Gain and loss of multiple functionally related, horizontally transferred genes in the reduced genomes of two microsporidian parasites

Jean-François Pombert[a,1], Mohammed Selman[b,1], Fabien Burki[a], Floyd T. Bardell[a], Laurent Farinelli[c], Leellen F. Solter[d], Douglas W. Whitman[e], Louis M. Weiss[f,g], Nicolas Corradi[b,2], and Patrick J. Keeling[a,2]

[a]Department of Botany, Canadian Institute for Advanced Research, University of British Columbia, Vancouver, BC, V6T 1Z4 Canada; [b]Department of Biology, Canadian Institute for Advanced Research, University of Ottawa, Ottawa, ON, K1N 1H7 Canada; [c]Fasteris SA, CH-1228 Plan-les-Ouates, Geneva, Switzerland; [d]Illinois Natural History Survey, Prairie Research Institute, University of Illinois, Champaign, IL 61820; [e]School of Biological Sciences, Illinois State University, Normal, IL 61790; and [f]Division of Parasitology, Department of Pathology, and [g]Division of Infectious Diseases, Department of Medicine, Albert Einstein College of Medicine, Bronx, NY 10461

Edited by David M. Hillis, University of Texas, Austin, TX, and approved June 25, 2012 (received for review March 26, 2012)

Microsporidia of the genus *Encephalitozoon* are widespread pathogens of animals that harbor the smallest known nuclear genomes. Complete sequences from *Encephalitozoon intestinalis* (2.3 Mbp) and *Encephalitozoon cuniculi* (2.9 Mbp) revealed massive gene losses and reduction of intergenic regions as factors leading to their drastically reduced genome size. However, microsporidian genomes also have gained genes through horizontal gene transfers (HGT), a process that could allow the parasites to exploit their hosts more fully. Here, we describe the complete sequences of two intermediate-sized genomes (2.5 Mbp), from *Encephalitozoon hellem* and *Encephalitozoon romaleae*. Overall, the *E. hellem* and *E. romaleae* genomes are strikingly similar to those of *Encephalitozoon cuniculi* and *Encephalitozoon intestinalis* in both form and content. However, in addition to the expected expansions and contractions of known gene families in subtelomeric regions, both species also were found to harbor a number of protein-coding genes that are not found in any other microsporidian. All these genes are functionally related to the metabolism of folate and purines but appear to have originated by several independent HGT events from different eukaryotic and prokaryotic donors. Surprisingly, the genes are all intact in *E. hellem*, but in *E. romaleae* those involved in de novo synthesis of folate are all pseudogenes. Overall, these data suggest that a recent common ancestor of *E. hellem* and *E. romaleae* assembled a complete metabolic pathway from multiple independent HGT events and that one descendent already is dispensing with much of this new functionality, highlighting the transient nature of transferred genes.

evolution | genomics | parasitology

Microsporidia are highly derived relatives of fungi that are obligate intracellular parasites of virtually all animal lineages and which lead to a number of economically and medically important diseases, particularly in sericulture and apiculture (1). To date, more than 1,200 microsporidian species have been described, and at least 13 of these species infect humans; many are opportunistic pathogens found in immune-compromised patients (2, 3). The group is distinguished by a number of cellular characteristics, including the presence of a specialized host-invasion apparatus (the polar tube), an unconventional Golgi apparatus, and highly reduced mitochondria called "mitosomes" (4, 5). Many other cellular features common to other eukaryotes are missing; because microsporidia now are recognized as being related to fungi (4), this simplicity is interpreted as extreme reduction that also extends to the molecular and genomic levels. Genome reduction in microsporidia has followed a number of routes, including losses of entire metabolically relevant pathways, the shortening of proteins, and in some species the shrinking of intergenic regions (6). In the most extreme cases these trends have resulted in major effects on cellular function (7) or genome evolution (8).

The smallest (nonorganellar) nuclear genomes currently known are those of microsporidian species in the genus *Encephalitozoon*, making them a model for extreme reductive forces in nuclear genome evolution. Complete genome sequences from two species (9, 10) were found to encode about 2,000 genes making up a reduced set of sometimes simplified molecular and biochemical pathways. Their high degree of host dependence also is reflected in the relatively large number of transporters encoded in these genomes (e.g., ATP transporters), which allow them to acquire essential energy and nutrients from their hosts. Some of these transporters are thought to have originated by horizontal gene transfer (HGT), possibly from coexisting bacterial pathogens (11), and the recent finding of an animal-derived gene in both *Encephalitozoon romaleae* and *Encephalitozoon hellem* (12) also raised the intriguing possibility that microsporidia can acquire genes from their hosts.

Here we describe the complete nuclear genomic sequences from *E. hellem* and *E. romaleae* to examine the extent of HGT in these lineages. With these two genomes we now have complete sequences for four of the five described species of *Encephalitozoon*, so relatively detailed analyses of gene presence/absence can be performed. Although both genomes are extremely similar in form and content to those of other *Encephalitozoon* species, we identified a number of genes coding for products that are involved in metabolic pathways that either are absent or are substantially reduced in all other members of the group. Overall, these genes make up nearly complete pathways for de novo folate and purine biosynthesis. Their absence in all other sequenced members of the group suggests that these genes have been acquired by HGT. In most cases this hypothesis is supported by robust phylogenetic evidence, suggesting that the genes have been acquired from different donors including both prokaryotes and eukaryotes. Curiously, the majority of these genes now are mutated to render them nonfunctional in *E. romaleae*, suggesting that the gain of function made possible through multiple

Author contributions: J.-F.P., N.C., and P.J.K. designed research; J.-F.P., M.S., F.B., L.F., and N.C. performed research; L.F.S., D.W.W., and L.M.W. contributed new reagents/analytic tools; J.-F.P., M.S., F.B., F.T.B., N.C., and P.J.K. analyzed data; and J.-F.P., N.C., and P.J.K. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database, www.ncbi.nlm.nih.gov/genbank/ [accession nos. CP002713–CP002724 (*Encephalitozoon hellem*) and CP003518–CP003530 (*Encephalitozoon romaleae*)].

[1]J.-F.P. and M.S. contributed equally to this work.

[2]To whom correspondence may be addressed. E-mail: ncorradi@uottawa.ca or pkeeling@mail.ubc.ca.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1205020109/-/DCSupplemental.

independent HGT events has been maintained in one branch of this lineage but was transient in another.

Results

General Characteristics of the *E. hellem* and *E. romaleae* Genomes. *E. hellem* [American Type Culture Collection (ATCC) 50504] DNA was isolated from spores grown in RK13 (rabbit kidney) cells. *E. romaleae* (SJ-2008) DNA was isolated from spores purified from infected captive male and female *Romalea microptera* grasshoppers. In both cases, total DNA was used for Illumina sequencing, and the resulting reads were assembled de novo. For *E. hellem* this process resulted in an assembly of 2,251,784 bp distributed among 12 contigs (53× average coverage) and for *E. romaleae* resulted in an assembly of 2,138,148 bp distributed among 13 contigs (300× average coverage). In both cases, single contigs corresponded to chromosomes I–VIII, X, and XI, whereas chromosome IX contained one gap in *E. hellem* and two gaps in *E. romaleae* (Fig. 1). The subtelomeric regions of *Encephalitozoon* chromosomes typically encode the ribosomal RNA (rRNA) operons, and seven of these operons were linked physically to the ends of *E. hellem* chromosome assemblies (compared with five in *Encephalitozoon intestinalis* and three in *Encephalitozoon cuniculi*). Both the *E. hellem* and *E. romaleae* rRNA operon sequences are represented 25-fold in excess of the rest of their respective genomes, suggesting that all their 22-chromosome ends also are likely capped by rRNA operon subtelomeres. We tested whether the gaps in chromosome IX could represent actual fragmentation by long-range PCR from the assembly ends to rRNA operons. No such a link could be demonstrated, and therefore we conclude that this region is problematic for assembly (it encodes multiple paralogous genes) until direct evidence indicates otherwise.

The *E. hellem* and *E. romaleae* genomic GC contents are different from those of *E. intestinalis* and *E. cuniculi* (Table 1). This variation is correlated with different codon-use biases among the four species, with *E. romaleae* favoring AT bases and *E. cuniculi* favoring GC bases in third codon positions (Dataset S1). This bias is uniform throughout all *Encephalitozoon* genomes with one notable exception: In all species chromosome I displays an increase of about 1.5–3% in GC content relative to the rest of the genome (Fig. 1). The four *Encephalitozoon* genomes share a total of 38 introns that are inserted at cognate sites in orthologous genes. This number includes two tRNA introns, 34 previously reported spliceosomal introns (13), and two additional spliceosomal introns identified in the present study (Table S1). The orthologous microsporidian introns are relatively well conserved, with an average nucleotide identity of 72.1% across the four species. Gene order also is extremely conserved across all *Encephalitozoon* genomes, with 1,824 colinear genes located within 55 completely syntenic blocks. Only three of these blocks were found to be inverted or translocated in one or more genomes (Dataset S2).

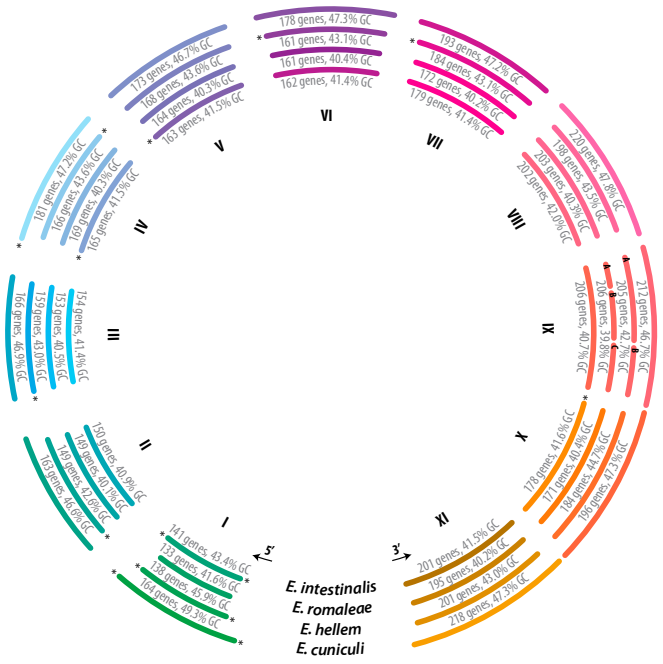

Fig. 1. Physical characteristics of *E. hellem*, *E. romaleae*, *E. intestinalis*, and *E. cuniculi* genomes. Chromosomes I–XI are numbered according to their respective sizes in *E. cuniculi* (9) and are shown to scale. Chromosome IX is fragmented in assemblies of *E. hellem* and *E. romaleae* (as indicated by IXa and IXb in *E. hellem* and IXa–IXc in *E. romaleae*), but there is no evidence supporting an actual physical fragmentation of this chromosome in either species.

***E. hellem* and *E. romaleae* Contain Functionally Related Genes That Likely Are Derived from Multiple HGT Events from Different Donors.** The gene contents of the *E. hellem* and *E. romaleae* genomes are virtually identical to those of *E. intestinalis* and *E. cuniculi*: They contain an identical set of tRNA genes, and only a handful of genes varies between species (Fig. S1). However, one suite of genes shared by *E. hellem* and *E. romaleae* stands out. These six genes from four chromosomal regions are absent in all other microsporidia, including other *Encephalitozoon* species (Fig. 2). They are significant because they are functionally related: Other microsporidia encode a few genes related to folate salvage and purine metabolism, but in *E. hellem* and *E. romaleae* these six genes contribute to making up intact pathways for folate salvage, folate de novo biosynthesis, and purine metabolism (Fig. 3).

The extremely narrow distribution of these genes within microsporidia is highly suggestive that they were acquired by HGT: They are found only in the two closely related sister-species, *E. hellem* and *E. romaleae*, and are absent from all other microsporidia examined to date, including other members of the same genus, *E. intestinalis* and *E. cuniculi* (Figs. 2 and 3). Phylogenetic analyses support the conclusion that these genes were derived by HGT but, surprisingly, not from a single source. Instead, the genes appear to be derived from multiple donor lineages, including both prokaryotes and eukaryotes (Fig. 4, Fig. S2, and ref. 12). The phylogenetic relationships in these trees frequently are complex or poorly resolved (especially within bacteria), but in several cases the overall position of microsporidia is resolved sufficiently to conclude that their evolutionary histories are not congruent. Specifically, the GTP cyclohydrolase I (GTPCH) is related to Gammaproteobacteria, folic acid synthase (FASP) is related to Firmicutes, phosphoribosyltransferase (PRT) is related to spirochetes, folylpolyglutamate synthase (FPGS) is related to either Metazoa or fungi, dihydrofolate synthase (DHFS) is related to fungi, and a previously reported purine nucleotide phosphorylase (PNP) is related to animals (12). The phylogenetic support for the horizontal acquisition of FPGS and DHFS is weaker, because these genes potentially are related to fungi, and microsporidia also are related to fungi; however, a horizontal acquisition of these genes still is favored by their distribution, because neither has been found in any other microsporidian genome, and their vertical transmission followed by independent losses in all lineages except *E. hellem* and *E. romaleae* would require a large number of independent losses.

In contrast to the diverse phylogenetic origins of the six potentially horizontally transferred genes described above, the phylogenies of all folate- and purine-related genes that are common to other microsporidia are consistent with vertical transmission (Fig. S3).

EVOLUTION

Table 1. General characteristics of *Encephalitozoon* genomes

Characteristic	*E. intestinalis* ATCC 50506	*E. hellem* ATCC 50504	*E. romaleae* SJ-2008	*E. cuniculi* GB-M1
Chromosomes (no.)	11	11	11	11
Genome size (Mbp)	2.3	2.5	2.5	2.9
Assembled (Mbp)	2.2	2.3	2.2	2.5
Genome coverage (%)	96	92	88	86
G+C content (%)	41.4	43.4	40.3	47
Gene density (gene/kbp)	0.86	0.86	0.84	0.83
Mean gene length	1,041 bp	1,080 bp	1,061 bp	1,041 bp
Mean intergenic length*	120 bp	124 bp	130 bp	166 bp
Overlapping genes present	Yes	Yes	Yes	Yes
SSU-LSU rRNA genes	22	22†	22†	22
5S rRNA genes	3	3	3	3
tRNAs	46	46	46	46
tRNA synthetases	21	21	21	21
tRNA introns (size)	2 (12, 41 bp)	2 (12, 41 bp)	2 (12, 41 bp)	2 (12, 41 bp)
Splic. introns (size)	36 (23–76 bp)	36 (23–76 bp)	36 (23–76 bp)	36 (23–76 bp)
Predicted ORFs	1,848	1,848	1,835	2,010

LSU, large subunit; SSU, small subunit.
*The values presented here were recalculated after correction of the start methionines; therefore the intergenic lengths differ slightly from the values presented in Corradi et al. (10).
†Minimum estimated from coverage.

***E. romaleae* Folate Biosynthesis Genes Are Nonfunctional Pseudogenes.** All six HGT-acquired folate- or purine-related genes are intact and appear functional in *E. hellem* but, surprisingly, not in *E. romaleae*. Indeed, three of the six genes show obvious signs of pseudogenization: GTP cyclohydrolase contains multiple frameshift mutations resulting in premature stop codons, whereas FASP and DHFS both sustained large-scale deletions, all of which were verified by PCR and sequencing. Strikingly, the functional distribution of these genes within the folate and purine pathways is not random: The three genes involved in the salvage of purines and in folate synthesis are all intact, whereas none of the genes functioning in the de novo synthesis of folate are functional, and no intact copy of any of these genes exists elsewhere in the genome.

Discussion

Horizontal gene transfer is known to have had an impact on a variety of eukaryotic genomes and on the functional versatility of their proteomes, even in the otherwise highly reduced and metabolically simple microsporidia (11, 14, 15). In most cases, this impact comes from either a single gene or a single donor, but *E. hellem* and *E. romaleae* apparently have acquired several functionally related genes from several distantly related donors. This mode of acquisition raises several questions about how and why this pathway was assembled and why it then was partially lost in *E. romaleae*.

In most organisms in which they are found, these folate-related pathways play an essential housekeeping role by producing tetrahydrofolate (THF), a compound used as cofactor by thymidylate synthase (TS) for the synthesis of DNA (16). Folate is composed of linked pterin and p-aminobenzoate rings attached to a glutamate moiety and feeds into the one carbon core (C1) metabolism (17–20). Although plants, most fungi, and most protists are capable of synthesizing folate, animals are not, and nearly all microsporidia infect animals. So even if a microsporidian can take up folate from its host (all microsporidia have a single folate transporter gene), infecting a folate-deficient animal could have serious deleterious effects on the parasite, and facultative folate synthesis could be beneficial. All four *Encephalitozoon* species investigated are capable of scavenging folate in the form of THF because of the presence of the ubiquitous dihydrofolate reductase (DHFR) gene and folate transporters (Fig. 3). Both *E. hellem* and *E. romaleae* also contain FPGS, an enzyme used for folate homeostasis, but apparently only *E. hellem* can synthesize folate from its most basic constituents (e.g., GDP, GMP, and guanine, deoxyguanosine). There is one caveat: We did not identify an alkaline phosphatase (ALP), which removes the pyrophosphate in the second step of the pathway (21). However, other parasites bypass the need for a specific ALP in various ways (22), and it is possible that *E. hellem* does so similarly, using a different enzyme or by using a host ALP.

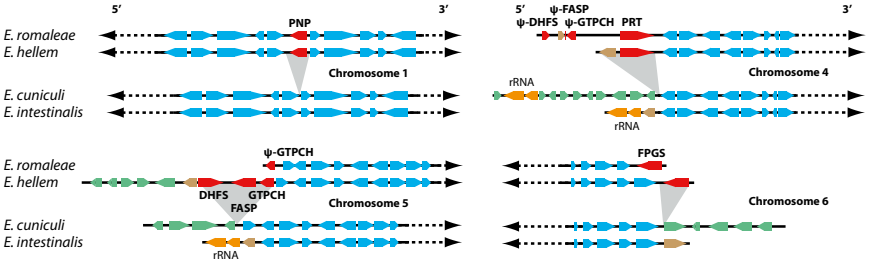

Fig. 2. Genomic context of HGT-acquired genes in *E. hellem* and *E. romaleae*. The genes in question that are unique to *E. hellem* and *E. romaleae* are shown in red, and syntenic genes that are common to all four completed *Encephalitozoon* genomes are shown in cyan. The rRNA genes are shown in orange, unknown ORFs displaying no synteny are shown in beige, and subtelomeric genes unique to *E. cuniculi* are shown in green. Decayed folate-related pseudogenes in *E. romaleae* are indicated by a ψ-sign.

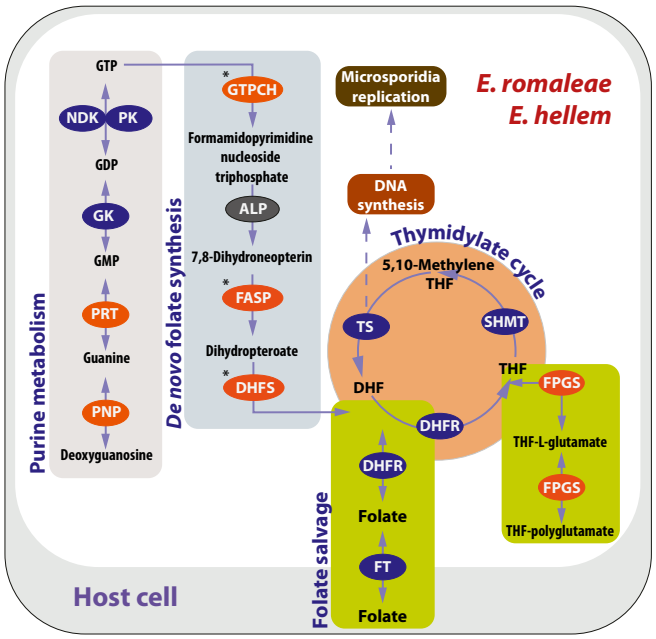

Fig. 3. Hybrid origin of the folate metabolic pathways in *E. hellem* and *E. romaleae*. Genes vertically inherited in *E. hellem*, *E. romaleae*, *E. intestinalis*, and *E. cuniculi* are shown in blue. Genes acquired by HGT in the lineage leading to *E. hellem* and *E. romaleae* are shown in orange. Functions for which specific genes have not been attributed in *E. hellem* or *E. romaleae* are shown in dark gray. Arrows denote the directions of the enzymatic reactions. Asterisks indicate genes that underwent pseudogenization in *E. romaleae*. FT, folate transporter; GK, guanylate kinase; PK, pyruvate kinase; SHMT, serine hydroxymethyltransferase. Note that the putative FT function attributed to homologs of ECU11_1600 (unique to *Encephalitozoon* species) is uncertain, and that folate transport often is carried out by miscellaneous ATP transporters. The presence of all these genes in the *E. hellem* and *E. romaleae* genomes was confirmed using PCR and sequencing.

The de novo biosynthesis of folate requires at least four proteins—GTPCH, FASP, DHFS, and FPGS—in addition to generic ALPs and/or hydrolases. Given the genes common to all *Encephalitozoon* species, the addition of FPGS alone would have a clear beneficial role in folate retention, but the immediate benefits accrued through the acquisition of any other of these genes by itself is not so obvious. The lack of an individual benefit raises questions about how the pathway could have been assembled from multiple independent HGT events, because a stepwise construction would result in intermediates encoding partial pathways without an obvious functional advantage. One possibility is that the pathway was assembled in some other genome and then was acquired as a whole by a single HGT in the ancestor of *E. hellem* and *E. romaleae*. This possibility limits the otherwise seemingly rare HGT to microsporidia but ultimately does not explain how the pathway was assembled. Alternatively, a complete pathway (e.g., a bacterial operon) may have been acquired by a single HGT in the ancestor of *E. hellem* and *E. romaleae*, and then individual genes were replaced by subsequent HGT events. This process requires no intermediates with partial pathways but requires a lot of HGT to microsporidia. Perhaps the most likely explanation is a combination: A complete pathway existed in a donor lineage in which HGT is relatively common, leading to the replacement of several individual genes, and this mosaic pathway was acquired by the ancestor of *E. hellem* and *E. romaleae* through a single HGT.

Other important questions are why this pathway was partially lost in *E. romaleae* and whether this loss is related to host range. *E. hellem* and *E. romaleae* are closely related sisters, and the overall phylogeny of *Encephalitozoon* suggests that their ancestors were vertebrate parasites (although the arthropod origin of PNP also could suggest that their ancestor infected arthropods). If *E. romaleae* moved to insects recently, it is possible that some aspects of biochemistry of the new host made synthesizing folate less advantageous or, alternatively, that this pathway is less

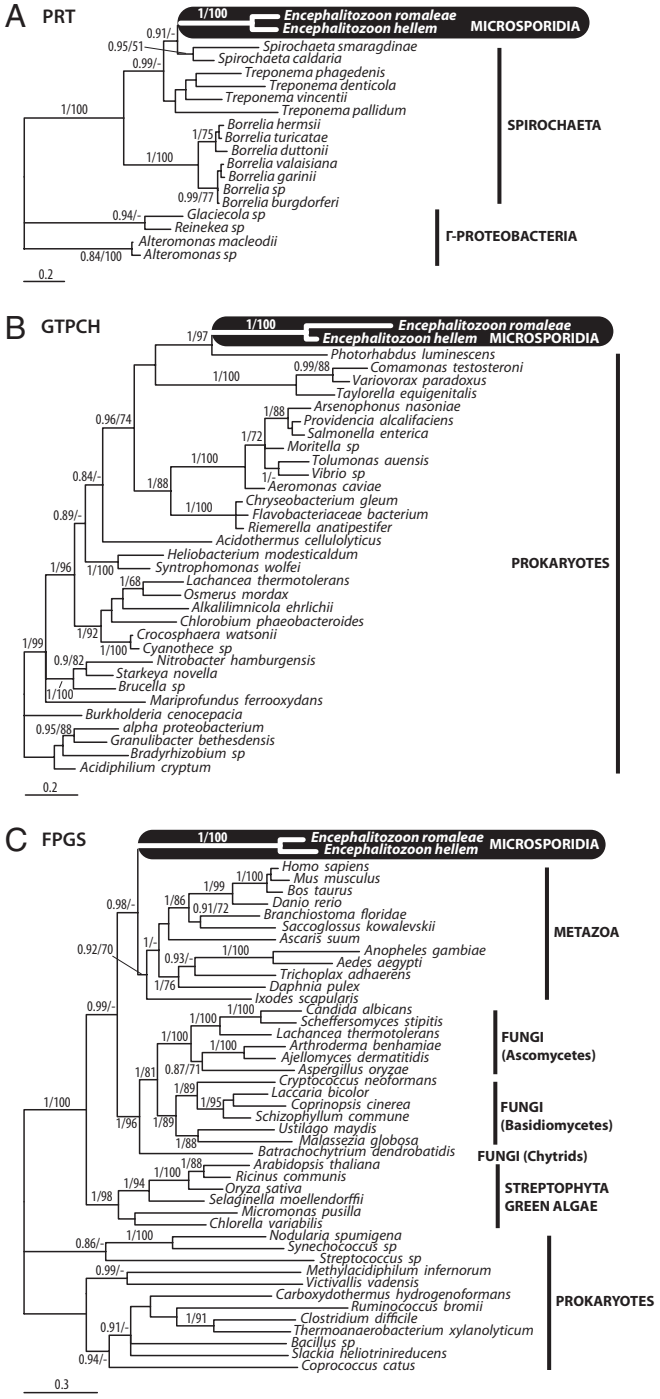

Fig. 4. Bayesian phylogenetic trees of three proteins, PRT (*A*), GTPCH (*B*), and FPGS (C), involved in the folate metabolic pathways in *E. hellem* and *E. romaleae* (shown in white on black) but absent from other microsporidians. Numbers at nodes representing Bayesian posterior probabilities (*Left*) and bootstrap proportions (*Right*) are indicated when higher than 0.8 and 70%, respectively. The scale bar corresponds to the estimated number of amino acid substitutions per site. All trees are shown unrooted.

EVOLUTION

sustainable in insects. For example, if ALP activity is supplied by the host, then a lower or more tissue-specific ALP activity in insects could make the parasite folate synthesis impossible (23, 24). Conversely, the new host environment might have made de novo folate synthesis unnecessary, leading to its loss. Additional data from *Encephalitozoon* parasites of vertebrate and invertebrate hosts doubtless would be very helpful in distinguishing these possibilities, but expanding the catalog of *Encephalitozoon* genomes will require additional sampling of their natural diversity, because genomes for all four cultivated species are now available.

It also is noteworthy that all but one of these genes is found in subtelomeric regions of *E. hellem* and *E. romaleae*. These poorly studied genomic regions usually evolve faster than the chromosome cores and often are associated with rapid biological innovation, such as parasites coevolving with their hosts' immune systems (25, 26), and in other cases have been observed to be rich in genes derived from HGT (27). Given the compactness of *Encephalitozoon* genomes, it also is more likely that recombination into the chromosome cores would disrupt vital sequence information (e.g., any random insertion is about 10 times more likely to hit a gene than to hit an intergenic region). Subtelomeric regions harbor many repeated genes and gene families, so any one copy can be disrupted with less likelihood of a negative effect. The apparently high rate of recombination between subtelomeres also might generate multiple copies of a new gene and consequently increase its chances of being fixed in the genome. Indeed, the folate-related genes acquired by *E. hellem* and *E. romaleae* that are located in the subtelomeric regions are present at other chromosome ends, as indicated by their 2× to 4× sequencing coverage relative to the single-copy genes from the chromosomal cores; therefore it is likely that they were multiplied in this way following their acquisition by HGT.

Materials and Methods

Cultivation and Collection of *E. hellem* and *E. romaleae* Material. Spores from *E. hellem* [ATCC 50504; originally isolated from humans (28)] were grown in the rabbit kidney fibroblast cell line RK 13 (ATCC CCL-37) with RPMI 1640 (Sigma-Aldrich) supplemented with 5% (vol/vol) foetal bovine serum (FBS), 2 mM L-glutamine, and antibiotics (penicillin, 100 U mL, and streptomycin, 100 μg/mL). T75 flasks were incubated at 37 °C with 5% CO_2, and culture medium was replaced two or three times per week. Supernatants containing spores were stored at 4 °C until extraction of DNA. To enrich spores from host-cell debris, the collected culture supernatants were subjected to sequential washes at 400 × *g* each with distilled H_2O, Tris-buffered saline (TBS)-Tween 20 (0.3%), and TBS. The final pellet was resuspended in TBS and mixed with an equal volume of 100% Percoll (Sigma-Aldrich), followed by centrifugation at 400 × *g* for 45 min at 4 °C. Host-cell debris in the top 75% volume of Percoll was removed. The lower 25% volume of Percoll and the pellet were transferred to a new tube, resuspended in TBS, and washed several times. Because of continued adherence of the host cell (that is, rabbit) nucleic acid to the spores, an additional series of washes was performed with TBS-SDS (0.1%) followed by three washes with TBS.

Spores from *E. romaleae* [SJ-2008; originally isolated from Lubber Grasshopper, *R. microptera* (29)] were produced in infected captive male and female *R. microptera* grasshoppers. Spores were collected, as previously described (30), from the alimentary canal by homogenizing the midgut and gastric caeca with sterile distilled water. The homogenate was filtered through nylon mesh cloth (sieve size ~200 μm). The collected spores were washed with sterile distilled water five times by centrifugation at 2,700 × *g* for 10 min and were purified on a 1:1 sterile water:Ludox HS-40 colloidal silica gradient (Sigma-Aldrich), totaling 35 mL, in 50-mL plastic centrifuge tubes. The purified spores were cleaned again by centrifugation in sterile water. A 1-mL suspension of spores was pelleted by centrifugation, resuspended in 150 mL of buffer (40 mM Tris-Acetate, 1 mM EDTA) in 1.5-mL microfuge tubes, and shaken with 150 mg of 0.5-mm glass beads in a Mini-Beadbeater (Biospec Products). Supernatants containing spores were stored at 4 °C until the extraction of DNA.

DNA Extraction. Genomic DNA for each species was extracted from spores with the MasterPure DNA purification kit (Epicentre). For each sample, spores were pelleted by centrifugation, resuspended in 300 μL of lysis solution (Epicentre) containing Proteinase K, and mixed thoroughly using a vortex. Glass beads (200 μL, 150–212 μm in diameter) were added to the samples, which were incubated immediately at 65 °C for 15 min and bead-beaten at 2,500 rpm in a Mini-Beadbeater (Biospec Products) for 30 s every 5 min. The samples then were cooled to 37 °C and incubated for 30 min at the same temperature with the addition of RNase A (10 μg total) (Epicentre). After treatment with RNase, the samples were placed on ice for 5 min, 150 μL of MPC Protein Precipitation Reagent (Epicentre) was added to each sample, and the solutions were vortexed vigorously for 10 s. Protein debris were pelleted at 4 °C for 10 min at a speed of ≥10,000 × *g*, and the supernatants were transferred to clean microcentrifuge tubes. DNA then was precipitated using isopropanol, rinsed twice using 70% (vol/vol) ethanol, and finally was suspended in Tris-EDTA buffer.

Genome Sequencing and de Novo Assembly. The *E. hellem* and *E. romaleae* deep-sequencing shotgun libraries, averaging insert sizes of 327 and 337 bp, respectively, were prepared as described by Corradi et al. (10) with Fasteris-modified bar-coded adapters (AAAGT and ACTTGA, respectively) added to the beginning of the forward and reverse reads for multiplexing. The *E. hellem* library was subjected to two rounds of deep sequencing, each using half a channel on the GA-IIx instrument (Illumina). In the first round, a total of 3,487,666 paired-end reads (101 bp) were generated, resulting in 334,815,936 bp of unique sequences. The 101 bp reads were trimmed to remove the barcodes and were assembled using Velvet 0.7.54 (31) with a hash value of 19, generating a total of 260 contigs with an average size of 8,510 bp and coverage of 53×. The reads were mapped on contigs using the addSolexaReads Perl script from Consed 20 (32), and the internal breaks within the scaffolds were polished using Consed in combination with PCR, cloning and Sanger sequencing. Subtelomeric regions were linked to the chromosome internal segments by PCR as described by Corradi et al. (10). A second round of sequencing (38-bp-long reads, 36,754,350 reads, 1,212,893,550 bp total) was performed in parallel to resolve potentially ambiguous regions from the first round. The resulting reads were mapped on the *E. hellem* contigs with Consed, and the ambiguous regions were curated manually. De novo assemblies also were performed independently with Ray 1.4.0 (33) using a k-mer of 21 and reads from both sequencing rounds for cross-validation. The *E. romaleae* library was subjected to one round of deep sequencing using half a channel of the GA-IIx instrument (Illumina), resulting in 1,901,807,856 bp of unique DNA sequence. Reads were assembled using Velvet with a hash value of 27, resulting in 165 scaffolds with an average size of 13,355 bp and an average coverage of 40×. The resulting contigs were polished as described for the *E. hellem* genome.

Genome Annotation and Analysis. ORFs in the *E. hellem* and *E. romaleae* genomes were predicted using Artemis 12.0 (34) and were identified by BLAST homology searches (35) against the National Center for Biotechnology Information nonredundant database; tRNAs were mapped on the chromosomes with tRNAscan-SE 1.21 (36). The start codons of the *E. hellem*, *E. intestinalis*, and *E. cuniculi* ORFs were assessed/reassessed by (*i*) orthologous alignment between the three *Encephalitozoon* species, (*ii*) cross-checking with the predicted translation initiation codons from Peyretaillade et al. (37), and (*iii*) adding the intron positions identified in Lee et al. (13). The start codons in the later-sequenced *E. romaleae* genome were annotated afterward. The introns previously identified in the *E. cuniculi* and *E. intestinalis* genomes were positioned on the *E. hellem* and *E. romaleae* genomes with DREG from the EMBOSS 6.3.1 package (38). Putative novel introns were searched for with DREG using the regular expression GTA[AG]GT[ACGT]{5,30}TT[ACGT]{0,3}AG derived from the microsporidian introns reported in Lee et al. (13) and were curated manually. Nucleotide sequence identity between orthologous introns inserted at the cognate site was calculated from their L-INS-i alignment computed with MAFFT 6.847b (39). The two additional introns reported in this study were ascertained to be spliced at the mRNA level in the sister species *E. cuniculi* using the 5′Race dataset from Lee et al. (13).

The overall GC content of each chromosome was determined using Artemis built-in tools. Codon use tables for the complete set of ORFs from each *Encephalitozoon* species were calculated with Artemis. Metabolic pathways were investigated using the KEGG PATHWAY database (40) and a biochemical atlas (41). The *E. hellem* and *E. romaleae* chromosomes are available in GenBank under the individual accession numbers CP002713–CP002724 and CP003518–CP003530, respectively.

Phylogenetic Analyses. Homologs of the *E. hellem* and *E. romaleae* proteins were identified by BLASTP searches against GenBank and were retrieved and automatically aligned with the L-INS-i method of the MAFFT package

(39). Poorly aligned positions were eliminated with Gblocks 0.91b (42), with half the gapped positions allowed, the minimum number of sequences for a conserved and a flank position set to 50% of the number of taxa plus one, the maximum of contiguous nonconserved positions set to 12, and the minimum length of a block set to 4, followed by manual inspection of the alignments using SeaView 4 (43). Bayesian analyses using the WAG $+\Gamma$ +F model (four gamma categories) were performed with MrBayes 3.2 (44). Each inference consisted of two independent runs starting from a random tree and four Metropolis-coupled Markov Chain Monte Carlo (MCMCMC), initially for 1,000,000 generations with sampling every 100 generations. The average SD of split frequencies was used to assess the convergence of the two runs after the initial 1,000,000 generations (< 0.01) and proved sufficient in all genes except nucleoside-diphosphate kinase (NDK) and TS, for which 1,000,000 additional generations were required to reach convergence. Bayesian posterior probabilities were calculated from the majority rule consensus of the tree sampled after the initial burn-in period, which corresponded to 25% of the total generations. Maximum Likelihood (ML) analyses were performed using RAxML 7.2.8 (45), with the rapid hill-climbing algorithm and the LG $+\Gamma$ +F model of evolution (-m PROTGAMMALGF, four discrete rate categories). The best-scoring ML trees were determined in multiple searches using 20 randomized, stepwise-addition, parsimony starting trees. Statistical support was evaluated with nonparametric bootstrapping using 100 replicates.

ACKNOWLEDGMENTS. We thank N. Fast for kindly giving us access to the *E. cuniculi* 5′Race dataset and C. Grisdale for verifying the splicing of the new introns. This work was supported by Grant MOP-42517 from the Canadian Institute for Health Research (to P.J.K.), a Discovery Grant from the Natural Sciences and Engineering Council of Canada (to N.C.), and Grant AI31788 from the National Institute of Allergy and Infectious Diseases of the National Institutes of Health (to L.M.W.). J.-F.P. was supported by a Fonds Québécois de la Recherche sur la Nature et les Technologies/Génome Québec Louis-Berlinguet postdoctoral fellowship. F.B. was supported by a prospective researcher postdoctoral fellowship from Swiss National Science Foundation and by a grant to the Centre for Microbial Diversity and Evolution from the Tula Foundation. P.J.K. and N.C. are a Fellow and a Scholar, respectively, of the Canadian Institute for Advanced Research.

1. Texier C, Vidau C, Viguès B, El Alaoui H, Delbac F (2010) Microsporidia: A model for minimal parasite-host interactions. *Curr Opin Microbiol* 13:443–449.
2. Weber R, Bryan RT (1994) Microsporidial infections in immunodeficient and immunocompetent patients. *Clin Infect Dis* 19:517–521.
3. Corradi N, Slamovits CH (2011) The intriguing nature of microsporidian genomes. *Brief Funct Genomics* 10:115–124.
4. Keeling PJ (2009) Five questions about microsporidia. *PLoS Pathog* 5:e1000489.
5. Didier ES, Weiss LM (2008) Overview of microsporidia and microsporidiosis. *Protistology* 5(4):243–255.
6. Peyretaillade E, et al. (2011) Extreme reduction and compaction of microsporidian genomes. *Res Microbiol* 162:598–606.
7. Keeling PJ, et al. (2010) The reduced genome of the parasitic microsporidian *Enterocytozoon bieneusi* lacks genes for core carbon metabolism. *Genome Biol Evol* 2: 304–309.
8. Corradi N, Gangaeva A, Keeling PJ (2008) Comparative profiling of overlapping transcription in the compacted genomes of microsporidia *Antonospora locustae* and *Encephalitozoon cuniculi*. *Genomics* 91:388–393.
9. Katinka MD, et al. (2001) Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature* 414:450–453.
10. Corradi N, Pombert J-F, Farinelli L, Didier ES, Keeling PJ (2010) The complete sequence of the smallest known nuclear genome from the microsporidian *Encephalitozoon intestinalis*. *Nat Commun* 1:77.
11. Lee SC, Weiss LM, Heitman J (2009) Generation of genetic diversity in microsporidia via sexual reproduction and horizontal gene transfer. *Commun Integr Biol* 2:414–417.
12. Selman M, et al. (2011) Acquisition of an animal gene by microsporidian intracellular parasites. *Curr Biol* 21:R576–R577.
13. Lee RCH, Gill EE, Roy SW, Fast NM (2010) Constrained intron structures in a microsporidian. *Mol Biol Evol* 27:1979–1982.
14. Tsaousis AD, et al. (2008) A novel route for ATP acquisition by the remnant mitochondria of *Encephalitozoon cuniculi*. *Nature* 453:553–556.
15. Keeling PJ, Palmer JD (2008) Horizontal gene transfer in eukaryotic evolution. *Nat Rev Genet* 9:605–618.
16. Carreras CW, Santi DV (1995) The catalytic mechanism and structure of thymidylate synthase. *Annu Rev Biochem* 64:721–762.
17. Henderson GB, Huennekens FM (1986) Membrane-associated folate transport proteins. *Methods Enzymol* 122:260–269.
18. Cossins EA, Chen L (1997) Folates and one-carbon metabolism in plants and fungi. *Phytochemistry* 45:437–452.
19. Hyde JE (2005) Exploring the folate pathway in *Plasmodium falciparum*. *Acta Trop* 94: 191–206.
20. Bermingham A, Derrick JP (2002) The folic acid biosynthesis pathway in bacteria: Evaluation of potential for antibacterial drug discovery. *Bioessays* 24:637–648.
21. Suzuki Y, Brown GM (1974) The biosynthesis of folic acid. *J Biol Chem* 249:2405–2410.
22. Hyde JE, et al. (2008) *Plasmodium falciparum*: A paradigm for alternative folate biosynthesis in diverse microorganisms? *Trends Parasitol* 24:502–508.
23. Jurat-Fuentes JL, et al. (2011) Reduced levels of membrane-bound alkaline phosphatase are common to lepidopteran strains resistant to Cry toxins from *Bacillus thuringiensis*. *PLoS ONE* 6:e17606.
24. Eguchi M (1995) Alkaline phosphatase isozymes in insects and comparison with mammalian enzyme. *Comp Biochem Physiol B Biochem Mol Biol* 111:151–162.
25. Hall AR, Scanlan PD, Morgan AD, Buckling A (2011) Host-parasite coevolutionary arms races give way to fluctuating selection. *Ecol Lett* 14:635–642.
26. Allen DE, Little TJ (2009) Exploring the molecular landscape of host-parasite coevolution. *Cold Spring Harb Symp Quant Biol* 74:169–176.
27. Gladyshev EA, Meselson M, Arkhipova IR (2008) Massive horizontal gene transfer in bdelloid rotifers. *Science* 320:1210–1213.
28. Didier ES, et al. (1991) Isolation and characterization of a new human microsporidian, *Encephalitozoon hellem* (n. sp.), from three AIDS patients with keratoconjunctivitis. *J Infect Dis* 163:617–621.
29. Lange CE, Johny S, Baker MD, Whitman DW, Solter LF (2009) A new *Encephalitozoon* species (Microsporidia) isolated from the Lubber grasshopper, *Romalea microptera* (Beauvois) (Orthoptera: Romaleidae). *J Parasitol* 95:976–986.
30. Johny S, Larson TM, Solter LF, Edwards KA, Whitman DW (2009) Phylogenetic characterization of *Encephalitozoon romaleae* (Microsporidia) from a grasshopper host: Relationship to *Encephalitozoon* spp. infecting humans. *Infect Genet Evol* 9:189–195.
31. Zerbino DR, Birney E (2008) Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18:821–829.
32. Gordon D, Abajian C, Green P (1998) Consed: A graphical tool for sequence finishing. *Genome Res* 8:195–202.
33. Boisvert S, Laviolette F, Corbeil J (2010) Ray: Simultaneous assembly of reads from a mix of high-throughput sequencing technologies. *J Comput Biol* 17:1519–1533.
34. Rutherford K, et al. (2000) Artemis: Sequence visualization and annotation. *Bioinformatics* 16:944–945.
35. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410.
36. Lowe TM, Eddy SR (1997) tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25:955–964.
37. Peyretaillade E, et al. (2009) Identification of transcriptional signals in *Encephalitozoon cuniculi* widespread among Microsporidia phylum: Support for accurate structural genome annotation. *BMC Genomics* 10:607.
38. Rice P, Longden I, Bleasby A (2000) EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet* 16:276–277.
39. Katoh K, Toh H (2010) Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics* 26:1899–1900.
40. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res* 38(Database issue):D355–D360.
41. Michal G ed. (1999) *Biochemical Pathways. An Atlas of Biochemistry and Molecular Biology* (John Wiley and Sons, New York) 277 pp.
42. Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 17:540–552.
43. Gouy M, Guindon S, Gascuel O (2010) SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* 27:221–224.
44. Ronquist F, et al. (2012) MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 61:539–542.
45. Stamatakis A (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.