



Draft Genome Sequences of *Thelohania contejeani* and *Cucumispora dikerogammari*, Pathogenic Microsporidia of Freshwater Crustaceans

Alexandre Cormier,^a Rémi Wattier,^b Isabelle Giraud,^a Maria Teixeira,^b Frédéric Grandjean,^a Thierry Rigaud,^b  Richard Cordaux^a

^aLaboratoire Ecologie et Biologie des Interactions, Université de Poitiers, UMR CNRS 7267, Poitiers, France

^bLaboratoire Biogéosciences, Université Bourgogne Franche-Comté, UMR CNRS 6282, Dijon, France

ABSTRACT We announce the draft genome sequences of two pathogenic microsporidia of European freshwater crustaceans, *Thelohania contejeani* (the causative agent of porcelain disease) and *Cucumispora dikerogammari*. Both species are implicated in mass mortalities in natural populations of their crayfish and amphipod hosts, respectively.

Microsporidia are unicellular eukaryotes related to Fungi, specialized in intracellular parasitism (1). They are abundant in aquatic organisms, in which they can cause important disease, including crustaceans (2). Here, we sequenced the genomes of two major microsporidian pathogens of freshwater crustaceans, *Thelohania contejeani* (3) and *Cucumispora dikerogammari* (4). *T. contejeani* is the causative agent of porcelain disease in marine and freshwater decapods (5, 6). It is the most frequently encountered microsporidian parasite in European crayfish species. In the advanced phase of the disease, the abdominal muscles of infected individuals have an opaque white color. It is a chronic disease with low levels of infection in healthy populations (<2%), but it can reach up to 30% and cause population crashes (5, 6). *C. dikerogammari* is among the most recently discovered

TABLE 1 Assembly statistics and accession numbers of the genome sequences of *Thelohania contejeani* and *Cucumispora dikerogammari*

Metric	Data for:	
	<i>Thelohania contejeani</i>	<i>Cucumispora dikerogammari</i>
Assembly size (bp)	10,381,894	32,414,047
No. of contigs	1,391	7,783
N_{50} (bp)	27,037	9,513
Sequencing depth (median, bp)	3,663	1,129
G+C content (%)	26.97	26.08
Proportion of repeats (%)	32.64	36.80
No. of genes	2,865	4,599
Gene density (genes/kb)	0.28	0.14
Mean CDS ^a length (bp)	1,012	875
BUSCO ($n = 518$)		
No. (%) of complete genes	488 (94.2)	268 (51.8)
No. (%) of complete and single-copy genes	485 (93.6)	261 (50.4)
No. (%) of complete and duplicated genes	3 (0.6)	7 (1.4)
No. (%) of fragmented genes	21 (4.1)	31 (6.0)
No. (%) of missing genes	9 (1.7)	219 (42.2)
GenBank accession no.	SBIQ00000000	SBJP00000000
SRA accession no.	SRR8476225 , SRR8476226	SRR8495097 , SRR8495098

^aCDS, coding DNA sequence.

Citation Cormier A, Wattier R, Giraud I, Teixeira M, Grandjean F, Rigaud T, Cordaux R. 2021. Draft genome sequences of *Thelohania contejeani* and *Cucumispora dikerogammari*, pathogenic microsporidia of freshwater crustaceans. Microbiol Resour Announc 10:e01346-20. <https://doi.org/10.1128/MRA.01346-20>.

Editor Christina A. Cuomo, Broad Institute

Copyright © 2021 Cormier et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Richard Cordaux, richard.cordaux@univ-poitiers.fr.

Received 25 November 2020

Accepted 17 December 2020

Published 14 January 2021

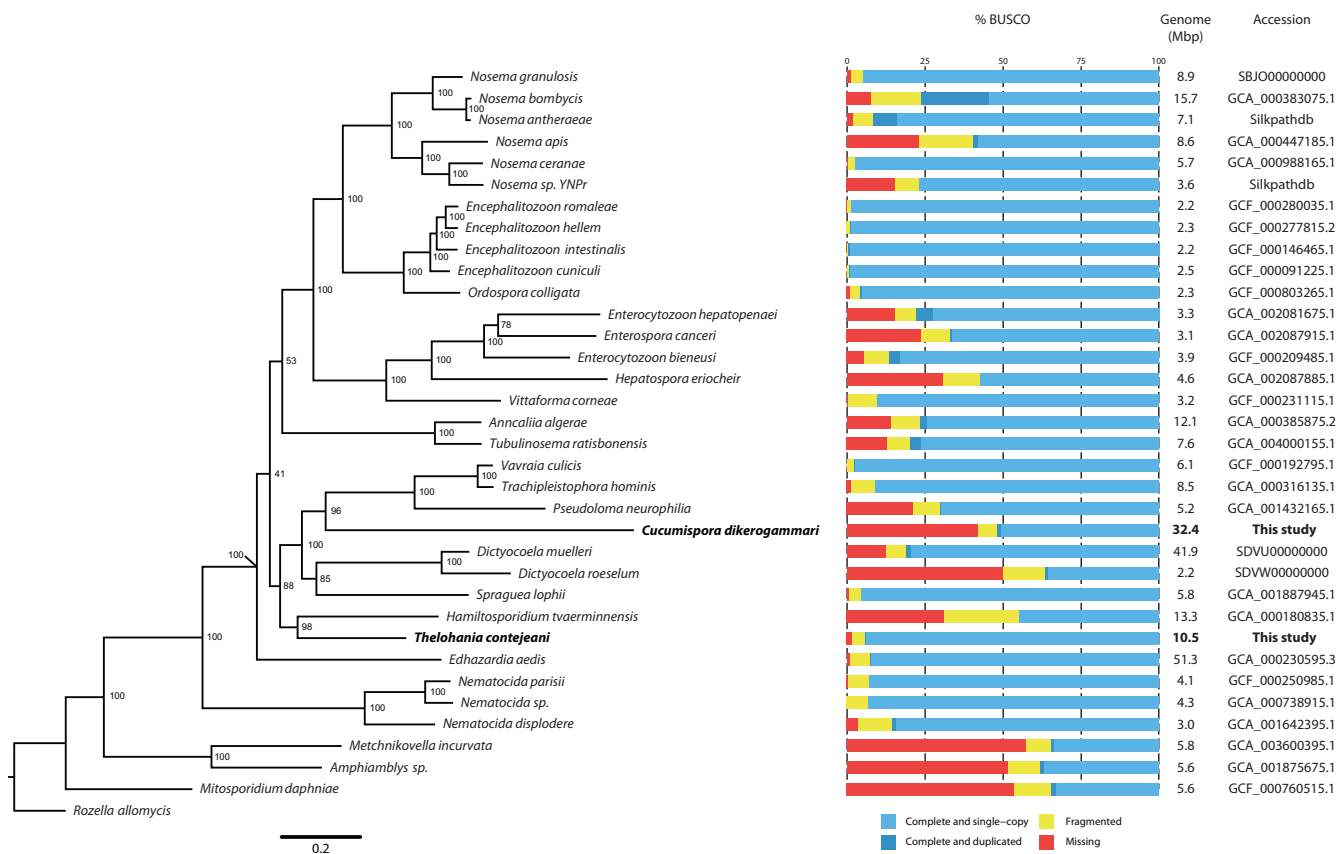


FIG 1 Phylogenomic analysis of microsporidia (left) along with genome assembly completeness (BUSCO, middle) and genome size and accession numbers (right). The tree was reconstructed using maximum likelihood with RAxML v8.2.9 under the LG+GAMMA model (19), based on the 68 single-copy orthologous genes (18,018 amino acids) identified in at least 31 species using OrthoFinder v2.2.7 (20). Protein sequences were aligned using MAFFT v7.299b in automatic mode (21). Individual alignments were trimmed using Gblocks (22) (-t=p; -p=n; -b3=8; -b4=2; -b5=h) and then concatenated into a single alignment using FASconCAT v1.0 (23). *Thelohania contejeani* and *Cucumispora dikerogammari* are bolded. Bootstrap values (percent) are indicated at each node (1,000 replicates). The scale bar indicates changes per site.

microsporidian parasites of crustaceans (4). It was originally described in the Ponto-Caspian amphipod *Dikerogammarus villosus*, and related parasites have since been reported in other amphipod species (7, 8). It causes severe disease in infected hosts through muscle invasion, resulting in an opaque white color (4), ultimately inducing high mortality levels (9). It is considered an emerging disease in European rivers (10).

Specimens of the crayfish *Austropotamobius pallipes* infected with *T. contejeani* and of the amphipod *D. villosus* infected with *C. dikerogammari* were collected in the Veude River (46.863192°N, 0.410296°E) and the Rhine River (47.814543°N, 7.546054°E), respectively. Abdominal muscles were punctured with a needle to collect microsporidian spores. Genomic DNA was directly extracted using the Qiagen DNeasy blood and tissue kit, according to the protocol for animal tissues. Paired-end libraries were constructed using the NEB Ultra II library prep kit and sequenced on an Illumina HiSeq 3000 instrument (2 × 150-bp reads) by Genome Québec, producing 232,008,634 and 252,071,596 raw paired-end reads for *T. contejeani* and *C. dikerogammari*, respectively. The raw reads were *de novo* assembled using SPAdes v3.12.0 (11) with default parameters, except for -k 21,33,55,77,99,127. The assemblies were checked for potential contamination with BlobTools v1.0 (12), MaxBin v2.2.4 (13), and PhylOligo v1.0 (14, 15). Gene structures and repeats were identified using GeneMarkS v4.30 (intronless eukaryotic parameter) (16) and RepeatModeler v1.0.8 (<http://www.repeatmasker.org>), respectively. Further details are available in reference 15.

The *T. contejeani* assembly was composed of 1,391 contigs with an N_{50} value of 27,037 bp and a length of 10.4 Mbp (Table 1). Genome completeness was assessed

using Benchmarking Universal Single Copy Orthologs (BUSCO) v3.0 (microsporidian set) (17), revealing that 98.3% of genes were present in the assembly. *T. contejeani* is among the most complete microsporidian genomes sequenced to date (Fig. 1). A phylogenomic analysis indicated that *T. contejeani* is related to the crustacean-borne *Hamiltosporidium tvaerminnensis* (18) (Fig. 1). The *C. dikerogammari* assembly was composed of 7,783 contigs with an N_{50} value of 9,513 bp and a length of 32.4 Mbp (Table 1). A phylogenomic analysis indicated that *C. dikerogammari* is highly divergent from the most closely related species (Fig. 1), which may help explain its BUSCO score of 57.7%.

Data availability. The genome projects and raw sequence data sets are available at DDBJ/EMBL/GenBank and the SRA under the accession numbers provided in Table 1.

ACKNOWLEDGMENTS

We thank Bouziane Moumen for assistance with the bioinformatic analyses and the Roscoff Bioinformatics platform ABiMS (<http://abims.sb-roscoff.fr>) for providing additional computational resources.

This work was funded by Agence Nationale de la Recherche grant ANR-15-CE32-0006 (CytoSexDet) to R.C. and T.R., the 2015 to 2020 State-Region Planning Contract and European Regional Development Fund, and intramural funds from the CNRS and the University of Poitiers.

REFERENCES

- Vávra J, Lukeš J. 2013. Microsporidia and “the art of living together.” *Adv Parasitol* 82:253–319.
- 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. <https://doi.org/10.1016/j.pt.2013.08.005>.
- Henneguy F, Thélohan P. 1892. Myxosporidies parasites des muscles chez quelques crustacés décapodes. *Ann Microgr* 4: 617–641. (In French.)
- Ovcharenko MO, Bacela K, Wilkinson T, Ironside JE, Rigaud T, Wattier RA. 2010. *Cucumispora dikerogammari* n. gen. (Fungi: Microsporidia) infecting the invasive amphipod *Dikerogammarus villosus*: a potential emerging disease in European rivers. *Parasitology* 137:191–204. <https://doi.org/10.1017/S0031182009991119>.
- Souty-Grosset C, Holdich DM, Noel PY, Reynolds JD, Haffner P (ed). 2006. Atlas of crayfish in Europe. University of Chicago Press, Chicago, IL.
- Grandjean F, Cornuault B, Archambault S, Bramard M, Otrebsky G. 2000. Life history and population biology of the white-clawed crayfish, *Austropotamobius pallipes pallipes*, in a brook from the Poitou-Charentes region (France). *BFPF Bull Fr Peche Prot Milieux Aquat* 356:55–70. (In French.) <https://doi.org/10.1051/kmae:2000004>.
- Bojko J, Dunn AM, Stebbing PD, Ross SH, Kerr RC, Stentiford GD. 2015. *Cucumispora ornata* n. sp. (Fungi: Microsporidia) infecting invasive “demon shrimp” (*Dikerogammarus haemobaphes*) in the United Kingdom. *J Invertebr Pathol* 128:22–30. <https://doi.org/10.1016/j.jip.2015.04.005>.
- Bojko J, Bacela-Spychalska K, Stebbing PD, Dunn AM, Grabowski M, Rachalewski M, Stentiford GD. 2017. Parasites, pathogens and commensals in the “low-impact” non-native amphipod host *Gammarus roeselii*. *Parasit Vectors* 10:193. <https://doi.org/10.1186/s13071-017-2108-6>.
- Bacela-Spychalska K, Wattier RA, Genton C, Rigaud T. 2012. Microsporidian disease of the invasive amphipod *Dikerogammarus villosus* and the potential for its transfer to local invertebrate fauna. *Biol Invasions* 14:1831–1842. <https://doi.org/10.1007/s10530-012-0193-1>.
- Wattier RA, Haine ER, Beguet J, Martin G, Bollache L, Muskó IB, Platvoet D, Rigaud T. 2007. No genetic bottleneck or associated microparasite loss in invasive populations of a freshwater amphipod. *Oikos* 116:1941–1953. <https://doi.org/10.1111/j.0030-1299.2007.15921.x>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
- Laetsch DR, Blaxter ML. 2017. BlobTools: interrogation of genome assemblies. *F1000Res* 6:1287. <https://doi.org/10.12688/f1000research.12232.1>.
- Wu Y-W, Tang Y-H, Tringie SG, Simmons BA, Singer SW. 2014. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome* 2:26. <https://doi.org/10.1186/2049-2618-2-26>.
- Mallet L, Bitard-Feidel T, Cerutti F, Chiappello H. 2017. PhylOligo: a package to identify contaminant or untargeted organism sequences in genome assemblies. *Bioinformatics* 33:3283–3285. <https://doi.org/10.1093/bioinformatics/btx396>.
- Cormier A, Chebbi MA, Giraud I, Wattier R, Teixeira M, Gilbert C, Rigaud T, Cordaux R. 20 November 2020. Comparative genomics of strictly vertically transmitted, feminizing microsporidia endosymbionts of amphipod crustaceans. *Genome Biol Evol* <https://doi.org/10.1093/gbe/evaa245>.
- Besemer J, Lomsadze A, Borodovsky M. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res* 29:2607–2618. <https://doi.org/10.1093/nar/29.12.2607>.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>.
- Haag KL, Larsson JIR, Refardt D, Ebert D. 2011. Cytological and molecular description of *Hamiltosporidium tvaerminnensis* gen. et sp. nov., a microsporidian parasite of *Daphnia magna*, and establishment of *Hamiltosporidium magnivora* comb. nov. *Parasitology* 138:447–462. <https://doi.org/10.1017/S0031182010001393>.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>.
- Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* 16:157. <https://doi.org/10.1186/s13059-015-0721-2>.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>.
- Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* 56:564–577. <https://doi.org/10.1080/10635150701472164>.
- Kück P, Longo GC. 2014. FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Front Zool* 11:81. <https://doi.org/10.1186/s12983-014-0081-x>.