

ORIGINAL ARTICLE

Lineage-specific molecular probing reveals novel diversity and ecological partitioning of haplosporidians

Hanna Hartikainen¹, Oliver S Ashford¹, Cédric Berney¹, Beth Okamura¹, Stephen W Feist², Craig Baker-Austin², Grant D Stentiford^{2,3} and David Bass¹

¹Department of Life Sciences, The Natural History Museum, London, UK; ²Centre for Environment, Fisheries and Aquaculture Science (Cefas), The Nothe, UK and ³European Union Reference Laboratory for Crustacean Diseases, Centre for Environment, Fisheries and Aquaculture Science (Cefas), The Nothe, UK

Haplosporidians are rhizarian parasites of mostly marine invertebrates. They include the causative agents of diseases of commercially important molluscs, including MSX disease in oysters. Despite their importance for food security, their diversity and distributions are poorly known. We used a combination of group-specific PCR primers to probe environmental DNA samples from planktonic and benthic environments in Europe, South Africa and Panama. This revealed several highly distinct novel clades, novel lineages within known clades and seasonal (spring vs autumn) and habitat-related (brackish vs littoral) variation in assemblage composition. High frequencies of haplosporidian lineages in the water column provide the first evidence for life cycles involving planktonic hosts, host-free stages or both. The general absence of haplosporidian lineages from all large online sequence data sets emphasises the importance of lineage-specific approaches for studying these highly divergent and diverse lineages. Combined with host-based field surveys, environmental sampling for pathogens will enhance future detection of known and novel pathogens and the assessment of disease risk.

The ISME Journal (2014) 8, 177–186; doi:10.1038/ismej.2013.136; published online 22 August 2013

Subject Category: Microbial ecology and functional diversity of natural habitats

Keywords: Asctosporaea; Haplosporidia; *Haplosporidium*; *Minchinia*; *Bonamia*; *Urosporidium*

Introduction

Characterising the diversity and distribution of parasites remains a challenge due to their patchy spatial and temporal distributions, host-restricted occurrence and poorly known life cycles. This is particularly true for microbial endoparasites, which are typically undersampled relative to macroparasites (for example, helminths). However, many parasites disperse and complete their life cycles via release of infectious stages into the environment or they infect small hosts that are incorporated in such samples. Environmental DNA detection techniques may therefore offer an untapped and powerful means of characterising parasite assemblages, particularly in view of how such approaches have revolutionised our understanding of microbial diversity in a wide range of habitat types (Moreira and López-García, 2002; Richards and Bass, 2005; Bik *et al.*, 2012).

The notable lack of information on haplosporidian diversity revealed by environmental sampling to date may reflect absence or low representation in environmental sequencing databases (due to, for example, short-lived zoospores), artefacts of sampling strategy and/or the use of broadly targeted PCR primers that are unsuitable for the highly divergent genes that may characterise many parasitic taxa (for example, Canning and Okamura, 2004).

Haplosporidia include the causative agents of the commercially significant oyster diseases MSX (caused by *Haplosporidium nelsoni* (Burrison and Ford, 2004)), bonamiosis (caused by various *Bonamia* spp. (Pichot *et al.*, 1980; Carnegie *et al.*, 2006)) and agents of several decapod crustacean diseases (Stentiford *et al.*, 2004; Utari *et al.*, 2012). The reticulate amoebae, *Filoreta* and *Gromia*, are the closest known free-living relatives; phytomyxids, vampyrellids and other Cercozoa are more distantly related (Cavalier-Smith and Chao, 2003; Bass *et al.*, 2009). There are currently four recognised haplosporidian genera: (1) *Bonamia*: four described species, all economically important oyster parasites in temperate waters; (2) *Minchinia*: a sparsely recorded genus of five species with sequence data from molluscs from

Correspondence: H Hartikainen, Department of Life Sciences, The Natural History Museum, London SW7 5BD, UK.
E-mail: h.hartikainen@nhm.ac.uk

Received 15 March 2013; revised 12 June 2013; accepted 9 July 2013; published online 22 August 2013

the US Atlantic coast, Western Australia and Europe; (3) *Urosporidium*: one described species with sequence data (*U. crescens*; Flores *et al.*, 1996) and another sequenced specimen (Reece *et al.*, 2004), both occurring as hyperparasites of trematode worms; and (4) *Haplosporidium*: the most intensively studied and diverse genus (approximately 40 species to date) and perhaps paraphyletic (Burreson and Ford, 2004). Most known members of the genus infect molluscs, but some infect crustaceans (Burreson and Ford, 2004; Stentiford *et al.*, 2004) and annelids (Siddall and Aguado, 2006). A minority of lineages infect freshwater molluscs (Molloy *et al.*, 2012). Additionally, some 'haplosporidians' that have not been taxonomically characterised and/or do not branch robustly with characterised taxa on molecular phylogenies have been sequenced (Reece *et al.*, 2004).

To date, methods of haplosporidian discovery have been largely based on direct histological and molecular characterisation of infected host tissues. Although this approach is appropriate for many studies, screening a wide range of potential hosts is not generally feasible and is therefore not a good way to estimate the diversity or size of the group. We hypothesised that there are many undetected haplosporidian lineages, as has been demonstrated in free-living protists (Bass and Cavalier-Smith, 2004; Bråte *et al.*, 2010; Howe *et al.*, 2011; Jones *et al.*, 2011). Because haplosporidians are important pathogens with global impacts on aquaculture, it is important to appreciate their diversity, particularly as changing climates and intensifying farming activities may provoke previously non-problematic parasites to emerge, causing outbreaks, high mortalities and threats to aquaculture (Okamura and Feist, 2011). Knowledge of spatial and temporal abundances of infectious stages in the environment may contribute to food security and aquaculture sustainability, for instance, by providing early warning systems of pathogen presence and planning tools, for example, for the siting of mollusc culture in low disease-risk regions.

Our aims were: (1) to characterise the spatial and temporal patterns of occurrence of known and novel haplosporidian lineages in a range of planktonic and sediment samples, (2) to examine haplosporidian diversity in a variety of marine and freshwater habitats and thereby improve the internal phylogeny of the group, and (3) to screen existing amplicon and metagenomic data sets for haplosporidians in order to determine whether such databases provide reliable representation of haplosporidians. Our results revealed a previously unappreciated and large diversity of haplosporidian lineages, including lineage-rich novel clades and divergent sequences not closely related to known lineages. We show decisively that most haplosporidian lineages are not present in eukaryote-wide surveys and thus a specific primer approach is necessary for their detection. Such an approach combined with targeted sampling schemes will ultimately provide

the basis for prediction and mitigation of diseases caused by haplosporidians and other economically important pathogens.

Materials and methods

Sampling protocol

Environmental samples were collected from two localities in Weymouth (WEY), SW England: an intertidal rock pool in the Newton's Cove (35 p.p.t. salinity; 50°34'N, 2°22'W) and a muddy, tidal channel in the Fleet Estuary (<10–30 p.p.t. salinity; 50°35'N, 2°28'W). Both sites were sampled during low tide in July 2011, October 2011 and April 2012. Approximately 150 l of water was passed serially through 100-, 50- and 20- μ m meshes, the material collected from the meshes using spatulas, kept cool and frozen at -80°C on return. Samples were thawed as part of the lysis step of the DNA extraction, thus avoiding degradation caused by enzymes released from freeze-thaw-lysed cells. The filtered water was further processed using pressure filtration (Sartorius, Stedim, Göttingen, Germany) onto 142-mm diameter, 0.45- μ m polycarbonate filter papers. The papers were scraped using razor blades, scrapings frozen in liquid nitrogen and stored at -80°C . Sixteen \sim 2 g sediment samples were collected from each site in July and October. To provide a wider geographic and ecological representation, five further sample sets were analysed: (1) South Africa (SA): 120 water (\sim 2 l) and sediment (\sim 1–2 g) samples collected in December 2011 from freshwater and marine sites in South Africa, (2) BioMarKs (BM): matched DNA and cDNA from water column and sediment samples from eight coastal marine sites from the North Sea, the Mediterranean, the Baltic and the Black Sea (collected as part of the BioMarKs consortium, <http://www.biomarks.eu>) (Bass *et al.*, 2012; Logares *et al.*, 2012; Pawlowski *et al.*, 2013), (3) Exeter (EXE): 118 DNA extractions enriched in small metazoans (by filter fractionation, picking of specimens and so on) from freshwater and marine water samples from southern England and the Mediterranean (K Hamilton and B Williams, personal communication), (4) Estuarine gradient (EST): DNA from 32 marine-to-brackish sediment samples (\sim 1–2 g), collected along the River Colne estuary in southeast England (Hawkins and Purdy, 2007; Dong *et al.*, 2009) and (5) Panama (PAN): 12 marine sediment samples (\sim 1–2 g) from sandy beach sites on the Pacific and Caribbean coasts around Panama City and Portobelo (Colón). All the samples tested are listed in Supplementary Material 1.

Sequencing

Samples from the WEY and SA sets were freeze dried for 12 h or until dry at -56°C and DNA extracted using the UltraClean Soil DNA Extraction kit (MoBio Laboratories, Carlsbad, CA, USA). The samples from PAN were extracted using the

same kit but without freeze-drying, the EXE samples were processed using a standard CTAB extraction protocol, followed by polyethylene glycol clean-ups, and the extraction methods for EST are available in Hawkins and Purdy (2007) and for BM in Logares *et al.* (2012). A nested primer set was designed to target all known haplosporidians (Supplementary Material S2). The target amplicon based on *H. nelsoni* (X74131) sequence was 667-bp long and included the variable regions V7 (~1324–1383), V8 (~1471–1530) and V9 (~1642–1743) of the 18 S rRNA gene. The amplification conditions are shown in Supplementary Table S2. Clone libraries were created using the Stratagene cloning kit (Stratagene (Agilent Technologies), Santa Clara, CA, USA). A minimum of 32 clones from each sample were sequenced in one direction using M13r primer. Further lineages for which highly targeted primers were designed are marked as α , β , γ , δ and ϵ in Figure 1a; the primer sequences and positions are given in Supplementary Table S2.

Sequence alignment and phylogenetic analyses

Haplosporidian sequences were identified by blastn searches and by building preliminary trees. Non-specific sequences amplified by the nested primer set included mainly bacterial, diatom and some metazoan (*Daphnia*) sequences, largely in freshwater samples where no true haplosporidians sequences were detected. Of the original 832 sequences, 488 reads were aligned in MAFFT using the l-ins-I algorithm, and alignments ordered according to preliminary RAXML tree topology. Sequence reads were collapsed into 'molecular operational taxonomic units' (MOTUs) (Blaxter *et al.*, 2005) where there were <3 sequence differences in each of the three variable regions in our amplicons—V7, V8 and V9), using the 50% majority rule consensus in BioEdit (Hall, 1999). Sequences with >3 differences were classified as separate MOTUs.

When an small subunit (SSU)-type is found in only one library, it is difficult to rule out the possibility that it is the result of a PCR chimera. In such cases, the singleton sequences were assessed in the alignment by eye to look for inconsistent signals, and trees were constructed using fractions of the alignment to confirm concordant signal along the length (Berney *et al.*, 2004). Six MOTUs were found to be chimeric, the remainder are presented in Figure 1 (KF208555–KF208603). Accumulation curves (Mao Tau index) for novel sequence types were generated in EstimateS v.8.2 (Colwell *et al.*, 2004) for the Newton's Cove, Fleet and combined data set with clone libraries treated as replicate samples. Non-metric Multi Dimensional Scaling and analysis of similarities were conducted in the Community Analysis Package v. 4.0 (Seaby and Henderson, 2007).

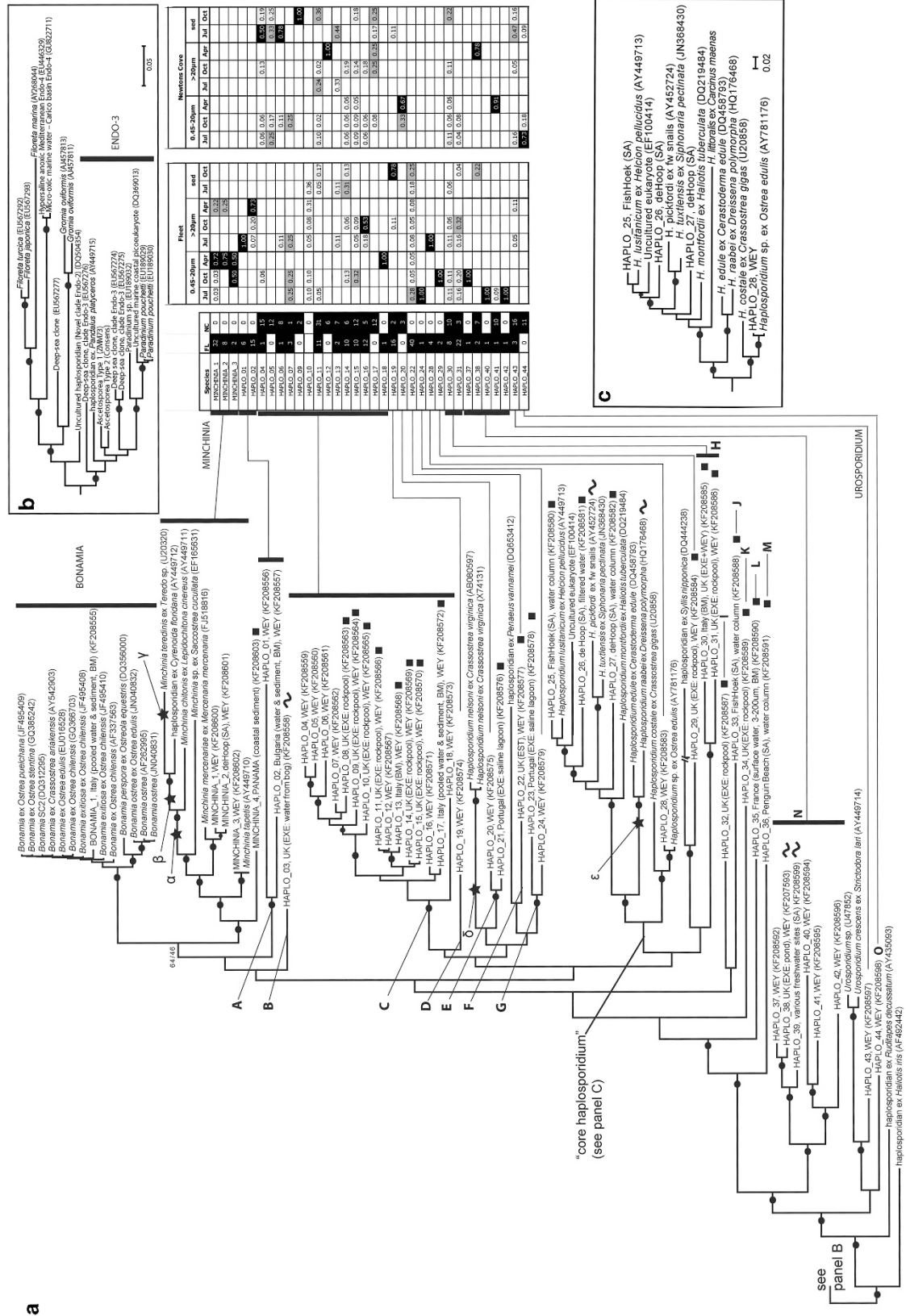
Sequences of known haplosporidians and their closest relatives were blastn searched against the Genbank nr/nt database to identify further sequences that group with known haplosporidians. These were aligned together with our newly generated sequences using the l-ins-i algorithm in MAFFT (Kato *et al.*, 2005) and refined by eye in MacGDE (Linton, 2005). Non-ascetosporean sequences were removed after preliminary analyses using the RAXML BlackBox (v. 7.3.1) (Stamatakis, 2006; Stamatakis *et al.*, 2008) on the Cipres Science Gateway (Miller *et al.*, 2010). The refined alignment was analysed in RAXML BlackBox (GTR model with CAT approximation (all parameters estimated from the data); average of 10 000 bootstrap values was mapped onto the tree with the highest likelihood value). A Bayesian consensus tree was constructed using MrBayes v 3.1.2 (Ronquist and Huelsenbeck, 2003) in parallel mode (Altekar *et al.*, 2004). Two separate MC³ runs with randomly generated starting trees were carried out for 6M generations each with one cold and three heated chains. The evolutionary model applied included a GTR substitution matrix, a four-category autocorrelated gamma correction and the covarion model. All parameters were estimated from the data. Trees were sampled every 100 generations. 2M generations were discarded as 'burn-in' (trees sampled before the likelihood plots reached a plateau) and a consensus tree was constructed from the returning sample. The sequence of *Haplosporidium littoralis* (JX185413) has not been placed in a phylogenetic context before, and to allow for maximum alignment positions, a separate RAXML analysis was run and results reported in Figure 1c.

Sequence data set mining

The known haplosporidian sequences and novel MOTUs were blastn searched against the following databases: (1) NCBI GenBank nr/nt, (2) 'All Metagenomic 454 Reads (N)' in the CAMERA database (Seshadri *et al.*, 2007; Sun *et al.*, 2011), and (3) BioMarKs V4 and V9 SSU rDNA sequences generated using eukaryote-wide primers as described in Bass *et al.* (2012) and Logares *et al.* (2012). Seed sequences were generated for each sequence types by isolating the V9 and, when available, the V4 regions, as these are the only SSU regions in the BioMarKs data, and the V4 region is generally the most variable and taxonomically informative region of the eukaryotic SSU rRNA gene (Wuyts *et al.*, 2000). The closest match sequences from CAMERA and BioMarKs were re-blasted against GenBank to confirm their phylogenetic affinities.

Results and discussion

Topologies retrieved by the Bayesian and Maximum Likelihood methods were concurrent apart from the



relative order of the branches weakly supported in both analysis methods (Figure 1a). Relationships between the four haplosporidian genera and their ascetosporean relatives were concordant with previous studies (Reece *et al.*, 2004; Bass *et al.*, 2009), in that *Minchinia* and *Bonamia* are resolved as monophyletic sister clades with moderate support. Together with novel lineages A and B, these are weakly sister to a large assemblage, including all *Haplosporidium* spp. The phylogenetic analysis in Figure 1a suggests that *Haplosporidium* could be monophyletic if lineages C–H are shown to be *Haplosporidium* when they are phenotypically characterised and more sequence data are added. In the meantime, we will conservatively adopt the view that *Haplosporidium* is paraphyletic and this radiation includes at least three or four distinct taxa. Two undescribed haplosporidian-like sequences from marine mollusc hosts (AF492442, AY435093), HAPLO_44 and *Urosporidium*+HAPLO_43, were the earliest branching haplosporidian lineages.

Ascetosporean diversity and ecology

The newly identified 49 unique MOTUs branched across the whole ascetosporean phylogeny (Figure 1a, labelled as HAPLO, MINCHINIA or BONAMIA). The majority of MOTUs were detected independently in multiple clone libraries, providing strong evidence that they are genuinely novel lineages present in our samples. The MOTU accumulation curves from the clone libraries showed that the majority of SSU-types that occurred in >1 library were sequenced, but not if those only found in one library (singletons) were included (Supplementary Material S3). The remaining MOTUs found in only one library did not show any detectable signs of chimeras (see Materials and methods) but await further verification by independent detection from other surveys.

In a Non-metric Multi Dimensional Scaling ordination plot of the Weymouth samples (not shown), muddy and brackish (Fleet) and rocky shore (Newton's Cove) assemblages clustered separately from each other on the primary axis. No other obvious clustering was apparent. These observations were corroborated by analysis of similarities, which showed a significant ($P=0.001$) difference between Fleet and Newton's Cove

samples but not between any pairs of planktonic size fractions or between sediment and all water column samples considered together. Samples taken in April and October were characterised by significantly different assemblages ($P=0.033$), providing evidence for seasonality in assemblage composition. With equivalent sampling effort, the April samples were less diverse than the other two dates, with the notable exception of three *Minchinia* lineages discussed below.

All lineages found in sediments were also found in the water column, suggesting that the haplosporidia detected may be parasitic of planktonic hosts or possess planktonic dispersal stages. The latter scenario is supported by a predominance of positive samples from the small (0.45–20 µm) size fraction, although some reads could be derived from fragmented host parts. Interestingly, we detected planktonic signal for well-studied haplosporidia, previously unknown in the plankton (for example, *Haplosporidium edule*, *Minchinia tapetis*, *Bonamia exitiosa*). Although transmission mode of most haplosporidians is unknown (Perkins 2000), sporulation is observed in some species. Thus free spore stages in the environment would be expected to occur and enables environmental sampling as a specimen-independent molecular approach to illuminate the life cycle and the diversity of protist parasites.

Few Ascetosporea are known from freshwater, but our findings suggest that they are more diverse in these habitats than previously realised. Novel lineage N (Figure 1) includes HAPLO_39, found exclusively in seven freshwater habitats in South Africa, and HAPLO_38, which was detected in a freshwater pond in the UK and also in the Weymouth coastal sites. Lineage B was found in a freshwater sample (water from a bog, EXE sample set). Two previously known freshwater lineages, *Haplosporidium pickfordi* AY452724 (Reece *et al.*, 2004) isolated from freshwater snails and *Haplosporidium raabei* HQ176468 (Molloy *et al.*, 2012) from zebra mussels (Figure 1), belong to the core haplosporidium clade comprised largely of marine lineages. Thus we demonstrate four marine-to-freshwater transitions (highlighted in Figure 1 with wavy lines). The presence of the same sequence type in marine and freshwater libraries may be the result of run-off from freshwater sources, or it may be that

Figure 1 (a) Haplosporidian phylogenetic tree generated using Bayesian Inference. Filled circles indicate posterior probabilities > 80 and bootstrap support (from Maximum Likelihood analysis) > 60. The filled squares indicate novel SSU types that were found in samples other than the WEY data set. The novel lineages are identified by lettering A–O. WEY = Weymouth, EXE = Exeter, SA = South Africa, EST = Estuarine gradient. The wavy line indicates SSU types found in freshwater samples and the lettering α , β , γ , δ and ϵ indicates the lineages for which more specific primers were designed. Occurrence of novel MINCHINIA and HAPLO MOTUs in the Weymouth samples (Fleet = muddy estuary and Newton's Cove = rock pool) are shown in the associated grid. The first two columns give the number of reads found in all clone libraries and the percentage of reads from each site, size fraction and season combination are shown in the main grid with colours indicating prevalence of the MOTU clones. Number of reads from clone libraries is intended to give an indication of the ease of detecting the particular MOTU and is subject to many PCR-induced biases. (b) Outgroup of other Ascetosporea for the main tree. (c) 'Core *Haplosporidium*' lineage with the addition of previously unpublished sequence of *Haplosporidium littoralis* (JX185413).

this haplosporidian lineage is active in both the habitat types.

Phylogenetic patterns

Haplosporidium and novel lineages. We reveal several highly distinct novel lineages branching between *Urosporidium* and *Minchinia* + *Bonamia* (labelled C–N in Figure 1a). All the confirmed *Haplosporidium* sequences also branch in this region of the tree, although statistical support does not confirm monophyly of *Haplosporidium*. Known *Haplosporidium* species comprise parasites of marine molluscs, decapod crustaceans, a crinoid echinoderm, an ascidian, polychaetes, platyhelminths, freshwater oligochaetes and snails (Perkins, 2000). If *Haplosporidium* is shown to be monophyletic, the host range is exceptionally broad.

The majority of the novel environmental sequences revealed by our study (notably the highly distinct lineages C–H) are more closely related to the confirmed *Haplosporidium* lineages than any other genus. Therefore, although *Haplosporidium* is currently the most extensively known ascetosporian genus, it also apparently harbours a proportionally greater diversity of unknown, probably parasitic lineages, than any other part of the ascetosporian radiation. A few of our novel lineages were closely related to known haplosporidia, for example, HAPLO_25 from South Africa to *H. lusitanicum*, reported only once previously from the limpet *Helcion pellucidus* in France (Reece *et al.*, 2004), and HAPLO_28 to an unnamed *Haplosporidium* from the oyster *Ostrea edulis*, previously known only from the Netherlands (unpublished). The clade labelled ‘core *Haplosporidium*’ contains numerous characterised *Haplosporidium* sequences mostly from molluscs but now also including the first haplosporidian formally described from a crustacean (*Carcinus maenas*): *H. littoralis* n. sp. (Stentiford *et al.*, 2013; Figure 1c). The ‘core’ *Haplosporidium* excluded *H. nelsoni* as well as several other undescribed haplosporidian-like sequences with unresolved branching order. As the sequence of the type specimen *Haplosporidium scolopli* (Caullery and Mesnil, 1899) is not available, it is impossible to assess which of the lineages retrieved could be assigned as the ‘true’ *Haplosporidium* (Siddall and Aguado, 2006).

Novel lineages C–O greatly increase the known haplosporidian diversity, with some lineages showing clear preference for one habitat type. Lineage C showed an especially rich radiation of novel MOTUs, with representatives in estuarine or rocky shore samples from Weymouth (for example, HAPLO_05, 12, 17 only on the rocky shore and HAPLO10, 22 only in the muddy estuary). HAPLO_22 was also present in the Colne Estuary (EST). Novel lineages I–M were only found in one clone library each and our checks (see Materials and

methods) did not suggest chimeric origins. Novel lineage N also possibly represents a new genus, with the subclade (HAPLO_39 (HAPLO_37, 38)) detected in eutrophic freshwater habitats in South Africa, UK (EXE) and three of the (saline) Weymouth samples. The degree of intragenomic variation in haplosporidians is not known, and it is possible that some of the novel MOTUs represent divergent rRNA gene copies from the same parasite and can only be fully resolved once the novel genotypes are linked with specimen-based surveys. However, sequences derived directly from individual haplosporidian infections are normally clean, suggesting at least in these cases that intragenomic SSU rDNA variation is low (G Ward, H Hartikainen, personal observation).

Minchinia, *Bonamia* and *Urosporidium*. *Minchinia* are parasites of molluscs and crabs (Newman *et al.*, 1976; Stokes *et al.*, 1995; Bearham *et al.*, 2008, 2009). Previous records of *Minchinia* spp. in the UK are from a histopathology survey of cockles (Longshaw and Malham, 2012), which suggested that *Minchinia mercenariae* and *M. tapetis* may be implicated in host population crashes. Our results provide the first molecular records of *Minchinia* in the UK.

Three new *Minchinia*-affiliated SSU-types were found exclusively in the Fleet but never in Newton's Cove. MINCHINIA_1 and 2 were closely related to *M. mercenariae* and MINCHINIA_3 identical to *M. tapetis* (both parasites of American clams and more recently associated with Welsh cockle mortalities but not previously detected in Weymouth). Strikingly, *Minchinia*-affiliated SSU-types were only detected in water column samples, strongly indicating a planktonic life-cycle stage (predominantly in the 0.45–20- μ m size fraction) and mostly in the April samples, suggesting a periodicity that may be related to host reproductive cycle. Novel lineage A, weakly basal to *Bonamia* + *Minchinia*, was detected only in the >20- μ m fraction. This implicates infection of a planktonic host and possibly of host larvae, as has been previously found for *Bonamia* (Arzul *et al.*, 2011).

Bonamia spp. cause diffuse or focal haemocyte infiltration of oysters (Carnegie and Cochenne-Laureau, 2004) and can lead to catastrophic mortalities. All available data indicate low SSU rRNA gene diversity in the genus. *B. exitiosa* has once been found in the southwest UK (<http://www.cefas.defra.gov.uk/idaad/>) but never in the vicinity of Weymouth, concordant with our findings. Lack of spore stages suggests that infection may be by direct transmission (Elston *et al.*, 1987). We detected only one *Bonamia* SSU-type (probably *B. exitiosa*, Figure 1a) in water column/sediment samples from Naples (Italy; (BioMarks DNA)) which has also been reported from oysters in the Adriatic coastline of Italy in 2010 (Nardsi *et al.*, 2011).

We found no new SSU-types close to the hyperparasite *Urosporidium*, although two lineages (HAPLO_43 and 44) formed a clade with the two known urosporidians. These were frequently found in Newton's Cove, although, depending on their life cycle, hyperparasites could be relatively difficult to detect using 'blind' environmental sampling.

Occurrence patterns, detection and species-specific probing

Several striking observations can be made in relation to occurrence patterns. First, of the 49 novel SSU-types and 37 pre-existing haplosporidian sequences, in only four cases an environmental and characterised sequence were sufficiently similar to infer that the environmental sequence represents the same or very closely related species. These were: *B. exitiosa* (infects *Ostrea* spp.), *M. mercenariae* and *M. tapetis* (from the clams *Mercenaria mercenariae* and *Ruditapes decussatus*, respectively), and an unnamed *Haplosporidium* from *O. edulis*. These mollusc hosts are recorded from Weymouth, but the diseases caused have not been recorded there. PCR-based assays of environmental DNA are sensitive approaches for detecting rare sequence types and the nested protocol used here further increases detection efficiency and improves amplification specificity. Consequently, some of the haplosporidians represented by our novel MOTUs may be very rare in the environment. Thus, this study demonstrates the potential of environmental sequencing approaches for detecting haplosporidian presence where other approaches do not and for assessing the effectiveness of biosecurity measures designed to protect against incursions by exotic pathogens as defined under the current EU Directives.

Second, although our primers were demonstrably inclusive enough to cover most known haplosporidian diversity, they necessarily include some mismatches to more divergent known haplosporidian types (for example, the *H. louisiana* sequence was deliberately excluded due to its extreme divergence). Furthermore, some potentially detectable amplicons may be missed in incompletely sampled libraries if they are relatively difficult to amplify (because of secondary structure characteristics, length variation and so on), perhaps compounded by rarity or ephemerality. Such amplifications will be outcompeted by more easily amplified templates in a mixed reaction, and much higher sampling levels and/or more highly specific primers may be required to detect them. In addition to primer bias, a very tight association of parasite with a relatively large (>5 mm) host would theoretically be difficult to detect without sampling potential hosts directly. Periodic occurrence and dispersal may also pose difficulties in detection. For example, *H. littoralis* (Figure 1c) is present and monitored in *C. maenas* in Newton's Cove, yet was not detected in our libraries, possibly due to the fact

that transmission stages are not released from infected hosts (Stentiford *et al.*, 2004, 2013) or because two nucleotide mismatches are present between the primers and the *H. littoralis* sequence. It is a priority for future work to determine to what extent molecular detection of haplosporidian lineages can be associated with actual infections and potential for disease.

It is highly likely that some haplosporidian lineages present in our samples were missed, either because they were too rare or were subject to PCR biases. This was tested with primers stringently targeted to α , β , γ , δ and ϵ (Figure 1a). Only ϵ primer set gave a positive result, with repeated detection of the *H. edule* SSU-type. A total of 36% of the samples from both the Weymouth sites, including sediment and water column samples, were found to include an SSU-type with 99–100% sequence identity with *H. edule*. This SSU-type was not recovered using the more broadly haplosporidian-specific primers. This indicates that lineage-targeted primers might reveal even further diversity and is important to consider when screening samples for particular lineages. That we did not detect any lineages within the target groups α , β , γ and δ could simply be explained by their general absence of: *H. raabei* (ϵ) because of its freshwater affiliation, *H. nelsoni* (δ) because it appears to be absent from UK/northern European waters and the inter-related *Minchinia*-lineages α , β and γ (*M. teredinis*, *M. chitonis* and a strain isolated from marsh clam *Cyrenoidea floridana* (east coast US), and the rock oyster *Saccostrea cucullata* (NW Australia)) have no previous European records (although note that we did detect the related *M. tapetis* and a close relative of *M. mercenariae* from Weymouth using our main primer set). The fact that the independent yet phylogenetically nested primer sets α , β and γ did not produce a positive result provides additional evidence that the corresponding *Minchinia* lineages were not present in our samples. Thus a highly targeted approach may offer a powerful tool for disease risk monitoring and tracking anthropogenic introductions and host shifts, along with pathogen surveys in potential hosts.

Mining next-generation sequencing data sets for Ascetosporea

There are relatively few environmentally derived haplosporidian sequences in NCBI Genbank's nr/nt database. Those that exist are shown in Figure 1. We therefore investigated the two sources of next-generation sequencing data from marine environments—the BioMark's amplicon data set and the PCR-independent meta-genomic and -transcriptomic libraries held in CAMERA. It is possible that the intense sampling power of the next-generation sequencing technologies captures a wider taxonomic range of protists than comparatively low-coverage environmental cloning

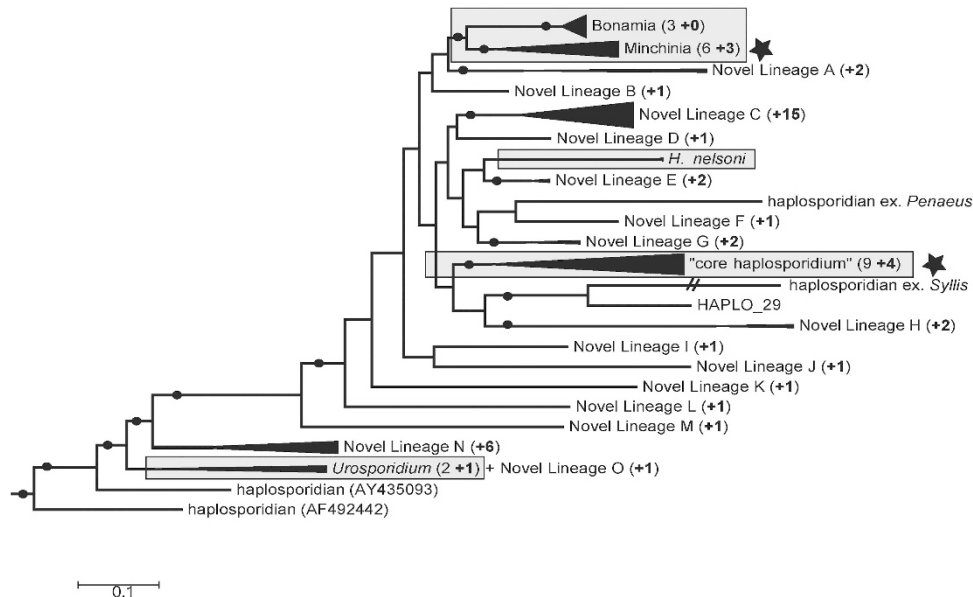


Figure 2 Summary of the novel lineages and known haplosporidians (in grey boxes). Numbers in brackets indicate known SSU sequence types and in bold the number of novel MOTUs discovered in this study. Stars indicate known lineages that were first detected at Weymouth, despite the lack of previous records of these parasites in the vicinity.

methods. The BioMarKs data comprise amplicons generated from general eukaryote primers targeting the V4 and V9 SSU rRNA gene variable regions (Wuyts *et al.*, 2002) and comprise c. 3M sequence types from subsurface and the deep chlorophyll maximum water column layer in coastal marine waters and associated benthic samples from nine sites around Europe. However, no confirmed haplosporidian sequences were recovered, showing that 'universal' primer sets can miss important elements of biodiversity. The corresponding BioMarKs DNA samples did yield haplosporidian sequences when amplified using our specific primers (Figure 1a), although the diversity revealed was far lower than in Weymouth—either because the offshore sampling location, where incidence of haplosporidian cells and/or their hosts may be lower, or due to the lower sampling volumes. The CAMERA metagenomic databases were almost as depauperate with respect to haplosporidia. With shotgun-sequenced metagenomic data, this cannot be attributed to PCR biases but more likely to the relative abundance or absence of haplosporidians in the samples analysed. Nonetheless, three apparently haplosporidian SSU V4 sequence types were recovered from the CAMERA data, two quite distantly related to *Bonamia perspora* (93% sequence similarity match to DQ356000) and *M. mercenariae* (92% sequence similarity match to FJ518816), and one more closely related to *H. pickfordi* (97% sequence similarity match to AY452724). These three sequences were distinct from others in our clone libraries and merit further investigation. One SSU V9-type returned a 92% match to *M. tapetis* AY449710, but the shortness of the V9 region makes taxonomic inference at this level uncertain.

The sequence databases (Genbank nr/nt, BioMarKs and CAMERA) have much higher representation of non-haplosporidian Ascomycota than true haplosporidians themselves, that is, sequences related to the copepod parasite *Paradinium* (clade ENDO-3; Figure 1b) (Skovgaard and Daugbjerg, 2008; Bass *et al.*, 2009) and the phylogenetically unstable environmental clone DQ504354 and Endo-2 (López-García *et al.*, 2006). This may be because the sequences in these clades are less divergent than most haplosporidia (they were first detected with general or cercozoan-specific primers that never amplified true haplosporidia despite extensive sampling (Bass and Cavalier-Smith, 2004; Bass *et al.*, 2009)). Additionally it may be because some of their few known hosts (copepods and prawn in the case of Endo-3) are more widely distributed at global and local scales than hosts of many haplosporidia.

Conclusion

The use of PCR primer combinations specifically targeted to protist groups of interest can provide a powerful insight into their biology. The approach is especially helpful when target groups are difficult to study by other means, for example, because of physical access to samples, cryptic and/or specialised life styles, or when gene-related unorthodoxy confounds standard research methods. Our targeted study has greatly enhanced our understanding of haplosporidian diversity and distributions, improved the phylogeny of the group and provided clues about their obscure life cycles. Figure 2 schematically summarises the number and

phylogenetic positions of novel and previously known lineages detected in this study. By definition, none of the novel environmental lineages has previously been detected at any site. Figure 2 also indicates known disease-causing agents detected, for the first time, by our approach at well-monitored sites. Thus we lay the groundwork for new research directions, including modelling the changing distributions and disease risks posed by haplosporidians and the consequent implications for the aquaculture sector and food security.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

This research was supported by The NHM Special Funds Grant and Expedition Funds (South Africa) to BO and DB and NERC New Investigator (NE/H000887/1) and Standard Research (NE/H009426/1) Grants to DB. We thank Kristina Hamilton and Bryony Williams from the University of Exeter, Kevin Purdy and Mark Osborn from the University of Warwick and the BioMarKs consortium for access to samples. We acknowledge the Western Cape Nature Conservation Board for permission to sample in South Africa (Permit number 0035-AAA007-00375). GS and SF were supported by the Department for Environment, Food and Rural Affairs project no. C5485 (to GS).

References

- Altekar G, Dwarkadas S, Huelsenbeck JP, Ronquist F. (2004). Parallel metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics* **20**: 407–415.
- Arzul I, Langlade A, Chollet B, Robert M, Ferrand S, Omnes E *et al.* (2011). Can the protozoan parasite *Bonamia ostreae* infect larvae of flat oysters *Ostrea edulis*? *Vet Parasitol* **179**: 69–76.
- Bass D, Cavalier-Smith T. (2004). Phylum-specific environmental DNA analysis reveals remarkably high global biodiversity of Cercozoa (Protozoa). *Int J Syst Evol Microbiol* **54**: 2393–2404.
- Bass D, Chao EEY, Nikolaev S, Yabuki A, Ishida K-i, Berney C *et al.* (2009). Phylogeny of novel naked filose and reticulose Cercozoa: Granofilosea cl. n. and Proteomyxidea revised. *Protist* **160**: 75–109.
- Bass D, Yabuki A, Santini S, Romac S, Berney C. (2012). Reticulamoeba is a long-branched Granofilosean (Cercozoa) that is missing from sequence databases. *PLoS One* **7**: e49090.
- Bearham D, Spiers Z, Raidal S, Jones JB, Nicholls PK. (2008). Detection of *Minchinia* sp., in rock oysters *Saccostrea cucullata* (Born, 1778) using DNA probes. *J Invertebr Pathol* **97**: 50–60.
- Bearham D, Spiers ZB, Raidal SR, Jones JB, Nicholls PK. (2009). Detection of *Minchinia occulta* in samples of pearl oysters *Pinctada maxima* infected by *Haplosporidium hinei*. *Aust Vet J* **87**: 430–437.
- Berney C, Fahrni J, Pawlowski J. (2004). How many novel eukaryotic 'kingdoms'? Pitfalls and limitations of environmental DNA surveys. *BMC Biol* **2**: 13.
- Bik HM, Porazinska DL, Creer S, Caporaso JG, Knight R, Thomas WK. (2012). Sequencing our way towards understanding global eukaryotic biodiversity. *TREE* **27**: 233–243.
- Blaxter M, Mann J, Chapman T, Thomas F, Whitton C, Floyd R *et al.* (2005). Defining operational taxonomic units using DNA barcode data. *Phi Trans R Soc B* **360**: 1935–1943.
- Bråte J, Logares R, Berney C, Ree DK, Klaveness D, Jakobsen KS *et al.* (2010). Freshwater Perkinsea and marine-freshwater colonizations revealed by pyrosequencing and phylogeny of environmental rDNA. *ISME J* **4**: 1144–1153.
- Burreson EM, Ford SE. (2004). A review of recent information on the Haplosporidia, with special reference to *Haplosporidium nelsoni* (MSX disease). *Aquat Living Resour* **17**: 499–517.
- Canning EU, Okamura B. (2004). Biodiversity and evolution of the Myxozoa. *Adv Parasitol* **56**: 43–131.
- Carnegie RB, Burreson EM, Mike Hine P, Stokes NA, Audemard C, Bishop MJ *et al.* (2006). *Bonamia perspora* n. sp. (Haplosporidia), a parasite of the oyster *Ostrea equestris*, is the first *Bonamia* species known to produce spores. *J Eukaryot Microbiol* **53**: 232–245.
- Carnegie RB, Cochenne-Laureau N. (2004). Microcell parasites of oysters: Recent insights and future trends. *Aquat Living Resour* **17**: 519.
- Cavalier-Smith T, Chao EEY. (2003). Phylogeny of Choanozoa, Apusozoa, and other Protozoa and early eukaryote megaevolution. *J Mol Evol* **56**: 540–563.
- Colwell RK, Mao CX, Chang J. (2004). Interpolating, extrapolating, and comparing incidence-based species accumulation curves. *Ecology* **85**: 2717–2727.
- Dong LF, Smith CJ, Papaspyrou S, Stott A, Osborn AM, Nedwell DB. (2009). Changes in benthic denitrification, nitrate ammonification, and anammox process rates and nitrate and nitrite reductase gene abundances along an estuarine nutrient gradient (the Colne Estuary, United Kingdom). *Appl Environ Microbiol* **75**: 3171–3179.
- Elston RA, Kent ML, Wilkinson MT. (1987). Resistance of *Ostrea edulis* to *Bonamia ostreae* infection. *Aquaculture* **64**: 237–242.
- Flores BS, Siddall ME, Burreson EM. (1996). Phylogeny of the Haplosporidia (Eukaryota: Alveolata) based on small subunit ribosomal RNA gene sequence. *J Parasit* **82**: 616–623.
- Hall TA. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* **41**: 95–98.
- Hawkins RJ, Purdy KJ. (2007). Genotypic distribution of an indigenous model microorganism along an estuarine gradient. *FEMS* **62**: 187–194.
- Howe AT, Bass D, Chao EE, Cavalier-Smith T. (2011). New genera, species, and improved phylogeny of Glissomonadida (Cercozoa). *Protist* **162**: 710–722.
- Jones MDM, Forn I, Gadelha C, Egan MJ, Bass D, Massana R *et al.* (2011). Discovery of novel intermediate forms redefines the fungal tree of life. *Nature* **474**: 200–203.

- Katoh K, Kuma K-i, Toh H, Miyata T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res* **33**: 511–518.
- Linton EW. (2005). MacGDE: genetic data environment for MacOS X. Software available at <http://www.msu.edu/~linton/macgde>.
- Logares R, Audic S, Santini S, Pernice MC, de Vargas C, Massana R. (2012). Diversity patterns and activity of uncultured marine heterotrophic flagellates unveiled with pyrosequencing. *ISME J* **6**: 1823–1833.
- Longshaw M, Malham SK. (2012). A review of the infectious agents, parasites, pathogens and commensals of European cockles (*Cerastoderma edule* and *C. glaucum*). *J Mar Biol Assoc UK* **1**: 1–21.
- López-García P, Vereshchaka A, Moreira D. (2006). Eukaryotic diversity associated with carbonates and fluid-seawater interface in Lost City hydrothermal field. *Environ Microbiol* **9**: 546–554.
- Miller MA, Pfeiffer W, Schwartz T. (2010). Gateway Computing Environments Workshop (GCE), 2010. *IEEE* pp 1–8.
- Molloy DP, Giambardino Rini L, Stokes NA, Burrenson EM, Ovcharenko MA. (2012). *Haplosporidium raabei* n. sp. (Haplosporidia): a parasite of zebra mussels, *Dreissena polymorpha* (Pallas, 1771). *Parasitology* **139**: 463–477.
- Moreira D, López-García P. (2002). The molecular ecology of microbial eukaryotes unveils a hidden world. *Trends Microb* **10**: 31–38.
- Nardis V, Arzul I, Cargini D, Mosca F, Calzetta A, Traversa D *et al.* (2011). Detection of *Bonamia ostreae* and *B. exitiosa* (Haplosporidia) in *Ostrea edulis* from the Adriatic Sea (Italy). *Dis Aquat Org* **89**: 79–85.
- Newman MW, Johnson CA, Pauley GB. (1976). A *Minchinia*-like haplosporidan parasitizing blue crabs, *Callinectes sapidus*. *J Invertebr Pathol* **27**: 311–315.
- Okamura B, Feist S. (2011). Emerging diseases in freshwater systems. *Freshwater Biol* **56**: 627–637.
- Pawlowski J, Audic S, Adl S, Bass D, Belbahri L, Berney C *et al.* (2013). CBOL protist working group: barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. *PLoS Biol* **10**: e1001419.
- Perkins FO. (2000). An Illustrated Guide to the Protozoa. Vol. 1. Lee JJ, Leedale GF, Bradbury P (eds). Society of Protozoologists: Lawrence, Kansas, USA. pp 1328–1341.
- Pichot Y, Comps M, Tige G, Grizel H, Rabouin MA. (1980). Recherches sur *Bonamia ostreae* gen. n. sp. n. parasite nouveau de l'huître plate *Ostrea edulis* L. *Revue des Travaux de l'Institut des Pêches Maritimes* **43**: 131–140.
- Reece KS, Siddall ME, Stokes NA, Burrenson EM. (2004). Molecular phylogeny of the Haplosporidia based on two independent gene sequences. *J Parasitol* **90**: 1111–1122.
- Richards TA, Bass D. (2005). Molecular screening of free-living microbial eukaryotes: diversity and distribution using a meta-analysis. *Curr Opin Microbiol* **8**: 240–252.
- Ronquist F, Huelsenbeck JP. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.
- Seaby RMH, Henderson PA. (2007). *Community Analysis Package 4.0*. PISCES Conservation Ltd: Lymington, UK. p 154.
- Seshadri R, Kravitz SA, Smarr L, Gilna P, Frazier M. (2007). CAMERA: a community resource for meta-genomics. *PLoS Biol* **5**: e75.
- Siddall ME, Aguado MT. (2006). Molecular phylogenetic evidence of a haplosporidian parasite infecting the polychaete *Syllis nipponica* (Imajima, 1966). *Parasitol Res* **99**: 309–312.
- Skovgaard A, Daugbjerg N. (2008). Identity and systematic position of *Paradinium poucheti* and other paradinium-like parasites of marine copepods based on morphology and nuclear-encoded SSU rDNA. *Protist* **159**: 401–413.
- Stamatakis A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688–2690.
- Stamatakis A, Hoover P, Rougemont J. (2008). A rapid bootstrap algorithm for the RAxML Web servers. *Syst Biol* **57**: 758–771.
- Stentiford GD, Bateman KS, Stokes NA, Carnegie RB. (2013). *Haplosporidium littoralis* sp. nov.: a crustacean pathogen within the Haplosporida (Cercozoa, Ascetosporea). *Dis Aquat Org*. Doi:10.3354/dao02619.
- Stentiford GD, Feist SW, Bateman KS, Hine PM. (2004). Haemolymph parasite of the shore crab *Carcinus maenas*: pathology, ultrastructure and observations on crustacean haplosporidians. *Dis Aquat Org* **59**: 57–68.
- Stokes NA, Siddall ME, Burrenson EM. (1995). Small subunit ribosomal RNA gene sequence of *Minchinia teredinis* (Haplosporidia: Haplosporidiidae) and a specific DNA probe and PCR primers for its detection. *J Invertebr Pathol* **65**: 300–308.
- Sun S, Chen J, Li W, Altintas I, Lin A, Peltier S *et al.* (2011). Community cyberinfrastructure for advanced microbial ecology research and analysis: the CAMERA resource. *Nucleic Acids Res* **39**: D546–D551.
- Utari HB, Senapin S, Jaengsanong C, Flegel TW, Kruatrachue M. (2012). A haplosporidian parasite associated with high mortality and slow growth in *Penaeus (Litopenaeus) vannamei* cultured in Indonesia. *Aquaculture* **366-367**: 85–89.
- Wuyts J, De Rijk P, Van de Peer Y, Pison G, Rousseeuw P, De Wachter R. (2000). Comparative analysis of more than 3000 sequences reveals the existence of two pseudoknots in area V4 of eukaryotic small subunit ribosomal RNA. *Nucleic Acids Res* **28**: 4698–4708.
- Wuyts J, Van de Peer Y, Winkelmans T, De Wachter R. (2002). The European database on small subunit ribosomal RNA. *Nucleic Acids Res* **30**: 183–185.

Supplementary Information accompanies this paper on The ISME Journal website (<http://www.nature.com/ismej>)