S, Dietrich FS, Keeling PJ, Heitman J (2008). Microsporidia evolved from ancestral sexual fungi. Curr. Biol. 18(21): 1675-1679.
Liu H, Pan G, Song S, Xu J, Li T, Deng Y, Zhou Z (2008). Multiple rDNA units distributed on all chromosomes of *Nosema* bombycis. J. Invertebr. Pathol. 99(2): 235-238.
Matheson AT, Auer J, Ramíerez C, Böck A (1990). Structure and evolution of archaebacterial ribosomal proteins. In Hill WE, Dahlberg A, Garrett RE, Moore PB., D. Schlessinger, J.R. Warner, eds, The Ribosome, Structure, Function and Evolution. American Society of Microbiologists, Washington, DC. USA. pp. 617-633.
Meyuhas O (2000). Synthesis of the translational apparatus is regulated at the translational level. Eur. J. Biochem. 267: 6321-6330.
Meyuhas O, Hornstein E (2000). Translational control of TOP mRNAs. In Sonenberg N, Hershey JWB, Mathews MB, eds, Translational Control of Gene Expression. Cold Spring Harbor Press, eds, Cold Spring Harbor, NY. USA. pp. 671-693,
Nissen P, Hansen J, Ban N, Moore PB, Steitz TA (2000). The structural basis of ribosome activity in peptide bond synthesis. Science, 289: 920-930.
Nomura M, Gourse R, Banghman G (1984). Regulation of the synthesis of ribosomes and ribosomal components. Annu. Rev. Biochem. 53: 75-117.
Ogle JM, Brodersen DE, Clemons .Jr. WM, Tarry MJ, Carter AP, Ramakrishnan V (2001). Recognition of cognate transfer RNA by the 30S ribosomal subunit. Science, 292: 897-902.
Palmer JD, Delwiche CF (1996). Second-hand chloroplasts and the case of the disappearing nucleus. Proc. Natl. Acad. Sci. USA. 93: 7432-7435.
Peyretaillade E, Biderre C, Peyret P, Duffieux F, Metenier G, Gouy M, Michot B, Vivares CP (1998). Microsporidian *Encephalitozoon cuniculi*, a unicellular eukaryote with an unusual chromosomal dispersion of ribosomal genes and a LSU rRNA reduced to the universal core. Nucleic Acids Res. 26(15): 3513-3520.
Peyretaillade E, Broussolle V, Peyret P, Metenier G, Gouy M, Vivares CP (1998). Microsporidia, amitochondrial protists, possess a 70-kDa heat shock protein gene of mitochondrial evolutionary origin. Mol. Biol. Evol. 15(6): 683-689.
Peyretaillade E, Gonçalves O, Terrat S, Dugat-Bony E, Wincker P, Cornman RS, Evans JD, Delbac F, Peyret P (2009). Identification of transcriptional signals in *Encephalitozoon cuniculi* widespread among Microsporidia phylum: support for accurate structural genome annotation. BMC Genomics, 10: 607.
Philippe H, Germot A (2000). Phylogeny of eukaryotes based on ribosomal RNA: long- branch attraction and models of sequence evolution. Mol. Biol. Evol. 17: 830-834.
Planta RJ, Mager WH (1998). The list of cytoplasmic ribosomal proteins of *Saccharomyces cerevisiae*. Yeast, 14: 471-477.
Reese MG (2001). Application of a time delay neural network to promoter annotation in the *Drosophila melanogaster* genome. Comput. Chem. 26: 51-56.
Roger JA (1999). Reconstructing early events in eukaryotic evolution. Am. Nat. 154: 146-163.
Saitou N, Nei M (1987). The neighbor joining method: A new 331 method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4: 406-425.
Schnare MN, Damberger SH, Gray MW, Gutell RR (1996). Comprehensive comparison of structural characteristics in eukaryotic cytoplasmic large subunit (23S-like) ribosomal RNA. J. Mol. Biol. 256: 701-719.
Slamovits CH, Fast NM, Law JS, Keeling PJ (2004). Genome compaction and stability in microsporidian intracellular parasites. Curr. Bilo. 14: 891-896.
Sogin ML, Gunderson JH, Elwood HJ, Alonso RA, Peattie DA (1989). Phylogenetic meaning of the kingdom concept: an unusual ribosomal RNA from *Giardia lamblia*. Science, 243: 75-77.
Tamura K, Dudley J, Nei M, Kumar S (2007). MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24: 1596-1599.
Van de Peer Y, Ben Ali A, Meyer A (2000). Microsporidia: accumulating molecular evidence that a group of amitochondriate and suspectedly primitive eukaryotes are just curious fungi. Gene, 246: 1-8.
Verschoor A, Srivastava S, Grassucci R, Frank J (1996). Native 3D structure of eukaryotic 80S ribosome: morphological homology with *E.*

coli 70S ribosome. J. Cell Biol. 133: 495-505.
Veuthey AL, Bittar G (1998). Phylogenetic relationships of fungi Plantae and Animalia inferred from homologous comparison of ribosomal proteins. J. Mol. Evol. 47: 81-92.
Vossbrinck CR, Maddox JV, Friedman S, Debrunner-Vossbrinck BA, Woese CR (1987). Ribosomal RNA sequence suggests microsporidia are extremely ancient eukaryotes. Nature, 326: 411-414.
Weiss LM, Vossbrinck CR (1999). Molecular biology, molecular phylogeny, and molecular diagnostic approaches to the microsporidia. In: Wittner M, Weiss LM, eds. The Microsporidia and Microsporidosis. Washington, DC: ASM Press. USA. pp. 129-171,
Williams BA, Lee RC, Becnel JJ, Weiss LM, Fast NM, Keeling PJ (2008). Genome sequence surveys of *Brachiola algerae* and *Edhazardia aedis* reveal microsporidia with low gene densities. BMC Genomics, 9: p. 200.
Wittmann HG (1982). Components of bacterial ribosomes. Ann. Rev. Biochem. 51: 155-183.
Wittmann-Leibold B, Kopke AKE, Arndt E, Kromer W, Hatakeyama T, Wittmann HG (1990). Sequence comparison and evolution of ribosomal proteins and their genes. *In* WE Hill, A Dahlberg, RE Garrett, PB Moore, D Schlessinger, JR Warner, eds, The Ribosome, Structure, Function and Evolution. Am. Soc. Microbiologists, Washington, DC. USA. pp. 598-616.
Wool IG, Chan YL, Gluck A (1995). Structure and evolution of mammalian ribosomal proteins. Biochem. Cell. Biol. 73: 933–947.
Xu J, Pan G, Fang L, Li J, Tian X, Li T, Zhou Z, Xiang Z (2006). The varying microsporidian genome: Existence of long terminal repeat retrotransposon in domesticated silkworm parasite *Nosema bombycis*. Int. J. Parasitol. 36: 1049-1056.
Yamaguchi K, Subramanian AR (2000). The plastid ribosomal proteins: identification of all the proteins in the 50S subunit of an organelle ribosome (chloroplast). J. Biol. Chem. 275: 28466-28482.
Yamaguchi K, Von Knoblauch K, Subramanian AR (2000). The plastid ribosomal proteins: identification of all the proteins in the 30S subunit of an organelle ribosome (chloroplast). J. Biol. Chem. 275: 28455-28465.