Journal of Invertebrate Pathology 198 (2023) 107908

Contents lists available at ScienceDirect

Journal of Invertebrate Pathology

journal homepage: www.elsevier.com/locate/jip

Comparative genomics for *Agmasoma* sp. (Microsporidia) parasitising invasive *Carcinus aestuarii* and *Carcinus maenas* in Argentina

Jamie Bojko [a,b,*], Antonella Frizzera [c], Nuria Vázquez [c], Gillian Taylor [a,b], Vikki Rand [a,b], Florencia Cremonte [c]

[a] National Horizons Centre, Teesside University, Darlington DL1 1HG, UK
[b] School of Health and Life Sciences, Teesside University, Middlesbrough TS1 3BX, UK
[c] Laboratorio de Parasitología, Instituto de Biología de Organismos Marinos (CCT CONICET-CENPAT), Boulevard Brown 2915, U9120ACF Puerto Madryn, Argentina

ARTICLE INFO

Keywords:
Parasite
Enterocytozoonida
Invasion
Systematics
Genomics

ABSTRACT

Carcinus spp. are global aquatic invaders and carriers of several parasites, including a taxonomically unrecognised microsporidian recently observed from Argentina. We provide genome drafts for two parasite isolates, one from *Carcinus maenas* and one from *Carcinus aestuarii*, and use multi-gene phylogenetics and genome comparison methods to outline their similarities. Their SSU genes are 100 % similar and other genes have an average similarity of 99.31 %. We informally name the parasite *Agmasoma carcini*, terming the isolates *Ac.* var. *aestuarii* and *Ac.* var. *maenas*, following the wealth of genomic data available for each. This study follows on from Frizzera et al. (2021), where this parasite was first histologically identified.

The microsporidia are a group of spore-forming obligate parasites of Animalia and Protozoa across a range of environmentally diverse ecosystems (Bojko et al., 2022). Their diversity includes over 1600 potential species; however, ~290 species (~18 %) are formally described with available pathological, ultrastructural, and genetic data (Murareanu et al., 2021; Bojko et al., 2022). Fewer (~50 species) have genomic sequence data available (Bojko et al., 2022). Systematics surrounding the microsporidia has been changeable over the past century, but the suggestion of a standard naming system that follows classic taxonomic nomenclature introduces a straight-forward system to continue describing new species and their higher taxonomy, using pathology, ultrastructure, and genetic/genomic techniques (Bojko et al., 2022).

Many microsporidian parasites have been described from invasive species, non-native organisms that negatively impact native biodiversity and ecosystems at their invasion site(s). In some cases, invasive species have been found to carry microsporidia to new ranges (Burgess and Bojko, 2022) and in other cases their microsporidian parasites have had a direct influence on their host's invasiveness (Bojko et al., 2019). The Crustacea, a group of invertebrate animals that have many associated microsporidian parasites, includes multiple highly invasive species, such as the European shore crab (invasive green crab) *Carcinus maenas* (Brachyura) (Bojko et al., 2021). *Carcinus maenas* is associated with ~95 symbionts (Bojko et al., 2021), many of which are parasitic, and includes 5 microsporidian species: *Parahepatospora carcini* (Bojko et al., 2017); *Nadelspora canceri* (Stentiford et al., 2013); *Abelspora portucalensis* (Azevedo, 1987); *Ameson pulvis* (Sprague and Couch, 1971); and *Thelohania maenadis* (Sprague and Couch, 1971). A relative of *C. maenas*, *Carcinus aestuarii* is also a global invader associated with microsporidian parasites (*T. maenadis*; *Ormieresia carcini*) (Vivares et al., 1977) and has most recently been found to carry a microsporidian parasite to Argentina, with genetic similarity to the genus *Agmasoma* (Enterocytozoonida) (Frizzera et al., 2021). It is important to mention that microsporidians are obligate parasites that can affect host health, representing a potential danger for other crustaceans that may have important economic or ecological value in Argentina.

The Enterocytozoonida currently holds 5 species with draft genomic data: *Hepatospora eriocheir* (variants from *Cancer pagurus* and *Eriocheir sinensis*) (accession: GCA_002087885; size: 4.70254 Mb; proteins: 2871; BUSCO: 44.0 %); *Enterospora canceri* (accession: GCA_002087915; size: 3.09538 Mb; proteins: 2169; BUSCO: 53.5 %); *Enterocytozoon bieneusi* (accession: GCA_000209485; size: 3.86074 Mb; proteins: 3632; BUSCO: 54.5 %); *Enterocytozoon hepatopenaei* (accession: GCA_002081675; size: 3.03974 Mb; proteins: 2536; BUSCO: 56.7 %); and *Vittaforma corneae* (accession: NZ_AEYK00000000; size: 3.21352 Mb; proteins: 2239; BUSCO: 75 %) (Mittleider et al., 2002; Akiyoshi et al., 2009; Wiredu Boakye et al., 2017). In this study we provide a genome draft for the

* Corresponding author at: National Horizons Centre, Teesside University, Darlington DL1 1HG, UK.
E-mail address: J.Bojko@tees.ac.uk (J. Bojko).

https://doi.org/10.1016/j.jip.2023.107908
Received 20 October 2022; Received in revised form 25 February 2023; Accepted 1 March 2023
Available online 5 March 2023

Agmasoma parasite infecting invasive *C. aestuarii* and *C. maenas* in Argentina (Frizzera et al., 2021), and we sequence isolates from both hosts to conduct a comparison and determine if it is the same species of parasite infecting each host.

A single *C. aestuarii* and a single *C. maenas* collected by hand from the intertidal zone around Puerto Madryn (45°50′S, 64°54′W) exhibited microsporidian infections described by Frizzera et al. (2021), and were dissected to obtain infected tissues in 96 % ethanol for DNA extraction. DNA was extracted from hepatopancreas, muscle and gill tissue from both animals using a 'Wizard® genomic DNA extraction kit' (Promega). The DNA extracts were pooled for each host species, resulting in two extracts that were submitted for paired-end next generation sequencing with Novogene. The two samples were prepared into individual libraries using an 'NEBNext® Ultra™ DNA Library Prep Kit' (PE150; Illumina) and sequenced on an Illumina NovaSeq platform.

The sequence data for the *C. aestuarii* isolate included 3.5×10^6 raw forward reads and 3.7×10^6 raw reverse reads. The sequence data for the *C. maenas* isolate included 2.3×10^6 raw forward reads and 2.4×10^6 raw reverse reads. The data were trimmed (Trimmomatic [LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36]; Bolger et al., 2014) and assembled (SPADES v.3.15.3 [phred-offset 33]; Bankevich et al., 2012) separately. The resulting metagenomic assemblies (*C. aestuarii*; n = 211,925 contigs > 1000 bp; N50: 2,146; L50: 65,319) (*C. maenas*; n = 136,893 contigs > 1000 bp; N50: 1,549; L50: 63,999) were then screened for similarity against all available microsporidian proteins using blastx (NCBI), helping to isolate contiguous sequences derived from the microsporidian parasites. Metaxa2 was used to isolate the contigs that included the microsporidian ribosomal small-subunit RNA gene (SSU) (Bengtsson-Palme et al., 2015). Trimmed reads were then mapped to the microsporidian contigs in 'CLC genomics workbench 22′ (Qiagen) to confirm their contiguity. Microsporidian contigs were then annotated using GeneMarkS (intronless eukaryotic mode; Besemer et al., 2001) and each protein was checked against all available microsporidian proteins using blastp (NCBI) and InterProScan. Via this method, the microsporidian isolate from *C. aestuarii* was represented by 70 contigs accounting for 4,248,920 bp of the parasite genome and encoded a predicted 1,626 proteins (12X, average coverage) (Fig. 1). The isolate from *C. maenas* was present at a much lower burden in the host and was represented by a more fragmented assembly including 715 contigs accounting for 1,626,467 bp and encoded a predicted 769 proteins (22X, average coverage). The contigs from the *C. maenas* isolate were also mapped to the *C. aestuarii* microsporidian draft genome in CLC genomics workbench v.9 to capture any missing contigs (similarity: 0.8; coverage: 0.8). BUSCO (Simão et al., 2015) was used to determine the overall completeness of the two genomes by comparing their translated proteins to the microsporidia_odb10 database, resulting in values of 62.7 % (373 complete and single-copy BUSCOs; 3 complete and duplicated BUSCOs; 3 fragmented BUSCOs; 221 missing BUSCOs) for the *C. aestuarii* isolate and 20.3 % (122 complete and single-copy BUSCOs; 0 complete and duplicated BUSCOs; 25 fragmented BUSCOs; 453 missing BUSCOs) for the *C. maenas* isolate. Based on the BUSCO values for the *C. aestuarii* isolate and other Enterocytozoonida, this might be the second most complete genome of the Enterocytozoonida so far, after *V. corneae* (Mittleider et al., 2002).

Conducting comparative genomics between the two isolates and the broader Enterocytozoonida involved phylogenetic and sequence similarity analyses. OrthoFinder (Emms and Kelly, 2019) was used to isolate orthogroups shared between available Enterocytozoonida species with genomic data (and outgroup *Nosema granulosis*; Nosematida) as well as the two new isolates. The translated orthogroups (n = 784, with 75 % of species having single-copy genes in any orthogroup) were then aligned and used in a concatenated maximum-likelihood (ML) phylogeny using the OrthoFinfer's 'iqtree' and tree inference 'msa' options (Fig. 2). This revealed that the two *Agmasoma* isolates branched at the base of the Enterocytozoonida, together, with a branch distance of 0.0127 between them. The average nucleotide similarity between all protein coding genes was 99.31 %, with the most similar being 100 % and the least similar being gene_390 (predicted: histidinol phosphatase) at 80.55 %.

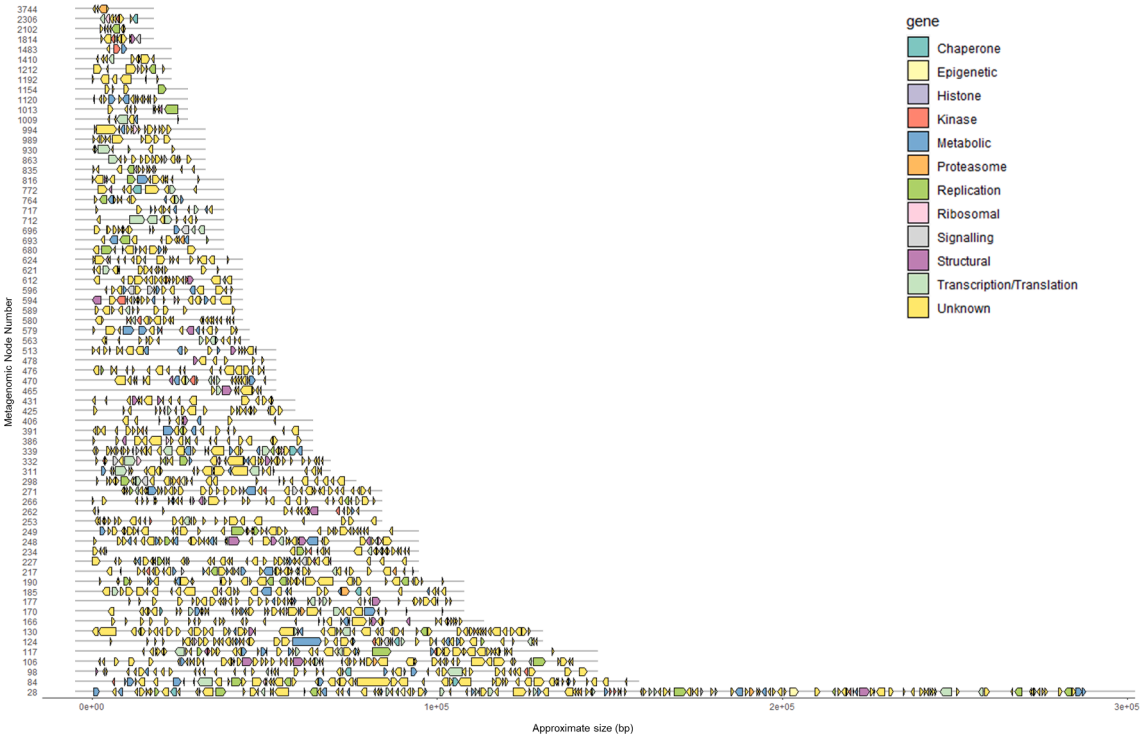

Fig. 1. A genome architecture plot of the contiguous sequences that encode protein coding genes, which compose the *Agmasoma carcini* var. *aestuarii* genome. The genes annotated correspond to the BUSCO identification of GeneMarkS annotated proteins. A colour key is provided to highlight the broad function of each hypothetical protein across 12 categories (see key). The plot was developed in RStudio (R Core Team, 2013) using the gggenes package (www.cran.r-project.org/web/packages/gggenes).

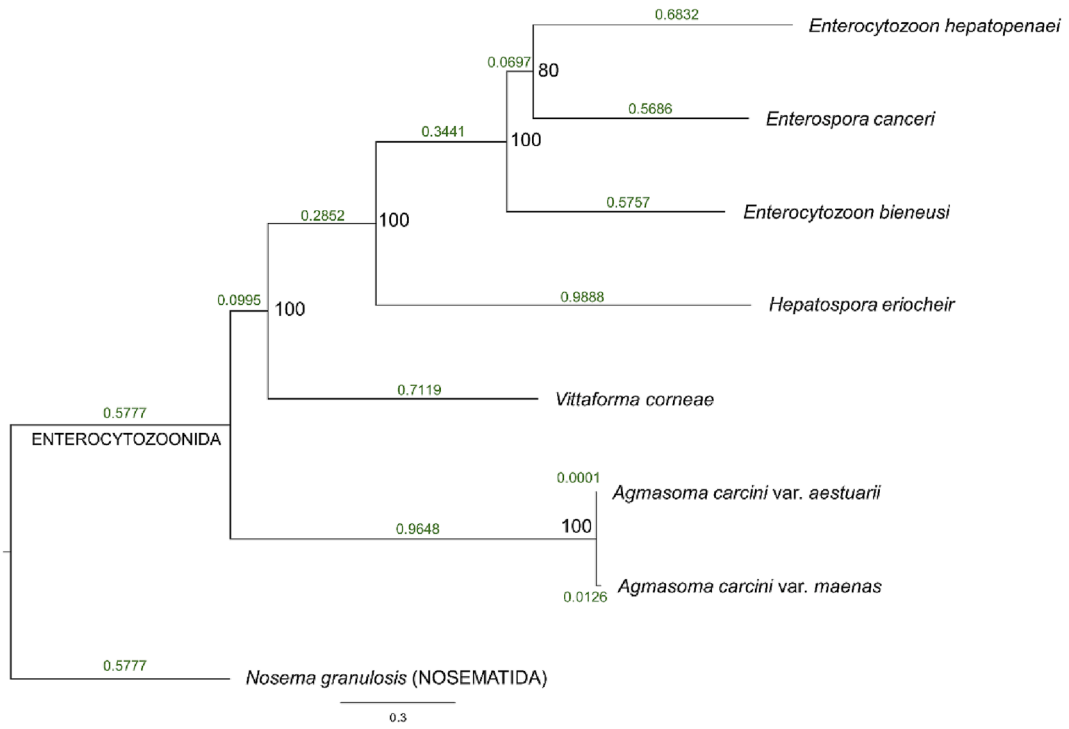

Fig. 2. A multi-protein concatenated tree using 793 orthogroups with a minimum of 75.0 % of species having single-copy genes in any orthogroups, based on a comparative genomics approach. The tree includes data from the genomes of *Hepatospora eriocheir* (GCA_002087885), *Enterospora canceri* (GCA_002087915), *Enterocytozoon bieneusi* (GCA_000209485), *Enterocytozoon hepatopenaei* (GCA_002081675), *Vittaforma corneae* (NZ_AEYK00000000), and the two new *Agmasoma* isolates, *Agmasoma carcini* var. *aestuarii* and *Agmasoma carcini* var. *maenas*. *Nosema granulosis* (Nosematida) is used as an outgroup to root the tree. Node support and branch lengths (in green) are noted on the tree, which was developed using OrthoFinder. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

A single-gene ML tree was produced from the MAFFT aligned SSU region of both *Agmasoma* isolates alongside 209 other microsporidian parasite SSU sequences, using IQ-Tree (Nguyen et al., 2015), allowing for broader comparison across microsporidian species from the Enterocytozoonida and other higher taxa (Fig. 3). This tree highlighted that the *Agmasoma* isolates form a well-supported branch alongside *Agmasoma penaei* from *Penaeus* spp., at the base of the Enterocytozoonida. Comparatively, the SSU and genome trees show the same topology; however, the branch distance is greater between the *Agmasoma* isolates when considering the multi-gene output from OrthoFinder.

A draft genome for a microsporidian at the base of the Enterocytozoonida may prove valuable for further metabolomic work. Wiredu Boakye et al. (2017) compared a series of metabolic pathways present across several microsporidian genomes, highlighting that the glycolysis pathway had been almost entirely lost from members of the Enterocytozoonidae, but was largely maintained in the *V. corneae* genome. Comparison between the *V. corneae* glycolysis proteins using blastp revealed that the *Agmasoma* isolate from *C. aestuarii* appears to have maintained the majority of genes, presenting putative homologous proteins to the *V. corneae* pyruvate kinase (gene_451), phosphoglycerate mutase (gene_78), hexokinase (glucose ATP phosphotransferase) (gene_1542), pyruvate DHase E1 beta (gene_559), phosphofructokinase (gene_1588), glyceraldehyde 3-phosphate dehydrogenase (gene_1110), phosphoglucose isomerase (gene_511), and triose phosphate isomerase (gene_216). However, homologous proteins to the *V. corneae* (or other microsporidian) enolase (phosphopyruvate hydratase), fructose-bisphosphate aldolase, pyruvate DHase E1 alpha, and phosphoglycerate kinase, were not identifiable from either *Agmasoma* isolate. Other than the potential evolutionary loss of these proteins, it may be more likely that these proteins are encoded on a chromosomal region we have not sequenced, or perhaps too divergent for us to determine using available protein comparison methods. Use of InterProScan on potential protein candidates with < 40 % similarity to the missing proteins did not determine any conserved region that might suggest that they may function appropriately.

Given the detailed phylogenetic and genomic detail provided for these two isolates, it draws into question whether the field of Microsporidiology and microsporidian systematics may begin to use genomic data to parameterise species, genera, and higher taxonomic units. Several recent taxonomic studies of microsporidian parasites have drawn genetic parameters into their taxonomic descriptions, such as the description of *Parahepatospora carcini* (Bojko et al., 2017). In the case of these two new *Agmasoma*-like isolates, we show that the two share 100 % similar SSU genes (a common comparative gene for Microsporidia) and that their comparable genomes are similar (overall, 99.31 %), and that they branch closely in our phylogenetic and phylogenomic assessments. Such information is used as a basis for viral taxonomy and promoted by the International Committee for the Taxonomy of Viruses (ICTV) to great effect.

It may be time for microsporidian taxonomy to take important steps towards accepting the wealth of taxonomic detail that genomic data provides. This suggests a movement away from reliance on developmental and ultrastructural measurements when considering systematic descriptions, in some cases. Conducting microscopy work would instead be necessary to forward our understanding of the microsporidian infection mechanism and host-parasite interactome. We (informally, at this time) suggest that the species we present here be termed *Agmasoma carcini* (with two variants: *Ac* var. *aestuarii* and *Ac* var. *maenas*) and use genomic data as the primary feature for this systematic identification. In this case, additional pathology data is provided by Frizzera et al. (2021) for this *Carcinus*-infecting parasite, supplying a more classical element of systematic description, but still lacking ultrastructure and developmental information.

We conclude that this microsporidian parasite is a basal member of

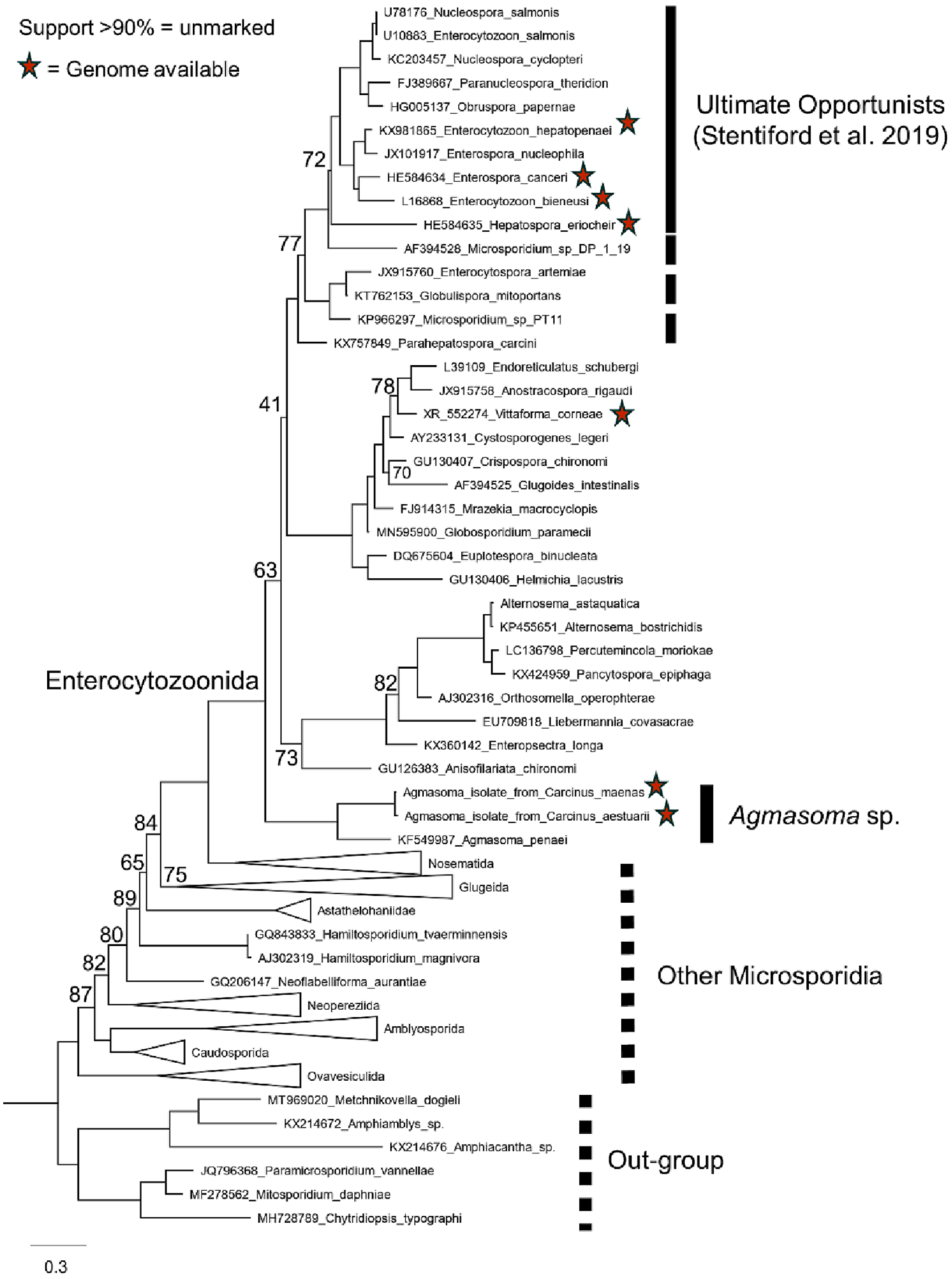

Fig. 3. A maximum-likelihood phylogenetic tree of microsporidian species focussing on the Enterocytozoonida, using SSU sequences. Bootstrap values are absent if the value was above 90 %, otherwise the value is stated at the node. Additional annotation is provided, including details from recent literature, identification of the different groups (*Agmasoma* sp.; Other Microsporidia; Out-group), and a star to represent those Enterocytozoonida that have genomic data available. The tree was developed using IQ-Tree.

the Enterocytozoonida, related to *Agmasoma penaei*, and a single microsporidian species (informally termed *Agmasoma carcini*) that has been determined to infect two related host crabs from the same brachyuran genus, based on close phylogenetic topology of the two isolates, their closely related SSU region, and similar protein coding gene sequences and protein products. The genome of the more complete *C. aestuarii* microsporidian isolate is the second largest of the Enterocytozoonida, with the first being *H. eriocheir*. We suggest that the two parasites we sequence could be referred to as variants (*Agmasoma carcini* var. *aestuarii* and *Agmasoma carcini* var. *maenas*), since there is noticeable, strain-level, variation in their genome sequences despite being largely similar. Further sequencing of isolates at highly variable sites in the genome could provide a range of valuable diagnostic methods for determining whether individual strains are more competent in either host species and therefor help to predict the influence of the parasite in these invasive brachyuran populations.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Contigs belonging to the *Agmasoma carcini* var. *aestuarii* isolate can be found in GenBank under accessions: JAOXFI000000000; PRJNA888958; and SAMN31227687. Contigs belonging to the *Agmasoma carcini* var. *maenas* isolate can be found in GenBank under accessions: JAOXFJ000000000; PRJNA888961; and SAMN31227688.

Acknowledgements

This work was conducted as part of a GCRF (UKRI) grant issued to all authors. JB would like to thank the Hipergator team at the University of Florida for use of the high-performance computer to analyse the genomic data.

References

Akiyoshi, D.E., Morrison, H.G., Lei, S., Feng, X., Zhang, Q., Corradi, N., Tzipori, S., 2009. Genomic survey of the non-cultivatable opportunistic human pathogen, *Enterocytozoon bieneusi*. PLoS Pathogens 5 (1), e1000261.

Azevedo, C., 1987. Fine structure of the microsporidian *Abelspora portucalensis* gen. n., sp. n. (Microsporida) parasite of the hepatopancreas of *Carcinus maenas* (Crustacea, Decapoda). J. Invertebrate Pathol. 49(1), 83–92.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19 (5), 455–477.

Bengtsson-Palme, J., Hartmann, M., Eriksson, K.M., Pal, C., Thorell, K., Larsson, D.G.J., Nilsson, R.H., 2015. METAXA2: improved identification and taxonomic classification of small and large subunit rRNA in metagenomic data. Mol. Ecol. Resour. 15 (6), 1403–1414.

Besemer, J., Lomsadze, A., Borodovsky, M., 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res. 29 (12), 2607–2618.

Bojko, J., Clark, F., Bass, D., Dunn, A.M., Stewart-Clark, S., Stebbing, P.D., Stentiford, G. D., 2017. *Parahepatospora carcini* n. gen., n. sp., a parasite of invasive *Carcinus maenas* with intermediate features of sporogony between the Enterocytozoon clade and other Microsporidia. J. Invertebr. Pathol. 143, 124–134.

Bojko, J., Stentiford, G.D., Stebbing, P.D., Hassall, C., Deacon, A., Cargill, B., Dunn, A.M., 2019. Pathogens of *Dikerogammarus haemobaphes* regulate host activity and survival, but also threaten native amphipod populations in the UK. Dis. Aquat. Organ. 136 (1), 63–78.

Bojko, J., Burgess, A.L., Baker, A.G., Orr, C.H., 2021. Invasive non-native crustacean symbionts: diversity and impact. J. Invertebr. Pathol. 186, 107482.

Bojko, J., Reinke, A.W., Stentiford, G.D., Williams, B., Rogers, M.S., Bass, D., 2022. Microsporidia: a new taxonomic, evolutionary, and ecological synthesis. Trends Parasitol. 38, 642–659.

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

Burgess, A., Bojko, J., 2022. Microsporidia are coming: *Cucumispora ornata* and *Dictyocoela berillonum* invade Northern Britain. BioInvasions Records 11.

Emms, D.M., Kelly, S., 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 20 (1), 1–14.

Frizzera, A., Bojko, J., Cremonte, F., Vázquez, N., 2021. Symbionts of invasive and native crabs, in Argentina: the most recently invaded area on the Southwestern Atlantic coastline. J. Invertebr. Pathol. 184, 107650.

Mittleider, D., Green, L.C., Mann, V.H., Michael, S.F., Didier, E.S., Brindley, P.J., 2002. Sequence survey of the genome of the opportunistic microsporidian pathogen, *Vittaforma corneae*. J. Eukaryot. Microbiol. 49 (5), 393–401.

Murareanu, B.M., Sukhdeo, R., Qu, R., Jiang, J., Reinke, A.W., 2021. Generation of a microsporidia species attribute database and analysis of the extensive ecological and phenotypic diversity of microsporidia. MBio 12 (3), e01490-21.

Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32 (1), 268–274.

R Core Team, 2013. R: A Language and Environment for Statistical Computing.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31 (19), 3210–3212.

Sprague, V., Couch, J., 1971. An annotated list of protozoan parasites, hyperparasites, and commensals of decapod Crustacea. J. Protozool. 18 (3), 526–537.

Stentiford, G.D., Bateman, K.S., Feist, S.W., Chambers, E., Stone, D.M., 2013. Plastic parasites: extreme dimorphism creates a taxonomic conundrum in the phylum Microsporidia. Int. J. Parasitol. 43 (5), 339–352.

Vivares, C.P., Bouix, G., Manier, J.F., 1977. Ormieresia carcini gen. n., sp. n., Microsporidie du Crabe Méditerranéan, *Carcinus mediterraneus* Czerniavsky, 1884: Cycle Évolutif et Étude Ultrastructurale. J. Protozool. 24(1), 83–94.

Wiredu Boakye, D., Jaroenlak, P., Prachumwat, A., Williams, T.A., Bateman, K.S., Itsathitphaisarn, O., Williams, B.A., 2017. Decay of the glycolytic pathway and adaptation to intranuclear parasitism within Enterocytozoonidae microsporidia. Environ. Microbiol. 19 (5), 2077–2089.