

REVIEW

Exploiting the architecture and the features of the microsporidian genomes to investigate diversity and impact of these parasites on ecosystems

E Peyretaille¹, D Boucher¹, N Parisot^{1,2}, C Gasc¹, R Butler³, J-F Pombert³, E Lerat⁴ and P Peyret¹

Fungal species play extremely important roles in ecosystems. Clustered at the base of the fungal kingdom are Microsporidia, a group of obligate intracellular eukaryotes infecting multiple animal lineages. Because of their large host spectrum and their implications in host population regulation, they influence food webs, and accordingly, ecosystem structure and function. Unfortunately, their ecological role is not well understood. Present also as highly resistant spores in the environment, their characterisation requires special attention. Different techniques based on direct isolation and/or molecular approaches can be considered to elucidate their role in the ecosystems, but integrating environmental and genomic data (for example, genome architecture, core genome, transcriptional and translational signals) is crucial to better understand the diversity and adaptive capacities of Microsporidia. Here, we review the current status of Microsporidia in trophic networks; the various genomics tools that could be used to ensure identification and evaluate diversity and abundance of these organisms; and how these tools could be used to explore the microsporidian life cycle in different environments. Our understanding of the evolution of these widespread parasites is currently impaired by limited sampling, and we have no doubt witnessed but a small subset of their diversity.

Heredity (2015) **114**, 441–449; doi:10.1038/hdy.2014.78; published online 3 September 2014

INTRODUCTION

Genome sequencing and phylogenetic inferences posit that the Microsporidian lineage is closely related to a recently erected, and highly diverse, phylum that sits at the base of the fungal tree, the Cryptomycota (James *et al.*, 2013). Members of this lineage have lost or simplified several typical eukaryotic features through a spectacular course of evolution. This adaptation has fuelled the diversification of a widespread lineage of obligate and often opportunistic intracellular parasites, which now comprise over 1500 described microsporidian species distributed amongst more than 187 different genera (Vavra and Lukes, 2013). Species from this group infect a range of eukaryotes from protists to most animal lineages (Lee *et al.*, 2008; Corradi and Selman, 2013; Vavra and Lukes, 2013) with increasingly evident medical, economic and environmental importance (Vavra and Lukes, 2013). Despite parasitising varied hosts, all microsporidian species present an original invasion strategy based on a highly specialised organelle: the polar tube. This represents one of the most sophisticated infection mechanisms among eukaryotes (Delbac and Polonais, 2008). After the germination process, the extruded polar tube penetrates the host's cell membrane to release an infective sporoplasma within the target cell (Peyretaille *et al.*, 2011).

The key to the success of Microsporidia lies in the flexibility of their transmission strategies and life cycles. Microsporidia exploit different strategies to maintain their presence within host populations, which

harness the host life cycles and make use of horizontal transmission (transfers between host individuals, including different species) and vertical inheritance (transmission of the parasite from parent to offspring(s)) (Dunn and Smith, 2001). In the latter process, some Microsporidia take advantage of the host reproduction system by distorting the sex ratio towards the transmitting sex (female), converting putative males into phenotypic functional females *via* a feminisation mechanism (Ironside *et al.*, 2003). They can also specifically eliminate males (male-killing process) during the embryonic stage, resulting in the production of all-female offspring (Dunn and Smith, 2001). A purely vertical transmission mode is believed to be accompanied by asexuality in some taxa (Dunn and Smith, 2001). However, some data suggest that sex may be present in numerous if not all microsporidian species (Ironside, 2013; Krebes *et al.*, 2014). The identification of gametes and the formation of diplokarya and mechanisms for haplosis *via* either meiosis or nuclear dissociation have been reported (Becnel *et al.*, 2005), as has the presence of diploid polymorphisms (Cuomo *et al.*, 2012; Selman *et al.*, 2013). Furthermore, a sex-related locus and some meiosis genes suggest that Microsporidia have a genetically controlled sexual cycle (Katinka *et al.*, 2001; Lee *et al.*, 2008). Thus, some members of this lineage are highly adapted to vertical transmission, whereas others may supplement this mode of transmission with a sexual phase on occasion (Krebes *et al.*, 2014).

¹Genomics, Clermont Université, Université d'Auvergne, EA 4678 CIDAM, Clermont-Ferrand, France; ²CNRS, UMR 6023, LMGE, Aubière, France; ³Illinois Institute of Technology, BCBS Biology Division, Chicago, IL, USA and ⁴Université de Lyon, Lyon, Université Lyon 1, CNRS, UMR 5558, Laboratoire de Biométrie et Biologie Évolutive, Villeurbanne, France

Correspondence: Dr E Peyretaille or P Peyret, Genomics, Clermont Université, Université d'Auvergne, EA 4678 CIDAM, 28 place Henri Dunant, Clermont-Ferrand F63001, France.

E-mail: eric.peyretaille@udamail.fr or pierre.peyret@udamail.fr

Received 31 March 2014; revised 16 July 2014; accepted 21 July 2014; published online 3 September 2014

Pathogens represent a large proportion of life on earth, but their ecological impact on ecosystems is not always appreciated (Marcogliese and Cone, 1997). Their role in population regulation is essential to maintaining ecosystem balance (Selakovic *et al.*, 2014). High-density species become the prime target for pathogens, leading to their population decline and maintenance of high biodiversity in the environment. Conversely, the combination of several parameters (such as competition, predation or pathogenicity) can lead to a loss of balance in the ecosystem, reducing the biodiversity. Furthermore, because Microsporidia are often assumed to be a negligible biomass in ecosystems because of their small size and their intracellular status, they are often omitted from studies of ecosystem trophic networks. Nevertheless, their presence makes more complex food webs by increasing the connectance (the proportion of realised links among the possible ones) in ecosystems (Lafferty *et al.*, 2006).

In summary, given their profound effects on trophic interactions through host population regulation, the overall function of the ecosystem may not be fully observed and understood unless Microsporidia are considered. To this end, metagenomics approaches (Handelsman *et al.*, 1998) are well adapted to deciphering the role of microsporidian species within trophic networks because of substantial cost reductions in sequencing technologies that have led to the rapid growth of this scientific field in recent years. The diversity and functionality of microorganisms can thus be revealed from complex environmental samples using the above-mentioned tools (Segata *et al.*, 2013), providing prime access to basic information about microbial traits such as physiology, epidemiology and evolutionary history that can be used to infer organism's ecological role in a given environment (Kim *et al.*, 2013).

MICROSPORIDIA IN TROPHIC NETWORKS

In pelagic aquatic environments, the flows of matter and energy are not organised only according to conventional linear trophic pathways based on photosynthetic assimilation where the primary producers

(phytoplankton) represent a food source for zooplankton which in turn act as an intermediate trophic link to the top of the food chain (fish, mammals). These flows can also pass through a microbial loop (composed of prokaryotes, fungi, heterotrophic and mixotrophic nanoflagellates, ciliates, and viruses) to form a true food web that is directly or indirectly connected to the classic food web (Figure 1). Almost half of the identified microsporidian species parasitise aquatic organisms, such as fish, arthropods, non-arthropod invertebrates and protists (Stentiford *et al.*, 2013; Vavra and Lukes, 2013), affecting the food web at various trophic levels (Figure 1). Specifically, the relationship between the phytoplankton, the prokaryotic biomasses and the zooplankton grazers can be modified by the presence of Microsporidia. Through the manipulation of their hosts, they alter the structure, stability and function of the entire ecosystem. For metazooplanktonic populations (for example, *Daphnia*), infection by Microsporidia leads to a reduced fecundity and longevity. An opaque appearance of tissues also makes them visually attractive to predators, such as zooplanktivorous fishes (Vizoso *et al.*, 2005). Because Microsporidia can exert a strong control over the host zooplanktonic population by decreasing its abundance, it affects the primary producer population by limiting grazing by zooplanktonic populations (Figure 1).

In some cases, infections from Microsporidia result in very large cysts packed with host spores and meronts, so it is possible that these could serve as an additional source of nutrients for the predators that inadvertently consume them. It is also noteworthy that, if the parasite is characterised by a sufficiently large host spectrum, the ingested Microsporidia may also impact the behaviour and the fitness of the predator population. In terrestrial ecosystems, the lifespan and egg production of the spined soldier bug (*Podisus maculiventris*) is significantly reduced after ingestion of prey infected with the Microsporidium *Vairimorpha necatrix* (Down *et al.*, 2004). Microsporidia infections similarly affect intraguild predation, where predation was observed between potential competitors, that is, species

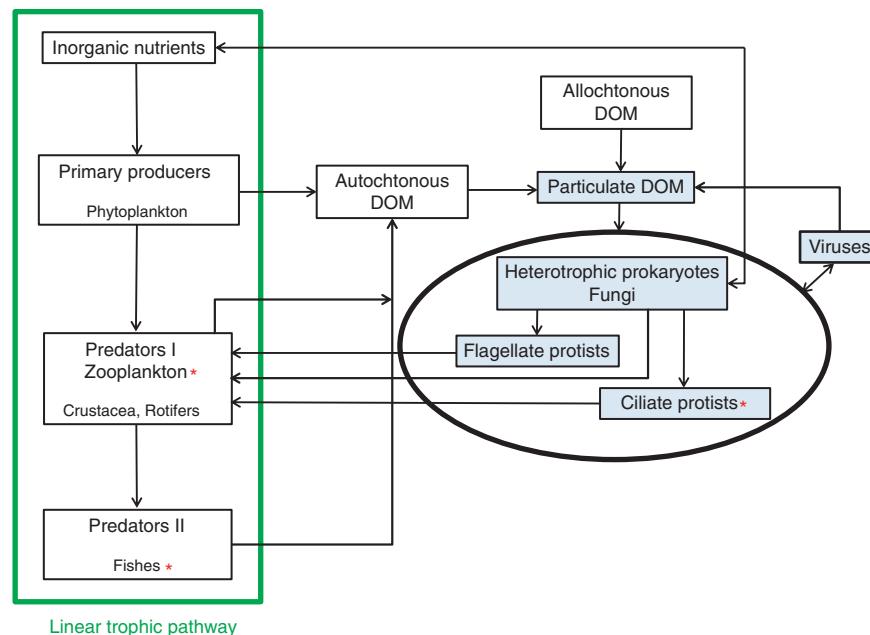


Figure 1 Simplified schematic representation of trophic networks integrating the microbial loop in pelagic aquatic environments (modified from (Amblard *et al.*, 1998)). All components of the microbial loop are shown in grey. The red asterisks indicate known organisms that are susceptible to Microsporidia parasitism. DOM, dissolved organic matter.

exploiting the same resources in similar ways (Polis *et al.*, 1989). The influence of microsporidian parasitism on predation hierarchy is also evident in freshwater environments. For example, microsporidian infection in the native amphipod *Gammarus duebeni celticus* results in an altered mobility because of abdominal musculature degeneration (MacNeil *et al.*, 2003b), which ultimately results in a reduced capacity to prey on smaller species (*G. tigrinus*, *Crangonyx pseudogracilis*) and an increased vulnerability to predation by the largest invading species (*G. pulex*) (MacNeil *et al.*, 2003a). Clearly, Microsporidia play hidden but crucial roles in predator–prey associations and competitor interactions.

Biodiversity is critical for an ecosystem, fuelling competition and natural selection, while reducing disease risk and vulnerability to pathogens. Losses in biodiversity can increase sensitivity to pathogens such as Microsporidia, which play a role in the success of biological invasions of new species threatening then ecosystem's functions. A host that is tolerant to a parasite can increase in relative abundance, affecting the populations of less tolerant species. For instance, microsporidium (*Nosema* spp) infections give the invasive harlequin ladybird *Harmonia axyridis* a competitive advantage (Vilcinskas *et al.*, 2013). In the North-American and European native ladybirds (*Coccinella septempunctata*, *Adalia bipunctata*), Microsporidia infection is lethal, but *H. axyridis* mounts an immune response, secreting large amounts of harmonin. This antimicrobial alkaloid may not only result in a greater resistance of *H. axyridis* to certain diseases but also act as a poison to native ladybirds that consume invasive ladybird eggs (Sloggett and Davis, 2010). Interestingly, such cases seem to reflect mutualism rather than parasitism as the relationship appears to be beneficial to both partners. Specifically, Microsporidia exploit host energetic resources for their growth and survival while the invasive harlequin ladybirds proliferate to new habitats following the activation of immune defences by the Microsporidia infection. Host *H. axyridis* has then adapted to resist Microsporidia, unlike other ladybird species. Other cases where Microsporidia may be redefined as mutualistic organisms include the amphipod *G. roeseli*, where infection by the microsporidian species *N. granulosis* results in increased survival (Haine *et al.*, 2007).

Life history traits are not only genotype-specific but also environmentally mediated. Environmental parameters may affect the coevolution of Microsporidia and their hosts. Environmental variation may influence the host population density, altering the transmission efficiency of the parasite. Thus, it has been suggested that large aggregations of harlequin ladybirds in winter can facilitate the transfer of *Nosema* spp. between individuals (Reynolds, 2013).

Parasitism can be particularly important when the hosts are keystone species with crucial functions in the environment, as it is the case of the honey bee (*Apis mellifera*) and the bumble bee (*Bombus terrestris*) pollinator species. Indeed, fecundity of these pollinators is affected by various *Nosema* species (*N. apis*, *N. ceranae* and *N. bombi*). Infection by Microsporidia also modifies honey bee behaviours, such as earlier and more intense foraging (Lin *et al.*, 2009) and higher sucrose consumption (Alaux *et al.*, 2010). These host–parasite interactions, combined with environmental parameters may have potentially contributed to the recent and steep global decline in populations of honey bee (Bromenshenk *et al.*, 2010) and bumble bees (Meeus *et al.*, 2011). The heterogeneity of the ecosystems, due in part to intensification of agriculture and landscape alterations, affects foraging areas by creating different floral resources. Consequently, different pollen diets (low in pollen or high in monofloral pollens) lead to a reduced lifespan of the honey bees parasitised by *N. ceranae* (Porrini *et al.*, 2011; Di Pasquale *et al.*, 2013). Similarly, pollutants

(for example, pesticides) may act synergistically with *N. ceranae* infections to further increase the mortality of *A. mellifera* populations worldwide (Vidau *et al.*, 2011; Pettis *et al.*, 2012). The diminution of this microsporidian host in turn inhibits lower trophic levels, namely the plants that have to be pollinated. Consequently, Microsporidia can be associated with biodiversity losses in both pollinator host populations and pollinated plants. It is clear that environmental factors, which can fluctuate through time and space, play an essential role in host–parasite interactions and coevolution. Such factors have to be taken into account in investigations of the exact role of Microsporidia in ecosystems.

Microsporidia, like pathogens in general, play a tremendous role in ecosystems worldwide, at all levels, and for this reason, they should be better integrated into studies aimed at understanding processes involved in the ecological networks. Hopefully, the recent acquisition of large scale genome data from several microsporidian species, coupled with advances in the field of metagenomics, will help reveal the true extent of microsporidian diversity in the field and the extent of their functionality in terrestrial and aquatic ecosystems.

STRUCTURE AND CONTENT OF MICROSPORIDIAN GENOMES—THE KEY FOR ENVIRONMENTAL SUCCESS OF MICROSPORIDIA

Both molecular karyotype analyses and genome sequencing of several microsporidian species indicate that their nuclear genomes range in size from only 2.3 Mbp for the human pathogen *Encephalitozoon intestinalis* (Corradi *et al.*, 2010) to 24 Mbp for the *Daphnia magna* pathogen *Hamiltosporidium tvaermanniensis*, formerly named *Octosporea bayeri* (Corradi *et al.*, 2009). Figure 2 describes the general features of sequenced microsporidian genomes and their phylogenetic relationships.

POTENTIAL IMPACTS OF TRANSPOSSABLE ELEMENTS ON MICROSPORIDIAN GENOME STRUCTURE

Transposable elements (TEs) and repeated sequences contribute to genome plasticity by enabling large-scale chromosomal rearrangements and by acting as buffer regions where transpositions and recombination events can occur without disrupting vital functions (Kidwell and Lisch, 2000; Biemont, 2009). In genomes without intergenic regions, paralogs and TEs, opportunities for chromosomal shuffling are limited. The smallest microsporidian genomes are found in the Encephalitozoonidae. The genomes of the *Encephalitozoon* spp. *E. cuniculi* (2.9 Mbp; (Katinka *et al.*, 2001)), *E. intestinalis*, (2.3 Mbp; (Corradi *et al.*, 2010)), *E. hellem* and *E. romaleae* (2.5 Mbp each; (Pombert *et al.*, 2012)) show an almost perfect content and synteny conservation (Peyret *et al.*, 2001; Corradi *et al.*, 2010; Pombert *et al.*, 2012). Their cores are extremely compact, with median intergenic lengths ranging from 75 to 82 bp (unpublished data, Figure 3a) and a few overlapping Coding DNA Sequences (CDSs) in head-to-head, tail-to-head or tail-to-tail configuration (Peyrettaillade *et al.*, 2009). This dense arrangement means few intergenic regions are available for potential recombination events. Furthermore, *Encephalitozoon* genomes lack simple sequence repeats, minisatellite arrays and TEs, features that all result from their extreme reductive evolution. The absence of RNA interference machinery in these genomes, whose function is involved in the repression of TE expression (Chung *et al.*, 2008), could in part explain their lack of TEs. Indeed, an asexual genome that would not be able to control the TE activity would be submitted to too many deleterious TE insertions, which would irremediably decrease its fitness (Dolgin and Charlesworth, 2006; Kraaijeveld *et al.*, 2012). Alternatively, the elimination of the RNA

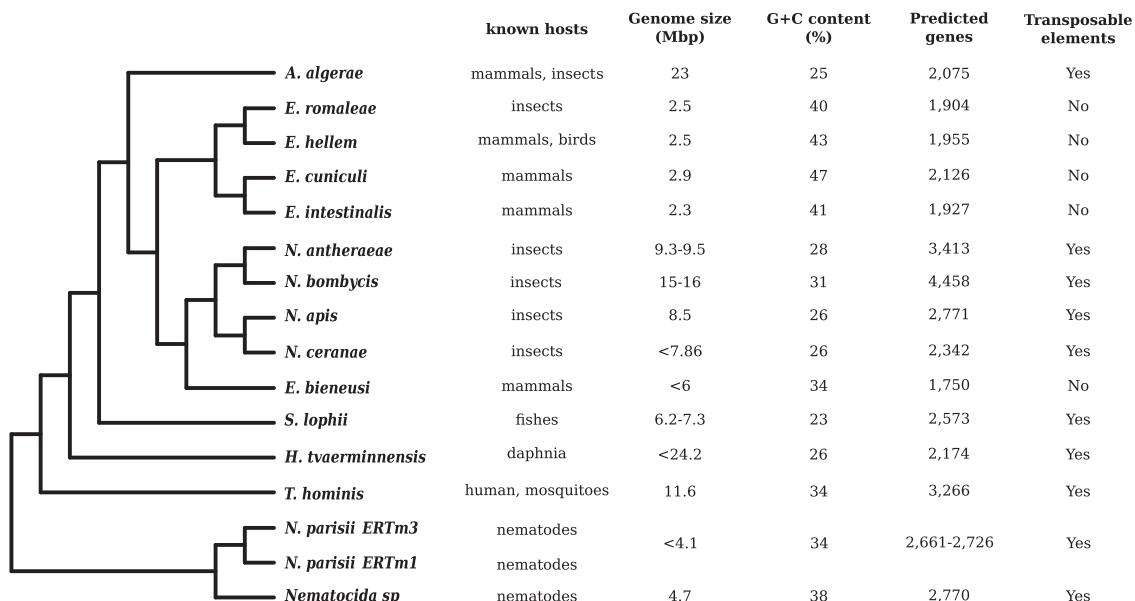


Figure 2 General features and phylogenetic relationships of sequenced microsporidian genomes. Phylogenetic relationships were established based on several analyses (Wang et al., 2006; Corradi et al., 2009; Cuomo et al., 2012; Suankratay et al., 2012).

interference pathway could instead have occurred subsequently to the loss of TEs, following which its machinery would have become superfluous.

In contrast, intermediate-sized microsporidian genomes display large variations in the numbers and types of TEs that they contain. Only a few TEs were initially reported in *N. apis* (Chen et al., 2013) whereas the genomes of *N. ceranae*, *N. antheraeae* and *N. bombycis* were shown to display around 1%, 6% and 20% of TEs, respectively (Pan et al., 2013). However, BLASTP (Altschul et al., 1997) and TransposonPSI (<http://transposonpsi.sourceforge.net/>) analyses allowed us to determine that many TEs were wrongly annotated as host genes in the four species (unpublished data), such that the real TE content is higher in these genomes and, respectively, closer to 2%, 7%, 10% and 25% in *N. apis*, *N. ceranae*, *N. antheraeae* and *N. bombycis*. In any case, the differences in TE content between *N. bombycis* and the other *Nosema* species are substantial and correlated with genome size variations, a trend also observed between *Nematocida* species (Cuomo et al., 2012). The unusually large expansion in *N. bombycis* has been attributed to the propagation of TEs, to the acquisition of genes by horizontal transfer and to large-scale and small-scale gene duplication events (Pan et al., 2013). All of these species feature LTR retrotransposons but differ in non-LTR retrotransposon, DNA transposon and *helitron* contents (Figure 3b; (Cornman et al., 2009; Peyretailade et al., 2012; Chen et al., 2013; Pan et al., 2013)). For example, *N. apis* lacks *Merlin* and *Mariner* DNA transposon families whereas *N. ceranae* lacks *piggyBac* DNA transposons. The distribution of TEs is not uniform in the *N. ceranae* genome and core genes are often clustered in islands that are either free of or rarely interrupted by TEs (Figure 3a). Although the intra-generic synteny observed between genomes from the *Nosema* and *Nematocida* lineages is relatively high, it is lower than that observed within the Encephalitozoonidae, which suggests that the presence of TEs could be responsible for numerous rearrangements.

Assessing the chromosomal architecture and TE content in large microsporidian genomes is currently limited by the dearth of sequenced representatives. None of the lineages investigated so far

featured more than one species per genus (this should be rectified soon with the incoming *Anncalia algerae* genomes) and those that were investigated were carried out at the draft level, focusing on core genes rather than structure. Both the *Hamiltosporidium tvaermennensis* (Corradi et al., 2009) and *A. algerae* (Peyretailade et al., 2012) drafts, respectively, covered 13.3 and 13.8 Mbp of their estimated ~24 Mbp genome size and corresponded to numerous small-to-mid-size contigs, whose assemblies were hindered by the presence of repeated elements. Indeed, with an estimated 34.2–37.2 and 4.7 coverage of *H. tvaermennensis* and *A. algerae* genomes, respectively, more than 50% of these genomes may correspond to TEs or repeated sequences artefactually assembled. This is especially true for *H. tvaermennensis*, whose draft was performed in the early stages of Illumina sequencing with 35-bp long short reads. TEs are found in both the *H. tvaermennensis* and *A. algerae* genomes (Figure 3b), with at least one-third of the *A. algerae* predicted CDSs found on contigs that contain one or more TE (Figure 3a, unpublished results), and we infer that more will be found in the missing regions upon further completion of the drafts.

Microsporidian genomes containing TEs are expected to be flexible and may thus be able to adapt faster to varying conditions. This in turn could occasionally yield genotypes that can jump to new host populations or even to new host species, as suggested for filamentous plant pathogens (Raffaele and Kamoun, 2012). Furthermore, recent study also suggests that *A. algerae* could use a TE as a lure strategy to escape the host innate immune system (Panek et al., 2014). Therefore, TEs may play an important role not solely in modulating the microsporidian genome architecture but also in the capacity of these parasites to infest multiple hosts and especially to adapt to new hosts.

GENE CONTENT AND COMMON TRANSCRIPTIONAL AND TRANSLATIONAL MICROSPORIDIAN FEATURES

Microsporidian species are minimalist parasites with an extremely reduced metabolic potential and fewer or derived cellular components compared with free-living fungi. They do not possess peroxisomes, their mitochondria have been reduced to mitosomes that are

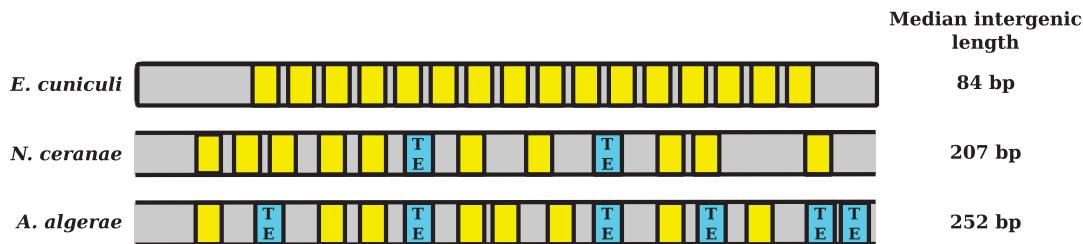
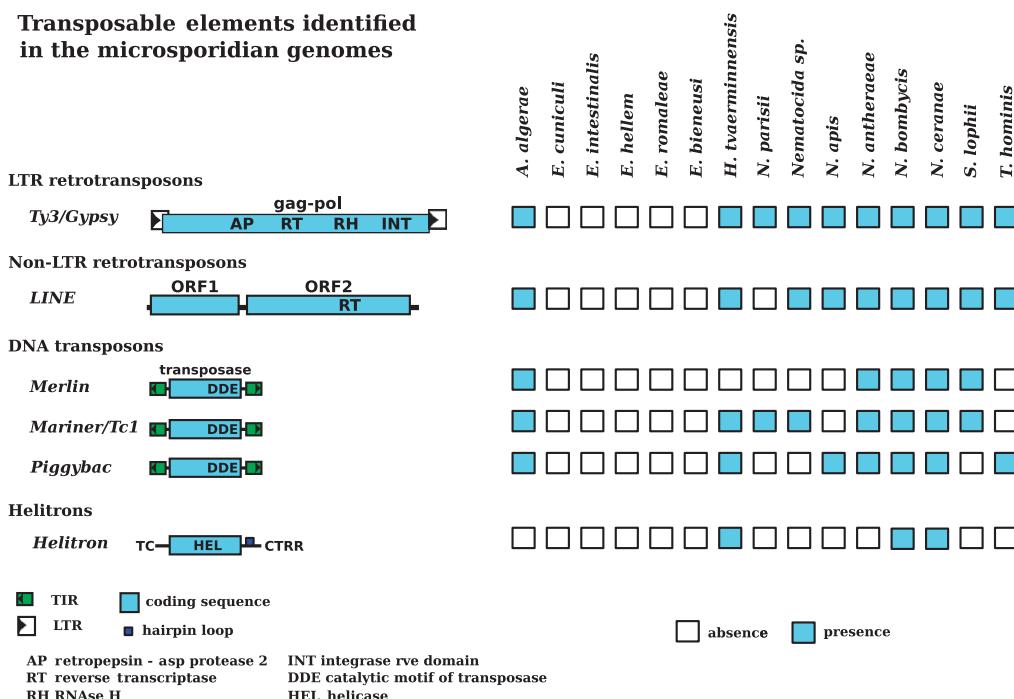
a Microsporidian genome architectures**b Transposable elements identified in the microsporidian genomes**

Figure 3 Impact of TEs on microsporidian genome architectures. (a) Schematic representation of three genome structures observed in the microsporidian species *Encephalitozoon cuniculi*, *Nosema ceranae* and *Anncalia algerae*. Genes are indicated in yellow boxes, and TEs are indicated in blue boxes. (b) Overview of TEs described in the microsporidian genomes. Information was obtained from both the literature and personal analysis with TransposonPSI (<http://transposonpsi.sourceforge.net/>).

incapable of oxygenic respiration and their Golgi apparatus is atypical and unstacked (Vavra and Lukes, 2013), the sum of which makes them highly host-dependent (Corradi and Selman, 2013; Corradi and Slamovits, 2011). Despite a 10-fold variation in genome size, Microsporidia seem to have conserved a similar gene repertoire ranging from ca. 1800 to 2600 protein coding genes. This array contains an irreducible set of core genes preserved throughout their evolution (Corradi and Slamovits, 2011; Vavra and Lukes, 2013) and variant subsets likely reflecting their adaptation to different hosts and environments (Vavra and Lukes, 2013). For instance, comparative genomics of four microsporidian species highlight a core genome of 932 CDSs (Peyretailade et al., 2012), from which 141 appear to be found exclusively in the microsporidium phylum (for example, polar tube proteins involved in the infestation process). Comparative analysis of 11 genomes also allowed to define a microsporidian-specific core including especially 32 gene families specific to Microsporidia phylum (Nakjang et al., 2013). However, the high sequence divergence of microsporidian genes makes comparison with other organisms difficult. This high divergence also complicates the identification of microsporidian sequences from complex environments

using only sequence similarity and base composition, leading to undetected contamination (Heinz et al., 2012; Peyretailade et al., 2012; Vavra and Lukes, 2013).

The number of predicted genes may also drastically fluctuate depending on the inclusion of small CDSs (<= 300 bp) defined by *ab initio* prediction software, and software-induced over-predictions produced a number of false positives in microsporidian annotations (Cheng et al., 2011). To ensure identification of such small genes, a re-evaluation of previous annotations of *Encephalitozoon* genomes performed using upstream transcriptional signals and synteny resulted in better delineated intergenic regions. Following these analyses, we propose that *E. cuniculi*, *E. intestinalis*, *E. romaleae* and *E. hellem* contain 2126, 1927, 1904 and 1955 CDSs, respectively (unpublished data). For these *Encephalitozoon* species, the mean CDS length is 991, 999, 996 and 1004 bp, respectively after reanalysis (unpublished data). Moreover, taking into account the new predictions, between 8 and 11% of the proteins are encoded by small CDSs for these four species (unpublished data). A high gene number has been identified for *N. antheraeae*, *N. bombycis* and *T. hominis* with mean CDS length 775, 741 and 871 bp, respectively (Heinz et al., 2012; Pan et al., 2013)

indicating a potential over-prediction of small genes. For example in *T. hominis*, approximately 23% (736) of the predicted genes encode proteins smaller than 100 amino acids. Protein homology has only been established for 147 of these 736 CDSs and many of the remaining CDSs may correspond to falsely predicted genes.

Genome sequencing has also revealed a variable proportion of duplicate genes from different species (Nassonova *et al.*, 2005; Akiyoshi *et al.*, 2009; Pan *et al.*, 2013). Nevertheless, in some cases, such gene duplication may in fact be related to polyploidy of the microsporidian genome as suggested for *N. ceranae* and *A. algerae* (Peyretailade *et al.*, 2012; Roudel *et al.*, 2013).

In microsporidian genomes, transcriptional and translational regulation signals are highly conserved. Regardless of the base composition, CCC-like or GGG-like signals located in close proximity to the start codon appear to play a key role in transcription initiation (Cornman *et al.*, 2009; Peyretailade *et al.*, 2009; Peyretailade *et al.*, 2012; Chen *et al.*, 2013) (Supplementary Figure S1). 5' RACE-PCR and RNAseq show that 5'UTRs may be highly reduced and in some cases absent in Microsporidia (Williams *et al.*, 2005; Corradi *et al.*, 2008; Peyretailade *et al.*, 2009; Gill *et al.*, 2010; Heinz *et al.*, 2012). Even in these cases, CCC-like or GGG-like signals are located upstream of the translational start site. In various genomes with a low G + C content, 5'UTRs are absent and the base composition upstream of the initiation start codon is strongly A + T-biased (that is, over 80%) (Peyretailade *et al.*, 2012). The analysis of highly expressed genes, such as those encoding ribosomal proteins, highlights a complementary AAAATTTT-like signal in close proximity to the CCC-like or GGG-like signals which seem universal in the microsporidian phylum (Peyretailade *et al.*, 2009; Peyretailade *et al.*, 2011; Peyretailade *et al.*, 2012) (Supplementary Figure S1). For the 3'-end mRNA processing, the consensus sequence AAUAAA or some other slightly degenerate signal (that is, only differing by a single nucleotide) is near the stop codon for all annotated genes in *E. cuniculi*, *E. intestinalis*, *N. ceranae* and *E. bieneusi* species (Peyretailade *et al.*, 2009; Peyretailade *et al.*, 2012). Finally, concerning the translational signals, no bias is detected upstream of the AUG codon, which is consistent with the existence of a highly reduced 5'UTR. Nevertheless, a strong bias towards adenine or guanine (~78%) in the +4 position and a significant bias for adenine (~45%) in the +5 position have been identified in these species (Peyretailade *et al.*, 2009).

GENOMIC FEATURES TO EXPLORE THE MICROSPORIDIAN LIFE CYCLE

Although Microsporidia are present in many environments, the first difficulty in exploiting genomic data from these species is to isolate their cells. For this, cell culture and single-cell approaches can be considered. Nevertheless, such techniques require to know culture conditions especially host cells and to be able to discriminate the microsporidian cells from other microbial populations (Figure 4). To study microbial diversity from complex environments, direct DNA extraction and sequencing remain the most promising and easy-to-use strategies to implement. However, current DNA procedures are not adapted to the different microsporidian forms, which are present as highly resistant spores in outside environments or as fragile meronts when inside their hosts (Vavra and Lukes, 2013). The release of DNA from the spores is difficult because of insufficient current lysis approaches. The methods applied to the infected host organisms to recover parasite DNA involve heating steps (Izquierdo *et al.*, 2011; Wang *et al.*, 2013; Sapir *et al.*, 2014) that may cause damage to A + T-rich microsporidian genomes (Figure 2).

Microsporidian taxa detection is also limited by the molecular identification methods used, that is, the sequencing of PCR products targeting small subunit ribosomal (rRNA) genes. This leads to an underestimation of their presence and diversity in different environments, because microsporidian rRNA genes are shorter than those of other eukaryotic organisms and feature extremely reduced variable regions (Peyretailade *et al.*, 1998). Thus, microsporidian diversity is investigated with available non-degenerate specific primers that target only few organisms (Fournier *et al.*, 2002; Ardila-Garcia and Fast, 2012; Ardila-Garcia *et al.*, 2013; Wang *et al.*, 2013; Sapir *et al.*, 2014) or a weakly degenerate primer set designed from the few known Microsporidia with available sequences in public databases (Vossbrinck *et al.*, 2004). Several studies show the presence of some known human microsporidian species in various environments (Fournier *et al.*, 2000; Kahler and Thurston-Enriquez, 2007; Izquierdo *et al.*, 2011). However, to our knowledge, only two studies highlight the microsporidian diversity in different ecosystems: soil, sand and compost (Ardila-Garcia *et al.*, 2013) or water from a mosquito larval habitat (Avery and Undeen, 1987). Thus, to unveil the microsporidian community structure as defined in microbial ecology (that is, richness, diversity and composition), new analytical tools must be developed. The definition of explorative primers to target either phylogenetic markers (small subunit rRNA) or functional genes (Delbac and Polonais, 2008; Dugat-Bony *et al.*, 2012; Peyretailade *et al.*, 2012) can be accomplished with recently developed design software such as HiSpOD (Dugat-Bony *et al.*, 2011) or KASpOD (Parisot *et al.*, 2012), and will be crucial for identifying uncharacterised microsporidian sequences and anticipating genetic variations (Figure 4).

Advances in high-throughput sequencing have led to a rapid growth of metagenomics studies to monitor complex ecosystems (Segata *et al.*, 2013). Binning, which represents the taxonomic classification of reads, is the most basic step in the characterisation of microbial communities from metagenomic samples. Classification can be performed either by *de novo* binning using intrinsic sequence features or by using extrinsic information provided in sequenced microbial genome databases (Segata *et al.*, 2013). Therefore, the identification of microsporidian sequences relies on a detailed knowledge of the genomic characteristics of these pathogens including specific gene repertoires and the particular organisation of their genomes (Figure 4). The unique transcriptional and translational regulation signals described herein could also confirm the presence of Microsporidia. Finally, identification of TEs specific to microsporidian species (unpublished data) provides sequence information to validate the presence of these organisms in complex environments. Encompassing all this genomic information into a comprehensive database dedicated to Microsporidia could improve both extrinsic and intrinsic approaches. Nevertheless, some horizontal gene transfers from different eukaryotic and prokaryotic donors can interfere with identification and should then be taken into account (Cuomo *et al.*, 2012; Heinz *et al.*, 2012; Pombert *et al.*, 2012).

By including metatranscriptomic data, a peculiar feature observed in some microsporidian species can also be utilised. For some species, the most gene-dense regions have increased levels of 'overlapping transcription'. This phenomenon is an atypical transcriptional process that produces mRNA encompassing multiple CDSs from different strands (Williams *et al.*, 2005; Corradi *et al.*, 2008; Peyretailade *et al.*, 2009). The presence of such mRNAs coupled with previously described genomic features can strengthen the taxonomic assignment to the Microsporidia phylum (Figure 4).

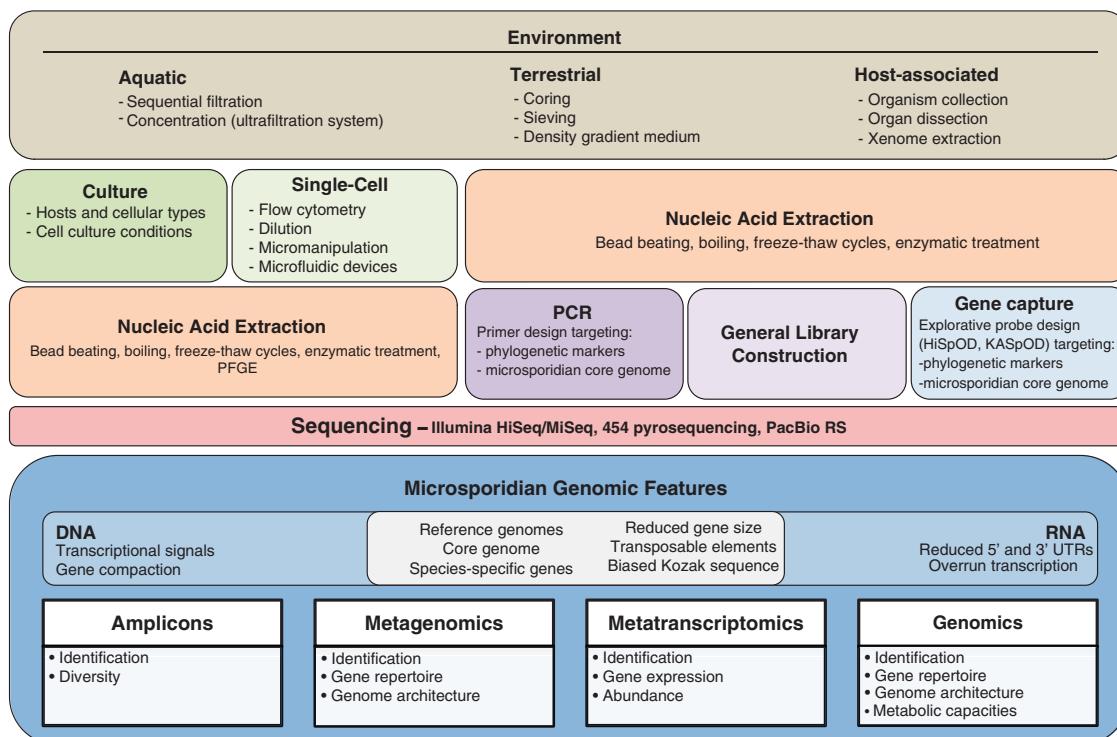


Figure 4 Schematic representation of the major approaches to appraise Microsporidia diversity in complex environments using microsporidian genomic features.

'Omics' studies require, however, a very large effort in both sequencing and analyses to assess the microsporidian species. Interesting alternatives rely on the reduction of sample complexity through methods such as single-cell genomics (Stepanauskas, 2012) or gene capture to specifically enrich for the nucleic acid sequences of interest. Gene capture targeting specific microsporidian genes can then be coupled to high-throughput sequencing for accurate and efficient identification of species (Denonfoux *et al.*, 2013) (Figure 4). Using these new tools will highlight the high diversity and distribution of environmental Microsporidia and will increase the understanding of host interactions and the roles of these organisms in the functioning of ecosystems.

DATA ARCHIVING

There were no data to deposit.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

NP and CG were funded by the French 'Direction Générale de l'Armement' (DGA). We thank Nicolas Corradi for providing comments on an earlier version of this manuscript.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W *et al.* (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.

Amblard C, Boisson JC, Bourdier G, Fontvieille D, Gayte X, Sime-Ngando T (1998). Ecologie microbienne en milieu aquatique: des virus aux protozoaires. *Rev.Sc.Eau, N° Spécial: Les Sciences de l'Eau: Bilan et perspectives* 145–162.

Ardila-Garcia AM, Fast NM (2012). Microsporidian infection in a free-living marine nematode. *Eukaryot Cell* **11**: 1544–1551.

Ardila-Garcia AM, Raghuram N, Sihota P, Fast NM (2013). Microsporidian diversity in soil, sand, and compost of the Pacific Northwest. *J Eukaryot Microbiol* **60**: 601–608.

Avery SW, Undeen AH (1987). The isolation of microsporidia and other pathogens from concentrated ditch water. *J Am Mosq Control Assoc* **3**: 54–58.

Becnel JJ, White SE, Shapiro AM (2005). Review of microsporidia-mosquito relationships: from the simple to the complex. *Folia Parasitol (Praha)* **52**: 41–50.

Biemont C (2009). Are transposable elements simply silenced or are they under house arrest? *Trends Genet* **25**: 333–334.

Bromenshenk JJ, Henderson CB, Wick CH, Stanford MF, Zulich AW, Jabbour RE *et al.* (2010). Iridovirus and microsporidian linked to honey bee colony decline. *PLoS ONE* **5**: e13181.

Chen Y, Pettis JS, Zhao Y, Liu X, Tallon LJ, Sadzewicz LD *et al.* (2013). Genome sequencing and comparative genomics of honey bee microsporidia, *Nosema apis* reveal novel insights into host-parasite interactions. *BMC Genomics* **14**: 451.

Cheng H, Chan WS, Li Z, Wang D, Liu S, Zhou Y (2011). Small open reading frames: current prediction techniques and future prospect. *Curr Protein Pept Sci* **12**: 503–507.

Chung WJ, Okamura K, Martin R, Lai EC (2008). Endogenous RNA interference provides a somatic defense against *Drosophila* transposons. *Curr Biol* **18**: 795–802.

Cornman RS, Chen YP, Schatz MC, Street C, Zhao Y, Desany B *et al.* (2009). Genomic analyses of the microsporidian *Nosema ceranae*, an emergent pathogen of honey bees. *PLoS Pathog* **5**: e1000466.

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.

Corradi N, Haag KL, Pombert JF, Ebert D, Keeling PJ (2009). Draft genome sequence of the *Daphnia* pathogen *Octosporea bayeri*: insights into the gene content of a large microsporidian genome and a model for host-parasite interactions. *Genome Biol* **10**: R106.

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**: 77.

Corradi N, Selman M (2013). Latest progress in microsporidian genome research. *J Eukaryot Microbiol* **60**: 309–312.

Akiyoshi DE, Morrison HG, Lei S, Feng X, Zhang Q, Corradi N *et al.* (2009). Genomic survey of the non-cultivable opportunistic human pathogen, *Enterocytozoon bieneusi*. *PLoS Pathog* **5**: e1000261.

Alaux C, Brunet JL, Dussaubat C, Mondet F, Tchamitchian S, Cousin M *et al.* (2010). Interactions between *Nosema* microspores and a neonicotinoid weaken honeybees (*Apis mellifera*). *Environ Microbiol* **12**: 774–782.

- Corradi N, Slamovits CH (2011). The intriguing nature of microsporidian genomes. *Brief Funct Genomics* **10**: 115–124.
- Cuomo CA, Desjardins CA, Bakowski MA, Goldberg J, Ma AT, Becnel JJ et al. (2012). Microsporidian genome analysis reveals evolutionary strategies for obligate intracellular growth. *Genome Res* **22**: 2478–2488.
- Delbac F, Polonais V (2008). The microsporidian polar tube and its role in invasion. *Subcell Biochem* **47**: 208–220.
- Denonfoux J, Parisot N, Dugat-Bony E, Biderre-Petit C, Boucher D, Morgavi DP et al. (2013). Gene capture coupled to high-throughput sequencing as a strategy for targeted metagenome exploration. *DNA Res* **20**: 185–196.
- Di Pasquale G, Salignon M, Le Conte Y, Belzunce LP, Decourtey A, Kretzschmar A et al. (2013). Influence of pollen nutrition on honey bee health: do pollen quality and diversity matter? *PLoS One* **8**: e72016.
- Dolgin ES, Charlesworth B (2006). The fate of transposable elements in asexual populations. *Genetics* **174**: 817–827.
- Down RE, Bell HA, Matthews HJ, Kirkbride-Smith AE, Edwards JP (2004). Dissemination of the biocontrol agent *Vairimorpha necatrix* by the spined soldier bug, *Podisus maculiventris*. *Entomologia Experimentalis et Applicata* **110**: 103–114.
- Dugat-Bony E, Missooui M, Peyretaille E, Biderre-Petit C, Bouzid O, Guinraud C et al. (2011). HiSpOD: probe design for functional DNA microarrays. *Bioinformatics* **27**: 641–648.
- Dugat-Bony E, Peyretaille E, Parisot N, Biderre-Petit C, Jaziri F, Hill D et al. (2012). Detecting unknown sequences with DNA microarrays: explorative probe design strategies. *Environ Microbiol* **14**: 356–371.
- Dunn AM, Smith JE (2001). Microsporidian life cycles and diversity: the relationship between virulence and transmission. *Microbes Infect* **3**: 381–388.
- Fournier S, Dubrou S, Liguyr O, Gaussen F, Santillana-Hayat M, Sarfati C et al. (2002). Detection of Microsporidia, cryptosporidia and *Giardia* in swimming pools: a one-year prospective study. *FEMS Immunol Med Microbiol* **33**: 209–213.
- Fournier S, Liguyr O, Santillana-Hayat M, Guillot E, Sarfati C, Dumoutier N et al. (2000). Detection of microsporidia in surface water: a one-year follow-up study. *FEMS Immunol Med Microbiol* **29**: 95–100.
- Gill EE, Lee RC, Corradi N, Grisdale CJ, Limpright VO, Keeling PJ et al. (2010). Splicing and transcription differ between spore and intracellular life stages in the parasitic microsporidia. *Mol Biol Evol* **27**: 1579–1584.
- Haine ER, Motreuil S, Rigaud T (2007). Infection by a vertically-transmitted microsporidian parasite is associated with a female-biased sex ratio and survival advantage in the amphipod *Gammarus roeseli*. *Parasitology* **134**: Pt 10 1363–1367.
- Handelman J, Rondon MR, Brady SF, Clardy J, Goodman RM (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* **5**: R245–R249.
- Heinz E, Williams TA, Nakjang S, Noel CJ, Swan DC, Goldberg AV 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.
- Ironside JE (2013). Diversity and recombination of dispersed ribosomal DNA and protein coding genes in microsporidia. *PLoS ONE* **8**: e55878.
- Ironside JE, Smith JE, Hatcher MJ, Sharpe RG, Rollinson D, Dunn AM (2003). Two species of feminizing microsporidian parasite coexist in populations of *Gammarus duebeni*. *J Evol Biol* **16**: 467–473.
- Izquierdo F, Castro Hermida JA, Fenoy S, Mezo M, Gonzalez-Warleta M, del Aguila C (2011). Detection of microsporidia in drinking water, wastewater and recreational rivers. *Water Res* **45**: 4837–4843.
- James TY, Pelin A, Bonen L, Ahrendt S, Sain D, Corradi N et al. (2013). Shared signatures of parasitism and phylogenomics unite cryptomycota and microsporidia. *Curr Biol* **23**: 1548–1553.
- Kahler AM, Thurston-Enriquez JA (2007). Human pathogenic microsporidia detection in agricultural samples: method development and assessment. *Parasitol Res* **100**: 529–538.
- Katinka MD, Duprat S, Cornillot E, Metenier G, Thomarat F, Prensier G et al. (2001). Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature* **414**: 450–453.
- Kidwell MG, Lisch DR (2000). Transposable elements and host genome evolution. *Trends Ecol Evol* **15**: 95–99.
- Kim M, Lee KH, Yoon SW, Kim BS, Chun J, Yi H (2013). Analytical tools and databases for metagenomics in the next-generation sequencing era. *Genomics Inform* **11**: 102–113.
- Kraaijeveld K, Zwanenburg B, Hubert B, Vieira C, De Pater S, Van Alphen JJ et al. (2012). Transposon proliferation in an asexual parasitoid. *Mol Ecol* **21**: 3898–3906.
- Krebse L, Zeidler L, Frankowski J, Bastrop R (2014). (Cryptic) sex in the microsporidian *Nosema granulosis*—Evidence from parasite rDNA and host mitochondrial DNA. *Infect Genet Evol* **21**: 259–268.
- Lafferty KD, Dobson AP, Kuris AM (2006). Parasites dominate food web links. *Proc Natl Acad Sci USA* **103**: 11211–11216.
- Lee SC, Corradi N, Byrnes EJ 3rd, Torres-Martinez S, Dietrich FS, Keeling PJ et al. (2008). Microsporidia evolved from ancestral sexual fungi. *Curr Biol* **18**: 1675–1679.
- Lin H, Sullivan J, Huang ZY (2009). Mechanisms through which *Nosema apis* affects onset of foraging in worker honeybees (*Apis mellifera* L.). In Proceedings Workshop ‘*Nosema* disease: lack of knowledge and work standardization’ (COST Action FA0803) Guadalajara, 19–22 October 2000 [WWW document]. URL <http://www.coloss.org/news/nosema-workshop-proceedings-online>.
- MacNeil C, Dick JT, Hatcher MJ, Fielding NJ, Hume KD, Dunn AM (2003a). Parasite transmission and cannibalism in an amphipod (Crustacea). *Int J Parasitol* **33**: 795–798.
- MacNeil C, Dick JTA, Hatcher MJ, Dunn AM (2003b). Differential drift and parasitism in invading and native *Gammarus* spp. (Crustacea: Amphipoda). *Ecography* **26**: 467–473.
- Marcogliese DJ, Cone DK (1997). Food webs: a plea for parasites. *Trends Ecol Evol* **12**: 320–325.
- Meeus I, Brown MJ, De Graaf DC, Smagghe G (2011). Effects of invasive parasites on bumble bee declines. *Conserv Biol* **25**: 662–671.
- Nakjang S, Williams TA, Heinz E, Watson AK, Foster PG, Sendra KM et al. (2013). Reduction and expansion in microsporidian genome evolution: new insights from comparative genomics. *Genome Biol Evol* **5**: 2285–2303.
- Nassonova E, Cornillot E, Metenier G, Agafonova N, Kudryavtsev B, Skarlato S et al. (2005). Chromosomal composition of the genome in the monomorphic diplokaryotic microsporidium *Paranosema grylli*: analysis by two-dimensional pulsed-field gel electrophoresis. *Folia Parasitol (Praha)* **52**: 145–157.
- Pan G, Xu J, Li T, Xia Q, Liu SL, Zhang G et al. (2013). Comparative genomics of parasitic silkworm microsporidia reveal an association between genome expansion and host adaptation. *BMC Genomics* **14**: 186.
- Panek J, El Alaoui H, Mone A, Urbach S, Demetre E, Texier C et al. (2014). Hijacking of host cellular functions by an intracellular parasite, the Microsporidian *Anncalia algerae*. *PLoS ONE* **9**: e100791.
- Parisot N, Denonfoux J, Dugat-Bony E, Peyret P, Peyretaille E (2012). KASpOD—a web service for highly specific and explorative oligonucleotide design. *Bioinformatics* **28**: 3161–3162.
- Pettis JS, vanEngelsdorp D, Johnson J, Dively G (2012). Pesticide exposure in honey bees results in increased levels of the gut pathogen *Nosema*. *Naturwissenschaften* **99**: 153–158.
- Peyret P, Katinka MD, Duprat S, Duffieux F, Barbe V, Barbazanges M et al. (2001). Sequence and analysis of chromosome I of the amitochondrial intracellular parasite *Encephalitozoon cuniculi* (Microspora). *Genome Res* **11**: 198–207.
- Peyretaille E, Brousolle V, Peyret P, Metenier G, Gouy M, Vivares CP (1998). Microsporidia, amitochondrial protists, possess a 70-kDa heat shock protein gene of mitochondrial evolutionary origin. *Mol Biol Evol* **15**: 683–689.
- Peyretaille E, El Alaoui H, Diogon M, Polonais V, Parisot N, Biron DG et al. (2011). Extreme reduction and compaction of microsporidian genomes. *Res Microbiol* **162**: 598–606.
- Peyretaille E, Goncalves O, Terrat S, Dugat-Bony E, Wincker P, Cornman RS et al. (2009). Identification of transcriptional signals in *Encephalitozoon cuniculi* widespread among Microsporidia phylum: support for accurate structural genome annotation. *BMC Genomics* **10**: 607.
- Peyretaille E, Parisot N, Polonais V, Terrat S, Denonfoux J, Dugat-Bony E et al. (2012). Annotation of microsporidian genomes using transcriptional signals. *Nat Commun* **3**: 1137.
- Polis GA, Myers CA, Holt RD (1989). The ecology and evolution of intraguild predation: potential competitors that eat each other. *Annu Rev Ecol Syst* **20**: 297–330.
- Pombert JF, Selman M, Burki F, Bardell FT, Farinelli L, Solter LF 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**: 12638–12643.
- Porrini MP, Sarlo EG, Medici SK, Garrido PM, Porrini DP, Damiani N et al. (2011). *Nosema ceranae* development in *Apis mellifera*: influence of diet and infective inoculum. *J Apicul Res* **50**: 35–41.
- Raffaele S, Kamoun S (2012). Genome evolution in filamentous plant pathogens: why bigger can be better. *Nat Rev Microbiol* **10**: 417–430.
- Reynolds SE (2013). Ecology, immunity and invasive success. *Science* **340**: 816–817.
- Roudel M, Aufauvre J, Corbara B, Delbac F, Blot N (2013). New insights on the genetic diversity of the honeybee parasite *Nosema ceranae* based on multilocus sequence analysis. *Parasitology* **140**: 1346–1356.
- Sapir A, Dillman AR, Connon SA, Gruppe BM, Ingels J, Mundo-Ocampo M et al. (2014). Microsporidia-nematode associations in methane seeps reveal basal fungal parasitism in the deep sea. *Front Microbiol* **5**: 43.
- Segata N, Boernigen D, Tickle TL, Morgan XC, Garrett WS, Huttenhower C (2013). Computational meta'omics for microbial community studies. *Mol Syst Biol* **9**: 666.
- Selakovic S, de Ruiter PC, Heesterbeek H (2014). Infectious disease agents mediate interaction in food webs and ecosystems. *Proc Biol Sci* **281**: 20132709.
- Selman M, Sak B, Kvac M, Farinelli L, Weiss LM, Corradi N (2013). Extremely reduced levels of heterozygosity in the vertebrate pathogen *Encephalitozoon cuniculi*. *Eukaryot Cell* **12**: 496–502.
- Sloggett JJ, Davis AJ (2010). Eating chemically defended prey: alkaloid metabolism in an invasive ladybird predator of other ladybirds (Coleoptera: Coccinellidae). *J Exp Biol* **213**: 237–241.
- Stentiford GD, Feist SW, Stone DM, Bateman KS, Dunn AM (2013). Microsporidia: diverse, dynamic, and emergent pathogens in aquatic systems. *Trends Parasitol* **29**: 567–578.
- Stepanauskas R (2012). Single cell genomics: an individual look at microbes. *Curr Opin Microbiol* **15**: 613–620.
- Suankratay C, Thiansukhon E, Nilaratanakul V, Putaporntip C, Jongwutiwes S (2012). Disseminated infection caused by novel species of Microsporidium, Thailand. *Emerg Infect Dis* **18**: 302–304.
- Vavra J, Lukes J (2013). Microsporidia and ‘the art of living together’. *Adv Parasitol* **82**: 253–319.
- Vidau C, Diogon M, Aufauvre J, Fontbonne R, Vigues B, Brunet JL et al. (2011). Exposure to sublethal doses of fipronil and thiacloprid highly increases mortality of honeybees previously infected by *Nosema ceranae*. *PLoS ONE* **6**: e21550.

- Vilcinskas A, Stoecker K, Schmidtberg H, Rohrich CR, Vogel H (2013). Invasive harlequin ladybird carries biological weapons against native competitors. *Science* **340**: 862–863.
- Vizoso DB, Lass S, Ebert D (2005). Different mechanisms of transmission of the microsporidium *Octosporea bayeri*: a cocktail of solutions for the problem of parasite permanence. *Parasitology* **130**(Pt 5): 501–509.
- Vossbrinck CR, Andreadis TG, Vavra J, Becnel JJ (2004). Molecular phylogeny and evolution of mosquito parasitic Microsporidia (Microsporidia: Amblyosporidae). *J Eukaryot Microbiol* **51**: 88–95.
- Wang LL, Chen KP, Zhang Z, Yao Q, Gao GT, Zhao Y (2006). Phylogenetic analysis of *Nosema antheraeae* (Microsporidia) isolated from Chinese oak silkworm, *Antheraea pernyi*. *J Eukaryot Microbiol* **53**: 310–313.
- Wang TC, Nai YS, Wang CY, Solter LF, Hsu HC, Wang CH et al. (2013). A new microsporidium, *Triwangia caridinae* gen. nov., sp. nov. parasitizing fresh water shrimp, *Caridina formosae* (Decapoda: Atyidae) in Taiwan. *J Invertebr Pathol* **112**: 281–293.
- Williams BA, Slamovits CH, Patron NJ, Fast NM, Keeling PJ (2005). A high frequency of overlapping gene expression in compacted eukaryotic genomes. *Proc Natl Acad Sci USA* **102**: 10936–10941.

Supplementary Information accompanies this paper on Heredity website (<http://www.nature.com/hdy>)