Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*

Michaël D. Katinka*, Simone Duprat*, Emmanuel Cornillot†, Guy Méténier†, Fabienne Thomarat‡, Gérard Prensier†, Valérie Barbe*, Eric Peyretaillade†, Philippe Brottier*, Patrick Wincker*, Frédéric Delbac†, Hicham El Alaoui†, Pierre Peyret†, William Saurin*, Manolo Gouy‡, Jean Weissenbach* & Christian P. Vivarès†

* *Genoscope, UMR CNRS 8030, CP 5706, 91057 Evry cedex, France*
† *Parasitologie Moléculaire et Cellulaire, Laboratoire de Biologie des Protistes, UMR CNRS 6023, Université Blaise Pascal, 63177 Aubière cedex, France*
‡ *Laboratoire de Biométrie et Biologie Evolutive, UMR CNRS 5558, Université Lyon I, 69622 Villeurbanne cedex, France*

Microsporidia are obligate intracellular parasites infesting many animal groups[1]. Lacking mitochondria and peroxysomes, these unicellular eukaryotes were first considered a deeply branching protist lineage[2] that diverged before the endosymbiotic event that led to mitochondria. The discovery of a gene for a mitochondrial-type chaperone[3–5] combined with molecular phylogenetic data[6–9] later implied that microsporidia are atypical fungi that lost mitochondria during evolution. Here we report the DNA sequences of the 11 chromosomes of the ~2.9-megabase (Mb) genome of *Encephalitozoon cuniculi* (1,997 potential protein-coding genes). Genome compaction is reflected by reduced intergenic spacers and by the shortness of most putative proteins relative to their eukaryote orthologues. The strong host dependence is illustrated by the lack of genes for some biosynthetic pathways and for the tricarboxylic acid cycle. Phylogenetic analysis lends substantial credit to the fungal affiliation of microsporidia. Because the *E. cuniculi* genome contains genes related to some mitochondrial functions (for example, Fe–S cluster assembly), we hypothesize that microsporidia have retained a mitochondrion-derived organelle.

Encephalitozoon cuniculi infects various mammals, including humans, and can cause digestive and nervous clinical syndromes in HIV-infected or cyclosporine-treated people[1]. Its reproduction proceeds as a sequence of two major stages: merogony, involving the multiplication of large, wall-lacking cells (meronts); and sporogony, leading to small, thick-walled spores. The sporal invasive apparatus is characterized by a long polar tube that can be quickly extruded then used for transferring the sporoplasm into the target cell. Consisting of 11 linear chromosomes ranging from 217 to 315 kb, the *E. cuniculi* genome is remarkably reduced (~2.9 Mb)[10]. The nucleotide sequence of the smallest chromosome has been recently reported[11]. The full sequencing of this minimal genome among eukaryotes was expected to provide insight into the metabolism and general biology of microsporidia and to help in the understanding of the evolutionary history of amitochondriate eukaryotes currently considered 'curious fungi'[9] (see Table 1).

The chromosome sequences were determined through a plasmid library (~3 kb inserts) and a miniBAC library (~20–25 kb inserts) totalling 15 genome equivalents (~46 Mb). All chromosomes possess a 'unique sequences' core region flanked by two 28-kb divergently oriented regions, each including one ribosomal DNA unit[11,12]. The subtelomeric repeats upstream from rDNA are mostly degenerated minisatellites, whereas downstream repeats consist essentially of non-polymorphic microsatellite arrays. The chromosome cores lack simple sequence repeats, minisatellite arrays and known transposable elements. No imprint of retrogenes or pseudoretrogenes of either polymerase (pol) II or pol III type was found. General features of genome organization are indicated in Table 1. The gradual increase in G + C content at the core centre described in chromosome I (chrI)[11] exists in all the chromosomes (maximum 51.0%). The 1,997 protein-coding DNA sequences (CDSs) represent about 90% of the chromosome cores, as a result of generally short intergenic regions (see Supplementary Information). Gene density is slightly lower than that observed in the nucleomorph genome of the cryptomonad *Guillardia theta*[13]. Only about 44% of the CDSs are assigned to functional categories and about 6% to conserved hypothetical proteins (Fig. 1). In contrast to the nucleomorph genome[13], no overlapping of CDSs with predicted functions was revealed. Structural or functional clusters are rare and never composed of more than two CDSs (for example, histones H3 and H4 on chrIX). Genome compaction can also be related to gene shortening, as indicated by the length distribution of all potential proteins (Fig. 2a). The mean and median lengths of all potential *E. cuniculi* proteins are only 359 and 281 amino acids, respectively. We compared the lengths of 350 proteins with *Saccharomyces cerevisiae* homologues (Fig. 2b). More than 85% of these proteins are shorter than in yeast, with a mean relative size difference of 14.6%. From the analysis of the protein size distributions derived from sequenced genomes, it has been suggested that the lengthening of proteins in eukaryotes (non-parasitic species) allows for more complex regulation networks[14]. Thus, protein shortening in *E. cuniculi* may reflect reduced protein–protein interactions as a result of various gene losses linked to the intracellular parasitic nature (Fig. 2a, b).

Perfect segmental duplications of 0.5–10-kb coding and non-coding sequences represent about 3.7% of the core region in average (from ~1% in chrII, III, IV and VI to ~7% in chrIX; see Supplementary Information). A segment carrying four enzyme-coding genes is perfectly duplicated in the extreme part of the chrI core[11] but partially duplicated near the end of chrVIII (truncated serine hydroxymethyltransferase gene). The genome core sequences exhibit a very low base polymorphism, restricted to a few positions in eight chromosomes. A duplicated gene homologous to CTP synthases (chrXI) revealed a rare polymorphic tract of 189 base pairs (bp) (116,847–117,036). Hybridization experiments indicate that each polymorphic sequence is present in about 50% of the DNA molecules (data not shown), suggesting that they are alleles and supporting the diploidy hypothesis[10,12].

A large proportion of CDSs is assigned to the conservation and transmission of genetic information as well as to protein modification and intracellular transport processes (Fig. 1), but with a significant degree of simplification, mainly related to the lack of DNA-containing organelles. For example, subunits for DNA pol I (δ), pol II (ε) and pol III (α but not β) are present while there is no candidate for the mitochondrial DNA polymerase γ. Apart from the telomerase catalytic subunit, no RNA-dependent reverse transcriptase was found. This and the lack of any retrotransposition elements may explain the absence of pseudoretrogenes. The potential transcription machinery includes RNA pol I, II and III, more than 70 messenger RNA transcription-associated proteins and a virtually

Table 1 General features of the *E. cuniculi* genome

Total sequenced length	2,507,519 bp
G + C content	
Protein-coding regions	47.6%
Intergenic regions	45.0%
Telomeric and subtelomeric regions	52.9%
No. of protein-coding sequences	1,997
Mean intergenic distance	129 bp
Gene density of chromosome cores	1 CDS per 1,025 bp
No. and sizes of spliceosomal introns	13 (23–52 bp)
No. of 16S–23S rRNA genes	22*
No. of 5S rRNA genes	3 (on chrV, VII, IX)
No. of tRNA genes	44†
No. and sizes of tRNA introns	2 (16, 42 bp)

* Two per chromosome.
† On all chromosomes.

complete set of genes for splicing, 5′ and 3′ processing. An (A + T)-rich consensus transcription initiation sequence is revealed in the 120-bp region upstream from the initiation codon of numerous genes, suggesting short 5′ leaders. Seventy different eukaryotic-type ribosomal proteins (40 from the large subunit, 30 from the small subunit) are predicted, which is slightly lower than in the amitochondriate protist *Giardia lamblia* (74)[15]. Compared with the cytoplasmic ribosome of *S. cerevisiae*, the missing proteins are LP1, L14, L29, L38, L40, L41, S27 and S27A. Small putative spliceosomal introns with usual GT–AG boundaries were detected in 11 ribosomal protein genes (S8, S17, S24, S26, L5, L19, L27A, L37, L37A and L39). They start either in the initiator ATG or the next codon, as often observed in yeast[16] or nucleomorph[13] genomes. Two more internal introns create frame shifts within a CDP-diacylglycerol serine phosphatidyltransferase gene (chrXI). Two of the 44 transfer RNA genes (tDNA^{Ile} and tDNA^{Tyr}) also harbour a small intron.

Reduced metabolic capacities and low diversity of transporters can be inferred from the genome sequence, as illustrated in Fig. 3. The repertoire for the biosynthesis of amino acids is restricted to asparagine synthetase and serine hydroxymethyltransferase genes. Genes for *de novo* biosynthesis of purine and pyrimidine nucleotides are absent but several nucleotide interconversions are predicted. Genes encoding a fatty acid synthase complex are lacking, which supports the uptake of host-derived fatty acids[17]. The *E. cuniculi* spore membrane contains cholesterol. This sterol might also be of host origin, as no gene for the conversion of farnesyl-PP into cholesterol was detected. In contrast, *E. cuniculi* seems to be capable of synthesizing usual membrane phospholipids. Genes for principal enzymes for the synthesis and degradation of trehalose confirm that this disaccharide could be the major sugar reserve in microsporidia[18], as in other fungi. A complete glycolytic glucose-to-pyruvate pathway is predicted. In contrast, genes required for the tricarboxylic acid cycle, fatty acid β-oxidation, respiratory electron-transport chain and the F_0F_1-ATPase complex are absent. Thus, ATP production in microsporidia would be possible by substrate-level phosphorylation only. As proliferating microsporidia recruit host mitochondria near their plasma membrane, it has been proposed that these parasites could use host-derived ATP[18]. This is reinforced by the finding of four genes encoding ADP/ATP carrier proteins that are homologous to ADP/ATP translocases from chloroplasts and obligate intracellular bacteria (*Rickettsia* and *Chlamydia*) capable of importing host ATP[19]. The fate of pyruvate remains difficult to predict because of the lack of genes for lactate and ethanol fermentation as well as for glycolysis reversal. A potential cytosolic glycerol-3P dehydrogenase (GPDH) might serve to reoxidize the NADH produced during glycolysis. Surprisingly, two CDSs have significant similarity to the subunits of the E1 component of the mitochondrial pyruvate dehydrogenase complex. Pyruvate decarboxylation could be inferred, but, in the absence of evidence for E2 and E3 components, a subsequent production of acetyl coenzyme A cannot be concluded.

Microsporidia have a presumably simplified Golgi apparatus in which a *cis–trans* polarity is not cytologically distinguishable but that is central in sporogony-specific secretion processes[1]. The spore wall protein SWP1 (ref. 20) is encoded by a unique gene on chrX whereas two genes for the polar tube proteins PTP1 and PTP2 are arranged in tandem on chrVI[21]. The set of chaperones for protein folding includes a complete oligomeric TCP-1 complex, four members of the HSP70 system but no chaperonin CPN60. The mitochondrial-type HSP70 has been previously characterized in three different microsporidian species[3–5]. Initial steps of N-glycosylation using UDP–*N*-acetylglucosamine (UDP-GNAc) and GDP-mannose (GDP-Man) may occur, but further trimming by mannosidases associated with endoplasmic reticulum or Golgi apparatus and formation of a complex N-linked oligosaccharide are not supported. The major sugar used for O-linked glycosylations would be mannose (two mannosyltransferases of the fungal PMT family). The lack of genes for the two enzymes involved in phosphorylation of mannose residues argues for the absence of sorting of lysosomes. Membrane fusion and recognition of some target membrane processes are sustained by a restricted set of potential proteins for Golgi and post-Golgi trafficking. The constitutive secretion pathway leading to the plasma membrane may involve some characteristic Rab proteins. Several potential partners for *trans*-Golgi and endosome transport include β1-adaptin, Vps1-like dynamin and vacuolar protein sorting-associated proteins, confirming that the Golgi-like apparatus is functionally polarized. Endocytosis of certain macromolecules was previously shown in *Spraguea lophii* sporoplasms[18], when maintained *in vitro* in a cell culture medium. Several genes are also suggestive of an endocytosis pathway in *E. cuniculi*. This might drive the internalization of macromolecular ligands representing sources of fatty acids, cholesterol or iron (see Fig. 4).

The evolutionary origin of microsporidia has been much debated

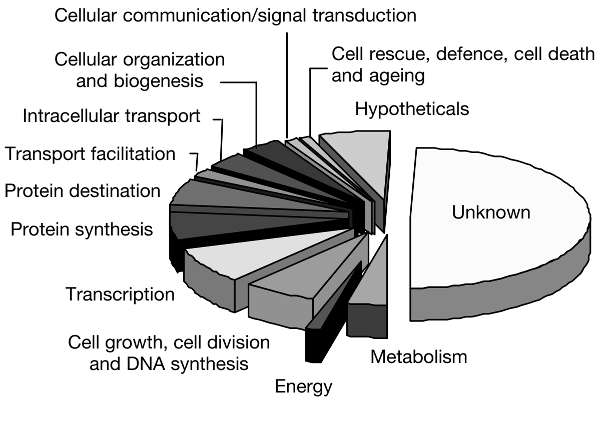

Figure 1 Distribution of predicted *E. cuniculi* proteins among functional groups. The corresponding gene list is given in the Supplementary Information.

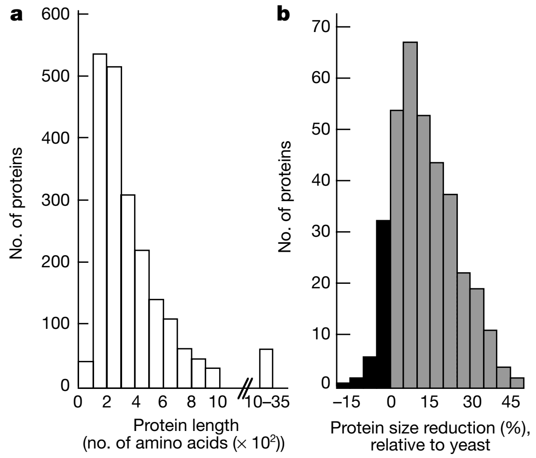

Figure 2 Sizes of *E. cuniculi* (*Ec*) proteins and comparison with *S. cerevisiae* (*Sc*) homologues. **a**, Distribution of the lengths of all the potential *E. cuniculi* protein chains. Only six have more than 2,000 amino acids (maximum 3,456 amino acids). **b**, Degrees of reduction in length of *E. cuniculi* proteins ($n = 350$) relative to those of *S. cerevisiae*, expressed as a percentage: 100(*Sc* protein length – *Ec* protein length)/(*Sc* protein length). The positive classes representative of shorter *E. cuniculi* proteins are in grey. The dynein heavy chain, the largest protein chain with a clearly predicted function (3,151 amino acids), is ~23% shorter than the yeast homologue (4,092 amino acids). Mean values associated with major functional categories are 7.3% for 'protein synthesis', 11.0% for 'protein destination', 11.4% for 'metabolism/energy', 14.4% for 'intracellular transport', 15.3% for 'transcription' and 20.1% for 'cell growth, cell division and DNA synthesis'.

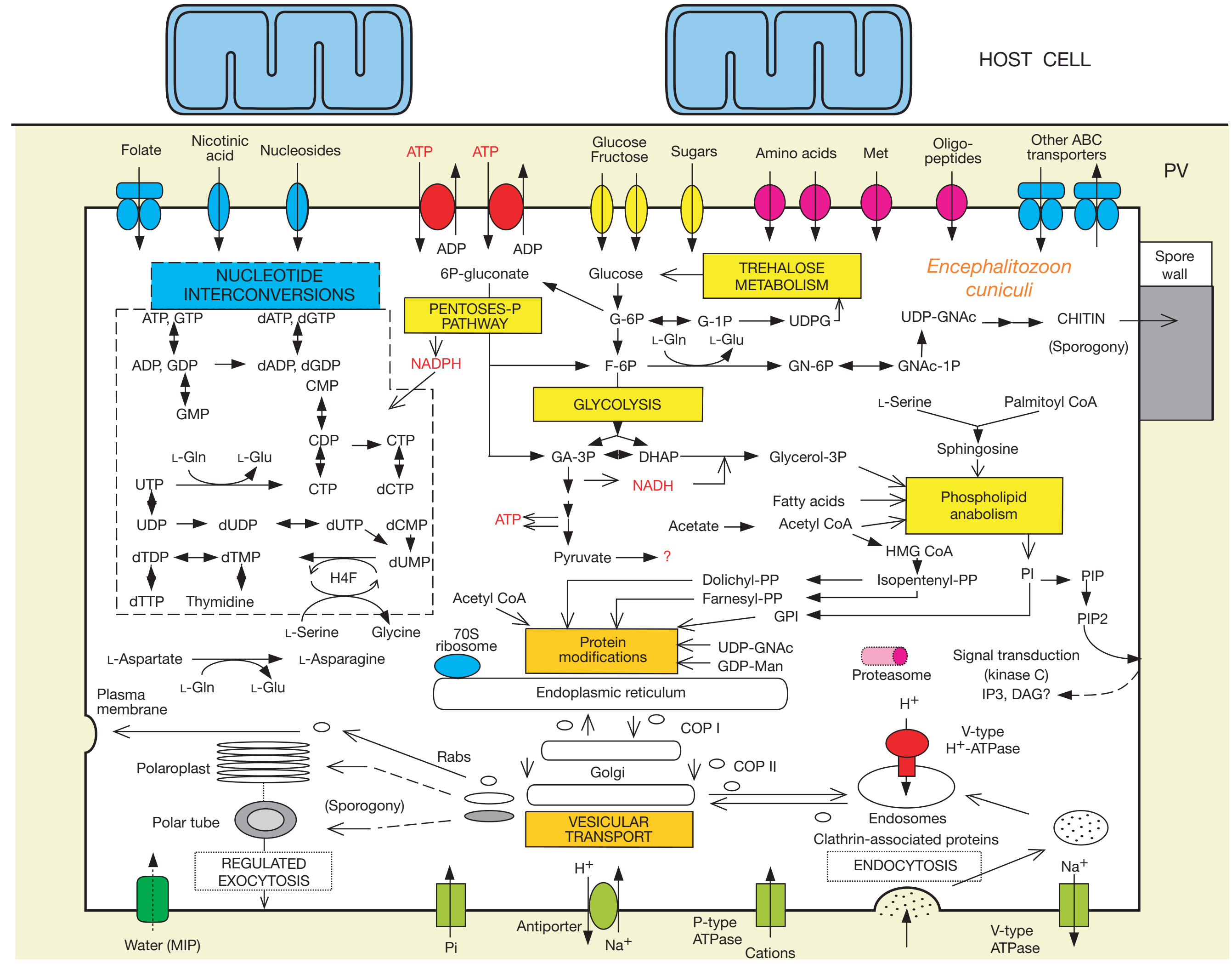

Figure 3 An overview of metabolism and transport in *E. cuniculi*, as deduced from genome sequence analysis. Pathways for nucleotide biosynthesis, energy production and chitin biosynthesis are indicated. Endocytosis and vesicular transport involving a *cis–trans* polarized Golgi apparatus are also illustrated. Potential transporters associated with the plasma membrane are shown with indications on substrate specificity. Question marks correspond to major uncertainties about the fate of pyruvate and the production of second messengers for signal transduction. The parasite is represented within a parasitophorous vacuole (PV) of the host cytoplasm.

but strong evidence supporting a fungal origin of these organisms has recently accumulated[6–9]. The present genome sequence extends this evidence: phylogenetic analyses of putative genes for seryl-tRNA synthetase, transcription initiation factor IIB, subunit A of vacuolar ATPase, and a GTP-binding protein place microsporidia as a sister group of fungi with bootstrap supports ranging from 70% to 92% (see Supplementary Information; a systematic phylogenetic analysis of the genome will be presented elsewhere). Genes of putative mitochondrial evolutionary origin in the *E. cuniculi* genome were systematically sought by comparison with the 423 recently surveyed yeast mitochondrial proteins encoded by the nuclear and the mitochondrial genomes[22]. Twenty-two genes with significant similarity to the yeast genes were identified and phylogenetic analysis showed that six of them are closely related to homologues from α-proteobacteria, the bacterial group from which mitochondria are believed to derive[23]. The yeast homologues of these six proteins are ATM1 (ABC transporter), ISU1/ISU2 (NIFU-like protein), NFS1 (unique homologue of bacterial ISC-S and NIF-S), SSQ1 (heat-shock protein of relative molecular mass 70,000), YAH1 (ferredoxin) and PDB1 (β-subunit of pyruvate dehydrogenase component E1). The first five proteins are typically involved in the Fe–S cluster assembly machinery, an essential function of mitochondria[24].

The presence of characteristic domains and key amino acids suggests that the potential mitochondrial-type proteins are functional. Moreover, PSORT analysis predicts amino-terminal presequences for the targeting of five of these proteins (see Supplementary Information). A common feature is an arginine residue at −2 relative to the cleavage site, similar to presequences of mitochondrial and hydrogenosomal proteins[25]. The amitochondriate protozoan *Entamoeba histolytica* has recently been shown to contain a residual mitochondrion-derived organelle[26–28] that some authors have called mitosome[28]. Considering the set of potential *E. cuniculi* proteins usually associated with mitochondria, we propose that a cryptic organelle is also present in microsporidia (Fig. 4). This mitosome would be significantly different from hydrogenosomes found in several anaerobic unicellular eukaryotes (type II anaerobes)[29]. Hydrogen production through pyruvate catabolism seems unlikely in microsporidia (because of a lack of a hydrogenase gene). The development of microsporidia in various aerobic host cells is suggestive of a rather high O_2 tolerance and therefore of an efficient protection against oxidative stress. No catalase gene is identified, in agreement with the lack of peroxisomes. Thus, in addition to glutathione and thioredoxin-based systems, *E. cuniculi* might use its unique manganese superoxide dismutase as an antioxidant.

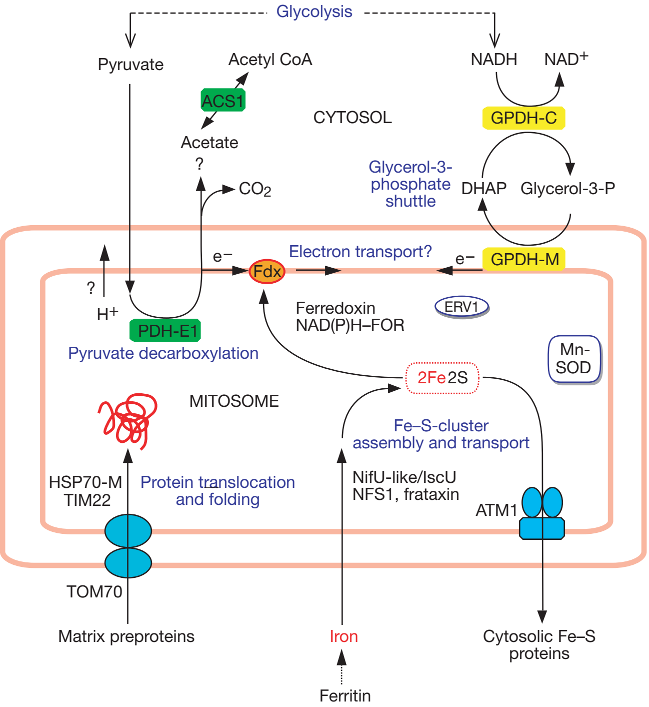

Figure 4 Conceptual scheme of a mitochondrion-derived organelle ('mitosome') in *E. cuniculi* suggested by the detection of several homologues of mitochondrial proteins. Under this hypothesis, pyruvate decarboxylation may occur through a heterotetrameric form (α2β2) of the pyruvate dehydrogenase E1 component (PDH-E1) and transfer of reducing power towards the organelle transits through a glycerol-3-phosphate shuttle involving both cytosolic (GPDH-C) and mitochondrial (GPDH-M) glycerol-3-phosphate dehydrogenases. By analogy with hydrogenosomal pyruvate:ferredoxin oxidoreductase, a system based on ferredoxin (Fdx) and NAD(P)H ferredoxin:oxidoreductase (FOR) is assumed to be used for acetate production. A cytosolic acetyl coenzyme A (CoA) synthetase (ACS1) may catalyse acetate activation. Considering an aerobic environment and the lack of hydrogenase, a simplified but specific electron transport towards molecular oxygen remains a possibility. The predicted manganese superoxide dismutase (Mn-SOD) would ensure protection against oxygen radicals. A homologue of yeast ERV1 (a small protein required for mitochondrial biogenesis) is indicated. In the lower part of the scheme are depicted some of the major potential factors required for targeting of proteins to the organelle and for biosynthesis of Fe–S clusters and transport of Fe–S proteins towards the cytosol. Iron is predicted to be essential for controlling the expression of mitosomal proteins.

This first report of the genome sequence of a eukaryotic parasite should stimulate proteomic approaches to identify gene products of interest for diagnosis and therapy of microsporidioses, as well as to test the mitosome hypothesis. In addition, we expect the *E. cuniculi* genome to provide a useful reference for comparative genomics of microbial eukaryotes, particularly to identify the relative importance of shared genes among various evolutionarily distant intracellular parasites, including major human-infecting parasites such as *Plasmodium* and *Leishmania*. □

Methods

The reference mouse isolate of *Encephalitozoon cuniculi* (GB-M1), cloning and library construction, nucleotide sequencing, sequence validation and sequence analysis have been described in detail previously[11] (see Supplementary Information). Further information on recombinant DNA preparation, sequencing and sequence analysis can be found elsewhere[30]. Annotation was manually performed with the help of AceDB and Artemis graphic interfaces. Genes were characterized by Glimmer prediction of coding DNA sequences, combined with BLAST all-homology results (BLAST X, BLAST N against 'nr', BLAST P against SwissProt, PSI-BLAST). Transfer RNA genes were detected with the tRNA Scan program. Spliceosomal-type introns were manually detected. Phylogenetic analyses were done as in Keeling *et al.*[8]. Gamma-corrected ML distances with eight rate categories and invariant sites were computed with TREE-PUZZLE version 5.0 and trees were derived with BioNJ. Bootstrapping was on 500 replicates with alpha parameters and the fraction of invariant sites estimated once from the original data. Sequences of individual chromosomes were submitted to EMBL under the accession numbers AL39173 for chrI and AL590442–AL590451 for chrII–chrXI, respectively.

Received 23 April; accepted 31 August 2001.

1. Wittner, M. & Weiss, L. M. *The Microsporidia and Microsporidiosis* (American Society of Microbiology, Washington DC, 1999).
2. Vossbrinck, C. R., Maddox, J. V., Friedman, S., Debrunner-Vossbrinck, B. A. & Woese, C. R. Ribosomal RNA sequence suggests microsporidia are extremely ancient eukaryotes. *Nature* **326,** 411–414 (1987).
3. Germot, A., Philippe, H. & Le Guyader, H. Evidence for loss of mitochondria in microsporidia from a mitochondrial-type HSP70 in *Nosema locustae*. *Mol. Biochem. Parasitol.* **87,** 159–168 (1997).
4. Hirt, R. P., Healy, B., Vossbrink, C. R., Canning, E. U. & Embley, T. M. A mitochondrial Hsp70 orthologue in *Vairimorpha necatrix*: molecular evidence that microsporidia once contained mitochondria. *Curr. Biol.* **7,** 995–998 (1997).
5. Peyretaillade, E. *et al.* Microsporidia, amitochondrial protists, possess a 70-kDa heat-shock protein gene of mitochondrial evolutionary origin. *Mol. Biol. Evol.* **15,** 683–689 (1998).
6. Hirt, R. P. *et al.* Microsporidia are related to fungi: evidence from the largest subunit of RNA polymerase II and other proteins. *Proc. Natl Acad. Sci. USA* **96,** 580–585 (1999).
7. Baldauf, S. L., Roger, A. J., Wenk-Siefert, I. & Doolittle, W. F. A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* **290,** 972–977 (2000).
8. Keeling, P. J., Luker, M. A. & Palmer, J. D. Evidence from β-tubulin phylogeny that microsporidia evolved from within the Fungi. *Mol. Biol. Evol.* **17,** 23–31 (2000).
9. Van de Peer, Y., Ben Ali, A. & Meyer, A. Microsporidia: accumulating evidence that a group of amitochondriate and suspectedly primitive eukaryotes are just curious fungi. *Gene* **246,** 1–8 (2000).
10. Biderre, C., Pagès, M., Méténier, G., Canning, E. U. & Vivarès, C. P. Evidence for the smallest nuclear genome (2.9 Mb) in the microsporidian *Encephalitozoon cuniculi*. *Mol. Biochem. Parasitol.* **74,** 229–231 (1995).
11. Peyret, P. *et al.* Sequence and analysis of chromosome I of the amitochondriate intracellular parasite *Encephalitozoon cuniculi* (Microspora). *Genome Res.* **11,** 198–207 (2001).
12. Brugère, J. F., Cornillot, E., Méténier, G., Bensimon, A. & Vivarès, C. P. *Encephalitozoon cuniculi* (Microspora) genome: physical map and evidence for telomere-associated rDNA units on all chromosomes. *Nucleic Acids Res.* **28,** 2026–2033 (2000).
13. Douglas, S. *et al.* The highly reduced genome of an enslaved algal nucleus. *Nature* **410,** 1091–1096 (2001).
14. Zhang, I. Protein-length distributions for the three domains of life. *Trends Genet.* **16,** 107–109 (2000).
15. Shirakura, T. *et al.* Characterization of the ribosomal proteins of the amitochondriate protist *Giardia lamblia*. *Mol. Biochem. Parasitol.* **112,** 153–156 (2001).
16. Spingola, M., Grate, L., Haussler, D. & Ares, M. Genome-wide bioinformatic and molecular analysis of introns in *Saccharomyces cerevisiae*. *RNA* **5,** 221–234 (1999).
17. El Alaoui, H., Bata, J., Bauchart, D., Doré, J.-C. & Vivarès, C. P. Lipids of three microsporidian species and multivariate analysis of the host–parasite relationship. *J. Parasitol.* **87,** 554–559 (2001).
18. Weidner, E., Findley, A. M., Dolgikh, V. & Skolova, J. in *The Microsporidia and Microsporidiosis* (eds Wittner, M. & Weiss, L. M.) 172–195 (American Society of Microbiology, Washington DC, 1999).
19. Wolf, Y. I., Aravind, L. & Koonin, E. V. Rickettsiae and Chlamydiae: evidence of horizontal transfer and exchange. *Trends Genet.* **15,** 173–175 (1999).
20. Böhne, W., Ferguson, D. J., Kohler, K. & Gross, U. Developmental expression of a tandemly repeated, glycine- and serine-rich spore wall protein in the microsporidian pathogen *Encephalitozoon cuniculi*. *Infect. Immun.* **68,** 2268–2275 (2000).
21. Delbac, F., Peuvel, I., Méténier, G., Peyretaillade, E. & Vivarès, C. P. Microsporidian invasion apparatus: identification of a novel polar tube protein and evidence for clustering of ptp1 and ptp2 genes in three *Encephalitozoon* species. *Infect. Immun.* **69,** 1016–1024 (2001).
22. Karlberg, O., Canback, B., Kurland, C. G. & Andersson, S. G. The dual origin of the yeast mitochondrial proteome. *Yeast* **17,** 170–187 (2000).
23. Andersson, S. G. *et al.* The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* **396,** 133–140 (1998).
24. Lill, R. & Kispal, G. Maturation of cellular Fe–S proteins: an essential function of mitochondria. *Trends Biochem. Sci.* **25,** 352–356 (2000).
25. Dyall, S. D. & Johnson, P. J. Origins of hydrogenosomes and mitochondria: evolution and organelle biogenesis. *Curr. Opin. Microbiol.* **3,** 404–411 (2000).
26. Rodriguez, M. A. *et al.* The pyruvate ferredoxin oxidoreductase is located in the plasma membrane and in a cytoplasmic structure in *Entamoeba*. *Microb. Pathog.* **25,** 1–10 (1998).
27. Mai, Z. *et al.* Hsp60 is targeted to a cryptic mitochondrion-derived organelle ('crypton') in the microaerophilic protozoan parasite *Entamoeba histolytica*. *Mol. Cell. Biol.* **19,** 2198–2205 (1999).
28. Tovar, J., Fischer, A. & Clark, C. G. The mitosome, a novel organelle related to mitochondria in the amitochondrial parasite *Entamoeba histolytica*. *Mol. Microbiol.* **32,** 1013–1021 (1999).
29. Martin, W. & Müller, M. The hydrogen hypothesis for the first eukaryote. *Nature* **392,** 37–41 (1998).
30. Artiguenave, F. *et al.* Genomic exploration of the hemiascomycetous yeasts: 2. Data generation and processing. *FEBS Lett.* **487,** 13–16 (2000).

Supplementary Information accompanies the paper on *Nature*'s website (http://www.nature.com).

Acknowledgements

We are grateful to everyone who provided computer resources and system management: D. Ferney and P. Bernard (LMA, UMR CNRS 6620), P. L. Reichstadt and B. Michel (LPC, IN2P3, UMR CNRS 6533). We thank B. Chebance and R. Guerry for technical assistance. L. Pereira, B. Duperrier, R. Méréry and E. Delorme are acknowledged for their contribution to the development of the annotation environment in the Université Blaise Pascal. We also express gratitude to C. White for critical reading of the manuscript. F.D., H.E.A and E.P were supported by a grant from the Ministère de la Recherche.

Correspondence and requests for materials should be addressed to C.P.V. (e-mail: christian.vivares@lbp.univ-bpclermont.fr).