Evolutionary Genomics of *Metchnikovella incurvata* (Metchnikovellidae): An Early Branching Microsporidium

Luis Javier Galindo[1,*], Guifré Torruella[1], David Moreira[1], Hélène Timpano[1], Gita Paskerova[2], Alexey Smirnov[2], Elena Nassonova[2,3], and Purificación López-García[1,*]

[1]Ecologie Systématique Evolution, CNRS, Université Paris-Sud, AgroParisTech, Université Paris-Saclay, Orsay, France

[2]Department of Invertebrate Zoology, Faculty of Biology, St Petersburg State University, Russia

[3]Laboratory of Cytology of Unicellular Organisms, Institute of Cytology Russian Academy of Sciences, St. Petersburg, Russia

*Corresponding authors: E-mails: puri.lopez@u-psud.fr; luis.galindo@u-psud.fr.

Accepted: September 12, 2018

Data deposition: This project has been deposited at DDBJ/ENA/GenBank under the BioProject number PRJNA477760, the SRA accession number SRP151413, and the Whole Genome Shotgun project QXFS00000000.

Abstract

Metchnikovellids are highly specialized hyperparasites, which infect and reproduce inside gregarines (Apicomplexa) inhabiting marine invertebrates. Their phylogenetic affiliation was under constant discussion until recently, when analysis of the first near-complete metchnikovellid genome, that of *Amphiamblys* sp., placed it in a basal position with respect to most Microsporidia. Microsporidia are a highly diversified lineage of extremely reduced parasites related to Rozellida (Rozellosporidia = Rozellomycota = Cryptomycota) within the Holomycota clade of Opisthokonta. By sequencing DNA from a single-isolated infected gregarine cell we obtained an almost complete genome of a second metchnikovellid species, and the first one of a taxonomically described and well-documented species, *Metchnikovella incurvata*. Our phylogenomic analyses show that, despite being considerably divergent from each other, *M. incurvata* forms a monophyletic group with *Amphiamplys* sp., and confirm that metchnikovellids are one of the deep branches of Microsporidia. Comparative genomic analysis demonstrates that, like most Microsporidia, metchnikovellids lack mitochondrial genes involved in energy transduction and are thus incapable of synthesizing their own ATP via mitochondrial oxidative phosphorylation. They also lack the horizontally acquired ATP transporters widespread in most Microsporidia. We hypothesize that a family of mitochondrial carrier proteins evolved to transport ATP from the host into the metchnikovellid cell. We observe the progressive reduction of genes involved in DNA repair pathways along the evolutionary path of Microsporidia, which might explain, at least partly, the extremely high evolutionary rate of the most derived species. Our data also suggest that genome reduction and acquisition of novel genes co-occurred during the adaptation of Microsporidia to their hosts.

Key words: Microsporidia, Metchnikovellidae, Holomycota, phylogenomics, phylogeny, comparative genomics.

Introduction

Microsporidia (Opisthokonta) are a highly specialized group of intracellular parasites. This phylum currently includes between 1,300 and 1,500 described species, which parasitize diverse animal groups and, less frequently, protists (Vávra and Lukeš 2013). Among their animal hosts, many have economic importance, such as silkworms, honey bees, and fish. They can also be opportunistic parasites of humans, being particularly harmful in immunosuppressed patients (Didier et al. 2004; Didier and Weiss 2006, 2011). Microsporidia harbor some of the most reduced genomes among eukaryotes (Corradi et al. 2010). During the course of evolution, members of this lineage have lost or drastically simplified several typical eukaryotic features, including canonical mitochondria (Embley and Martin 2006), the flagellum (James et al. 2006) and a conventional Golgi apparatus (Beznoussenko et al. 2007). Although their evolutionary history has been essentially reductive, Microsporidia have also developed key evolutionary innovations, such as a unique infection apparatus, the polar tube, which serves to penetrate the host (Wittner and Weiss 1999).

© The Author(s) 2018. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

For a long time Microsporidia were erroneously thought to be the deepest branching eukaryotes (Leipe et al. 1993; Kamaishi et al. 1996) until the discovery that this result was likely due to a long branch attraction artefact produced by the fast evolutionary rate of these organisms (Philippe et al. 2000). Recently, the position of Microsporidia as close fungal relatives has been recurrently substantiated (Capella-Gutiérrez et al. 2012; James et al. 2013). In particular, Microsporidia appeared to be related to a diverse clade of parasitic organisms known as rozellids—synonymic taxonomic designations: Rozellida (Lara et al. 2010), Cryptomycota (Jones et al. 2011), Rozellomycota (Corsaro et al. 2014), Rozellosporidia (Karpov et al. 2017). However, the precise reconstruction of their phylogenetic relationships and evolutionary traits remains problematic due to their high evolutionary rates, reduction or loss of cellular organelles, and loss of core metabolic routes (Williams et al. 2002; Thomarat et al. 2004; Corradi et al. 2010). Until recently, the only rozellid species with a sequenced genome was *Rozella allomyces*. In recent years, more sequence data for more or less distant relatives of *Rozella* and early-branching Microsporidia have been made available, opening up the possibility to carry out comparative genomic analyses and gain insights in the genome reduction process that seemingly occurred along the Microsporidia branch. These include the genomic surveys on the early-branching *Mitosporidium daphniae* (Haag et al. 2014), which is the only microsporidium with functional DNA-containing mitochondria described to date, and the rozellid *Paramicrosporidium saccamoebae*, branching at a somewhat intermediate position between *Rozella* and *Mitosporidium* (Quandt et al. 2017). Another key deeply branching lineage along the Microsporidia branch is that of metchnikovellids.

The metchnikovellids (taxonomically designated as the family Metchnikovellidae; Caullery and Mesnil, 1914) unites hyperparasites of gregarines (Apicomplexa) that inhabit the intestinal tract of marine annelids (Vivier 1975). Only a few genera have been described to date, including *Amphiamblys*, *Amphiacantha*, and *Metchnikovella*. Some members of the clade, such as *Metchnikovella incurvata*, have been known for >100 years (Caullery and Mesnil 1914). The phylogenetic affiliation of this long-standing *incertae sedis* group has been debated over time. Because of their morphological and ultrastructural characteristics, metchnikovellids were often thought to be related to Microsporidia (Sprague 1977). Indeed, like most Microsporidia, they lack canonical mitochondria. However, their spores do not exhibit some key microsporidian features, such as the coiled polar filament, the polaroplast and a merogonial proliferation in the life cycle (Sokolova et al. 2013). Phylogenomic analysis of the first available genome of a metchnikovellid, that of *Amphiamblys* sp. (Mikhailov et al. 2017) placed this lineage as the sister group of all derived Microsporidia with the exception of *M. daphniae*, which was placed close to the root of Microsporidia, thereby confirming the long-held suspicion that the metchnikovellids are early diverged Microsporidia. The analysis of the *Amphiamblys* sp. genome revealed some remarkable features, such as the absence of the ATP/ADP translocase family, which is ubiquitous in all derived Microsporidia (Tsaousis et al. 2008), and raised the question of how metchnikovellids obtain ATP without this transporter. However, although seemingly quite complete, the amplified *Amphiamblys* sp. genome is nonetheless partial and these peculiar features need to be verified in other members of the group. Obtaining novel metchnikovellid genome sequences might thus be very useful to determine synapomorphies for the clade and refine the evolutionary path to extreme genome reduction observed along the Microsporidia branch.

In this study, we have analyzed the genome of a second metchnikovellid species, *M. incurvata*. This is the first genomic and phylogenetic study of a taxonomically described and well-documented metchnikovellid species (Sokolova et al. 2013; Rotari et al. 2015). Our results confirm the monophyly of *Amphiamblys* and *Metchnikovella* and strongly support the notion that metchnikovellids branch deeply in the Microsporidia lineage, providing insights into the evolution of the Microsporidia proteome along the diversification of this lineage.

Materials and Methods

Biological Samples

Individual cells of the gregarine *Polyrhabdina* sp. infected with the metchnikovellid *M. incurvata* were isolated from the intestinal tract of the polychaete *Pygospio elegans*. Polychaetes were collected from the Levin Reach silt littoral zone in the Chupa Inlet of the Kandalaksha Gulf, located in the White Sea (66°17′52.68″N, 33°27′46.44″E). Polychaetes were dissected and infected gregarine cells were individually isolated, washed by successive passage into filtered seawater droplets, and sorted into separate tubes for further analyses. At light microscopic level the infected gregarines were easily recognized by the presence of rounded and oval inclusions (fig. 1*A* and *B*) or by elongated and slightly curved cysts inside the cytoplasm (fig. 1*C*). On the basis of the individual characters used for the species identification within the genus *Metchnikovella*, the super-host range, the host range, the size and shape of the observed cysts (Sokolova et al. 2013; Rotari et al. 2015) the microsporidium was identified as *M. incurvata*.

Single-Cell Genome Amplification and Sequencing

Total DNA extraction was performed on one single-isolated infected gregarine cell (containing proliferating cells of *M. incurvata*; see fig. 1*A*) using the PicoPure kit (Thermo Fisher Scientific) according to the manufacturer's protocol. Whole genome amplification (WGA) of the DNA purified from this single infected gregarine was carried out using two methodologies, either multiple displacement amplification (MDA)

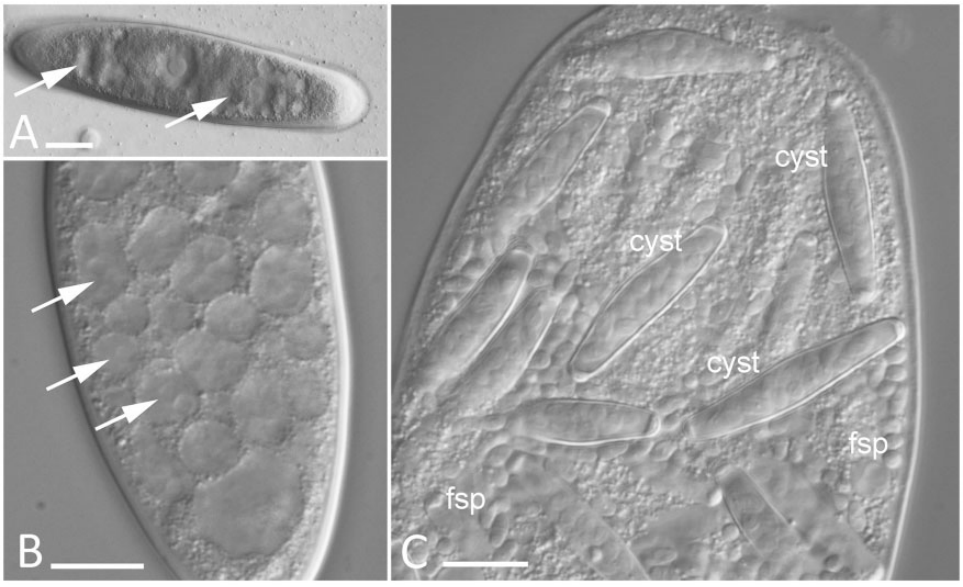

Fig. 1.—*Metchnikovella incurvata*, a hyperparasite of gregarines *Polyrhabdina* sp. from the polychaete *Pygospio elegans*. (*A*, *B*) Infected gregarine cells filled with rounded and oval inclusions (arrowed) corresponding to the early stages of hyperparasite proliferation. Panel *A* shows the cell from which DNA extraction and the subsequent whole genome amplification by MDA were done and which was further used for single-cell genome sequencing. (*C*) Infected gregarine cell filled with the cysts (cyst) and free spores (fsp). Scale bar: 5 μm.

with the REPLI-g kit (QIAGEN) or amplification with the MALBAC single cell WGA kit (Yikon Genomics), following the manufacturer's protocols. We then proceeded to exclude single amplified genomes (SAGs) which did not yield the expected metchnikovellid 18S rDNA amplicons when tested by PCR amplification. Three independent WGA reactions were done on DNA extracted from the individually isolated gregarine cell documented in figure 1*A*: Two MDA reactions (LNA5-MDA1 and LNA5-MDA2) and one MALBAC reaction (LNA5-MALBAC). DNA amplification was confirmed by assessing the DNA quantity using Qubit fluorometric quantification (Life Technologies), together with PCR amplification and Sanger sequencing of the 18S rRNA gene. From the two MDA reactions performed for the same sample, we obtained a DNA concentration of 60 ng/µl (LNA5-MDA1) and 129.2 ng/µl (LNA5-MDA2), and for the MALBAC reaction, we obtained 23.4 ng/µl (LNA5-MALBAC). Since the MALBAC sample did not yield a high DNA amount and since we failed to amplify the 18S rRNA gene, we proceeded to sequence one of the MDA samples (LNA5-MDA1). We prepared two WGA TruSeq Paired-end libraries from this sample and sequenced them on a HiSeq 2500 Illumina instrument (2 × 125 bp) chemistry v4. For each library, we obtained 102,700,329 reads for a total length of 25,675 Mbp and 99,095,593 reads for a total length of 24,774 Mbp, respectively.

Genome Decontamination, Assembly, and Annotation

The quality of the paired-end reads was assessed with FastQC (Andrews 2010) before and after quality trimming. We then trimmed the Illumina adapters with Trimmomatic v0.32 in Paired End mode (Bolger et al. 2014), with a minimal length of 100 bp, removing the first 15 bp, a minimum quality allowed of 20 at the beginning and end of the read. Trimmed pair-end reads were assembled using SPAdes 3.9.1 in single-cell mode (Bankevich et al. 2012), with four k-mer values (25, 77, 99, 117); two assembly rounds were performed, one before and one after decontamination. The first round resulted in an assembly of 6.45 Mb formed by 1,667 contigs of prokaryotic and eukaryotic origin. To decontaminate the assembly, we used BlobTools (Kumar et al. 2013; Laetsch and Blaxter 2017), generating taxon-annotated GC plots. We then eliminated abundant contaminant prokaryotic reads identified in the generated plots. We also removed a few reads that were suspected to belong to the apicomplexan or to the polychaetan hosts by the BlobTools taxonomic identity; these reads were manually inspected with BlastN (Altschul et al. 1990) prior to removal. Surprisingly, there were almost no contaminating reads coming from the apicomplexan host, likely due to the advanced stage of *M. incurvata* infection. After decontamination, the remaining reads underwent a second round of assembly, and were again analyzed with BlobTools to confirm the success of the decontamination procedure (supplementary fig. S1, Supplementary Material online). The final *M. incurvata* assembly had 5.39 Mb and 1,257 contigs. The statistics of the final assembled genome were assessed with QUAST 4.5 (Gurevich et al. 2013) and Qualimap v2.2.1 (Okonechnikov et al. 2015) for coverage estimation. De novo functional gene annotation for the *M. incurvata* genome was performed using two gene

prediction programs: Augustus 3.0.3 (Stanke and Morgenstern 2005) and GeneMarkS v3.26 (Besemer et al. 2001). A few potential introns were predicted by Augustus, but further exploration using BlastX (BLAST 2.6.0; Altschul et al. 1997) against a wide diversity of eukaryotic proteomes included in our local database, rejected them. We used GeneMarkS, an intronless gene prediction algorithm, for gene prediction in the *M. incurvata* genome. Repetitive elements in the genome of *M. incurvata* were searched using RepeatModeler 1.0.10 (Smit et al. 1996). We generated a custom library of repetitive families with RepeatModeler. The resulting sequences were subsequently aligned to the *M. incurvata* assembly and masked by RepeatMasker 4.0 (Smit et al. 2013) to provide a table of the distribution of repetitive element families for the group. This annotation of families of repetitive elements was done including several control genomes to confirm the consistency and quality of the annotation. To assess the completeness of the *M. incurvata* genome, we used BUSCO v2.0.1 (Simão et al. 2015) on the annotated genes with the Fungi and Microsporidia data sets of near-universal single-copy orthologs.

Phylogenetic Analyses

We reconstructed molecular phylogenetic trees of the 18S rRNA genes and phylogenomic analyses of a multigene data set using maximum likelihood (ML) methods (Felsenstein 1981) and Bayesian inference (BI; Huelsenbeck and Ronquist 2001). All alignments were performed using MAFFT v7.388 (Katoh and Standley 2013) with default parameters. Alignments were inspected manually using Geneious v6.0.6 (Kearse et al. 2012), and trimmed from ambiguously aligned regions and gaps using trimAl v1.2 in automated1 mode (Capella-Gutiérrez et al. 2009). For the 18S rRNA, ML inferences were done using IQ-TREE v1.6.2 (Nguyen et al. 2015) applying the TIM3 model with four gamma categories and empirical base frequencies (F+G4), which was the best fit model chosen by BIC (Posada, 2008). 18S rRNA BI analyses were performed using Phylobayes v1.5a (Lartillot and Philippe 2004, 2006; Lartillot et al. 2007), under the CAT-Poisson evolutionary model. Two independent MCMC chains for each data set were run for 10,000 cycles and summarized with a 25% burn-in. For the multigene phylogenomic analyses we used a previously assembled 56 Single-Copy Protein Domains (SCPD) data set (Torruella et al. 2012) containing 32 representatives of the Holomycota clade and 5 other Amorphea species as outgroup (2 Holozoa, 1 Apusomonadida, and 2 Amoebozoa). Proteome data were obtained from GenBank (http://www.ncbi.nlm.nih.gov/genbank, last accessed February 25, 2018), except for the proteomes of *Antonospora locustae*, *Mortierella alpina*, *Rhizopus oryzae*, and *Lichtheimia corymbifera*, which were obtained from the JGI Genome Portal (http://genome.jgi.doe.gov/, last accessed January 22, 2018). ML analyses were performed using IQ-TREE v1.6.2 (Nguyen et al. 2015) applying the LG evolutionary model with four gamma categories, empirical amino acid frequencies and a proportion of invariable sites (LG + F + I + G4), which was the best fit model according to BIC. The BI analyses of the SCPD data set were performed with Phylobayes v1.5a (Lartillot and Philippe 2004, 2006; Lartillot et al. 2007), under the CAT-Poisson evolutionary model. Two independent MCMC chains for each data set were run for 10,000 cycles and summarized with 25% burn-in. All trees were visualized using FigTree v1.4.3 (Rambaut 2016).

Functional Annotation of the Predicted Proteome

Assignment of Gene Ontology (GO) terms to the *M. incurvata* proteome was done using the eggNOG-mapper from the EggNOG v4.5 (Huerta-Cepas et al. 2017) database, using DIAMOND as mapping mode, and the taxonomic scope to adjust automatically. We did this for 12 Holomycota proteomes (1 Fungi, 2 Cryptomycota, and 9 Microsporidia). We used the native R heatmap function (R Development Core Team 2014) to plot a proteome comparison between 73 different core GO terms. We also compared GO terms of gene/pathways involved in DNA repair for 27 holomycotan proteomes, specifically GOs involved in homologous recombination, nonhomologous end joining, mismatch repair, nucleotide-excision repair, and base-excision repair. In addition, we used HMMER 3.1b2 software (Finn et al. 2011) to search for specific genes of interest in the *M. incurvata* proteome, including 15 genes reported to have been acquired by Microsporidia by horizontal gene transfer (HGT; Tsaousis et al. 2008; Marcet-Houben and Gabaldón 2010; Xiang et al. 2010; Heinz et al. 2012; Pombert et al. 2012; Nakjang et al. 2013; Alexander et al. 2016).

Gene Gain and Loss Analysis

We selected the 37 opisthokont proteomes used for the phylogenomic analysis, including that of *M. incurvata*, and carried out orthologue clustering with OrthoFinder v1.1.20 (Emms and Kelly 2015) with default parameters. We identified 12,448 orthogroups, of which 792 were genome-specific. To infer gene gain and loss in protein families along the Microsporidia line, we then applied the Dollo parsimony method implemented in the Count software (Csurös 2010) on the phylogenomic tree topology obtained by BI. Finally, specific metchnikovellid orthologs were later validated with HMMER search (Finn et al. 2011) on EggNOG v4.5 (Huerta-Cepas et al. 2017).

Data Availability

Data generated for this study has been deposited at DDBJ/ENA/GenBank under the BioProject number PRJNA477760,

Table 1
Key Statistics for Genome Assembly and Annotation

	R. allomycis	*P. saccamoebae*	*M. daphniae*	*M. incurvata*	*Amphiamblys* sp.	*N. parisii* (ERTm1)	*T. hominis*	*E. cuniculi*
Assembly size (Mb)	11.86	7.28	5.64	5.4	5.6	4.15	8.5	2.5
GC%	34.5	46.9	43	32.62	50.2	34.5	34.1	34.5
Number of contigs	1,150	216	612	1,257	1,843	65	1,632	11
N50	48,693	69,936	32,031	14,622	10,678	649,559	9,528	220,294
Protein-coding genes	6,350	3,750	3,331	2,803	3,647	2,726	3,212	1,996
Repetitive elements %	3.23%	5.91%	3.69%	17.53%	32.6%	10.59 %	7.43%	10.78%

the SRA accession number SRP151413, and the Whole Genome Shotgun project QXFS00000000.

Results and Discussion

Genome Organization and Repetitive Elements

After the amplification and Illumina sequencing of the *M. incurvata* genome from one single gregarine cell (fig. 1*A*), we generated an initial draft genome of 6.45 Mb, with an N50 of 14,687 bp. After decontamination of the identified nonmetchnikovellid sequences (supplementary fig. S1, Supplementary Material online), we obtained a genome of 5.4 Mb with a N50 of 14,622 bp and a GC content of 32.62%, encoding a total of 2,803 proteins (table 1). The sequence coverage distribution along the genome followed a normal distribution with a notably high mean of ~7000× (supplementary fig. S2, Supplementary Material online), a result that might be expected after the MDA amplification of a relatively small genome. To assess the completeness of this genome, we conducted a BUSCO analysis using as reference the Fungi and Microsporidia data sets of near-universal single-copy orthologs (supplementary table S1, Supplementary Material online). The fungal data set provided a better coverage of the early branching metchnikovellid genome, as it allowed us to identify 196 complete single-copy, 6 duplicated, 31 fragmented, and 57 missing core orthologs, from a total of 290. This allowed inferring ~80% genome completeness for the *M. incurvata* genome sequence, which is comparable to that for the *Amphiamblys* sp. genome (90%; Mikhailov et al. 2017).

Around 43% of the *M. incurvata* genome corresponded to coding regions (similar to the 46% seen in *Amphiamblys* sp.). Repetitive elements accounted for 17.53% of the whole genome, most of which were interspersed repeats (16.2% of the genome; supplementary table S2, Supplementary Material online). To compare our results to those of previous reports we calculated the percentage of repetitive elements in other members of the group (table 1). For *Amphiamblys* sp. we inferred that 32.6% of the genome corresponded to repetitive sequences; similar to the 30% calculated by Mikhailov et al. (2017). However, the percentage of repetitive elements that we found in *P. saccamoebae*, was 5.92%, ten times the percentage (0.53%) reported by Quandt et al. (2017). Interestingly, the percentage of repetitive elements in organisms with smaller genomes was higher than that in those with larger genomes. This might seem at odds with the tendency of Microsporidia towards genome reduction and compaction. Nonetheless, these repetitive elements could be involved in the genome reduction process, as their presence often accompanies pseudogenization and gene loss (Lynch and Conery 2000; Jurka 2004; Dewannieux and Heidmann 2005). At the same time, the presence and relative abundance of repetitive elements in Microsporidia could also play an important role in the adaptation and evolution of this lineage (Parisot et al. 2014). Repetitive elements contribute to genome plasticity in several organisms (Biémont 2010), including closely related fungi. Accordingly, they might contribute to the adaptation to different hosts (Dean 2005; Amyotte et al. 2012; Raffaele and Kamoun 2012). In the case of the microsporidium *Anncaliia algerae*, these elements might be speculated to help eluding the host immune system by acting as a lure (Panek et al. 2014).

The loss of introns and spliceosome activity is thought to have occurred independently in several Microsporidia lineages. Similarly to *Amphiamblys* sp. (Mikhailov et al. 2017), *M. incurvata* seems to lack introns and, consequently, the spliceosome machinery is practically inexistent. For example, neither *M. incurvata* nor *Amphiamblys* sp. seem to possess genes coding for Sf3b1 and Prp8 (two proteins that form the spliceosome catalytic core); these genes have been found in Microsporidia with active splicing (Desjardins et al. 2015). However, we found a RtcB-like ligase involved in tRNA splicing and repair (supplementary fig. S3, Supplementary Material online). This gene is ubiquitous in all eukaryote groups, except fungi, but is present in Microsporidia including *M. daphniae* (Haag et al. 2014). We have also found this gene in *P. saccamoebae*. Therefore, it was likely present in the last common holomycotan ancestor and retained during the evolution and diversification of Microsporidia as intracellular parasites. It has been hypothesized that, in bacteria, RctB may be involved in repair from stress-induced RNA damage; their homologs might catalyze tRNA repair or splicing reactions in archaea and eukaryotes (Tanaka et al. 2011; Tanaka and Shuman 2011). Consequently, RctB in Microsporidia possibly evolved to cope with RNA damage (Thomarat et al. 2004).

Microsporidia are classically thought to lack canonical mitochondria and to possess instead mitosomes, which are highly reduced mitochondrial-derived organelles without respiratory function, involved in the biosynthesis of iron–sulfur clusters essential for many proteins (Hirt et al. 1997; Tsaousis et al. 2008; Waller et al. 2009; Boniecki et al. 2017; Freibert et al. 2017). However, recent studies have shown that the early-branching microsporidium *M. daphniae* and the closely related *P. saccamoebae* still possess mitochondria. Since the phylogenetic position of Metchnikovellidae lies between Microsporidia with and without mitochondria, they might have retained intermediate mitochondria-related organelles important to understand the transition towards mitosomes. We found genes coding for both mitochondrial Hsp70 and the essential sulfur donor Nfs1 (which is of mitochondrial origin) (Emelyanov 2003) in the *M. incurvata* genome (supplementary fig. S4, Supplementary Material online). These genes were also present in *Amphiamblys* sp. and in all Microsporidia with a mitosome (Tsaousis et al. 2008). They seem to play a key role in mitosomes, being required for the maturation of diverse functional proteins (Kispal et al. 2005). The Hsp70 gene phylogeny was congruent with the known phylogeny of Microsporidia (supplementary fig. S4*A*, Supplementary Material online). We did not find any genes related to the main mitochondrial metabolic routes, such as a functional oxidative phosphorylation or the tricarboxylic acid cycle (supplementary table S3, Supplementary Material online). Therefore, the metchnikovellid mitochondrion-derived organelle seems to resemble more a microsporidian mitosome than the *M. daphniae* or *P. saccamoebae* mitochondria.

Phylogenomics of *M. incurvata*

We identified the 18S rRNA gene sequence in the *M. incurvata* assembled genome and reconstructed the corresponding phylogenetic tree (supplementary fig. S5, Supplementary Material online). As expected, *M. incurvata* formed a clade with the rest of metchnikovellids, and the Metchnikovellidae family formed a strongly supported group (with maximum support values for bootstrap and posterior probabilities) branching basally to the clade of canonical, long-branching Microsporidia (which we name here Core Microsporidia). These results support the monophyly of metchnikovellids and their early divergence in the lineage, even though many other nodes in the 18S rRNA gene tree remain unresolved.

To reconstruct a more robust phylogeny for metchnikovellids, we carried out a multigene phylogenomic analysis for several members of the Holomycota. Our ML and BI trees further confirmed that *M. incurvata* forms a solidly supported monophyletic lineage with *Amphiamblys*. Despite this, *Metchnikovella* is only distantly related to *Amphiamblys* (fig. 2), suggesting that the family Metchnikovellidae might encompass a wide diversity of fast-evolving parasites specialized in various hosts. Interestingly, the Metchnikovellidae branched at the base of the fast-evolving, more derived, Core Microsporidia, right after the basal-branching *M. daphniae* and *P. saccamoebae*. Their intermediate position between classical Microsporidia with mitosomes, and the basal mitochondriate members *M. daphniae* and *P. saccamoebae* and *R. allomycis*, makes this group interesting for studying the evolutionary path to the extreme genome reduction and specialization undergone by the long-branching Microsporidia.

Molecular analyses based on 18S rRNA gene amplification and sequencing from various environments have uncovered a wide diversity of eukaryotes along the lineage leading to Core Microsporidia, from the very basal rozellids to the metchnikovellid divergence (Lara et al. 2010; Bass et al. 2018). In addition to this uncharacterized environmental diversity, a few described genera occupy deep-branching positions in this broad lineage based on 18S rRNA gene phylogenies and their further study should shed some light in the evolutionary history of Microsporidia. An example is *Nucleophaga*, which branches between *Paramicrosporidum* and metchnikovellids plus the Core Microsporidia (Corsaro et al. 2016).

Genome Evolution and Gene Gain and Loss along the Microsporidia Line

We annotated the *M. incurvata* genome and 11 additional members of the Holomycota clade for 73 different GO terms, and then clustered the annotated genomes according to their gene content similarity (fig. 3). In agreement with their phylogeny, *Metchnikovella* and *Amphiamblys* clustered together also according to their gene content. However, the gene content of metchnikovellids has undergone the loss of many core function genes, resembling more that of the highly reduced derived Microsporidia (indicated as Core Microsporidia in fig. 3). This suggests that metchnikovellids rely on the assimilation of metabolites from their hosts, as long-branching Microsporidia do. Remarkably, the metabolic core of *M. daphniae* and *P. saccamoebae* clustered them with *R. allomycis* (fig. 3). Thus, although *Paramicrosporidium* and *Mitosporidium* are considered to be basal microsporidia by some authors (Bass et al. 2018), their functional gene content is more similar to that of rozellids. This, together with the fact that *P. saccamoebae* and *M. daphniae* possess functional mitochondria may question their classification as microsporidia and claims for a taxonomic revision of Microsporidia and Rozellida, the boundaries of which are blurry.

Despite their reduced genomes, *Amphiamblys* and *Metchnikovella* still conserve some basic core metabolic pathways, such as the pentose phosphate and glycolysis pathways and trehalose biosynthesis (supplementary table S3, Supplementary Material online). Nevertheless, like most Microsporidia, metchnikovellids cannot synthesize their own nucleotides, amino acids and lack practically all enzymes involved in fatty acid metabolism (supplementary tables S3, S4, Supplementary Material online).

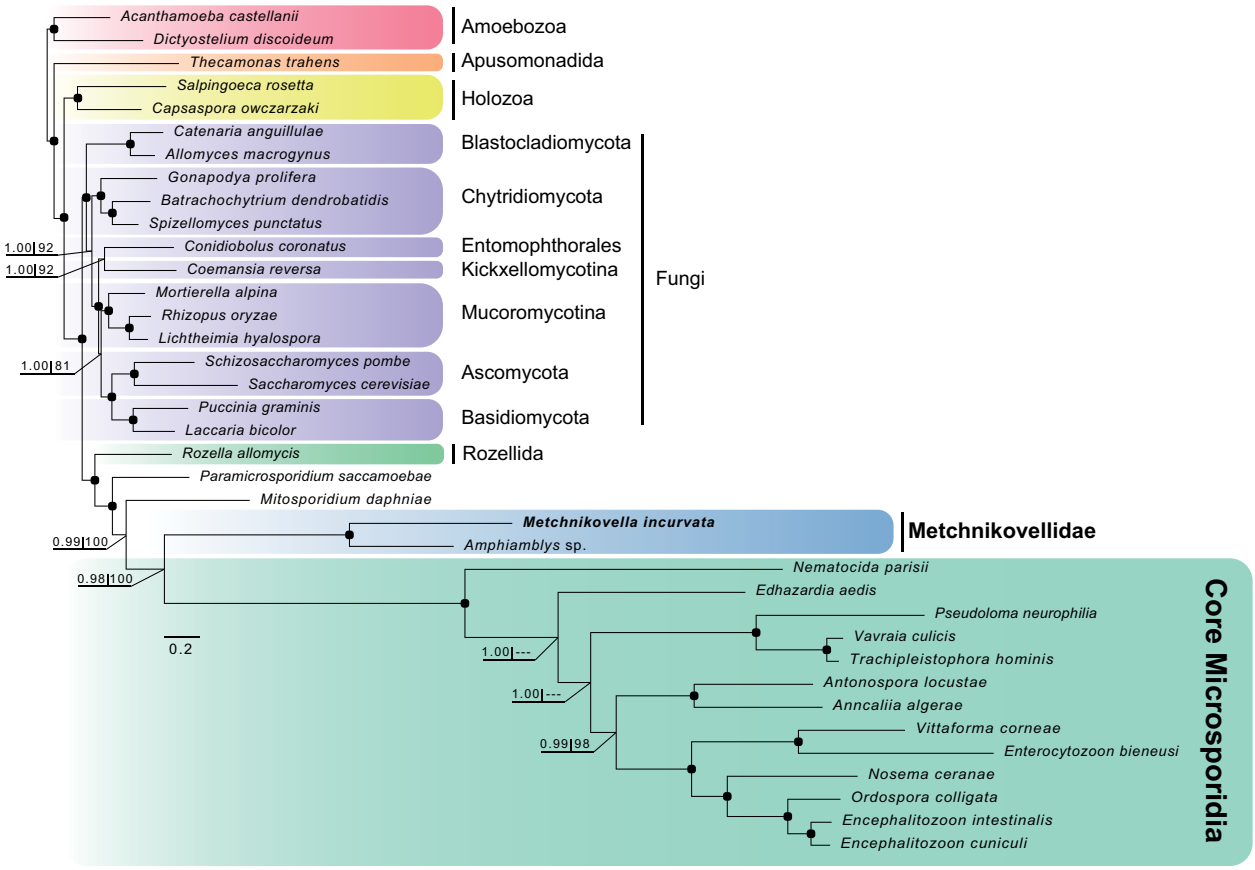

Fig. 2.—Bayesian phylogenomic tree showing the position of metchnikovellids. The tree was reconstructed using a concatenated alignment of 56 single-copy protein domain data set for 32 representatives of the Holomycota clade and 5 other Amorphea species as an outgroup (2 Holozoa, 1 Apusomonadida, and 2 Amoebozoa). Split supports are posterior probabilities (pp) (values on the left) and maximum likelihood (ML) bootstrap (bs) values (on the right). Sequences obtained in this study are highlighted in black. Support values >0.99 pp and >95% bs are indicated by a black bullet.

Interestingly, unlike both divergent Microsporidia and early branching members of this lineage, the two metchnikovellid species lacked an alternative oxidase and the mitochondrial glycerol-3-phosphate dehydrogenase, one of the two enzymes of the glycerol-3-phosphate shuttle. This may appear surprising, since both are thought to be essential for the viability of microsporidian energy metabolism (Dolgikh et al. 2009; Williams et al. 2010). However, metchnikovellids retain a cytosolic glycerol-3-phosphate dehydrogenase (supplementary fig. S6, Supplementary Material online). This enzyme seems to be important in their metabolism as it may allow the synthesis of glycerol 3-phosphate, the starting material for de novo synthesis of glycerolipids and NAD^+, necessary to maintain the adequate NAD^+ cellular levels.

The tricarboxylic acid cycle does not seem to be functional in metchnikovellids, although genes for some enzymes of the pathway are still present, namely the citrate synthase (only detected in *Amphiamblys* sp.) and the malate dehydrogenase (supplementary table S3, Supplementary Material online). The mitochondrial malate dehydrogenase (mMDH) found in *M. incurvata* and *Amphiamblys* sp. is absent in all derived Microsporidia. It has been shown that in some cases the function of this enzyme may shift to a lactate dehydrogenase (LDH; Wilks et al. 1988), which potentially constitutes the final step of anaerobic energy metabolism in metchnikovellids, by balancing the reducing potential of glycolysis. In *Amphiamblys* sp., this change has been reported to occur when an arginine in a key active site changed to a tyrosine (Mikhailov et al. 2017). In *M. incurvata* the change is from arginine to tryptophan, another hydrophobic amino acid (supplementary fig. S7, Supplementary Material online). However, it is likely that the enzyme works as a LDH, since the same substitution (Arg102 to a Trp107) functionally turned the MDH into LDH in Apicomplexa, another group of parasitic protists (Boucher et al. 2014).

One interesting example of differential gene retention in metchnikovellids as compared with core Microsporidia relates to clathrin-coated vesicle formation. Mikhailov et al. (2017) already noted that the *Amphiamblys* sp. genome encoded several proteins required for the formation of clathrin vesicles.

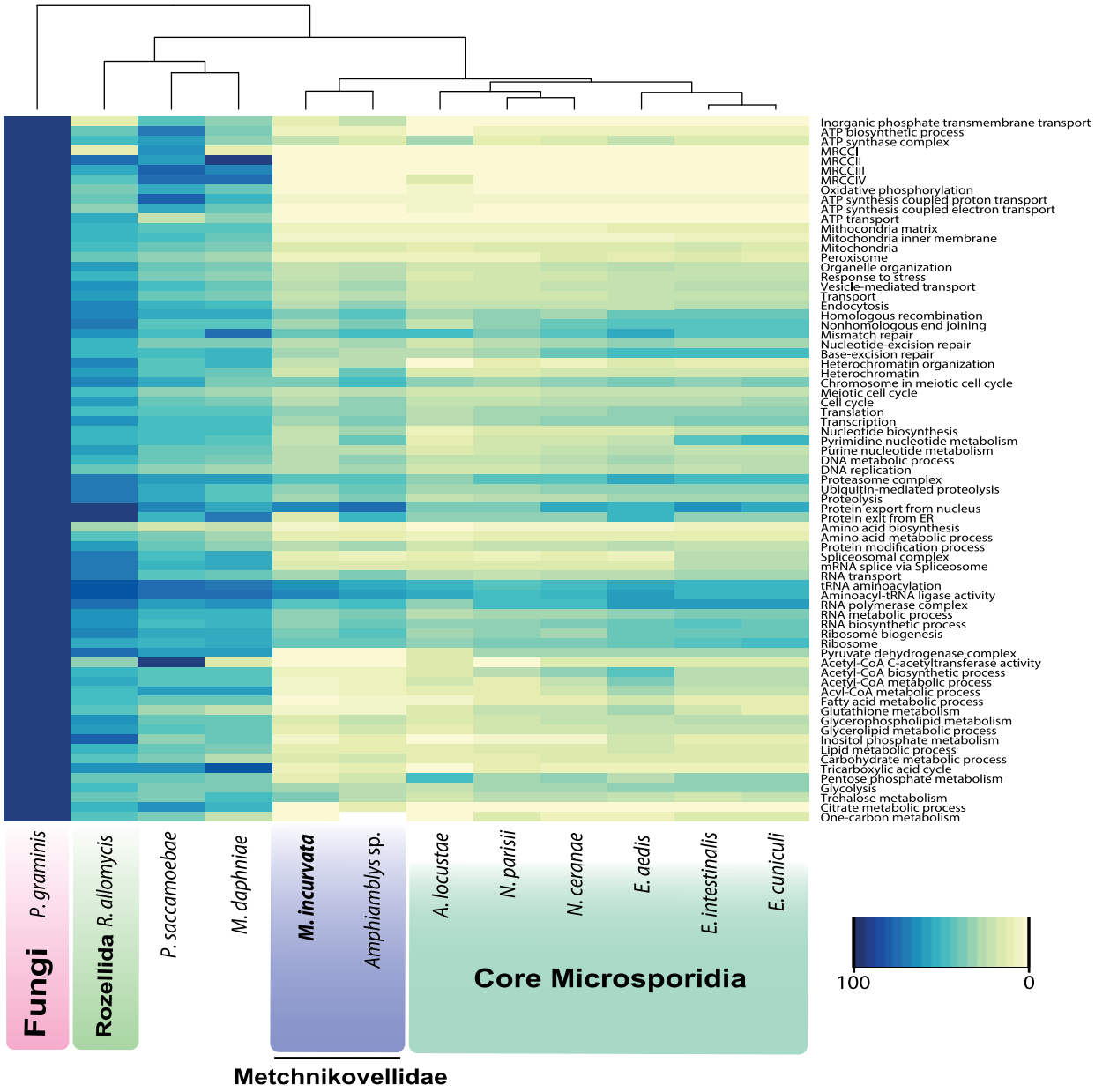

Fig. 3.—Heatmap illustrating the conservation of pathways and structures in metchnikovellids and neighbor lineages. It is based on 73 GO terms, identified using EggNOG (Huerta-Cepas et al. 2017) for 12 Holomycota representative proteomes. Sequences obtained in this study are highlighted in black. Colors indicate the percentage of annotated genes with a GO term.

These are lacking in more derived Microsporidia, which have a highly reduced endocytic machinery. We identified in *M. incurvata* 10 out of the 12 genes involved in clathrin-coated vesicle formation present in *Amphiamblys* sp. (supplementary table S5, Supplementary Material online). We failed to identify an actin-binding WH2 domain (PF02205) and an ARP2/3 complex subunit ARPC3 (PF04062). Since our genome is not complete, the presence of these two genes cannot be discarded. In addition, we do find fragments of other members of the two respective families: WH1 (PF00568) and 2 subunits of the complex ARP2/3 (PF04045 and PF05856). Therefore, with the data available for the two metchnikovellid genomes, the endocytic components of the clathrin vesicle-mediated transport seem to be conserved in metchnikovellids, making it one of the main distinctions of this group from more derived Microsporidia.

Microsporidian genomes are among the fastest-evolving eukaryotic genomes (Thomarat et al. 2004). This high evolutionary rates, which make this group prone to long-branch attraction artifacts, made it difficult for a long time to place

Microsporidia in the tree of life (Capella-Gutiérrez et al. 2012). Since Microsporidia have highly reduced genomes, we asked whether the loss of genes involved in DNA repair might have been responsible for an augmentation of mutation rates. We thus searched for genes in the five main GO terms involved in DNA repair (homologous recombination, nonhomologous end joining, mismatch repair, nucleotide-excision repair, and base-excision repair) in 27 opisthokont genomes. We found that Microsporidia, including metchnikovellids, do indeed have a lower number of genes involved in DNA repair (supplementary table S6; fig. S8, Supplementary Material online). Even in the case of nucleotide-excision repair, where gene loss might appear less important that in other DNA repair systems (supplementary fig. S8, Supplementary Material online), the two metchnikovellid genomes have a gene reduction of circa 50% as compared with more basal opisthokonts (*Amphiamblys*, 41 genes; *M. incurvata*, 46 genes; vs. e.g., *S. rosetta*, 98 genes; *T. trahens*, 78 genes; *R. allomycis*, 80 genes; supplementary table S6, Supplementary Material online). Therefore, this loss of genes involved in DNA repair is likely one of the causes leading to the increased evolutionary rates seen in Microsporidia. A high mutation and recombination rate may be at the origin of pseudogenization and gene loss but, at the same time, offers a powerful mean to successfully cope with the arms race established with the microsporidian hosts.

Microsporidia have acquired several genes through HGT (Alsmark et al. 2013; Alexander et al. 2016). In highly reduced genomes, these genes may play important adaptive roles. A paradigmatic example is the acquisition of a bacterial ATP/ADP translocase of probable chlamydial origin, which is responsible for the import of ATP from the host (Tsaousis et al. 2008). However, we did not find this ATP/ADP translocase family or any other paralogue in the *M. incurvata* genome. The family is also missing in *Amphiamblys* sp., *M. daphniae*, and *P. saccamoebae*. Although the two metchnikovellid genomes available are not fully complete, it is unlikely that this gene was missed in the both genomes. This observation might suggest that the ATP/ADP translocase family was acquired by HGT after the divergence of Metchnikovellidae. However, since this gene family is also present in the rozellid *R. allomycis* (James et al. 2013), an alternative explanation is that the gene was acquired by a common ancestor of *R. allomycis* and Microsporidia (Dean et al. 2018), and later lost in the Metchnikovellidae.

It has been proposed that, in *Amphiamblys* sp., a gene related to the mitochondrial carrier protein family (MCF) might have evolved for nucleotide transport, playing a role in mitosomal metabolism (Mikhailov et al. 2017). This MCF gene is found in fungi (e.g., *Saccharomyces cerevisiae*) and other holomycota, and originally transported inorganic phosphate into the mitochondrion. We found the same gene in *M. incurvata*, and in fact, it is the only MCF member still retained in the metchnikovellid genomes (supplementary fig. S9, Supplementary Material online). Although we observe that the gene is also found in early-branching members that produce ATP with active mitochondria (*M. daphniae*, *P. saccamoebae*, *R. allomycis*, and the aphelid *Paraphelidium tribonemae*), it is unrelated to MCF genes found in the late-branching microsporidium *A. locustae*.

However, the MCF gene previously identified in *A. locustae* (Williams et al. 2008) derived from an EST and is not actually encoded in the available genome sequence for *A. locustae*, such that it represents a potential contaminant gene. If this is indeed the case, the MCF gene would have been retained only in metchnikovellids and basal members of the lineage and later lost in Core Microsporidia. The presence of MCF gene homologues in both *M. incurvata* and *Amphiamblys* sp. supports the idea that MCF is retained in metchnikovellids. Moreover, the *M. incurvata* contig in which the MCF gene is located has a total length of 19,629 bp and encodes 10 eukaryotic proteins. When blasted, all proteins except one (which is fragmented) have homologs in *Amphiamblys* sp. This further confirms that this MCP gene belongs to the *M. incurvata* genome and is not a contaminant. Metchnikovellids are the only Microsporidia with no functional aerobic mitochondria still conserving this gene. Although experimental and cellular localization evidence are still lacking, we might hypothesize that this MCF has become an ATP transporter in metchnikovellids, evolving its function to pump ATP from their hosts.

In addition to the canonical ATP/ADP translocase family, we have studied other 15 published cases of HGT that are widespread in different Microsporidia. We searched for these markers in 19 genomes of derived Microsporidia, metchnikovellids and basal members of the group (supplementary table S7, Supplementary Material online). In the *M. incurvata* genome we found only one out of the 15 transferred genes, the manganese superoxide dismutase (MnSOD; Xiang et al. 2010). The ML tree of this protein including members of bacteria, Microsporidia, and various other eukaryotes supports several independent acquisitions of the MnSOD gene by members of Microsporidia and Rozellida from various bacterial donors (supplementary fig. S10, Supplementary Material online). These transfers affect the Metchnikovellidae (both *M. incurvata* and *Amphiamblys* sp.), the microsporidium *A. locustae*, and the remaining derived Microsporidia. Independent acquisitions of the MnSOD gene occur also in aphelids (*P. tribonemae*; Torruella et al. submitted) and the two parasitic anaerobic gut fungi *Piromyces finnis* and *Anaeromyces robustus* (supplementary fig. S10, Supplementary Material online). These multiple acquisitions likely reflect an important adaptive function. In fact, the MnSOD gene seems to play a key role in protecting anaerobic life from the well-known deleterious effects of oxygen (Holley et al. 2011). Several examples, mainly in bacteria and yeast, have shown that cells expressing MnSOD as a result of stimulation with high oxygen levels were more resistant to hyperbaric oxygen concentrations (Gregory

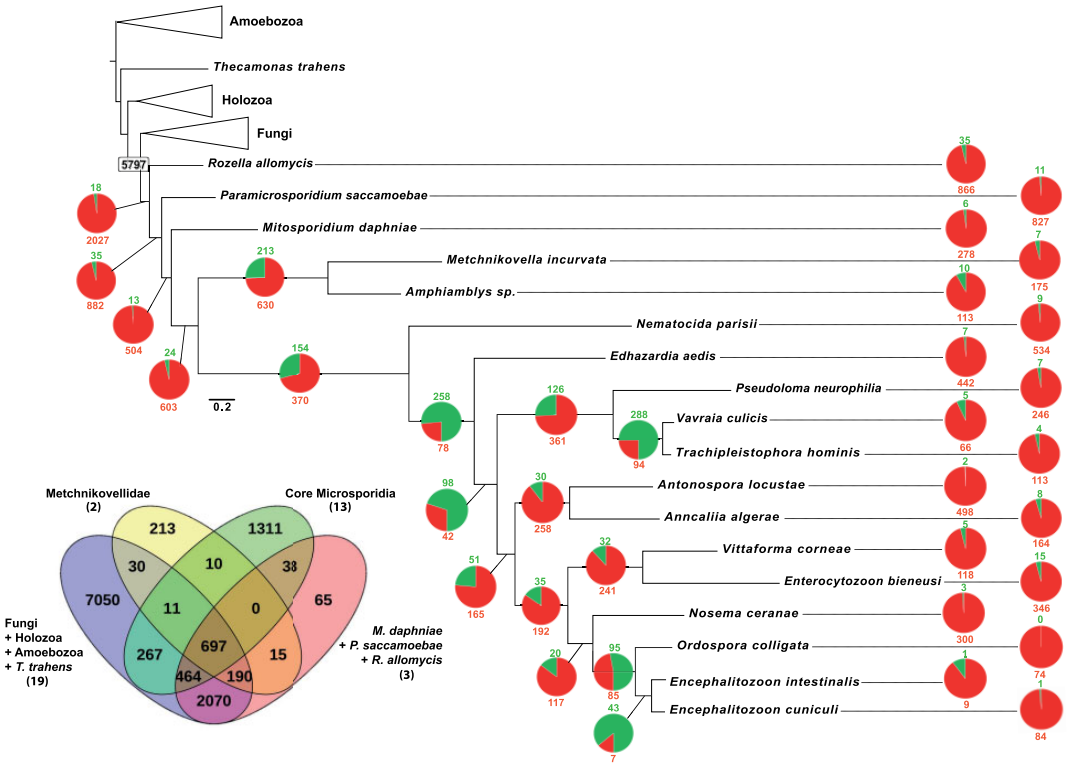

Fig. 4.—Gain and loss of protein orthogroups along the evolutionary lineage of Microsporidia, based on a BI phylogeny from the multigene data set. Pie charts at nodes represent the total gains (green) and losses (red) shown respectively in numbers above and below the pie charts. The deepest node indicates the estimated number of ancestral protein orthogroups (5, 797). The Venn diagram represents the total number of protein orthogroups shared between 37 opisthokont proteomes grouped into metchnikovellids (2 proteomes), Microsporidia (13 proteomes), Fungi + Holozoa + Amoebozoa + *T. trahens* (19 proteomes) and *M. daphniae* + *P. saccamoebae* + *R. allomycis* (3 proteomes).

and Fridovich 1973; Gregory et al. 1974). The recurrent acquisition of the MnSOD gene in Microsporidia might be therefore linked to cell protection against reactive oxygen species.

To have a global overview on the genome evolution in metchnikovellids and in other Microsporidia, we carried out a gain and loss analysis using the reconstructed phylogeny of Microsporidia as a backbone and applying Dollo parsimony on protein orthogroups (fig. 4; supplementary tables S8–S10, Supplementary Material online). As shown in figure 4, an important proportion of orthogroups were lost at the beginning of the diversification of a clade comprising *R. allomycis*, *P. saccamoebae* and the mitochondriate microsporidium *M. daphniae*. This early genome reduction undergone by Microsporidia might relate with the transition towards an obligatory intracellular parasite lifestyle (Heinz et al. 2012; Nakjang et al. 2013). Metchnikovellids occupy an intermediate position along the microsporidian branch. We found 213 gains of specific orthogroups in these organisms, including 561 proteins, of which 251 correspond to *Amphiamblys* sp. and 310 to *M. incurvata*. HMMR searches using EggNOG v4.5 for the 561 metchnikovellid proteins retrieved homologues only for 11 *Amphiamblys* sp. and 15 *M. incurvata* proteins in this database. These proteins included 17 orthogroups comprising 37 proteins (11 of them did not return positive hits; supplementary table S11, Supplementary Material online), which likely correspond to genes retained from the common Microsporidia ancestor. However, we found 524 proteins present in both *M. incurvata* and *Amphiamblys* sp. grouped in 196 orthogroups without relatives in other lineages. These genes were probably gained by metchnikovellids (supplementary table S120, Supplementary Material online), implying that soon after the important gene loss experienced by the microsporidian ancestors, gene gain started to overcome gene loss, coinciding with the diversification of long-branch Microsporidia. Indeed, there is also a remarkable process of gene gain at the node leading to one of the most derived lineages of Microsporidia, comprising *Ordospora colligata* and the *Encephalitozoon* clade. This clade comprises Microsporidia with some of the most simplified genomes (Corradi et al. 2010) suggesting that genome reduction and evolution of new markers have co-occurred during the adaptation of Microsporidia to their hosts. However, we cannot completely exclude the possibility that gene losses and gains actually correspond to genes that have evolved beyond recognition (even if some are likely true new genes and clearly adaptive). Collectively, these observations of gene gain reflect

specialization to different hosts, likely achieved by the acquisition of some adaptive genes. Thus, as derived Microsporidia diversified, gene losses started to overcome gains and they became progressively more adapted to their parasitic lifestyle.

Conclusions

Phylogenomic analyses using data from the *M. incurvata* genome have confirmed that the Metchnikovellidae are a deep-branching group inside the Microsporidia, and the deepest of those without functional mitochondria. Comparative analysis of the two available metchnikovellid genomes confirmed that gene complement resembles more to those of typical derived Microsporidia than to those of the less derived mitochondriate *M. daphniae*, *P. saccamoebae*, and *R. allomycis*. We observed a reduction of DNA repair pathways, which seems to correlate with the high evolutionary rates seen in the clade. Interestingly, the typical microsporidian ATP/ADP translocase family does not seem to be present in any of the two metchnikovellid genomes. We hypothesize that this gene was replaced by a MCF gene that became an ATP transporter. Lastly, our gain and loss analysis suggests that reductive evolution is not the only ongoing process in Microsporidia and that the evolution of new genes has also taken place during the adaptation of Microsporidia to their hosts.

Both phylogenomic and comparative genomic analyses rise the need for taxonomic revision of Microsporidia and Rozellida, since the boundaries between them are blurry. For this, is essential to continue with the surveys and sequencing efforts for new members of the group, which will help to fill the gap in knowledge still present in the evolution of the clade.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

This work was funded by the European Research Council Advanced Grant "ProtistWorld" (No. 322669) and the Horizon 2020 research and innovation program under the Marie Skłodowska-Curie ITN project SINGEK (http://www.sin-gek.eu/; grant agreement No. H2020-MSCA-ITN-2015-675752). It also received support from the RFBR grant 18-04-01359. G.T. was financed by the European Marie Sklodowska-Curie Action (704566 AlgDates). We thank the Core Facility Centers "Development of molecular and cell technologies," "Culturing of microorganisms," and the Educational Research Station "Belomorskaia" of Saint-Petersburg State University for access to their equipment facilities.

Literature Cited

Alexander WG, Wisecaver JH, Rokas A, Hittinger CT. 2016. Horizontally acquired genes in early-diverging pathogenic fungi enable the use of host nucleosides and nucleotides. Proc Natl Acad Sci USA. 113(15):4116–4121.

Alsmark C, et al. 2013. Patterns of prokaryotic lateral gene transfers affecting parasitic microbial eukaryotes. Genome Biol. 14(2):R19.

Altschul SF, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25(17):3389–3402.

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

Amyotte SG, et al. 2012. Transposable elements in phytopathogenic *Verticillium* spp.: insights into genome evolution and inter- and intra-specific diversification. BMC Genomics. 13(1):314.

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/.

Bankevich A, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 19(5):455–477.

Bass D, et al. 2018. Clarifying the relationships between microsporidia and cryptomycota. J Eukaryot Microbiol. doi: 10.1111/jeu.12519. [Epub ahead of print]

Besemer J, Lomsadze A, Borodovsky M. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res. 29:2607–2618.

Beznoussenko GV, et al. 2007. Analogs of the Golgi complex in microsporidia: structure and avesicular mechanisms of function. J Cell Sci. 120:1288–1298.

Biémont C. 2010. A brief history of the status of transposable elements: from junk DNA to major players in evolution. Genetics 186(4):1085–1093.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114–2120.

Boniecki MT, Freibert SA, Mühlenhoff U, Lill R, Cygler M. 2017. Structure and functional dynamics of the mitochondrial Fe/S cluster synthesis complex. Nat Commun. 8(1):1287.

Boucher JI, Jacobowitz JR, Beckett BC, Classen S, Theobald DL. 2014. An atomic-resolution view of neofunctionalization in the evolution of apicomplexan lactate dehydrogenases. Elife 3:02304.

Capella-Gutiérrez S, Marcet-Houben M, Gabaldón T. 2012. Phylogenomics supports microsporidia as the earliest diverging clade of sequenced fungi. BMC Biol. 10(1):47.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25(15):1972–1973.

Caullery M, Mesnil F. 1914. Metchnikovellidae et autres Protistes parasites des Grégarines d'Annélides. CR Soc Biol. 77:527–532.

Corradi N, Pombert JF, Farinelli L, Didier ES, Keeling PJ. 2010. The complete sequence of the smallest known nuclear genome from the microsporidian *Encephalitozoon intestinalis*. Nat Commun. 1(6):1.

Corsaro D, et al. 2014. Microsporidia-like parasites of amoebae belong to the early fungal lineage Rozellomycota. Parasitol Res. 113(5):1909–1918.

Corsaro D, et al. 2016. Molecular identification of *Nucleophaga terricolae* sp. nov. (Rozellomycota), and new insights on the origin of the Microsporidia. Parasitol Res. 115(8):3003–3011.

Csurös M. 2010. Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. Bioinformatics 26(15): 1910–1912.

Dean RA, et al. 2005. The genome sequence of the rice blast fungus *Magnaporthe grisea*. Nature 434(7036):980–986.

Dean P, et al. 2018. Transporter gene acquisition and innovation in the evolution of Microsporidia intracellular parasites. Nat Commun. 9(1):1709.
Desjardins CA, et al. 2015. Contrasting host-pathogen interactions and genome evolution in two generalist and specialist microsporidian pathogens of mosquitoes. Nat Commun. 6:7121.
Dewannieux M, Heidmann T. 2005. LINEs, SINEs and processed pseudogenes: parasitic strategies for genome modeling. Cytogenet Genome Res. 110(1–4):35–48.
Didier ES, et al. 2004. Epidemiology of microsporidiosis: sources and modes of transmission. Vet Parasitol. 126(1–2):145–166.
Didier ES, Weiss LM. 2006. Microsporidiosis: current status. Curr Opin Infect Dis. 19(5):485–492.
Didier ES, Weiss LM. 2011. Microsporidiosis: not just in AIDS patients. Curr Opin Infect Dis. 24(5):490–495.
Dolgikh VV, et al. 2009. Heterologous expression of pyruvate dehydrogenase E1 subunits of the microsporidium *Paranosema* (Antonospora) *locustae* and immunolocalization of the mitochondrial protein in amitochondrial cells. FEMS Microbiol Lett. 293(2):285–291.
Embley TM, Martin W. 2006. Eukaryotic evolution, changes and challenges. Nature 440(7084):623–630.
Emelyanov VV. 2003. Phylogenetic affinity of a *Giardia lamblia* cysteine desulfurase conforms to canonical pattern of mitochondrial ancestry. FEMS Microbiol Lett. 226(2):257–266.
Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16:157.
Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol. 17(6):368–376.
Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 39(Suppl.):W29.
Freibert SA, et al. 2017. Evolutionary conservation and in vitro reconstitution of microsporidian iron-sulfur cluster biosynthesis. Nat Commun. 8:13932.
Gregory EM, Fridovich I. 1973. Oxygen toxicity and the superoxide dismutase. J Bacteriol. 114(3):1193–1197.
Gregory EM, Goscin SA, Fridovich I. 1974. Superoxide dismutase and oxygen toxicity in a eukaryote. J Bacteriol. 117(2):456–460.
Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29(8):1072–1075.
Haag KL, et al. 2014. Evolution of a morphological novelty occurred before genome compaction in a lineage of extreme parasites. Proc Natl Acad Sci USA. 111(43):15480–15485.
Heinz E, et al. 2012. The genome of the obligate intracellular parasite *Trachipleistophora hominis:* new insights into microsporidian genome dynamics and reductive evolution. PLoS Pathog. 8:e1002979.
Hirt RP, Healy B, Vossbrinck CR, Canning EU, Embley TM. 1997. A mitochondrial Hsp70 orthologue in Vairimorpha necatrix: molecular evidence that microsporidia once contained mitochondria. Curr Biol. 7(12):995–998.
Holley AK, Bakthavatchalu V, Velez-Roman JM, St. Clair DK. 2011. Manganese superoxide dismutase: guardian of the powerhouse. Int J Mol Sci. 12(10):7114–7162.
Huelsenbeck JP, Ronquist F. 2001. MR BAYES: bayesian inference of phylogenetic trees. Bioinformatics 17(8):754–755.
Huerta-Cepas J, et al. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Mol Biol Evol. 34(8):2115–2122.
James TY, et al. 2006. Reconstructing the early evolution of Fungi using a six-gene phylogeny. Nature 443(7113):818–822.
James TY, et al. 2013. Shared signatures of parasitism and phylogenomics unite cryptomycota and microsporidia. Curr Biol. 23(16):1548–1553.
Jones MDM, Richards TA, Hawksworth DL, Bass D. 2011. Validation and justification of the phylum name Cryptomycota phyl. nov. IMA Fungus. 2(2):173–175.
Jurka J. 2004. Evolutionary impact of human Alu repetitive elements. Curr Opin Genet Dev. 14(6):603–608.
Kamaishi T, et al. 1996. Complete nucleotide sequences of the genes encoding translation elongation factors 1 and 2 from a microsporidian parasite, *Glugea plecoglossi*: implications for the deepest branching of eukaryotes. J Biochem. 120(6):1095–1103.
Karpov SA, Torruella G, Moreira D, Mamkaeva MA, López-García P. 2017. Molecular phylogeny of *Paraphelidium letcheri* sp. nov. (Aphelida, Opisthosporidia). J Eukaryot Microbiol. 64(5):573–578.
Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780.
Kearse M, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28(12):1647–1649.
Kispal G, et al. 2005. Biogenesis of cytosolic ribosomes requires the essential iron–sulphur protein Rli1p and mitochondria. EMBO J. 24(3):589–598.
Kumar S, Jones M, Koutsovoulos G, Clarke M, Blaxter M. 2013. Blobology: exploring raw genome data for contaminants, symbionts, and parasites using taxon-annotated GC-coverage plots. Front Genet. 4:1–12.
Laetsch DR, Blaxter ML. 2017. BlobTools: interrogation of genome assemblies. F1000Research 6:1287.
Lara E, Moreira D, López-García P. 2010. The environmental clade LKM11 and Rozella form the deepest branching clade of fungi. Protist 161(1):116–121.
Lartillot N, Brinkmann H, Philippe H. 2007. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. BMC Evol Biol. 7(Suppl. 1):S4.
Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol Biol Evol. 21(6):1095–1109.
Lartillot N, Philippe H. 2006. Computing Bayes factors using thermodynamic integration. Syst Biol. 55(2):195–207.
Leipe DD, Gunderson JH, Nerad TA, Sogin ML. 1993. Small subunit ribosomal RNA+ of Hexamita inflata and the quest for the first branch in the eukaryotic tree. Mol Biochem Parasitol. 59(1):41–48.
Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. Science (80-) 290(5494):1151–1155.
Marcet-Houben M, Gabaldón T. 2010. Acquisition of prokaryotic genes by fungal genomes. Trends Genet. 26(1):5–8.
Mikhailov KV, Simdyanov TG, Aleoshin VV. 2017. Genomic survey of a hyperparasitic microsporidian *Amphiamblys* sp. (Metchnikovellidae). Genome Biol Evol. 9:454–467.
Nakjang S, et al. 2013. Reduction and expansion in microsporidian genome evolution: new insights from comparative genomics. Genome Biol Evol. 5(12):2285–2303.
Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 32(1):268–274.
Okonechnikov K, Conesa A, García-Alcalde F. 2015. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics 32(2):292–294.
Panek J, et al. 2014. Hijacking of host cellular functions by an intracellular parasite, the microsporidian *Anncaliia algerae*. PLoS ONE. 9(6):e100791.
Parisot N, et al. 2014. Microsporidian genomes harbor a diverse array of transposable elements that demonstrate an ancestry of horizontal exchange with metazoans. Genome Biol Evol. 6(9):2289–2300.

Philippe H, et al. 2000. Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions. Proc R Soc B Biol Sci. 267(1449):1213–1221.
Pombert J-F, et al. 2012. Gain and loss of multiple functionally related, horizontally transferred genes in the reduced genomes of two microsporidian parasites. Proc Natl Acad Sci USA. 109(31):12638–12643.
Posada D. 2008. jModelTest: phylogenetic model averaging. Mol Biol Evol. 25(7):1253–1256.
Quandt CA, et al. 2017. The genome of an intranuclear parasite, *Paramicrosporidium saccamoebae*, reveals alternative adaptations to obligate intracellular parasitism. Elife 6:pii: e29594.
R Core Team (2014). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. URL http://www.R-project.org/.
Raffaele S, Kamoun S. 2012. Genome evolution in filamentous plant pathogens: why bigger can be better. Nat Rev Microbiol. 10(6):417–430.
Rambaut A. 2016. FigTree v1.4.3. Mol. Evol. phylogenet ics Epidemiol.
Rotari YM, Paskerova GG, Sokolova YY. 2015. Diversity of metchnikovellids (Metchnikovellidae, Rudimicrosporea), hyperparasites of bristle worms (Annelida, Polychaeta) from the White Sea. Protistology 9:50–59.
Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19):3210–3212.
Smit A, Hubley R, Green P. 1996. RepeatMasker Open-3.0. RepeatMasker Open-3.0. www.repeatmasker.org.
Smit A, Hubley R, Green P. 2013. RepeatMasker Open-4.0. 2013–2015. http://repeatmasker.org.
Sokolova YY, Paskerova GG, Rotari YM, Nassonova ES, Smirnov AV. 2013. Fine structure of *Metchnikovella incurvata* Caullery and Mesnil 1914 (microsporidia), a hyperparasite of gregarines Polyrhabdina sp. from the polychaete *Pygospio elegans*. Parasitology 140(7): 855–867.
Sprague V. 1977. In: Bulla LA, Cheng TC, editors. Annotated list of species of microsporidia BT—comparative pathobiology: volume 2 systematics of the Microsporidia. Springer US: Boston, MA. p. 31–334.
Stanke M, Morgenstern B. 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 33(Web Server):W465.
Tanaka N, Meineke B, Shuman S. 2011. RtcB, a novel RNA ligase, can catalyze tRNA splicing and HAC1 mRNA splicing in vivo. J Biol Chem. 286(35):30253–30257.
Tanaka N, Shuman S. 2011. RtcB is the RNA ligase component of an *Escherichia coli* RNA repair operon. J Biol Chem. 286(10):7727–7731.
Thomarat F, Vivarès CP, Gouy M. 2004. Phylogenetic analysis of the complete genome sequence of Encephalitozoon cuniculi supports the fungal origin of microsporidia and reveals a high frequency of fast-evolving genes. J Mol Evol. 59(6):780–791.
Torruella G, et al. 2012. Phylogenetic relationships within the Opisthokonta based on phylogenomic analyses of conserved single-copy protein domains. Mol Biol Evol. 29(2):531–544.
Tsaousis AD, et al. 2008. A novel route for ATP acquisition by the remnant mitochondria of *Encephalitozoon cuniculi*. Nature 453(7194):553–556.
Vávra J, Lukeš J. 2013. Microsporidia and 'the art of living together'. Adv Parasitol. 82:253–319.
Vivier E. 1975. The microsporidia of the protozoa. Parazitologica 11:345–361.
Waller RF, et al. 2009. Evidence of a reduced and modified mitochondrial protein import apparatus in microsporidian mitosomes. Eukaryot Cell. 8(1):19–26.
Wilks HM, et al. 1988. A specific, highly active malate dehydrogenase by redesign of a lactate dehydrogenase framework. Science 242(4885):1541–1544.
Williams BAP, et al. 2010. A broad distribution of the alternative oxidase in microsporidian parasites. PLoS Pathog. 6(2):e1000761.
Williams BAP, Haferkamp I, Keeling PJ. 2008. An ADP/ATP-specific mitochondrial carrier protein in the microsporidian *Antonospora locustae*. J Mol Biol. 375(5):1249–1257.
Williams BAP, Hirt RP, Lucocq JM, Embley TM. 2002. A mitochondrial remnant in the microsporidian *Trachipleistophora hominis*. Nature 418(6900):865–869.
Wittner M, Weiss LM. eds. 1999. The microsporidia and microsporidiosis. Washington, D.C: American Society of Microbiology.
Xiang H, et al. 2010. A tandem duplication of manganese superoxide dismutase in *Nosema bombycis* and its evolutionary origins. J Mol Evol. 71(5–6):401–414.

Associate editor: Martin Embley