Nucleic Acids Research, 1998, Vol. 26, No. 15 **3513–3520**

Microsporidian *Encephalitozoon cuniculi*, a unicellular eukaryote with an unusual chromosomal dispersion of ribosomal genes and a LSU rRNA reduced to the universal core

Eric Peyretaillade, Corinne Biderre, Pierre Peyret, Francis Duffieux, Guy Méténier, Manolo Gouy[1], Bernard Michot[2] and Christian P. Vivarès*

Laboratoire de Protistologie Moléculaire et Cellulaire des Parasites Opportunistes, LBCP, UPESA CNRS 6023, Université Blaise Pascal, 24 Avenue des Landais, 63177 Aubière Cedex, France, [1]Laboratoire de Biométrie, Génétique et Biologie des Populations, Université Claude Bernard, 43 Boulevard du 11 Novembre 1918, 69622 Villeurbanne Cedex, France and [2]Laboratoire de Biologie Moléculaire Eucaryote du CNRS, Université Paul Sabatier, 118 Route de Narbonne, 31062 Toulouse Cedex, France

Received April 20, 1998; Revised and Accepted May 21, 1998

DDBJ/EMBL/GenBank accession nos AJ005581 and AJ005582

ABSTRACT

Microsporidia are eukaryotic parasites lacking mitochondria, the ribosomes of which present prokaryote-like features. In order to better understand the structural evolution of rRNA molecules in microsporidia, the 5S and rDNA genes were investigated in *Encephalitozoon cuniculi*. The genes are not in close proximity. Non-tandemly arranged rDNA units are on every one of the 11 chromosomes. Such a dispersion is also shown in two other *Encephalitozoon* species. Sequencing of the 5S rRNA coding region reveals a 120 nt long RNA which folds according to the eukaryotic consensus structural shape. In contrast, the LSU rRNA molecule is greatly reduced in length (2487 nt). This dramatic shortening is essentially due to truncation of divergent domains, most of them being removed. Most variable stems of the conserved core are also deleted, reducing the LSU rRNA to only those structural features preserved in all living cells. This suggests that the *E.cuniculi* LSU rRNA performs only the basic mechanisms of translation. LSU rRNA phylogenetic analysis with the BASEML program favours a relatively recent origin of the fast evolving microsporidian lineage. Therefore, the prokaryote-like ribosomal features, such as the absence of ITS2, may be derived rather than primitive characters.

INTRODUCTION

In contrast to prokaryotes, the four RNAs contained in the cytoplasmic ribosomes of eukaryotic cells result from transcription of two types of units which are generally not linked and occur in tandemly repeated clusters. The largest unit, termed the rDNA, is transcribed by RNA polymerase I as a single precursor in which the small subunit (SSU) rRNA, 5.8S rRNA and large subunit (LSU) rRNA are surrounded by spacer sequences. The second unit yields 5S rRNA after transcription by RNA polymerase III. Also, eukaryotes have larger ribosomes containing larger SSU and LSU rRNAs and more proteins than prokaryotes. Surprisingly, in microsporidia, unicellular eukaryotes lacking mitochondria and displaying obligate intracellular parasitic life cycles, the low sedimentation coefficients for ribosomes (70 S), ribosomal subunits (30 S and 50 S) and the SSU (16 S) and LSU (23 S) rRNAs bring them closer to prokaryotes than eukaryotes (1,2). Our current knowledge of rDNA sequences from microsporidia derive mainly from analysis of a segment extending from the 5′-end of the SSU region to ~500 nt inside the LSU 5′-region (3,4). The SSU rRNA sequence revealed a length of ~1300 nt, as against ~1600 in prokaryotes and 1800 in eukaryotes (3,5), and the secondary structure model has been established (5). The known microsporidian rDNA sequence, which encompasses the 5.8S region, displayed another striking feature. The internal transcribed spacer two (ITS2), separating the 5.8S from the 28S region in all eukaryotes, is absent and a covalent link joins the 5.8S with the 23S region, as in all prokaryotic species (6). To date, the phylum Microspora is the only eukaryotic group exhibiting such a prokaryote-like feature. It is thus tempting to deduce that microsporidia have retained some molecular features of the primitive eukaryotic translation apparatus. This seems to be supported by molecular phylogenies based on SSU rRNA and translation elongation factor genes, indicating that Microspora was one of the earliest of the eukaryotic lineages to diverge (3,7). However, other investigations on microsporidian gene sequences are required to test the ‘primitive ribosome’ hypothesis.

In previous studies, we showed that *Encephalitozoon cuniculi*, known to infect various mammals, including man, possesses a very small nuclear genome (haploid genome size 2.9 Mbp)

*To whom correspondence should be addressed. Tel: +33 473 40 7457; Fax: +33 473 40 7455; Email: vivares@cicsun.univ-bpclermont.fr

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

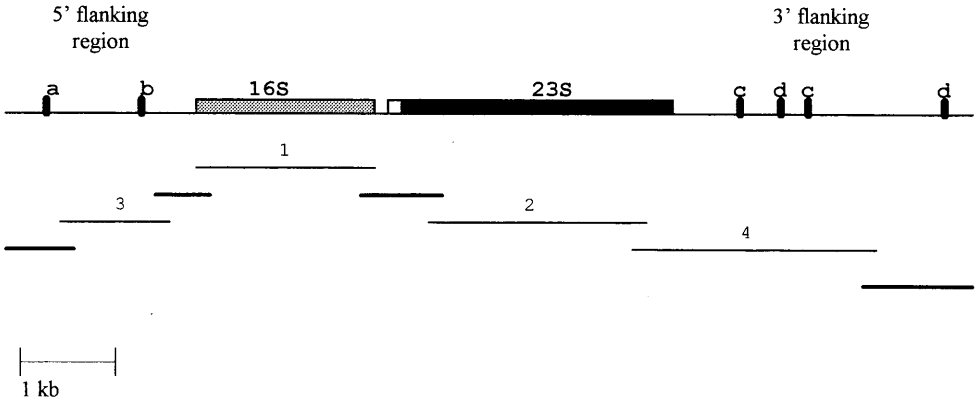


Figure 1. Genetic map of the *E.cuniculi* rDNA gene and flanking regions. Mature rRNA domains are boxed. The locations of the clones used for the sequencing strategy are indicated. The clones represented as bold lines were obtained by screening of a partial genomic plasmid DNA library and those numbered 1–4 by PCR amplification. The short repeats are marked in lower case letters (a–d).

divided into 11 chromosomes partially characterized through hybridization experiments with specific DNA probes (8,9). As part of investigations on some conserved genes of this 'minigenome', we have studied the chromosomal location of the rRNA transcription unit and 5S unit, determined their respective sequences and built secondary structures of the LSU and 5S rRNAs. The LSU rRNA of *E.cuniculi* appears to be the smallest identified so far in eukaryotes. Phylogenetic analysis with BASEML supports a late evolutionary origin for microsporidia and therefore argues against the primitiveness of the prokaryote-like features of their ribosomes.

MATERIALS AND METHODS

EMBL accession nos: AJ005581 (rDNA unit) and AJ005582 (5S unit).

Strain and cell culture

An *E.cuniculi* isolate from a mouse was kindly provided by Prof. E.U.Canning (Imperial College of Science, Technology and Medicine, London, UK). Spores were produced in MDCK cells and harvested from the culture supernatant as described (10).

DNA extraction

Spores were treated with 300 mM Tris–HCl, pH 9.0, 100 mM EDTA, 1% SDS to lyse the host cells and then incubated with 500 U DNase I (Gibco BRL) to remove residual host cell DNA. *Encephalitozoon cuniculi* DNA was released by boiling the spores as described elsewhere (11).

PCR amplification, cloning and sequencing of the rRNA genes

The entire coding region for SSU rRNA was amplified from DNA using two primers (CAGGTTGATTCTGCCTG and GACTCAGACCTTCCGATC) situated at the extreme 5′- and 3′-termini of a SSU rRNA published sequence (5). The PCR product was analysed on a 1% agarose gel and the DNA band was excised and then melted in 2 vol water. This PCR product was used as a probe to screen a partial library of *E.cuniculi* genomic plasmid DNA (9). The extracted plasmids were slotted on a Nylon membrane (BioTrans™; ICN) and then hybridized as described (12). The positive clones were sequenced by the Sanger technique using a Perkin-Elmer ABI 377 apparatus.

To amplify the 3′-end region of the LSU rDNA region, we used a direct specific primer (ACAGTGGGAATCTCGTTGC) in the nearest positive clone and a reverse primer (GTTCCTCTCGTACT) corresponding to a highly conserved region close to the 3′-end of LSU rRNA genes (GenBank/EMBL). The PCR product was cloned into the pGEM-T Easy Vector System I (Promega). Recombinant plasmids were sequenced as indicated above. Completion of the sequence of LSU rDNA was obtained by a modified single specific primer PCR (SSP-PCR) technique (13). The *E.cuniculi* genomic DNA was completely digested with *Xho*I and ligated in the pBluescript II SK+ vector (Stratagene). A specific primer determined in the previous PCR product (TAGA-CCGTCGTGAGACAGG) and the reverse vector primer were used successfully to amplify a DNA fragment downstream of the known sequence. This fragment was used as a probe to screen the partial library and another clone corresponding to the 3′-flanking region of the LSU rRNA was displayed. The detailed physical map obtained is shown in Figure 1.

The nucleotide sequence of the 5S gene 3′-end was obtained using the SSP-PCR technique with an oligodeoxynucleotide primer (CATCAGAACTCCGAAGTTAA) matching a conserved part of 5S eukaryotic sequences, including that of the microsporidian *Nosema bombycis* (14). The 5′-region was obtained with the same technique using a primer (TTGCAGCACCCGGTATTC) located in the fragment previously sequenced. The PCR products were cloned and then sequenced as described above.

Contour-clamped homogeneous electric field (CHEF) gel electrophoresis and restriction analyses

CHEF gel electrophoresis of agarose blocks including intact chromosomal DNA of microsporidian species was performed as described elsewhere (8). The *E.cuniculi* chromosomes were separately electroeluted from the electrophoretic gels and each then digested with the restriction endonucleases *Bam*HI, *Pst*I, *Sac*II and *Xho*I according to the manufacturer's instructions (Eurogentec). DNA fragments were electrophoresed on a 0.8% agarose gel. Southern blotting of molecular karyotypes and restriction analysis of *E.cuniculi* chromosomes were performed on Nylon membrane (BioTrans™; ICN) and hybridization was achieved.

Sequence alignment and secondary structure construction

The *E.cuniculi* rDNA sequence was manually aligned to a database of 48 eukaryotic LSU rDNA sequences representing all currently known evolutionary diversity. The 5S and 23S rRNA secondary structure models were produced by comparison with available models (15,16). Secondary structure was drawn using the ESSA program (17).

Phylogenetic analysis

Two distinct maximum likelihood methods were employed for phylogenetic analyses, the fastDNAml program (global branch swapping option set to 13) (18) and the BASEML program of the PAML package (19). These methods differ mainly in that the latter allows sequence sites to display distinct rates of evolution while the former expects all sites to evolve at the same rate. BASEML was used with eight categories of site rates and Kimura's two parameter model (20), allowing for unequal transition and transversion rates.

RESULTS AND DISCUSSION

Sequence, secondary structure and chromosomal location of the 5S rRNA gene in *E.cuniculi*

Using the SSP-PCR technique, a fragment of 287 bp was isolated and sequenced. A comparison with a compilation of eukaryotic 5S rRNA sequences shows a strong sequence homology over 120 nt, which is the expected size for a typical 5S rRNA, a highly conserved molecule in evolution.The *E.cuniculi* 5S rRNA shows the five helices and loops organizing all 5S rRNA secondary structure (21). All structural hallmarks differentiating eukaryotes from prokaryotes display the eukaryotic version, making this molecule clearly more related to eukaryotes than prokaryotes. Autoradiographic signals provided by hybridization experiments with a 5S DNA probe were associated with two chromosomes (V and IX) of *E.cuniculi*.

Chromosomal location and copy number of the rDNA unit

The *E.cuniculi* chromosomes separated by CHEF electrophoresis were hybridized with SSU and LSU rDNA probes. The rDNA distribution is unusual in that the copies are scattered over all chromosomes (Fig. 2A). It was of particular interest to make a comparison with the two other known *Encephalitozoon* species, *E.hellem* and *E.intestinalis*, typically detected in AIDS patients. Initial chromosomal separation experiments in these species also disclosed a reduced chromosomal size range. Assuming that the karyotype was fully resolved, the sum of the sizes corresponding to 10 bands identified so far in *E.intestinalis* provides an estimate of haploid genome size close to only 2.3 Mbp. As in *E.cuniculi*, the rDNA probes hybridized with all the *E.intestinalis* bands (Fig. 2B). Concerning *E.hellem*, one band representative of chromosome VIII (256 kbp) was the exception (Fig. 2C). This contrasts with *N.bombycis*, an insect microsporidian in which rDNA location is restricted to a single large chromosome (22). Intermediate cases also exist, as shown here for two fish microsporidians containing larger genomes: *Glugea atherinae* (19.6 Mbp) and *Spraguea lophii* (6.2 Mbp) (23). Half of their chromosomal bands were labelled with rDNA probes, i.e. eight of 16 in *G.atherinae* and six of 12 in *S.lophii* (Fig. 2D–E). The localization of rDNA on every chromosome has been previously found only in the genome of the vestigial nucleus (nucleomorph) of a eukaryotic endosymbiont within certain unicellular algae. The three chromosomes of this very reduced genome (660 kbp) included two rDNA copies in symetrical sub-telomeric positions (24).

Given the unusually high chromosomal dispersion of rDNA units in the three *Encephalitozoon* species, we undertook determination of the copy number in *E.cuniculi* by digestion experiments using three different restriction enzymes. Total genomic DNA, seven well-resolved individual chromosomal bands and remaining doublets (II/III and IV/V) were treated. For every digested DNA preparation, Southern blotting revealed a single labelled band with a size depending on the enzyme used (data not shown). This indicates the absence of a head-to-tail tandem repeat of the rDNA. The rDNA copy number per *E.cuniculi* haploid genome could therefore be estimated as 11. It should be noted that tandem organization of rRNA genes does not exist in several intracellular parasitic protozoans from the phylum Apicomplexa, which display a very low copy number: only two copies in *Theileria* (25), two to four in *Babesia* (26) and four to eight in *Plasmodium* (27). A sub-telomeric position can be excluded, as indicated by preliminary *Bal*31 nuclease digestion experiments.

Full sequence of the rDNA unit and major features of flanking regions

A combination of amplification, cloning and sequencing strategies allowed complete sequencing of a final contig of 8871 bp in length which encompasses the whole rDNA transcription unit with flanking regions. The sequence submitted to the EMBL database corresponds to the unit located on the smallest chromosome (chromosome I), which is currently subject to systematic sequencing in our laboratory (28). Analysis of several clones derived from total DNA supports an interchromosomal sequence homogeneity of the SSU and LSU rDNA sequences. The overall GC content of the contig is close to 50%. No open reading frame and no sequence homologous to either tRNAs or 5S rRNA were detected outside the coding regions of SSU and LSU rRNAs. In agreement with this, PCR experiments (for units located on chromosomes V and IX) did not reveal the presence of a neighbouring 5S rRNA or tRNA gene, unlike in bacteria and some fungal or protozoan species (see for example 29,30).

In the 5′-flanking region, two perfect copies of a 29 bp element and eight imperfect repeats of a 14 bp element were identified. They are located ~240 and 1300 bp upstream of the 5′-end of the SSU rRNA coding sequence respectively (Fig. 1, zones a and b). These short tandemly repeated elements within the 5′-flanking region, commonly found in promoters of rRNA genes, may be involved in control of transcription. However, the promoter region cannot be identified only on the basis of comparative sequence analysis. Likewise, the real length of the 5′-External Transcribed Spacer cannot be estimated because of the lack of sequence conservation with other eukaryotes. Various interspersed repeats were also identified in the 3′-flanking region, in particular, two elements of 51 and 43 bp, the location of which is shown in Figure 1 (zones c and d). The 51 bp element includes a high frequency of two 4 bp motifs differing by a single base pair (CTTC and CTGC). A few copies of larger elements were tandemly arranged, similarly to minisatellites. Despite the presence of numerous common short motifs (2–7 bp) containing GT in various combinations, a sequence comparison between the 5′- and the 3′-flanking regions failed to disclose significant

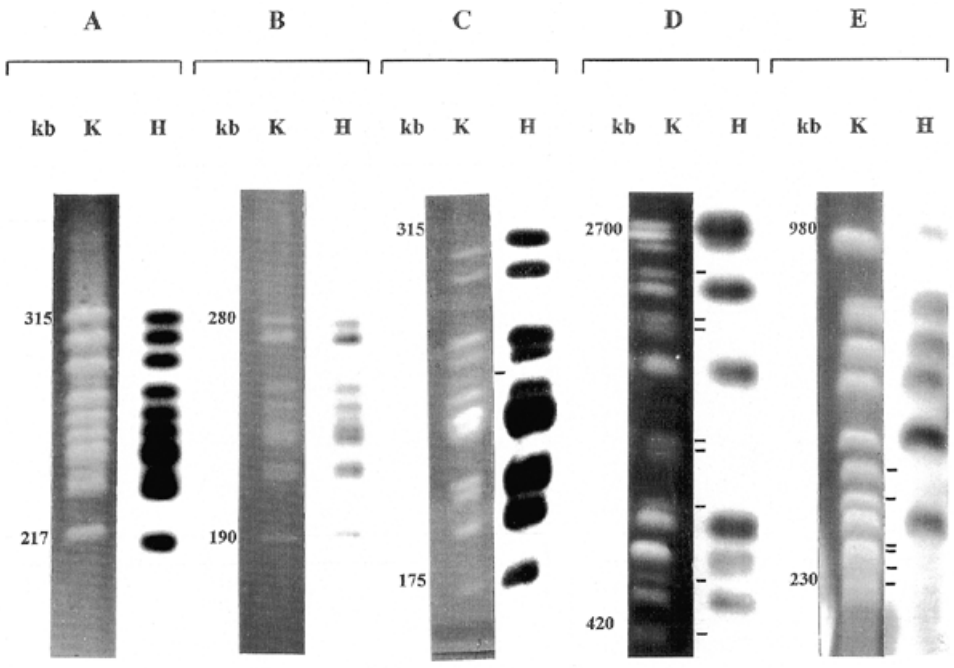


Figure 2. Southern hybridizations with the SSU rDNA probe revealing the chromosomal location of rDNA genes in the genome of the three human microsporidia *E.cuniculi* (**A**), *E.intestinalis* (**B**) and *E.hellem* (**C**) and of the two fish microsporidia *G.atherinae* (**D**) and *S.lophii* (**E**). The sizes of the smallest and the largest chromosomes are indicated in kilobases (kb) on the left of each molecular karyotype. K, molecular karyotype; H, hybridization pattern. Unlabelled chromosomes are marked on the right with dark lines.

homology. This absence of homology and the estimated copy number argue against a head-to-tail arrangement of rDNA units in *E.cuniculi*.

Table 1. Length comparison of shortened regions in *E.cuniculi*

Variable region	*E.cuniculi*	*Giardia*	Other eukaryotes	Eubacteria
B13_1–B17	106	107–114	142–215	99–143
C1	13	8–29	205–865	25–44
D13–D14[a]	9	4–6	5–29	3–6
D20	15	14	24–199	17–45
E15	0	0	15–22	0–26
E20	20	20–28	46–718	45
B21	0	12–14	12	12–13
D4	17	84–95	53–127	54–60 (73)
D14_1	8	37	35–88	3–6
E9_1	2	20–22	27–74	3
E11_1	2	19–22	23–83	0–14 (130)
G4[b]	0	6	8–65	35
G5	0	14	64–287	24
H1	4	20–42	81–235	22

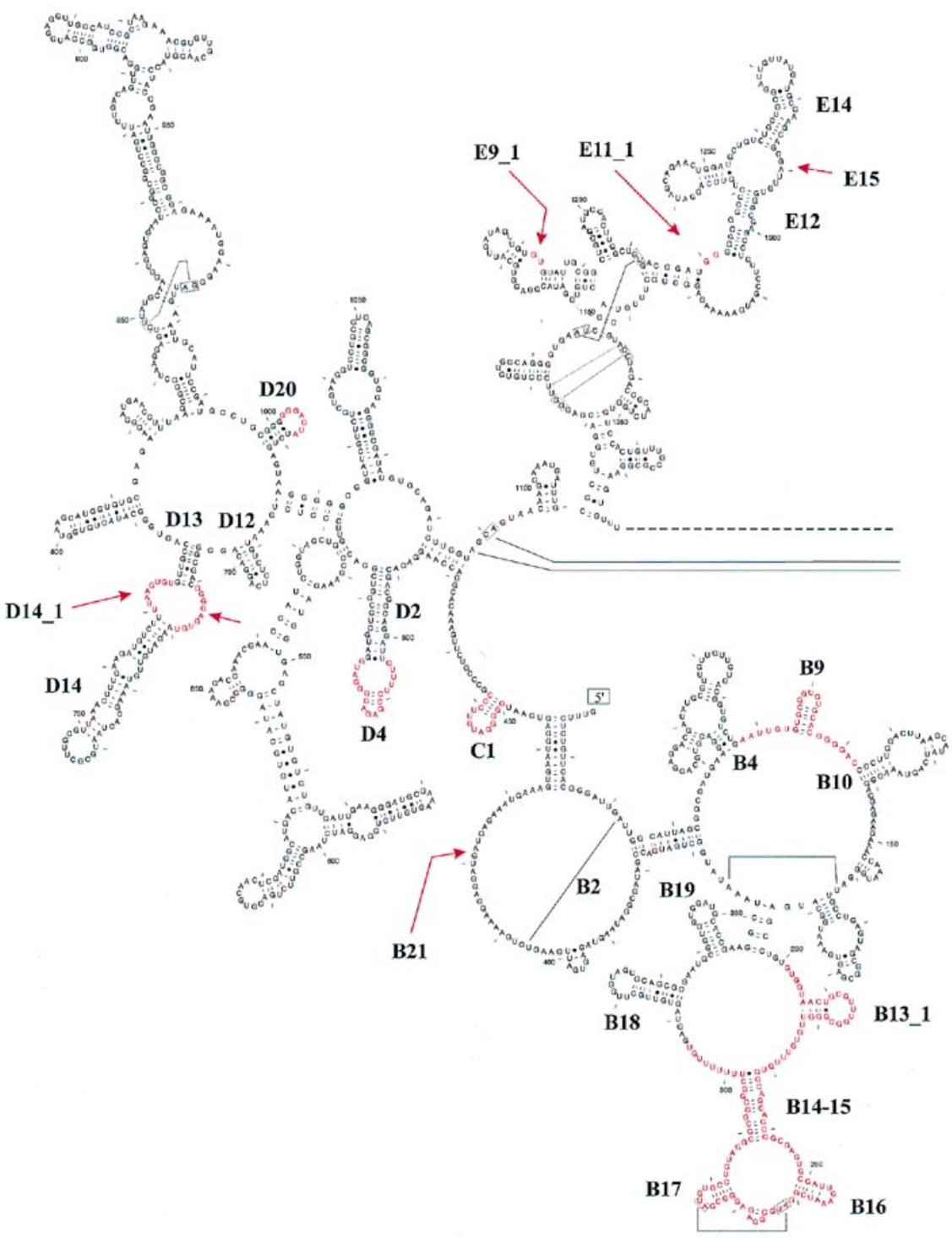
Length indications given for *E.cuniculi* correspond to the red regions in Figure 5.
[a]Unpaired nucleotides of the 5′-strand of the internal loop separating the two stems D13 and D14.
[b]The apical part of stem G4.

Secondary structure of the LSU rRNA

Sequence comparison with eukaryotic 5.8S and LSU rRNA sequences localized the 5′-end of the *E.cuniculi* LSU. In contrast, identification of the 3′-end involved a comparison at the secondary structure level because of the lack of significant homology from *E.cuniculi* position 2300 with any other species. This gives a length of 2487 nt for *E.cuniculi* LSU rRNA, with the 3′-end lying immediately upstream of the first cluster of tandem repeats. The considerable length reduction as compared with eukaryotes, including the early diverging diplomonad *Giardia* with a compact rDNA unit (31–33), and prokaryotes is accommodated over the regions which vary in length and sequence among species (Fig. NO TAGa and b), in particular within the previously identified divergent domains (34). These *E.cuniculi* regions are either in the size range of their shortest known version or even removed (Table 1b). Thus, variable regions C1 and E20, which are restricted in *Giardia* and *E.cuniculi* to two unbranched stems, make these species more related to eubacteria than to any other species, whereas G5 and H1_1–3, which generally present a complex folding in eukaryotes, have no equivalent in *E.cuniculi*. Therefore, *E.cuniculi* LSU rRNA is likely to be reduced to the structural features pertaining only to the universal core as defined by De Rijk *et al*. (35).

The variable region surrounding stem D4 is certainly the best example of this evolution. This region shows profound differences in its folding according to the three major kingdoms, in which three to five linear branches emerge from a central multiple loop (Fig. 3a). The structural diversification is not limited to the presence or absence of one or more structural features.Variations within each stem can give rise to patterns that are phylogenetically

a



preserved and specific to large phylogenetic groups. Despite its very small size, the LSU rRNA from three *Giardia* species folds according to the eukaryotic model and presents an equivalent for each stem (31). In contrast, the *E.cuniculi* LSU rRNA retains a single secondary structure feature from the universally conserved core: the 2 bp long hairpin stem closed by a GNRA tetraloop. These deletions of all regions containing group-specific structural features which were predicted to carry diversified functions (36–39) suggest that *E.cuniculi* LSU rRNA encodes only for universal functions involving the basic molecular mechanisms of translation.

A few punctual differences within the core structure have been considered as distinguishing large phylogenetic groups such as the three major kingdoms. In *E.cuniculi*, four pairings are more particularly interesting (Table 2). As in other eukaryotes, a non-canonical pair can be formed in the basal part of variable region E20 (1440–1453) and in stem G19 (2156–2164); C-A and

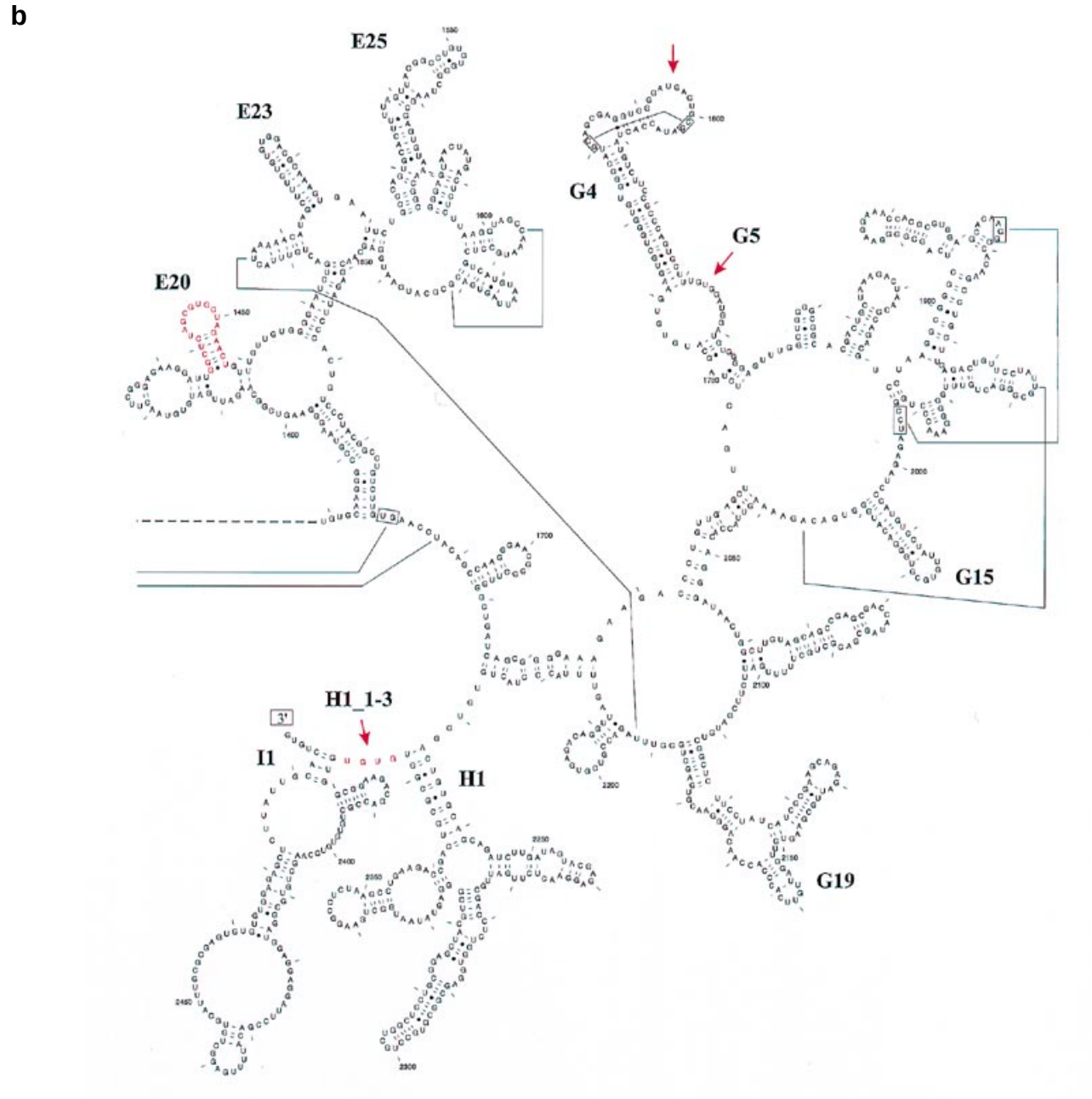


Figure 3. Secondary structure of the *E.cuniculi* 23S rRNA 5′-half (**a**) and 3′-half (**b**). The ends of the molecule were identified by homology with other eukaryotic species. Regions which vary in length and sequence in *E.cuniculi* are red and/or located by red arrows. All stems are named according to De Rijk *et al.* (35). Tertiary interactions (38) are shown by continuous lines.

A-C respectively. In contrast, within stem D12 (689–702) a canonical pair is present, as in eubacteria, whereas all other eukaryotes have U-U. At the tertiary structure level, although most phylogenetically preserved potential interactions are possible, the six Watson–Crick pair long complementarity between loops B19 and G15 is not supported in *E.cuniculi*.

Our comparative analysis performed at the secondary structure level of LSU rRNA also reveals that *E.cuniculi* does not share any common structural feature with prokaryotes, neither in variable regions nor in the conserved core (Fig. NO TAGa and b). Actually, the sole prokaryote-like feature is the absence of an ITS2-like sequence. This dramatic and surprising *E.cuniculi* feature could be interpreted as a primitive character. This region surrounding stem B9 is generally folded in three unbranched stems (B7–B9), in eukaryotes as in bacteria (Fig. 3b). Stems B7 and B8 are lacking in *E.cuniculi*, as in one *Giardia* species (*G.muris*). In bacteria, deletion of one of the two stems B7 and B8 was found only in extreme halophiles. As shown by others in *Vairimorpha* (6), stem B9, which can be unambiguously identified by sequence homology with eukaryotes, does not contain an ITS2 sequence and cleavage sites. Because helix regions can be inserted/deleted in a relatively short evolutionary time, as suggested by B7/B8 stem evolution in *Giardia* species, the absence of the ITS2 sequence might be a relatively recent event. A possible correlation with a relatively recent shortening of the *E.cuniculi* genome involving two steps can be postulated. Firstly,

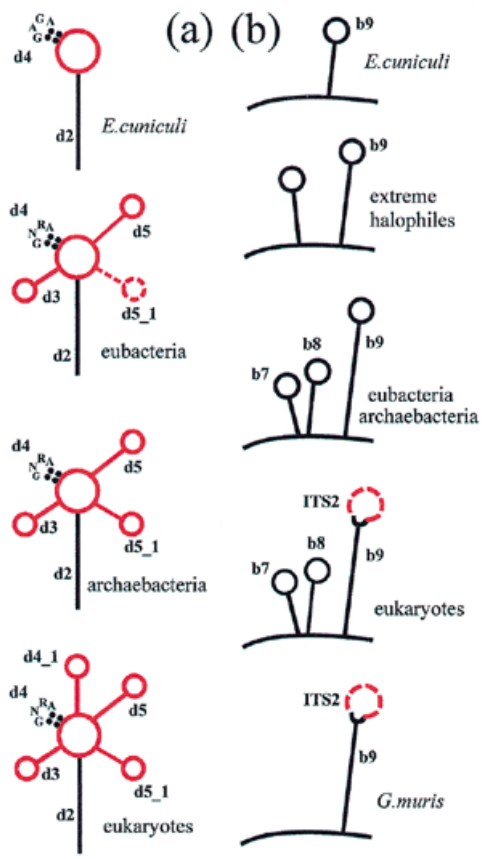


Figure 3. Diversification of constraints during evolution in the B9 and D4 regions. A backbone view is presented respectively for the region surrounding stem D4 (**a**) and for the region located between stems B4 and B10 (**b**). In (a) the variable regions are shown in red. Within the eubacterial model, stem d5_1, which is present in only a few species, is drawn as a red dotted line. In (b) the location of the internal spacer ITS2 is indicated by a red dotted line.

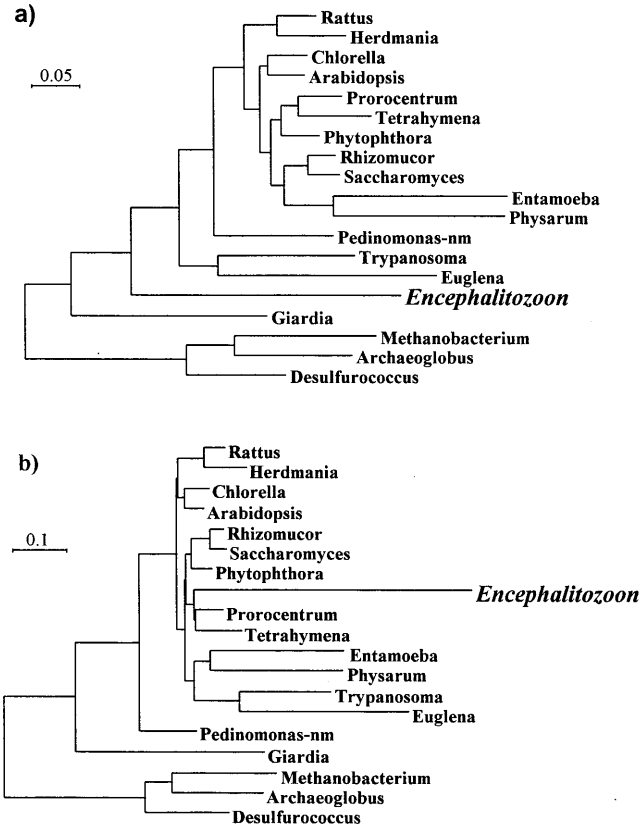


Figure 4. Phylogenetic relationships between 16 eukaryotic and three archaeal LSU rRNA sequences inferred by the fastDNAml method (**a**) and the BASEML method (**b**) rooted between eukaryotic and archaeal sequences. Horizontal branches are drawn proportional to inferred evolutionary distances (see scales). Species names and sequence accession numbers: *Rattus norvegicus*, X01069; *Herdmania momus*, X53538; *Chlorella ellipsoidea*, D17810; *Arabidposis thaliana*, X52320; *Rhizomucor racemosus*, M26190; *Saccharomyces cerevisiae*, J01355; *Phytophtora megasperma*, X75631; *Encephalitozoon cuniculi*, this work; *Prorocentrum micans*, X16108; *Tetrahymena thermophila*, X54512; *Entamoeba histolytica*, X65163; *Physarum polycephalum*, V01159; *Trypanosoma brucei*, X14553; *Euglena gracilis*, X53361; nucleomorph of *Pedinomonas minutissima*, U58510; *Giardia ardeae*, X58290; *Methanobacterium thermoautotrophicum*, X15364; *Archaeoglobus fulgidus*, M64487; *Desulfurococcus mobilis*, X05480. To accelerate computation of (b), the following well-established groupings were fixed: (*Rhizomucor*, *Saccharomyces*); (*Trypanosoma*, *Euglena*); (*Rattus*, *Herdmania*); (*Chlorella*, *Arabidopsis*); (*Methanobacterium*, *Archaeoglobus*), *Desulfurococcus*.

an essential *trans*-acting factor (an enzyme or small RNA) of the processing machinery was lost, then, in the absence of cleavage, the ITS2 sequence, which becomes a real new divergent domain, was eliminated like any other variable region. It should be stressed that ITS2 exists in *Pneumocystis carinii* and *Theileria parva* (38), which are, respectively, fungal and apicomplexan obligate intracellular parasites. Whether the intracellular parasitic way of life plays a significant role in rDNA reduction therefore remains an open question.

Table 2. Comparison of group-specific canonical base pairs, including the case of *E.cuniculi* 23S rRNA

Position[a]	*E.cuniculi*	Other eukaryotes	Eubacteria
22–406	A-U	C-G	C-G
689–702	G-C	U U	C-G
1440–1453	C A	C A	U-A
2156–2164	A C	A C	G-C

[a]Position in *E.cuniculi* 5.8S–23S.

Phylogenetic trees

Evolutionary relationships between 19 eukaryotic and archaeal LSU rRNA sequences representing all major eukaryotic and archaeal phyla with known LSU rRNA sequences were reconstructed using two maximum likehihood methods. A set of 1723 reliably aligned and gap-free sites was used. The program fastDNAml (Fig. 4a) produced a tree in which three fast evolving lineages represented by *Giardia* (Diplomonada), *Encephalitozoon* (Microspora), *Trypanosoma* and *Euglena* (Euglenozoa) are predicted to have diverged early from other eukaryotic lineages. The BASEML tree-building method differs from the fastDNAml method mainly in using a more realistic model of molecular evolution, allowing for rate heterogeneity among sequence sites,

and was shown to fit the data significantly better (40). The BASEML method produced a different tree from fastDNAml (Fig. 4b), where the *Encephalitozoon* sequence is placed in a relatively late emerging position with a very long terminal branch, indicating a high rate of evolution. The two fast evolving euglenozoan sequences also emerge later with BASEML than with fastDNAml. We therefore conclude in favour of the relatively late evolutionary origin of microsporidia. This is supported by several recent results. First, genes encoding mitochondrial HSP70 have been discovered in *Nosema locustae* (41), *Vairimorpha necatrix* (42) and *Encephalitozoon* spp. (43). Therefore, the hypothesis of a primitive lack of mitochondria in microsporidia cannot be taken to corroborate an early evolutionary origin of these organisms. Second, β-tubulin sequences indicated a late evolutionary origin of microsporidia, close to that of fungi (44), supported by a few common structural and developmental features. Third, microsporidia seem to have a high rate of molecular evolution (41–43), as found in several other intracellular parasites (45). Together, these data suggest that the early evolutionary origin of microsporidia deduced from SSU rRNA analyses and from Figure 4 may result from the long branch attraction artefact known to bias phylogenetic inferences in the presence of evolutionary rate heterogeneity among sequences (46). The BASELM tree-building method is too computation-intensive to allow evaluation of tree reliability using the bootstrap approach. As the internal branches are very short, clustering of *Encephalitozoon* with the dinoflagellate *Prorocentrum* and the ciliate *Tetrahymena* may not be significant. It should be noted that parasitic dinoflagellate species are frequent and some of them are endoparasites able to form plasmodes and spores, like microsporidia. Parasitism by dinoflagellates is often associated with the loss of organelles such as flagella and chloroplasts (47). It is therefore conceivable that microsporidian radiation has been characterized by secondary loss of various organelles, as argued for the lack of mitochondria.

The evidence presented here that the rDNA copies are highly dispersed throughout the *E.cuniculi* nuclear genome should stimulate further work toward a precise mapping of chromosomal regions surrounding each copy. Transcription controlling sequences, the precursor transcript and maturation intermediates also remain to be identified. Further phylogenetic analyses with protein coding genes should be done to test the placement of microsporidia within the terminal crown.

ACKNOWLEDGEMENTS

We thank B.Chebance and R.Guerry for excellent technical assistance. E.P. is supported by a grant from the French Ministry for higher Education and Research.

REFERENCES

1 Ishihara,R. and Hayashi,Y. (1968) *J. Invert. Pathol*., **11**, 243–248.
2 Curgy,J., Vavra,J. and Vivarès,C.P. (1980) *Biol. Cell*, **38**, 49–52.
3 Vossbrinck,C.R., Maddox,J.V., Friedman,S., Debrunner-Vossbrinck,B.A. and Woese,C.R. (1987) *Nature*, **326**, 411–414.
4 Vossbrinck,C.R., Baker,M.D., Didier,E.S., Debrunner-Vossbrinck,B.A. and Shadduck,J.A. (1993) *J. Eukaryote Microbiol*., **40**, 354–362.
5 Hartskeerl,R.A., Schuitema,A.R.J. and De Wachter,R. (1993) *Nucleic Acids Res*., **21**, 1489.
6 Vossbrinck,C.R. and Woese,C.R. (1986) *Nature*, **320**, 287–288.
7 Kamaishi,T., Hashimoto,T., Nakamura,Y., Masuda,Y., Nakamura,F., Okamoto,K.-I., Shimizu,M. and Hasegawa,M. (1996) *J. Biochem*., **120**, 1095–1103.
8 Biderre,C., Pagès,M., Méténier,G., Canning,E.U. and Vivarès,C.P. (1995) *Mol. Biochem. Parasitol*., **74**, 229–231.
9 Biderre,C., Duffieux,F., Peyretaillade,E., Glaser,P., Peyret,P., Danchin,A., Pagès,M., Méténier,G. and Vivarès,C.P. (1997) *Gene*, **191**, 39–45.
10 Beauvais,B., Sarfati,C., Challier,S. and Dérouin,F. (1994) *Antimicrobial Agents Chemother*., **38**, 240–248.
11 Ombrouck,C., Ciceron,L. and Desportes-Livage,I. (1996) *Parasite*, **3**, 85–86.
12 Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
13 Shyamala,V. and Ames,G.F.L. (1993) *Methods Enzymol*., **217**, 436–446.
14 Kawakami,Y., Inoue,T., Kikuchi,M., Takayanagi,M., Sunairi,M., Ando,T. and Ishihara,R. (1992) *J. Seric. Sci. Jap*., **61**, 321–327.
15 Szymansky,M., Specht,T., Barciszewska,M.Z., Barciszewska,J. and Erdmann,V.A. (1998) *Nucleic Acids Res*., **26**, 156–159.
16 Gutell,R.R, Gray,M.W. and Schnare,M. (1993) *Nucleic Acids Res*., **21**, 3055–3074.
17 Chetouani,F., Monestié,P., Thébault,P., Gaspin,C. and Michot,B. (1997) *Nucleic Acids Res*., **25**, 3514–3522.
18 Olsen,G.J., Matsuda,H., Hagstrom,R. and Overbeek,R. (1994) *Comput. Appl. Biosci*., **10**, 41–48.
19 Yang,Z. (1997) *Comput. Appl. Biosci*., **13**, 555–556.
20 Kimura,M. (1980) *J. Mol. Evol*. **16**, 111–120.
21 DeRijk,P., Neefs,J.-M., VandePeer,Y. and De Wachter,R. (1992) *Nucleic Acids Res*., **20**, 2075–2089.
22 Kawakami,Y., Inoue,T., Ito,K., Kitamizu,K., Hanawa,C., Ando,T., Iwano,H. and Ishihara,R. (1994) *J. Invert. Pathol*., **64**, 147–148.
23 Biderre,C., Pagès,M., Méténier,G., David,D., Bata,J., Prensier,G. and Vivarès,C.P. (1994) *C.R. Acad. Sci*., **317**, 399–404.
24 Eschbach,S., Hofmann,C.J.B., Maier,U.G., Sitte,P. and Hansmann,P. (1991) *Nucleic Acids Res*., **19**, 1779–1781.
25 Kibe,M.K., Ole-MoiYoi,O.K., Nene,V., Khan,B., Allsopp,B.A., Collins,N.E., Morzaria,S.P., Gobright,E.I. and Bishop,R.P. (1994) *Mol. Biochem. Parasitol*., **66**, 249–259.
26 Dalrymple,B.P., Dimmock,C.M., Parrodi,F. and Wright,I.G. (1992) *Int. J. Parasitol*., **22**, 851–855.
27 McCutchan,T.F., Li,J., McConkey,G.A., Rogers,M.J. and Waters,A.P. (1995) *Parasitol. Today*, **11**, 134–143.
28 Duffieux,F., Peyret,P., Roe,B.A. and Vivares,C.P. (1998) *Microbial Comp. Genomics*, **3**, 1–11.
29 Rubin,G.M. and Sulston,J.E. (1973) *Mol. Biol*., **79**, 521–530.
30 Curtis,S.E. and Rawson,J.R.Y. (1981) *Gene*, **15**, 237–247.
31 Van Keulen,H., Gutell,R.R., Campbell,S.R., Erlandsen,S.L. and Jarroll,E.L. (1992) *J. Mol. Evol*., **35**, 318–328.
32 Healey,A., Mitchell,R., Upcroft,J.A., Boreham,P.F.L. and Upcroft,P. (1990) *Nucleic Acids Res*., **18**, 4006.
33 Van Keulen,H., Horvat,S., Erlandsen,S.L. and Jarroll,E.L. (1991) *Nucleic Acids Res*., **19**, 6050.
34 Michot,B., Hassouna,N. and Bachellerie,J.-P. (1984) *Nucleic Acids Res*., **12**, 4259–4279.
35 De Rijk,P., Caers,A., Van De Peer,Y. and De Wachter,R. (1998) *Nucleic Acids Res*., **26**, 183–186.
36 Michot,B. and Bachellerie,J.-P. (1987) *Biochimie*, **69**, 11–23.
37 Michot,B., Qu,L.-H. and Bachellerie,J.-P. (1990) *Eur. J. Biochem*., **188**, 219–229.
38 Schnare,M.N., Damberger,S.H., Gray,M.W. and Gutell,R.R. (1996) *J. Mol. Biol*., **256**, 701–719.
39 Chenuil,A., Solignac,M. and Michot,B. (1997) *Mol. Biol. Evol*., **14**, 578–588.
40 Yang,Z and Roberts,D. (1995) *Mol. Biol. Evol*., **12**, 451–458.
41 Germot,A., Philippe,H. and Le Guyader,H. (1997) *Mol. Biochem. Parasitol*., **87**, 159–168.
42 Hirt,R.P., Healy,B., Vossbrinck,C.R., Canning,E.U. and Embley,T.M. (1997) *Curr. Biol*., **7**, 995–998.
43 Peyretaillade,E., Broussole,V., Peyret,P., Méténier,G., Gouy,M. and Vivarès,C.P. (1998) *Mol. Biol. Evol*., **15**, 683–689.
44 Edlind,T.D., Li,J., Visvesvara,G.S., Vodkin,M.H., McLaughlin,G.L. and Katiyar,S.K. (1996) *Mol. Phylogenet. Evol*., **5**, 359–367
45 Moran,N. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 2873–2878.
46 Kuhner,M.K. and Felsenstein,J. (1994) *Mol. Biol. Evol*., **11**, 459–468.
47 Taylor,F.J. (1980) *Biosystems*, **64**, 311–329.