The highly repetitive genome of *Myxobolus* sp., a myxozoan parasite of fathead minnows

Viraj R. Muthye[1,2], Aralia Leon Coria[2,3], Hongrui Liu[2,3], Constance A. M. Finney[2,3], Cameron P. Goater[2,4], James D. Wasmuth[1,2,*ra]

1. Faculty of Veterinary Medicine, University of Calgary
2. Host-Parasite Interactions Research Training Network, University of Calgary
3. Department of Biological Sciences, University of Calgary
4. Department of Biological Sciences, University of Lethbridge
* Corresponding author: jwasmuth@ucalgary.ca

Keywords: myxozoa, parasite, genome

Abstract

Background

The Myxozoa is a group of at least 2,400 endoparasites within the phylum Cnidaria. All myxozoans have greatly reduced in size and morphology compared to free-living members of the phylum. They are best known for causing disease in economically important fish across the world; for example, *Myxobolus cerebralis* causes whirling disease, which can kill 90% of infected juvenile salmonid fish. In 2017, a new myxozoan species was identified in Alberta. *Myxobolus* sp. causes distinct lesions in fathead minnows, which are ultimately fatal. Here, we sequenced, assembled and analyzed the genome of *Myxobolus* sp. to understand how the parasite interacts with its fish host and identify potential strategies to counter this emerging threat.

Results

At 185 Mb, the *Myxobolus* sp. genome is the largest myxozoan genome sequenced so far. This large genome size is, in part, due to the high repetitive content; 68% of the genome was interspersed repeats, with the MULE-MuDR transposon covering 18% of the *Myxobolus* sp. genome. Similar to myxozoan genomes, the *Myxobolus* sp. genome has lost many genes well conserved in other eukaryotes. However, we also identified multiple expansions in gene families—serine proteases, hexokinases, and FLYWCH-domain containing proteins—which suggests their functional importance in the parasite. The mitochondrial genome of *Myxobolus* sp. encodes only five of the thirteen protein-coding genes typically found in animals. We found that the mitochondrial gene *atp6* was transferred to the nucleus and acquired a mitochondria-targeting signal in *Myxobolus* sp.

Conclusions

Our study provides valuable insights into myxozoan biology and identify promising avenues for future research. We also propose that *Myxobolus* sp. is promising myxozoan model to explore host-parasite interactions in these parasites.

1. Introduction

Myxozoa (phylum Cnidaria) are an enigmatic group of approximately 2,400 species of intracellular parasites, which have drastically simplified their morphology, compared to their free-living relatives: jellyfish, corals, and sea anemones (Alama-Bermejo & Holzer, 2021) (Figure 1A). Myxozoans are found in both freshwater and marine environments, and cycle between invertebrate (definitive) and vertebrate (intermediary) hosts (reviewed in (Eszterbauer et al., 2015)). While myxozoans have been identified in birds (Bartholomew et al., 2008), amphibians (Eiras, 2005), reptiles (Eiras, 2005), and even mammals (Friedrich et al., 2000), they are infamous for causing disease in fish, and pose significant ecological, conservation, and economic threats globally (Alama-Bermejo & Holzer, 2021; Hutchins et al., 2021; James et al., 2021; Koel et al., 2006; Sarker et al., 2015).

Myxobolus (Bütschli 1882) is the most species-rich genus in Myxozoa, with approximately 850 species described so far (Liu et al., 2019). Arguably the most well-known member of the genus is *Myxobolus cerebralis* (Hofer 1903), the causative agent of whirling disease in salmonid fish (Bartholomew & Reno, 2002). *M. cerebralis* attacks the cartilage of juvenile rainbow trout (*Oncorhynchus mykiss*), leading to skeletal deformities. Its definitive host is the oligochaete *Tubifex tubifex* (sludge worm). In 2016, whirling disease was first detected in Canada, in Johnson Lake in Banff National Park, Alberta (Canadian Food Inspection Agency, 2019), and *M. cerebralis* has now been found in four major Albertan water sheds, with approximately 85% of wild rainbow trout fingerlings in the Crowsnest River showing pathological signs of disease (James et al., 2021). This poses severe social, economic, and conservation concerns. The Crowsnest River rainbow trout are sympatric with the endangered westslope cutthroat trout, *Oncorhynchus clarkii lewisi*, which are also susceptible to *M. cerebralis*, raising concerns for their conservation (Koel et al., 2006). Further, rainbow trout is the third most valuable recreational fish species in Alberta, a sport that generates more than $1.1 billion/year for Alberta (Veillard & Clayton, 2020). Multiple other members of the genus cause disease in fish, for example *Myxobolus honghuensis* which infects the pharynx of the economically important gibel carp (*Carassius auratus gibelio*) and *Myxobolus holzerae,* which causes severe gill disease in an Indian major carp (*Labeo rohita*) (Liu et al., 2012; Gupta & Kaur, 2017).

In 2017, a potential new member of the genus, *Myxobolus* sp., was discovered in fathead minnows (*Pimephales promelas*) in southern Alberta (Tilley *et al*. submitted, 2024). As with *M. cerebralis, Myxobolus* sp. uses *T. tubifex* as a definitive host (Figure 1B). Fathead minnows infected with *Myxobolus sp.* have large lesions in the head, ocular, oral, and sinus cavities, leading to reduced olfactory acuity, and increased fatigue, hyperactivity and death compared to uninfected minnows. It is too soon to know the population-level effect on fathead minnows. A concern is that *Myxobolus sp.* could—or already has—spread to other sympatric cyprinid species, including the western silvery minnow (*Hybognathus argyritis*), which is already designated as threatened in Canada (COSEWIC Assessment and Status Report, 2017). Further, fathead minnows have their own important place in the aquatic ecosystem, as a food source for larger fish, and feeding on algae thereby limiting algal overgrowth.

Despite their environmental and economic importance, myxozoans remain under studied. The diagnoses, monitoring, and treatment of parasitic diseases has greatly benefited from knowledge of the genome—and other omics data—for the causative organisms. Myxozoans have been relatively slow to enter the genomic age. To date, genomes have been published for only eight myxozoan species, with transcriptomes available for a further eight species (reviewed in (Alama-Bermejo & Holzer, 2021)). These resources are assisting myxozoan researchers answer questions related to virulence and diversity. Further, of the ~2400 myxozoan species, only five have a laboratory-confirmed life cycles with just another fifty having partially describe life cycles (Eszterbauer et al., 2015). This knowledge gap has limited efforts to explore myxozoan-host interactions and develop strategies for myxozoan control.

In this study, we set out to sequence, assemble, and annotate the genome of *Myxobolus sp.*. This is only the second time that long-read sequencing technology has been used for the myxozoans, the first being *Myxobolus honghuensis* (Guo et al., 2022). The assembled genome is the largest of any myxozoan sequenced to date. The use of long-read sequencing allowed us to explore the repeat landscape of the parasite genome which showed that *Myxobolus sp.* has one of the most repetitive animal genomes. We compared the genome with other myxozoan and free-living cnidarians and identified multiple gene families that have undergone expansion in *Myxobolus sp.*, like serpins, peroxiredoxins, and hexokinases. In addition, we also present the mitochondrial genome of *Myxobolus sp.* and show that the parasite possesses a highly reduced mitochondrial genome and mitochondrial proteome.

Note: In a previous version of this preprint, we used a species epithet that is under peer-review (February 2024). Our inclusion of it in the preprint was not intended for the purposes of zoological nomenclature. We have removed it in this current version.

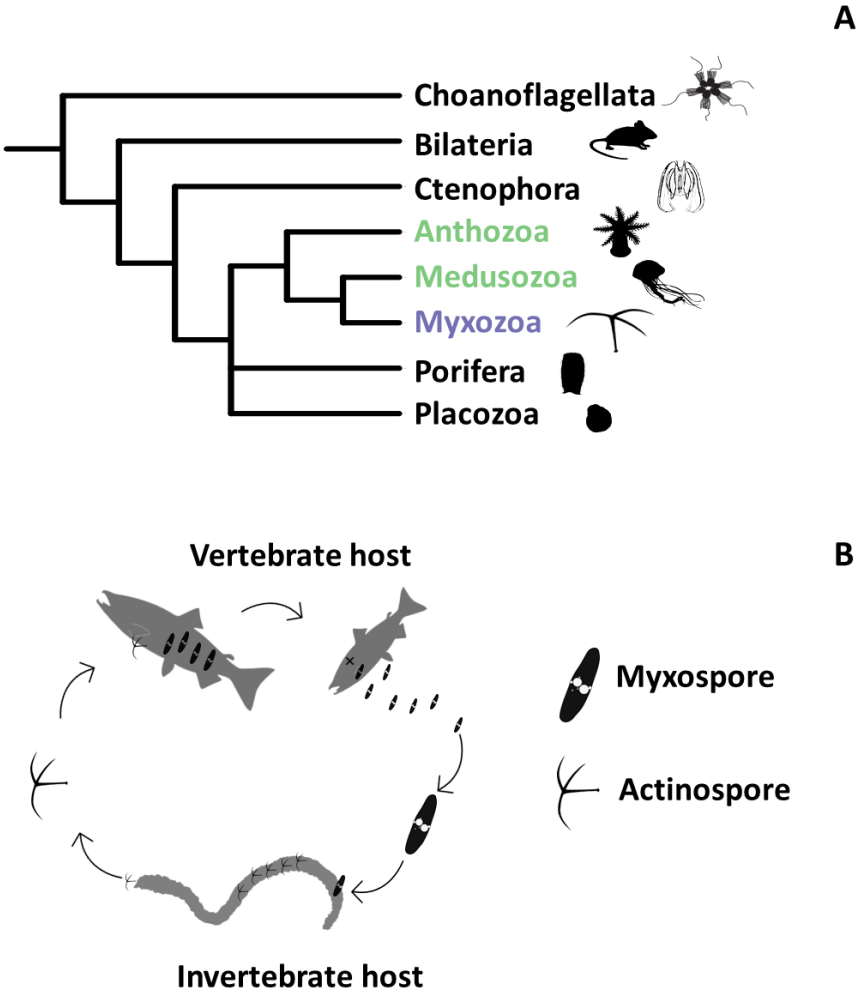


Figure 1. A) Phylogenetic relationships between the five metazoan groups (Porifera, Cnidaria, Ctenophora, Placozoa, and Bilateria) and the outgroup Choanoflagellata, where the free-living classes of phylum Cnidaria are colored in green and Myxozoa is colored in purple. Silhouettes for the animals are taken from PhyloPic (www.phylopic.org). B) A simplified life cycle of *Myxobolus sp.*.

2. Methods and Materials

2.1. Sample Collection, DNA extraction, and Genome Sequencing

Myxobolus sp. myxospores were isolated from lesions in fathead minnows following the protocols established by Dr. Cameron Goater and Molly Tilley (Tilley *et al.* 2024 *submitted*). High molecular weight was extracted using the MagAttract® HMW DNA Qiagen kit (Ref.67563). Clean-up used a combination of two kits: the beads from the Select-a-size Magbead kit (Zymo Cat# D4084-3-10) and the buffers from Genomic DNA Clean & Concentrator -25 (Zymo Cat# D4064). A detailed protocol is attached as Supplementary File S1. For sequencing, the SQK-LSK112 Ligation Sequencing Kit (Q20+) from Oxford Nanopore Technologies (ONT) was used to generate the libraries. No changes were made to the steps outlined in the manual for the SQK-LSK112 Ligation Sequencing Kit. The Mk1C MinION sequencer from ONT was used to sequence the genome and Guppy v2.1.7 was used for base-calling with default options.

2.2. Genome assembly and contamination removal

To remove the host (fathead minnow) reads, all the base-called reads were mapped to the fathead minnow draft genome (GCF_016745375.1) using minimap2 v2.24 (default options) (Li, 2018; Martinson et al., 2022). Three assemblies for *Myxobolus sp.* from the unmapped reads using different assembly software:

1. FLYE v2.9 (Kolmogorov et al., 2019) ("flye –nano-hq <READSFILE> -o flye -t 48 -g 50m"),
2. wtdbg2 (Ruan & Li, 2020) ("wtdbg2 -x ont -g 50m -t 14 -i <READSFILE> -fo mras; wtpoa-cns -t 14 -i mras.ctg.lay.gz -fo mras.ctg.fa"),
3. Shasta (Shafin et al., 2020) ("shasta-Linux-0.10.0 –input <READSFILE> --threads 14 –config Nanopore-May2022 –Reads.minReadLength 5000").

BUSCO v5.4.3 (Eukaryota lineage) and N50 was used to select the preferred genome assembly for subsequent annotation and downstream analysis (Simão et al., 2015). The selected assembly was polished with MEDAKA v1.6.0 (https://www.github.com/nanoporetech/medaka). A further round of decontamination was performed using BlobTools v1.1 (Laetsch & Blaxter, 2017), Uniprot Reference Proteomes (downloaded 01 May 2023) (The UniProt Consortium, 2019), and DIAMOND BLASTX (Buchfink et al., 2021). All assembly contigs assigned as chordate, prokaryote, or viral origin were discarded.

151 ## 2.3. RNA extraction and *de novo* transcriptome assembly

152 Samples of fish muscles containing myxospores were lysed by adding 200 µl of Trizol followed
153 by manual homogenization in dry ice, for a total of three freeze-thaw cycles in a final volume of
154 600 µl. The resulting mixture was centrifuged, and the supernatant was collected into a new tube
155 where RNA was extracted using the Direct-zol RNA Miniprep kit (Zymo cat. no. R2050)
156 following the manufacturer's instructions. The extracted RNA was then cleaned using the RNA
157 Clean and Concentrator-5 kit (Zymo cat. no. R1015) following the manufacturer's instructions.
158 The quantity and quality of the total RNA were assessed on an Agilent 2200 TapeStation RNA
159 ScreenTape. Libraries for sequencing were prepared using the NEBNext Ultra II Directional
160 RNA Library Prep Kit with PolyA capture following the manufacturer's instructions. Libraries
161 were multiplexed and paired-end sequenced on Illumina NextSeq 500 for 300 cycles (2 × 150 bp)
162 using a mid-output cell, following the manufacturer's instructions. FastQC v0.11.9 was used to
163 assess the quality of the RNAseq reads before and after filtering reads (Andrews & others, 2010).
164 Adapter sequences and low-quality reads were identified with Trimmomatic v0.39 and removed
165 (Bolger et al., 2014). A *de novo* transcriptome assembly was generated with SPAdes v3.15.4
166 using default parameters (Bushmanova et al., 2019).
167

168 ## 2.4. Identification of repeat content and genome annotation

169 Interspersed repeats and low complexity DNA sequences in the *Myxobolus sp.* genome were
170 masked using RepeatModeller v2.0.1 (Flynn et al., 2020) and RepeatMasker v4.1.4
171 (www.repeatmasker.org) with default parameters. This soft-masked genome was used as input
172 for BRAKER v2.1.6 to predict protein-coding genes with protein evidence (Brůna et al., 2021).
173 The OrthoDB v10 metazoa protein database was used as the protein evidence for the genome
174 annotation software BRAKER (Kriventseva et al., 2019). BRAKER was also used to annotate
175 the four publicly available myxozoan genome without genome annotations and to re-annotate the
176 four myxozoan genomes which had publicly-available genome annotations. For this, BRAKER
177 was run using protein evidence (OrthoDB v10 metazoa protein database) (Table S2). BUSCO
178 v5.4.3 was used to analyze the completeness of the myxozoan genomes. The lineage used was
179 'eukaryota_odb10' (Creation date: 2020-09-10, number of genomes: 70, number of BUSCO
180 genes: 255). For myxozoan species for whom BRAKER did not improve existing genome
181 annotations based on BUSCO analysis, previously published genome annotations were used for
182 downstream analysis.
183

184 ## 2.5. Analysis of the mitochondrial genome and proteome

185 A database of mitochondrial genomes from four myxozoan species, *Polypodium hydriforme,* two
186 cnidarian species, human, *Caenorhabditis elegans*, *Drosophila melanogaster*, and yeast were
187 used to identify whether *Myxobolus sp.* has a mitochondrial genome (Table 1). BLASTN was

188 used to align these genomes to the *Myxobolus sp.* genome (e-value ≤ $1e^{-5}$) (Camacho et al.,
189 2009). Two methods were used to identify genes in the mitochondrial genome. First, MITOS2
190 was used to annotate the contig containing the mitochondrial genome with parameters listed in
191 Table 2 (Bernt et al., 2013). MITOS2 was also used to annotate the five publicly available
192 myxozoan mitochondrial genomes. Second, an HMM model approach was used to identify
193 protein-coding genes in the mitochondrial genome. HMM models for the thirteen protein coding
194 genes found in the majority of animal mitochondrial genomes were downloaded from the
195 eggNOG protein database V5.0.0 (Cantalapiedra et al., 2021). HMMer (hmmer.org) was used to
196 search the mitochondrial genome of *Myxobolus sp.* and the publicly-available myxozoan
197 mitochondrial genomes. PROKSEE was used to visualize the mitochondrial genome (Grant et
198 al., 2023).
199
200 Most proteins that function in the mitochondria are encoded in the nuclear genome and then
201 imported into the mitochondria (Meisinger et al., 2008). To identify these proteins, human
202 mitochondrial proteins catalogued in MitoCarta v3 (Rath et al., 2021) were downloaded.
203 BLASTP was used to align these human mitochondrial proteins to the proteins from all
204 myxozoan and free-living cnidarian nuclear genomes (e-value ≤ $1e^{-5}$).
205
206 **Table 1** Publicly available mitochondrial genomes used in this analysis to identify the
207 mitochondrial genome in *Myxobolus* sp.

Species	**GenBank ID**
Polypodium hydriforme	MN794187.1 (Novosolov et al., 2022)
Myxobolus squamalis	MK087050.1 (Yahalomi et al., 2020)
Kudoa hexapunctata	LC009437.1 (Takeuchi et al., 2015)
Kudoa septempunctata	AB731753.1 (Takeuchi et al., 2015)
Kudoa iwatai	LC009438.1 (Takeuchi et al., 2015)
Nematostella vectensis	OW052000.1 (Fletcher & Pereira da Conceicoa, Lyndell, n.d.)
Dendronephthya gigantea	NC_013573.1 (Park Eun-Ji et al., 2010)
Homo sapiens	NC_012920.1 (Anderson et al., 1981)
Caenorhabditis elegans	NC_001328.1 (Wolstenholme et al., 1994)
Drosophila melanogaster	NC_024511.2 (Hoskins et al., 2015)
Saccharomyces cerevisiae	NC_027264.1 (Wolters et al., 2015)

208
209 **Table 2** The parameters used in MITOS2 to annotate the mitochondria genome of *Myxobolus sp.*

Property	**Value**
Reference	RefSeq 63 Metazoa
Genetic Code	4
Proteins	TRUE
tRNAs	TRUE
rRNAs	TRUE
OH	TRUE
OL	TRUE

Circular	TRUE
Use Al Arab et al.	FALSE
E-value Exponent	2
Final Maximum Overlap	50nt
Fragment Quality Factor	100
Standard Code	FALSE
Cutoff	50.00%
Clipping Factor	10
Fragment Overlap	20.00%
Local only	TRUE
Sensitive only	FALSE
ncRNA overlap:	50 nt

2.6. Genome Analysis

2.6.1. Identifying the function of protein-coding genes

If a gene had multiple transcripts, the longest transcript and associated protein per gene was selected. eggNOG-mapper v2.1.9 (Cantalapiedra et al., 2021) to annotate the predicted proteins using default parameters, except two (percentage identity was changed to 30% and all queries were realigned to the PFAM database instead of importing PFAM annotations from orthologs). The output of eggNOG-mapper was parsed to extract 1) PFAM domains, 2) Gene Ontology (GO) terms, and 3) KEGG pathways mapped to each protein.

2.6.2. Gene family analysis

OrthoFinder v2.5.4 was used, with default parameters, to identify groups of orthologous proteins (OGs) between the predicted protein sets of *Myxobolus sp.*, six publicly-available myxozoan species, and seven cnidarian species (listed in Table 3) (Emms & Kelly, 2019). The sets of predicted proteins for the free-living cnidarian species were downloaded from UniProt (one protein per gene). We omitted two myxozoan species (*K. iwatai* and *S. zaharoni*) as they had a high rate of duplication and therefore OrthoFinder identified considerably lower number of single-copy orthologs.

The single-copy orthologs identified by OrthoFinder were used to analyze phylogenetic relationships between these 14 species. First, all single-copy orthologs were aligned using MAFFT ("--auto" option) (Katoh et al., 2002) and concatenated into a single alignment using the script "catfasta2phyml.pl". Next, the alignment was trimmed using trimAI (Capella-Gutiérrez et al., 2009). The resulting alignment was used as input for IQ-TREE v2.2.5 (Nguyen et al., 2015). The appropriate substitution model was identified using ModelFinderPlus ("-MFP" option)

235 (Kalyaanamoorthy et al., 2017) based on the Bayesian information criterion (BIC). Branch
236 supports were assessed with 1,000 standard ultrafast bootstrap approximation (UFBoot)
237 replicates (Hoang et al., 2018; Minh et al., 2013). All phylogenetic trees in this study were
238 visualized using iTOL (Letunic & Bork, 2021). CAFE5 (Computational Analysis of gene Family
239 Evolution) was used to identify rapidly expanding and contracting gene families (p-values <
240 0.05) using the OrthoFinder results and the IQ-TREE-generated phylogenetic tree (Mendes et al.,
241 2021).
242
243 Further, 98 proteomes from the class Actinopterygii were downloaded from UniProt. These were
244 used to identify *Myxobolus sp.* proteins that lack sequence similarity to proteins from
245 Actinopterygii. For this, *Myxobolus sp.* proteins were aligned to a database of proteins from
246 these 98 species (e-value ≤ $1e^{-3}$).
247
248 **Table 3** Details about the seven publicly-available myxozoan genomes used in this study

Species	Source	Reference
Sphaeromyxa zaharoni	GCA_001455285.1	(Chang et al., 2015)
Kudoa iwatai	GCA_001407235.2	(Chang et al., 2015)
Enteromyxum leei	GCA_001455295.2	(Chang et al., 2015; Yahalomi et al., 2017)
Myxobolus squamalis	GCA_010108815.1	(Yahalomi et al., 2020)
Henneguya salminicola	GCA_009887335.1	(Yahalomi et al., 2020)
Thelohanellus kitauei	GCA_000827895.1	(Yang et al., 2014)
Myxobolus honghuensis	Harvard Dataverse	(Guo, 2021; Guo et al., 2022)
Ceratonova shasta	GCA_918559645.1	

249

250 2.6.3. Search for minicollagens and nematogalectins

251 Protein sequences for minicollagens and nematogalectins from the study done by (Shpirer et al.,
252 2014) were downloaded and BLASTP was used to align them to the myxozoan genomes (e-value
253 ≤ $1e^{-5}$).
254

255 2.6.4. Identification of proteases/inhibitors

256 The MEROPS Scan Sequences database was downloaded, which is a non-redundant subset of
257 the MEROPS database that is used for the MEROPS batch BLAST searches (Rawlings et al.,
258 2010). BLASTP was used to align the proteins from all species in this study to the MEROPS
259 Scan database (e-value ≤ $1e^{-5}$).
260

261 2.6.5. Searching for homologs of drug target proteins from ChEMBL

262 All 9,072 single protein drug targets from ChEMBL 33 (www.ebi.ac.uk/chembl) were
263 downloaded. BLASTP was used to align *Myxobolus* sp. proteins to the database of these
264 ChEMBL targets (e-value $\leq 1e^{-10}$).

265

3. Results

3.1. Genome assembly and repeat content analysis

3.1.1. Genome assembly statistics

We extracted DNA from *Myxobolus sp.* myxospores that were isolated from lesions on the heads of fathead minnows. Approximately 3.2 million sequence reads (11.38 Gb) were generated from a single Oxford Nanopore run and passed base-calling filters. Of these, ~2.1 million reads (65%) could be mapped to the available genome assembly of the fathead minnow, and so were flagged as contamination and removed. The remaining ~1.1 million reads were used to assemble the *Myxobolus sp.* genome. To achieve the best possible assembly, we generated assemblies from three algorithms: FLYE, wtdbg2, and Shasta. By reviewing assembly metrics—the most complete BUSCO complement and assembly continuity—we selected the FLYE assembly for the subsequent analysis (Table 4). We used BlobTools to remove further contamination by discarding all contigs that were either mapped to bacteria, eukaryote, or viruses (92 contigs). The final *Myxobolus sp.* genome was comprised of 1,503 contigs (185 Mb with N50 of 407.91 Kb). At 185 Mbp, the *Myxobolus sp.* genome assembly is the largest of any myxozoan species sequenced so far (Table 5).

Table 4 Comparison of the *Myxobolus sp.* genome assembled using three genome assemblers (FLYE, wtdbg2, and Shasta) based on their BUSCO scores (Eukaryota lineage in the genome mode) and N50 values.

Assembler	Genome size (nt)	# Contigs	BUSCO (Eukaryota) % where n=225	N50
FLYE	190,986,082	1,595	C:32.2 [S:31. 4, D:0.8], F:19.2, M:48.6	383 kb
wtdbg2	178,055,605	1,282	C:28.2 [S:27.8, D:0.4], F:18.8, M:53.0	331 kb
Shasta	165,736,712	3,742	C:27.5 [S:27.1, D:0.4], F:17.3, M:55.2	80 kb

3.1.2. Repeat content analysis

Approximately 68% of the genome is comprised of interspersed repeats (Table 6). To confirm that the large repeat content in *Myxobolus sp.* was not an artefact of the genome assembler, we also analyzed the repeat content of the genomes assembled using two other tools- wtdbg2 and Shasta. We found that these two genome assemblies were highly repetitive as well (67.2% for the wtdbg2 assembly and 68.2% for the Shasta assembly). In addition, we compared the types and prevalence of different types of repeats with *Myxobolus sp.*. At the time of this analysis, the *M. honghuensis* assembly was the only other myxozoan built from long-reads, giving us

295 confidence in the assembly of repetitive loci. Using the same tools, we found that just over half
296 the *M. honghuensis* genome (51.0%) was comprised of repetitive regions.
297
298 The *Myxobolus sp.* transposable element (TE) landscape was dominated by DNA transposons,
299 which occupied 43.8 % of the genome and retrotransposons occupied a further 9.3 % of the
300 genome (Table 6). In contrast, the *M. honghuensis* assembly is comprised of 14.7 % DNA
301 transposons and 1% retrotransposons. Of the *Myxobolus sp.* DNA transposons, MULE-MuDR
302 transposons were the most abundant and were identified in 18.1% of the total genomic sequence.
303 Compared to *Myxobolus sp.*¸ we found that just 2.5% of the genome was occupied by MULE-
304 MuDR elements. It is important to note that more repeats in *M. honghuensis* were unclassified
305 (35.4%) compared to *Myxobolus sp.* (15.1%).
306
307 **Table 5** Genome statistics of all publicly-available myxozoan genomes and *Myxobolus sp.*

Species	Platform	# Contigs	Size (Mb)	Contig N50 (Kb)	BUSCO (genome mode, eukaryote dataset) % where n=225
Myxobolus sp.	Oxford Nanopore	1503	184.9	407.9	C:31.4 [S:31.0, D:0.4], F:19.2, M:49.4
Myxobolus honghuensis	PacBio + Illumina HiSeq 2500	1118	161.1	1273.5	C:32.6 [S:31.8, D:0.8], F:20.8, M:46.6
Myxobolus squamalis	Illumina HiSeq 3000	37921	43.7	1.3	C:26.3 [S:26.3, D:0.0], F:17.3, M:56.4
Sphaeromyxa zaharoni	Illumina HiSeq 2000	70935	173.6	4.5	C:30.6 [S:11.0, D:19.6], F:23.1, M:46.3
Kudoa iwatai	Illumina HiSeq 2000 / Illumina Genome Analyzer IIx	1646	31.2	39.5	C:31.8 [S:16.1, D:15.7], F:17.3, M:50.9
Enteromyxum leei	Illumina HiSeq 2000	69178	68.2	1.0	C:22.8 [S:20.4, D:2.4], F:21.6, M:55.6
Henneguya salminicola	Illumina HiSeq 3000	18332	61.4	1.2	C:32.9 [S:32.9, D:0.0], F:20.8, M:46.3
Thelohanellus kitauei	454 GS FLX + Illumina Genome Analyzer II	15632	150.3	13.0	C:32.2 [S:31.4, D:0.8], F:18.0, M:49.8
Ceratonova shasta	Illumina HiSeq 2000	24903	69.8	9.3	C:27.1 [S:25.1, D:2.0], F:8.2, M:64.7

308
309
310
311
312
313
314

Table 6 Results of RepeatModeller and RepeatMasker analysis for genomes of *Myxobolus* sp. and *Myxobolus honghuensis*

Repeat Elements		***Myxobolus* sp.**			***Myxobolus honghuensis***		
		# elements	**length occupied**	**% of sequence**	**# elements**	**length occupied**	**% of sequence**
Retroelements		14495	17182784	9.29	1121	1469706	0.91
	SINEs	0	0	0	0	0	0
	Penelope	0	0	0	0	0	0
	LINEs	532	258231	0.14	0	0	0
	LTR elements	13963	16924553	9.15	1121	1469706	0.91
DNA transposons		112916	80978766	43.79	43041	23662456	14.69
	hobo-Activator	6092	6575235	3.56	18240	11145247	6.92
	Tc1-IS630-Pogo	37716	22841177	12.35	12193	4972777	3.09
	En-Spm	0	0	0	0	0	0.00
	MULE-MuDR	39574	33412948	18.07	6488	4008242	2.49
	PiggyBac	41	30174	0.02	733	508393	0.32
	Tourist/Harbinger	0	0	0	0	0	0.00
	Other_(Mirage,P-element, Transib)	0	0	0	0	0	0.00
Rolling-circles		2705	949873	0.51	0	0	0
Unclassified		116134	27904502	15.09	166939	57023818	35.4
Total interspersed repeats			126066052	68.17		82155980	51.00
Small RNA		0	0	0	0	0	0
						`	
Satellite		0	0	0	0	0	0
Simple repeats		18636	870644	0.47	36080	1727836	1.07
Low complexity		8293	422235	0.23	12949	668608	0.42

3.2. Genome annotation and analysis

We identified 10,106 protein-coding genes in the *Myxobolus sp.* genome, which is around 65% of the number of genes in the *M. honghuensis* genome. The average intron length, exon length and gene lengths in *Myxobolus sp.* and *M. honghuensis* were similar (Table 7). We reannotated

322 all eight publicly-available myxozoan genomes (with protein evidence) (Table 8). Four of these
323 eight genomes lacked publicly-available annotations and a further two species— *Myxobolus*
324 *squamalis* and *Thelohanellus kitauei*—received improved annotations, based on their BUSCO
325 scores (Table 8).
326
327 **Table 7** Distribution of intron lengths, exon lengths, and gene lengths for *Myxobolus sp.* and *M.*
328 *honghuensis,* from BRAKER results

	Intron lengths		Exon length		Gene length	
	Myxobolus sp.	*M. honghuensis*	*Myxobolus sp.*	*M. honghuensis*	*Myxobolus sp.*	*M. honghuensis*
values	32,234	52,165	43,272	69,808	10,106	16,867
min	5	4	3	3	196	108
max	59,944	24,000	12,697	12,542	89,608	113,913
range	59,939	23,996	12,694	12,539	89,412	113,805
sum	24,066,630	36,410,620	11,571,580	17,172,630	33,931,130	51,515,720
median	22	28	155	150	1,007	1,208
mean	746.62	697.99	267.41	246.00	3,357.52	3,054.23
std.dev	3,262.03	1,701.45	397.77	345.06	6,684.27	4,959.44

329
330 **Table 8** BUSCO analysis of predicted proteins from myxozoan genomes using the eukaryote
331 database. For myxozoans with publicly-available annotations, the predicted protein sets from
332 both the original annotation and from our BRAKER annotation were assessed using BUSCO.
333 N.A. not available. Note 1: We could not find the annotations file from the publication.

Species	**Original annotations**		**BRAKER2 annotations**	
	Protein-coding genes	**BUSCO (Eukaryota) % where n=225**	**Protein-coding genes**	**BUSCO (Eukaryota) % where n=225**
Sphaeromyxa zaharoni	N.A.	N.A.	18,437	C:35.7 [S:13.7, D:22.0], F:23.5, M:40.8
Kudoa iwatai	5,533	N.A.[1]	6,921	C:67.1 [S:31.0, D:36.1], F:7.5, M:25.4
Enteromyxum leei	N.A.	N.A.	20,185	C:22.8 [S:21.6, D:1.2], F:24.7, M:52.5
Myxobolus squamalis	5,710	C:22.0 [S:22.0, D:0.0], F:19.2, M:58.8	4,538	C:29.8 [S:29.8, D:0.0], F:16.9, M:53.3
Henneguya salminicola	8,187	C:35.3 [S:35.3, D:0.0], F:18.8, M:45.9	4,163	C:32.6 [S:31.0, D:1.6], F:17.3, M:50.1
Thelohanellus kitauei	15,020	C:29.4 [S:28.6,D:0.8], F:15.7, M:54.9	9,498	C:37.3 [S:34.9, D:2.4], F:14.9, M:47.8
Ceratonova shasta	N.A.	N.A.	15,032	C:27.1 [S:25.1, D:2.0], F:8.2, M:64.7
Myxobolus honghuensis	15,433	C:51.0 [S:49.4, D:1.6], F:14.1, M:34.9	16,867	C:44.4 [S:42.4, D:2.0], F:16.9, M:38.7
Myxobolus sp.	-	-	10,106	C:44.3 [S:41.6, D:2.7], F:16.5, M:39.2

Next, we assigned functional annotations to 57% (5,757 of 10,106) *Myxobolus sp.* proteins with: GO terms to 2,611 proteins; KEGG pathways to 1,873 proteins; and PFAM domains to 4,671 proteins. Nearly all KEGG pathways assigned to at least one *Myxobolus sp.* protein (357/361 KEGG pathways) were common between *Myxobolus sp.* and *M. honghuensis.* Further, 340 pathways contained at least one protein from all myxozoan species. The number of pathways with at least one protein was similar across free-living and parasitic cnidarians, ranging from 349 in *C. shasta* to 382 in *T. kitauei* in myxozoans and from 390 in *C. hemisphaerica* to 405 in *S. pistillata.* When we looked at the 339 KEGG pathways with at least one protein from all myxozoan and free-living cnidarian species, we observed that, on average, free-living cnidarians had more proteins assigned to the pathway compared to myxozoans.

We assigned 1,729 PFAM domains to at least one *Myxobolus sp.* protein, of which 1,522 domains (88%) were also found in the genome of *M. honghuensis.* Of these 1,729 domains, nine were not identified in any of the free-living cnidarian species (Table 9). Ten proteins had the "Arg_repressor" (Arginine repressor, DNA binding domain) protein domain in *Myxobolus sp.* and *C. shasta,* two in *T. kitauei* and none in any free-living cnidarians analyzed in this study.

Table 9 Number of myxozoan proteins containing the nine domains that were identified in at least one *Myxobolus sp.* protein but none of the free-living cnidarians used in the study

Pfam domain identifier	*Myxobolus sp.*	*M. honghuensis*	*M. squamalis*	*C. shasta*	*E. leei*	*H. salminicola*	*K. iwatai*	*S. zaharoni*	*T. kitauei*
Arg_repressor	10	0	0	10	0	0	0	0	2
Terminase_5	4	0	0	0	0	0	0	0	0
DUF1398	1	0	0	0	0	0	0	0	0
Lambda_Bor	1	0	0	0	0	0	0	0	0
Mononeg_RNA_pol	1	0	0	0	0	0	0	3	0
Phage_holin_3_1	1	0	0	0	0	0	0	0	0
Phage_lysis	1	0	0	0	0	0	0	0	0
RdRP_2	1	0	0	0	0	0	0	0	0
Saw1	1	0	0	0	0	0	0	0	0

A further 22 PFAM domains were present in more *Myxobolus sp.* proteins compared to free-living cnidarians (the number of proteins with a domain in *Myxobolus sp.* was at least three times the average number of proteins containing the domain in free-living cnidarian species). From these 22 domains, "RVP_2", "DDE_Tnp_IS1595", and "FLYWCH" were the top three most abundant protein domains. Proteins with the FLYWCH domain are a part of a superfamily of

DNA transposon-derived conserved families of transcription factors. The number of FLYWCH domains ranged from none in three myxozoans to 55 in *Myxobolus sp.* in myxozoans, and none in two free-living cnidarians to three in *S. pistillata* in free-living cnidarians. In *Myxobolus sp.,* the majority of proteins with the FLYWHC domain also contained the MULE transposase protein domain.

We found that both *Myxobolus sp.* and *M. honghuensis* had more proteins containing the "Hexokinase_1" and "Hexokinase_2" protein domains when compared to free-living cnidarians. *Myxobolus sp.* had six proteins containing both domains. Phylogenetic analysis of all proteins that contained the two domains showed that the family underwent duplication in these two species (Figure 2).

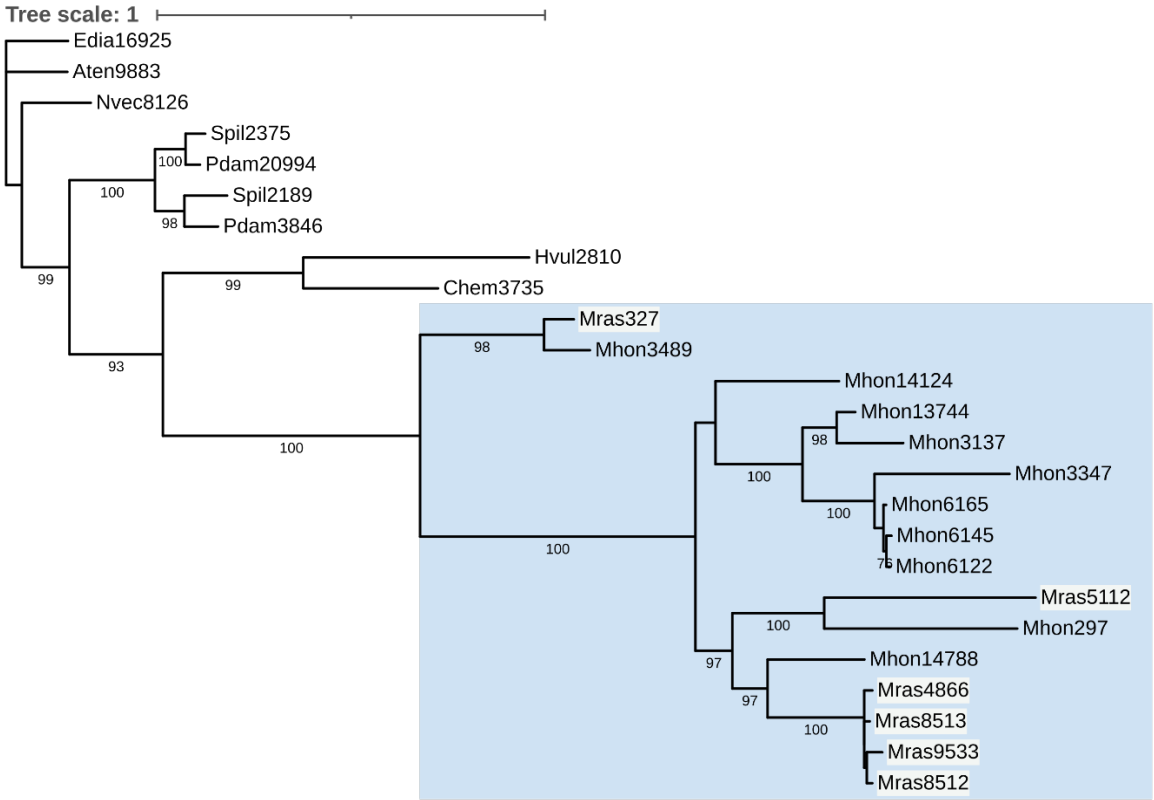


Figure 2 Phylogenetic tree of proteins containing the "Hexokinase_1" and "Hexokinase_2" protein domains in free living cnidarians and *Myxobolus* sp. and *M. honghuensis*, in which a blue box represents the clade containing all myxozoan proteins and *Myxobolus* sp. proteins have a white background. Bootstrap support values of 75% and above are annotated on the branches of the tree.

3.3. Gene family analysis

Using OrthoFinder, we identified 9,843 groups of orthologous proteins (OGs) that contained at least one myxozoan protein, of which 3,972 OGs contained at least one *Myxobolus sp.* protein. Of these, 3,095 OGs also contained protein from *M. honghuensis*. There were 877 OGs with at least one *Myxobolus sp.* protein but not *M. honghuensis,* and 1070 OGs vice versa. Less than one-third of all 9,843 OGs (2,727 OGs) contained a protein from the majority of myxozoan species used in the analysis (four of the seven). These 2,727 OGs represent a set of core conserved proteins in myxozoans.

3.3.1. Phylogenetic relationships between the myxozoan species

Using the 57 single-copy orthologs, we reconstructed a ML-based phylogeny of both myxozoans and free-living cnidarians and found that *M. honghuensis* is the sister species to *Myxobolus sp.* with the other relationships agreeing with previous studies (Figure 3).

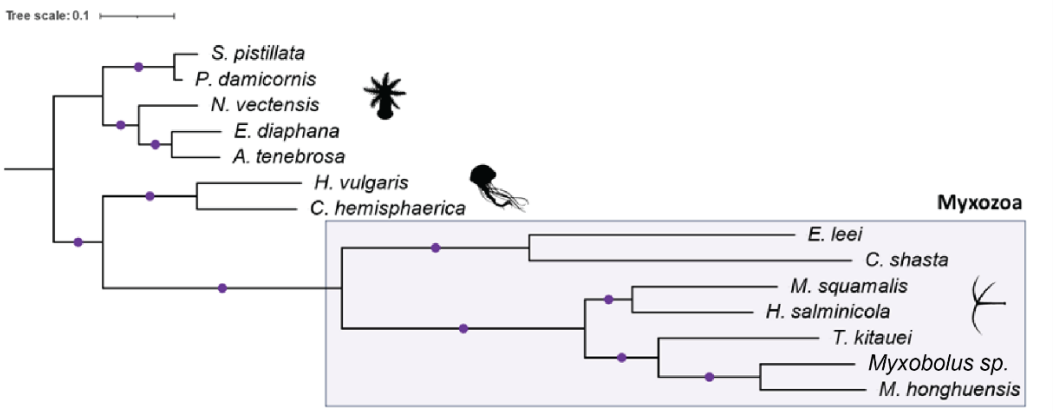


Figure 3 Phylogenetic relationships between free-living and parasitic cnidarians generated using IQTREE. Single-copy orthologs identified by OrthoFinder were used as input. All nodes with boot-strap values greater than 90% are annotated with purple circles. The silhouettes for Anthozoa, Medusozoa, and Myxozoa are taken from PhyloPic (www.phylopic.org).

3.3.2. Gene family expansions in *Myxobolus sp.*

Using the species tree and full set of orthology clusters results, we identified 10 gene families that were significantly expanded in myxozoans. We identified a further 19 gene families that were significantly expanded in only *Myxobolus sp.*, the majority of which were involved in repeat-related processes, and hence were not analyzed further. Here, we describe on three gene families of interest.

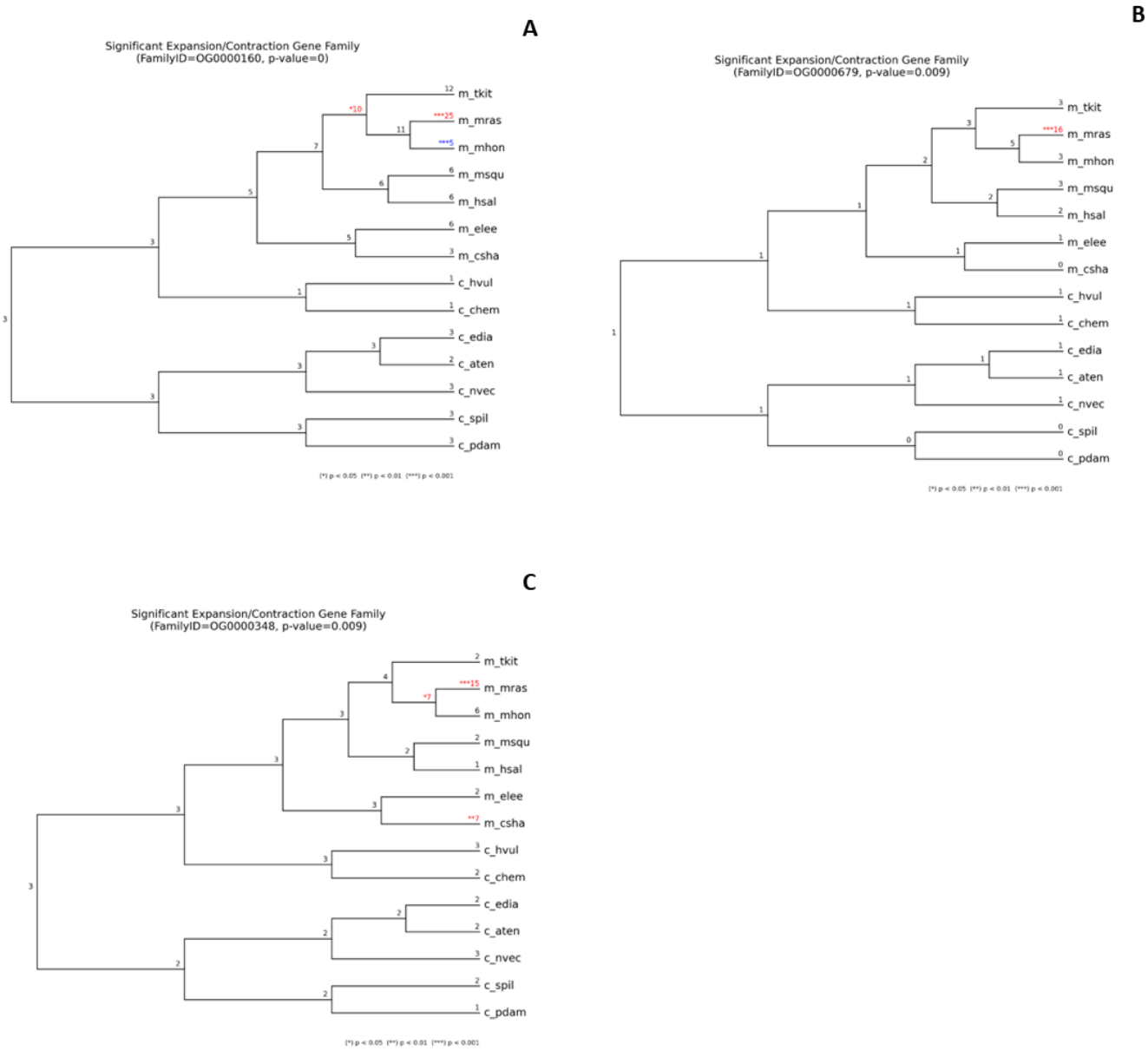


Figure 4 CAFE5 analysis of A) OG0000160, B) OG0000679, and C) OG0000348 with significant expansions and contractions colored in red and blue, respectively. Significant changes on the branch are labelled with asterisks, where red color indicates gene family expansion and blue color indicates gene family contraction. (*) indicates a p-value < 0.05, (**) indicates a p-value < 0.01, and (***) indicates a p-value < 0.001.

The first gene family of interest (OG0000160) was comprised of the orthologs of the *H. vulgaris* SERPINB6 (serine protease inhibitor B6). This family underwent expansion on the branch leading to *Myxobolus sp., M. honghuensis,* and *T. kitauei,* which was followed by a contraction

416 on the branch leading to *M. honghuensis* (Figure 4A). All myxozoan proteins in this OG
417 clustered together in the maximum likelihood (ML) phylogenetic tree (Figure 5). Since serpins
418 are considered as potential targets for myxozoan control and have been extensively studied in *M.*
419 *cerebralis*, we inspected this gene family further. We observed that only 11 of the 25 *Myxobolus*
420 *sp.* proteins from this OG contained the "Serpin" domain. Additionally, two more OGs contained
421 *Myxobolus sp.* proteins with the "Serpin" domain. Of these, one OG (OG0022393) had proteins
422 only from *Myxobolus sp.* and *M. honghuensis.* We analyzed the phylogenetic relationships
423 between all proteins containing serpin domains from *Myxobolus sp.*, *M. honghuensis, M.*
424 *cerebralis,* and free-living cnidarians. We observed that all the myxozoan serpins clustered
425 together (Figure 6). Further, myxozoan homologs of *M. cerebralis* serpin 1 underwent expansion
426 in *Myxobolus sp.* and *M. honghuensis*.
427
428 The second gene family of interest (OG0000679) underwent expansion in *Myxobolus sp.* and
429 was comprised of orthologs of the *H. vulgaris* protein "Uncharacterized protein LOC100211173
430 isoform X1". Proteins in this OG harbor the "Lupus La" protein domain. This gene family
431 experienced an expansion on the branch leading to both *Myxobolus sp.* and *M. honghuensis*
432 (Figure 4B, 7). Finally, the third gene family of interest that experienced expansions in
433 *Myxobolus sp.* consisted of orthologs of three peroxiredoxin proteins in *H. vulgaris*. This gene
434 family experienced an expansion on the branch leading to *Myxobolus sp., M. honghuensis,* and *T.*
435 *kitauei,* and also in *C. shasta* (Figure 4C, 8).
436

Tree scale: 1
Edia2955
Aten10535
Nvec15638
Spil3376
Pdam778
Spil9195
Pdam2899
Nvec9861
Edia11035
Aten18674
Spil22494
Pdam6495
Nvec6959
Edia21533
Hvul3395
Chem4511
Msqu3188
Mras3358
Mhon4906
Tkit1858
Tkit1857
Tkit1340
Tkit1850
Tkit8824
Tkit8400
Tkit5256
Tkit1849
Csha2287
Csha2284
Tkit8659
Tkit325
Tkit8304
Tkit1785
Msqu3156
Msqu894
Hsal6798
Hsal2206
Hsal1043
Mhon2907
Mhon14271
Mhon9969
Mhon2726
Mras2351
Mras341
Mras2350
Mras340
Mras6461
Mras3925
Mras6528
Mras3929
Mras3928
Mras1460
Mras1461
Mras2265
Mras2264
Mras6456
Mras6458
Mras376
Mras6463
Mras6459
Mras374
Mras3924
Mras6460
Mras3919
Mras3136

437

Figure 5 Phylogenetic tree of the gene family OG000160, in which the *Myxobolus sp.* proteins have a white background and a blue box represents the clade containing all myxozoan proteins. Bootstrap support values of 75% and above are annotated on the branches of the tree.

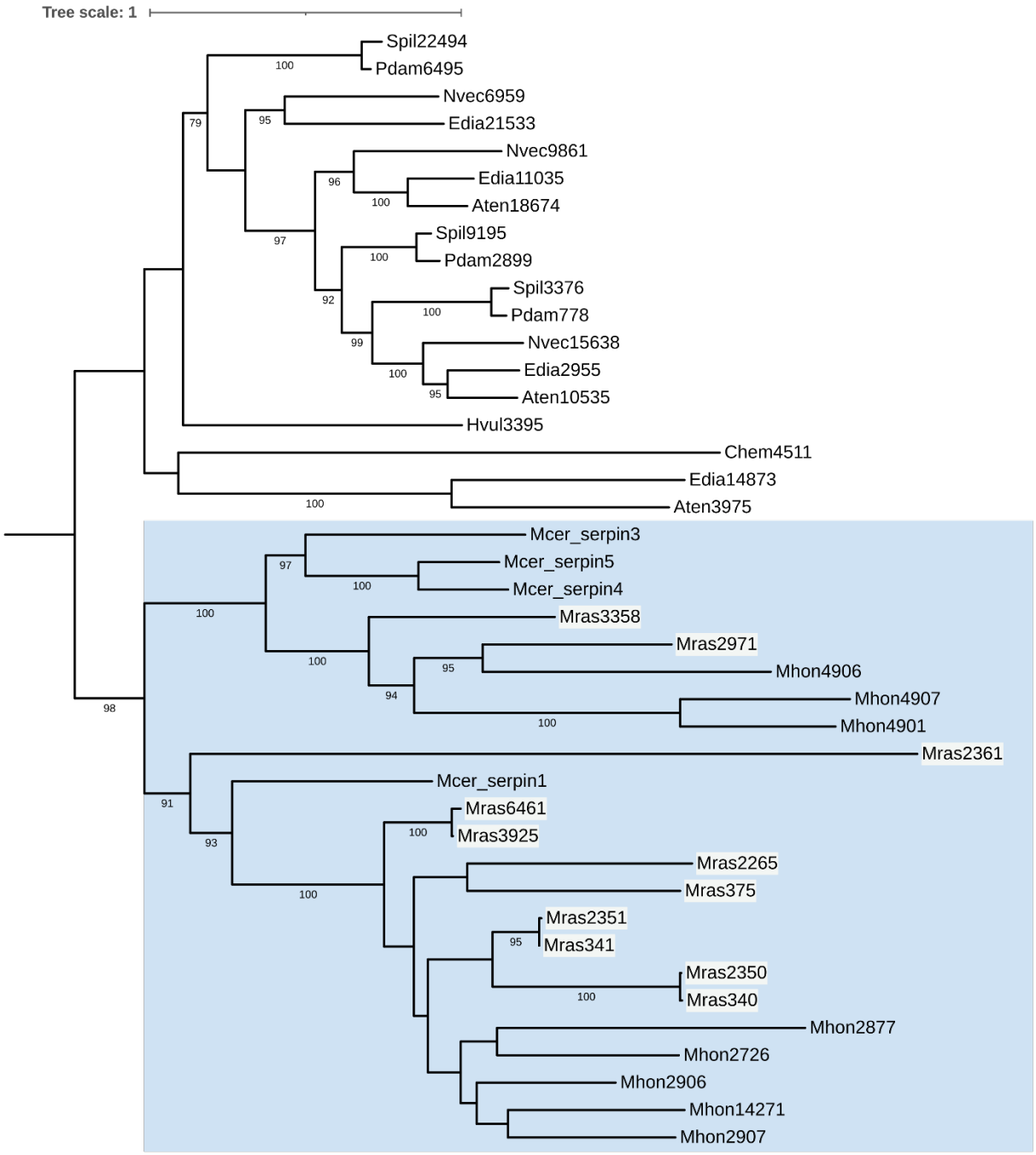
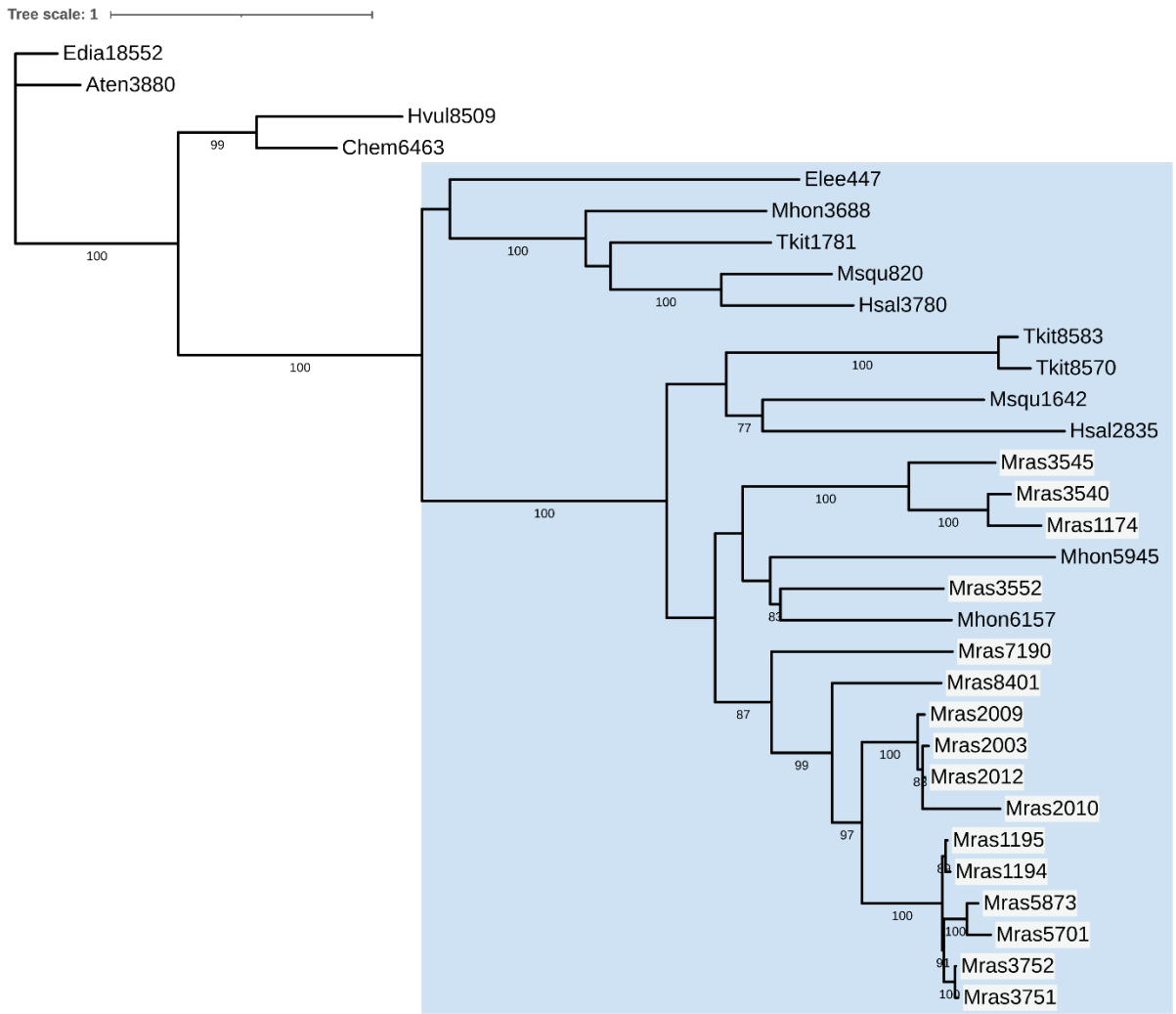


Figure 6 Phylogenetic tree of proteins containing the "Serpin" domain in free-living cnidarians, *Myxobolus sp.*, *M. honghuensis*, and *M. cerebralis* in which the *Myxobolus sp.* proteins have a

445 white background and a blue box represents the clade containing all myxozoan proteins.
446 Bootstrap support values of 75% and above are annotated on the branches of the tree.
447



448
449 **Figure 7** Phylogenetic tree of the gene family OG000679 in which the *Myxobolus sp.* proteins
450 have a white background and a blue box represents the clade containing all myxozoan proteins.
451 Bootstrap support values of 75% and above are annotated on the branches of the tree.
452
453

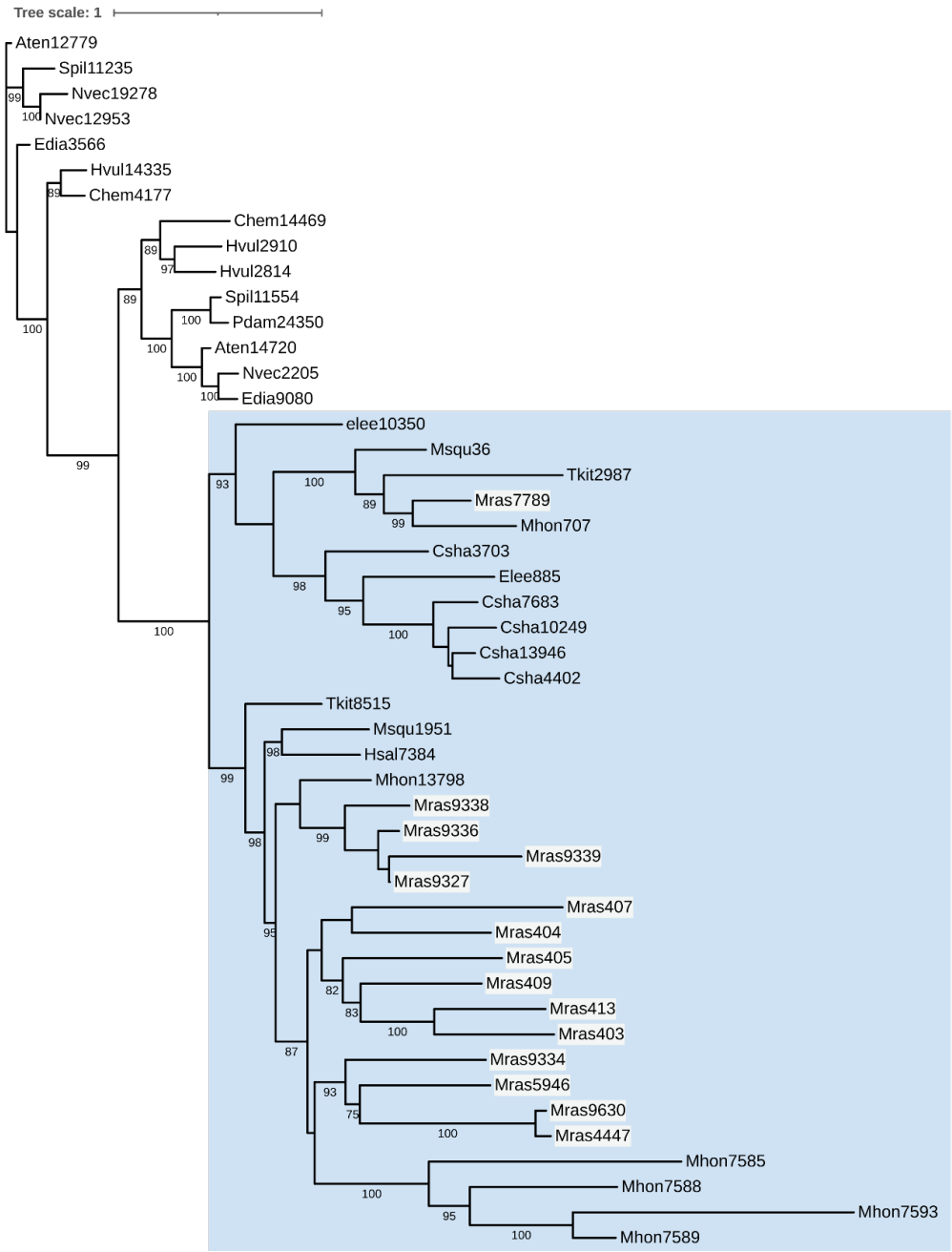


Figure 8 Phylogenetic tree of the gene family OG000348 (peroxiredoxin) in which the *Myxobolus sp.* proteins have a white background and a blue box represents the clade containing all myxozoan proteins. Bootstrap support values of 75% and above are annotated on the branches of the tree.

460 *3.3.3.* Gene families of interest in *Myxobolus sp.*

461 In addition to these three expanded gene families, we explored the evolution of two gene families
462 that have been of interest in myxozoan biology. First, we analyzed the complement of proteases
463 found in *Myxobolus sp.* since proteases have been proposed to be promising targets for
464 myxozoan control. We observed that 377 *Myxobolus sp.* proteins had a top hit to 155 proteases
465 from the MEROPS database: 39 metalloproteases, 38 cysteine proteases, 22 serine proteases, 19
466 threonine proteases, 17 aspartic proteases, and 20 protease inhibitors. The most abundant
467 category of proteases was A28B (the aspartic peptidase "skin SASPase"), followed by S01A
468 (serine peptidase "chymotrypsin A"), and C01A (cysteine peptidase "papain"). We also
469 identified 31 proteins from the I04 family of serine and cysteine endopeptidase inhibitors.
470
471 We aligned all but 16 of these 377 *Myxobolus sp.* proteases to proteins from the fathead minnow
472 and 13 of the remaining 16 proteins could be aligned to Actinopterygii (ray-finned fish) proteins
473 from the NR database (e-value $\leq 1e^{-3}$). We inspected the function annotation remaining three
474 proteins (mras5248, mras6483, and mras9733): protein mras5248 belonged to the same
475 homologous superfamily as the mouse retroviral-like aspartic protease 1 (A28B subfamily) and
476 also has the same PFAM protein domain ("gag-asp_protease" gag-polyprotein putative aspartyl
477 protease); protein mras6483 was assigned to the same homologous superfamily as the mouse
478 retroviral-like aspartic protease 1 (A28B subfamily) but did not have any predicted PFAM
479 protein domains; protein mras9733 could not be assigned any family membership or PFAM
480 domains.
481
482 Next, we analyzed the repertoire of minicollagen and nematogalectin proteins in *Myxobolus sp.*.
483 We searched for significant sequence similarity with minicollagen sequences from myxozoans
484 and free-living cnidarians, and analyzed phylogenetic relationships between them (Figure 9).
485 Previously, minicollagens were shown to belong to three different clades: I, II, and III. We found
486 one *Myxobolus sp.* protein in each of clades II and III, and none from clade I. In addition, we
487 identified one *Myxobolus sp.* and one *M. honghuensis* protein that grouped with free-living
488 cnidarian minicollagens from clade III. There were two *Myxobolus sp.* proteins and two *M.*
489 *honghuensis* proteins that clustered together but did not have sufficient phylogenetic support to
490 assign to a minicollagen clade. For the nematogalectins, we found three proteins in *Myxobolus*
491 *sp.*: one closely related to *M. bejeranoi* nematogalectin A; one to *M. bejeranoi* nematogalectin C;
492 and the third to a putative nematogalectin in *S. zaharoni* (Figure 10).
493
494

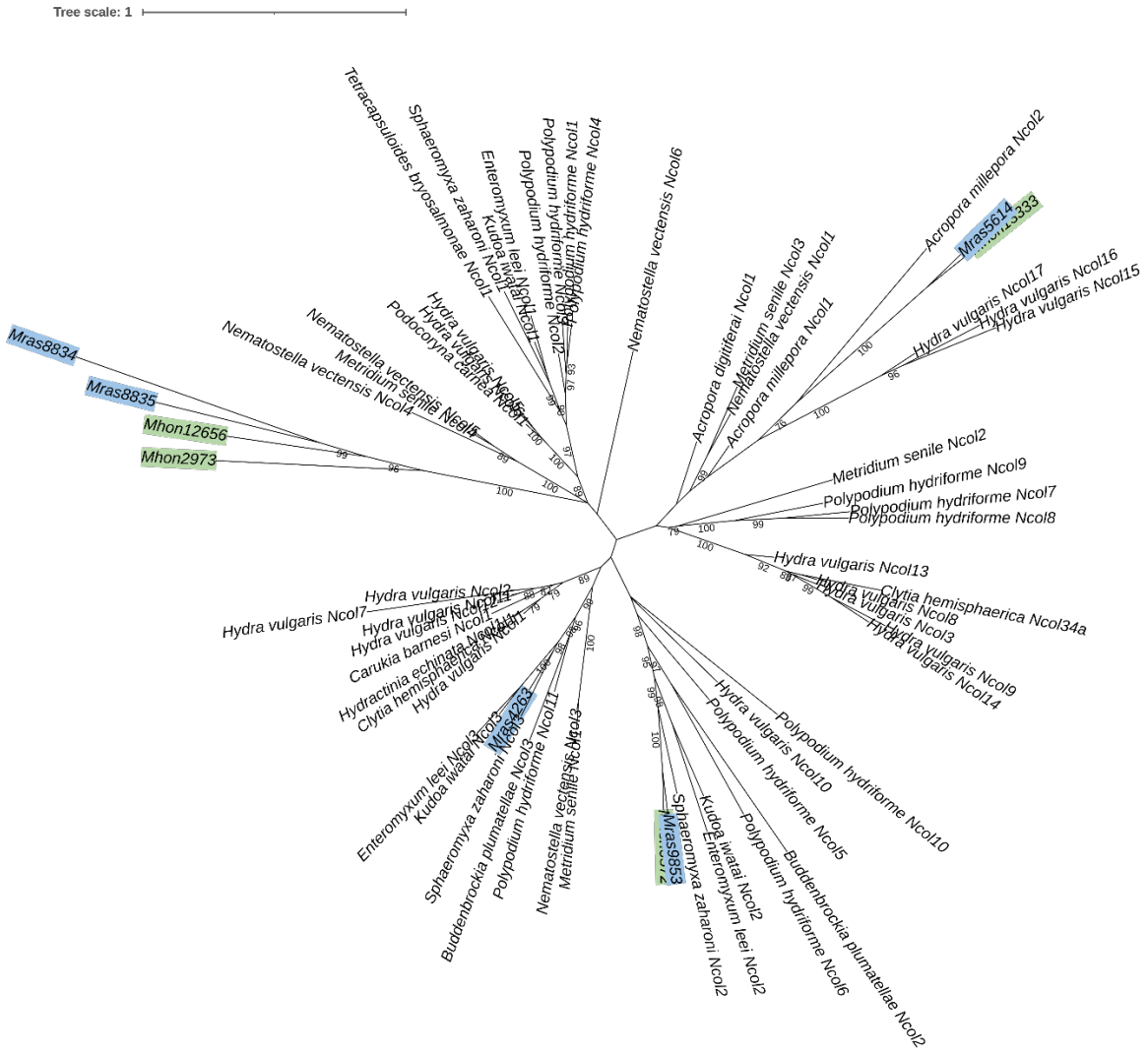


Figure 9 Phylogenetic tree of the minicollagens, in which *Myxobolus sp.* and *M. honghuensis* sequences are labelled with blue and green color respectively. Bootstrap support values of 75% and above are annotated on the branches of the tree.

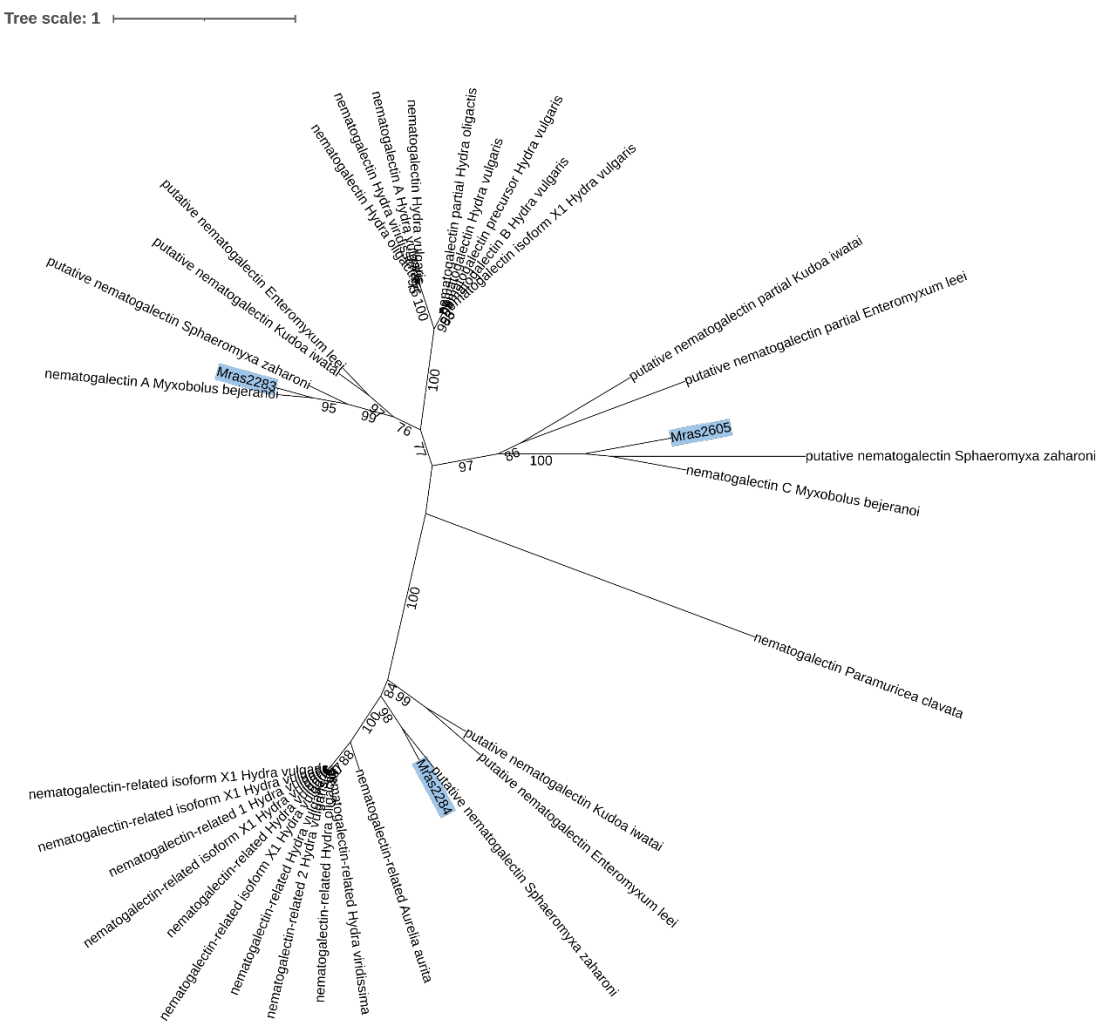


Figure 10 Phylogenetic tree of the nematogalectins in which *Myxobolus sp.* sequences are labelled with blue. Bootstrap support values of 75% and above are annotated on the branches of the tree.

3.3.4. Gene-loss in *Myxobolus sp.*

We identified 111 OGs with proteins from *M. honghuensis* and the majority of the remaining five myxozoan species, but none from *Myxobolus sp.*. We extracted the 178 *M. honghuensis* proteins from these 111 OGs and examined the protein domains and KEGG pathways associated with these proteins as determined by EggNOG-mapper. The top three most abundant protein domains were “Dimer_Tnp_hAT” (hAT family C-terminal dimerisation region), “Pkinase” (Protein kinase domain), and “UAA” (UAA transporter family). The top three most abundant

511 KEGG pathways were “Metabolic pathways” (map01100), “Cellular senescence” (map04218),
512 and “cGMP-PKG signaling pathway” (map04022).
513

514 ### *3.3.5. Myxobolus sp.* proteins with no sequence similarity to proteins from the fathead
515 minnow

516 We detected 3,444 *Myxobolus sp.* proteins that did not have sequence similarity to any fathead
517 minnow protein, as identified using BLASTP (e-value ≤ $1e^{-3}$), of which 639 proteins also had
518 BLASTP hits to proteins from Actinopterygii species (e-value ≤ $1e^{-3}$). The remaining 2,805
519 *Myxobolus sp.* proteins were queried against the ChEMBL database to search for potential drug
520 targets. To this end, we identified eight *Myxobolus sp.* proteins which had a BLASTP hit to a
521 protein from the ChEMBL database (e-value ≤ $1e^{-5}$). For a more stringent selection criteria, we
522 only looked at *Myxobolus sp.* proteins that did not have a BLASTP hit to proteins from the
523 fathead minnow (e-value ≤ 0.05). This yielded three proteins which we analyzed in detail.:
524 “mras18” was aligned to inositol phosphorylceramide synthase (CHEMBL1641343),
525 “mras2094” to gag-pol polyprotein (CHEMBL3638331), and “mras2163” to heat shock protein
526 hsp-16.2 (CHEMBL2146313).
527

528 ## 3.4. Characterization of the mitochondrial genome and mitochondrial
529 proteome

530 We identified a single 31,057 bp contig as the mitochondrial genome (mt-genome) of *Myxobolus*
531 *sp.* (Figure 11A). The *Myxobolus sp.* mt-genome was predicted to encode only five of the 13
532 protein-coding genes that are typically found in most animal mitochondrial genomes—*cox1*,
533 *cox2*, *nad1*, *nad5*, and *cob*—as well as two ribosomal RNA genes (*rrnS* and *rrnL*). The *rrnL*
534 gene was split into two fragments. We aligned both the fragments of the 16S RNA gene (*rrnL*) to
535 the NT database and found that the top hit to both fragments was from the myxozoan *Myxobolus*
536 *squamalis*. When we searched the nuclear genome of *Myxobolus sp.* with the same HMM
537 profiles of protein-coding genes encoded in the mt-genome, we found that *atp6* was transferred
538 to the nuclear genome and had acquired introns and an mitochondria-targeting signal (MTS). In
539 fact, we found that *atp6* was transferred to the nuclear genome and acquired an MTS in *M.*
540 *honghuensis*, *S. zaharoni*, *T. kitauei, K. iwatai*, and *M. squamalis*. None of the other protein-
541 coding genes in the mt-genome were transferred to the nuclear genome in myxozoans.
542

543 Next, we analyzed the mitochondrial proteome—the set of proteins encoded in the nuclear
544 genome and imported into the mitochondrion—of *Myxobolus sp.*. First, we identified homologs
545 of nuclear-encoded mitochondrial proteins from humans in the cnidarian genomes. For this, we
546 aligned the set of human MitoCarta v3 proteins to the proteomes from myxozoans and free-living
547 cnidarians. *Myxobolus sp.* had homologs of 542 of the 1,136 human MitoCarta genes (Figure
548 11B). On average, myxozoans had homologs of 526 MitoCarta genes, compared to 951 for the

549 free-living cnidarians. We identified 741 MitoCarta genes that had homologs in at least one
550 myxozoan species, of which 241 were present in all. We used ConsensusPathDB to find out the
551 functions of these 241 mitochondrial protein-coding genes. The top five enriched KEGG
552 pathways included “Valine, leucine and isoleucine degradation”, “TCA cycle”, “Fatty acid
553 degradation”, and “Aminoacyl-tRNA biosynthesis”. On the contrary, 395 mitochondrial protein-
554 coding genes did not have a significant sequence similarity hit to any myxozoan protein in this
555 study, representing a potential loss of mitochondrial functions in myxozoans. ConsensusPathDB
556 found that the top five enriched pathways in this missing set included “Oxidative
557 phosphorylation”, “Thermogenesis”, and “Apoptosis”.
558
559 Next, we identified potential mitochondria-targeted proteins in *Myxobolus sp.* which do not have
560 homologs to known mitochondrial proteins in humans. For this, we identified *Myxobolus sp.*
561 proteins that are predicted to possess a mitochondrial targeting signal (MTS) at their N-terminus
562 end using MitoFates. We observed that only 271 *Myxobolus sp.* proteins were predicted to harbor
563 an MTS, of which 36 were predicted to possess an MTS with a high probability (at least 90%).
564 Of these, 15 had a hit to a human protein, for which the top human hit was also predicted to
565 possess a MTS. A further seven *Myxobolus sp.* proteins had a hit to a human protein that did not
566 have MTS. These proteins represent potential instances of neolocalization in the parasite (an
567 instance where the homolog of a non-mitochondrial protein in human is targeted to the
568 mitochondria in *Myxobolus sp.*). These included the endogenous retrovirus group K member 113
569 Pol protein (P63132), NYNRIN (Q9P2P1), alpha-1,3-mannosyl-glycoprotein 2-beta-N-
570 acetylglucosaminyltransferase (P26572), adenosine 5'-monophosphoramidase HINT1 (P49773),
571 and cytoplasmic asparagine--tRNA ligase (O43776). The remaining 14/36 proteins did not have
572 BLASTP hit to any human protein, and we used BLASTP to search for proteins with sequence
573 similarity in the NR database. Of these, five did not have a BLASTP hit to any non-myxozoan
574 protein in the NR database.
575

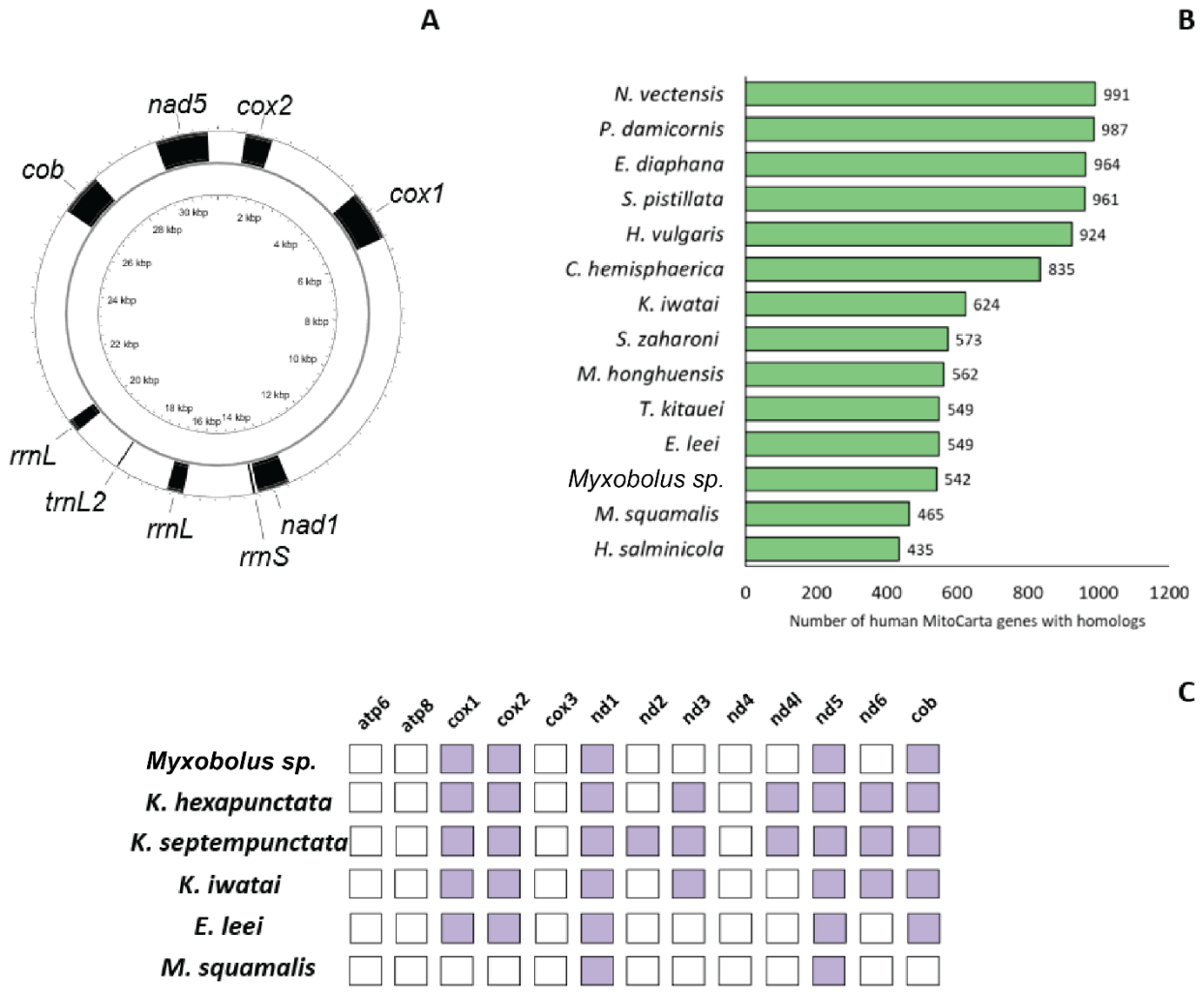


Figure 11 A) The mitochondrial genome of *Myxobolus sp.* visualized using PROKSEE, B) The number of human MitoCarta genes with Blast hits to myxozoan (purple) genomes and cnidarian (green) genomes, C) The presence/absence of protein-coding genes in the mitochondrial genomes of myxozoans.

4. Discussion

4.1. The high repeat content of the *Myxobolus* sp. genome contributes to genome expansion and necessitates the use of long-read sequencing of myxozoan genomes

In this study, we sequenced, annotated, and analyzed the genome of *Myxobolus* sp., a recently discovered myxozoan parasite of fathead minnows in southern Alberta. We used the Oxford

587 Nanopore long-read sequencing technology, and thus were able to provide the second long-read
588 genome in this species-rich group of parasites. While relatively short for an animal genome, it is
589 one of the most repetitive. The only other long-read-based myxozoan genome, *M. honghuensis*,
590 is also highly repetitive. Such repetitive sequences can negatively affect the assembly of a
591 genome, especially in genomes sequenced using short-read platforms where repetitive loci are
592 more likely to be misassembled, which likely explains the relatively high number of scaffolds in
593 both the *M. honghuensis* and our *Myxobolus sp.* assemblies (Tørresen et al., 2019). Since both
594 long-read genomes belong to a single genus in myxozoans, we cannot say with confidence
595 whether high repeat content is a feature of all myxozoan genomes.
596
597 The high repeat content in the *Myxobolus sp.* genome is a significant factor explaining the large
598 genome size in this species. We detected a proliferation of the MULE-MuDR elements in the
599 genome. A recent analysis of the diversity of these Mutator-like elements (MULEs) across
600 eukaryotes found that, outside of plants, many genomes containing these elements were from
601 pathogens (Dupeyron et al., 2019). Indeed, in fungal pathogens, such TE-expansions have been
602 proposed to account for genome expansion (Raffaele & Kamoun, 2012). Furthermore, the
603 increase in repeat content in parasites could result in new functions and genome arrangements
604 which benefit the parasite's efforts to escape host defenses (Dong et al., 2015). It is possible that
605 the small sizes of myxozoan genomes that were sequenced using short-read sequencing
606 platforms, and the smaller repeat content predicted in them, are an underestimation of the true
607 genome size of these parasites.
608
609 The increase in repeat content of the *Myxobolus sp.* genome can also further our understanding
610 of myxozoan biology. A good example of this is the expansion in proteins with the FLYWCH
611 domains identified in *Myxobolus sp.*. A recent analysis of DNA transposon-derived transcription
612 factors in animals found that FLYWCH-domain containing transcription factors are rare in
613 animals but have undergone expansion in one phylum of non-bilaterian animals – Ctenophora
614 (Mukherjee & Moroz, 2023). This study did not include myxozoans in their comparisons. We
615 show that *Myxobolus sp.* could contain the highest repertoire of these FLYWCH proteins in
616 animals.

617 ## 4.2. Gene family expansions

618 A prominent feature of myxozoan genome evolution is the loss of genes that are otherwise well-
619 conserved in free-living eukaryotes. Hence, gene families that defy this trend and instead
620 undergo expansion are interesting as they represent potential adaptations to the parasitic
621 myxozoan lifestyle.
622
623 In one such example of gene family expansion, serpins (serine proteases) play an important role
624 in parasitic organisms, like defense against host proteases in *Schistosoma mansoni* and acting as
625 immunomodulators and inhibit host immune responses in nematodes (Maizels et al., 2001;

626 Valdivieso et al., 2015). Recently, a study analyzed the expression-profiles of *M. cerebralis*
627 serpins and proposed that they represent good targets for inactivating the parasite at an early
628 stage in the fish host (Eszterbauer et al., 2021). We found a group of *Myxobolus sp.* proteins that
629 clustered together with serpins 3, 4, and 5 from *M. cerebralis.* Researchers have shown that the
630 expression of these three serpins is upregulated during fish invasion and early development, and
631 their functional predictions suggest that these three are caspase inhibitors. Thus, the *Myxobolus*
632 *sp.* proteins that cluster together with these three serpins represent promising targets for parasite
633 control. Additionally, serpin 1, predicted to be a putative chymotrypsin-like inhibitor, has
634 undergone extensive duplication in *Myxobolus sp..* While this suggests that the homologs of
635 serpin 1 are probably involved in important functions in *Myxobolus sp.,* developing targets
636 against them would be challenging.
637
638 Further, we also identified an expansion in a family of peroxiredoxins in *Myxobolus sp.* and
639 other myxozoans. Peroxiredoxins are ubiquitous enzymes with roles in antioxidation and
640 signaling systems (Gretes et al., 2012). These enzymes play important roles in parasitic defenses
641 against reactive oxygen and nitrogen species generated by the host immune system. These
642 enzymes have been proposed to be drug targets several parasites like trypanosomes and as
643 vaccine antigens in *L. major* (Bayih et al., 2014).
644
645 We also observe that both *Myxobolus sp.* and *M. honghuensis* have a higher number of
646 hexokinase domain containing proteins. Hexokinases catalyze the first step of glycolysis. The
647 increase in the hexokinase repertoire in the parasite could reflect the reliance of the parasite on
648 glycolysis to compensate for a reduced and inefficient oxidative phosphorylation system. Indeed,
649 we noticed that several human OXPHOS proteins did not have homologs in *Myxobolus sp.*.
650 These could represent a lead for future antiparasitic strategy.

651 ### 4.3. *Myxobolus* sp. has a reduced mitochondrial genome and proteome

652 Most animal mitochondrial genomes are single circular compact molecules that encode 37 genes
653 (13 protein coding genes, 2 rRNA genes, and 22 tRNA genes) (Lavrov & Pett, 2016). Myxozoan
654 mitochondrial genomes differ from this archetypal animal mitochondrial genome in several
655 aspects. For instance, all myxozoan genomes have a reduced set of genes (reviewed in (Lavrov
656 & Pett, 2016)). *Enteromyxum leei* has a multipartite mitochondrial genome, comprised of eight
657 megacircles (Yahalomi et al., 2017). The mitochondrial genome of *Myxobolus sp.* is no
658 exception. It represents just the second mitochondrial genome from the species-rich *Myxobolus*
659 genus. The only other mitochondrial genome from the genus belongs to *M. squamalis,* and only a
660 single protein coding gene can be identified in the genome. Further, while researchers produced a
661 high-quality nuclear genome of *M. honghuensis*, they were unable to identify a contig containing
662 the mitochondrial genome. Thus, here we present the most complete mitochondrial genome from
663 the genus *Myxobolus*. It is worth noting that while the mitochondrial genome of *Myxobolus sp.*
664 lacks several protein-coding genes, it is larger than the human mitochondrial genome, due to

large intergenic regions, instead of one large non-coding region in most animal mitochondrial genomes. Further, myxozoan mitochondrial genomes have some of the fastest rates of sequence evolution in animals (Lavrov & Pett, 2016), which could make identifying some mitochondrial genes difficult. In fact, some myxozoan mitochondrial genomes from the genus *Kudoa* have some unidentified open reading frames (ORFs) (Takeuchi et al., 2015), which could be standard animal mitochondrial genes that have evolved beyond recognition by sequence-based approaches.

We find that the mitochondrial gene *atp6* is transferred to the nuclear genome in multiple myxozoans and has acquired a mitochondrial targeting signal. This suggests that the protein encoded by this gene is imported into the organelle, however proteomic analyses need to be done to confirm this. Such a transfer of *atp6* to the nuclear genome and the acquisition of a mitochondrial targeting signal has been reported before, like in the ctenophore *Mnemiopsis leidyi* (Pett et al., 2011) and in the non-animal eukaryote *Chlamydomonas reinhardtii* (Funes et al., 2002). This transfer of *atp6* to the nuclear genome and its potential for its subsequent import back into the organelle make myxozoans potential models to study mitochondrial protein import for aiding efforts in mitochondrial gene therapy (Figueroa-Martínez et al., 2011; Manfredi et al., 2002), as was suggested for *M. leidyi* (Pett et al., 2011).

All proteins encoded by genes in the mitochondrial genome function in complexes that require proteins that are encoded in the nuclear genome and imported into the organelle. The majority of mitochondrial processes rely on just the import of nuclear-encoded proteins. This underscores the need to analyze the mitochondrial proteome for analyzing the evolution of mitochondrial processes in myxozoans. The analysis of mitochondrial proteomes outside of model animals is limited. Only a few studies have looked at the evolution of the mitochondrial proteome in myxozoan species, both showing that myxozoans lack several mitochondrial proteins found in mammals (Muthye & Lavrov, 2018; Yahalomi et al., 2020). Our analysis of the mitochondria proteomes from myxozoans align with these results. On average, just less than half the mitochondrial proteins from mammals had homologs in myxozoans, compared to ~85% in the three non-parasitic cnidarians. We found that around 35% of the human mitochondrial proteins were missing in all the myxozoan species analyzed, representing a set of core lost mitochondrial functions. These missing proteins were enriched in pathways such as oxidative phosphorylation and apoptosis. These findings are consistent with a recent study that found that several apoptosis proteins are lost in myxozoan genomes (Neverov et al., 2023).

4.4. *Myxobolus* sp. represents a potential model system to study host-parasite interactions in the genus *Myxobolus*

The lack of experimentally confirmed life cycles from myxozoans impedes efforts to analyze how these parasites infect the hosts, transmit between them, and interact with their hosts post

704 infection. Within the Myxozoa, only six life cycles have been established in the laboratory, of
705 which three are from the genus *Myxobolus*: *M. cerebralis* (El-Matbouli et al., 1999; Wolf &
706 Markiw, 1984), *Myxobolus pseudodispar* (Kallert et al., 2007; Marton & Eszterbauer, 2012;
707 Székely et al., 1999, 2001), and *Myxobolus parviformis* (Kallert et al., 2005). However, these
708 three species currently lack genomic resources; only a transcriptome is available only for *M.*
709 *cerebralis*. Thus, *Myxobolus sp.* represents a promising model to study infection and host-
710 parasite interactions within *Myxobolus* – since both hosts have been molecularly confirmed, a
711 good quality long-read genome has been sequenced, and a transcriptome is available. Further, the
712 maintenance of this life cycle involving fathead minnows in a laboratory setting is much easier
713 and cost-effective than doing the same with salmonid fish.
714

715 # 5. Conclusion

716 We sequenced, assembled, and annotated the genome of *Myxobolus sp.*, an emerging parasite of
717 fathead minnows in Alberta, using the long read Oxford Nanopore sequencing technology. The
718 *Myxobolus sp.* genome is the largest genome among myxozoan species sequenced till date, and
719 one of the most repetitive genomes among animals. We compare the *Myxobolus sp.* genome to
720 genomes from other myxozoan and free-living cnidarian species, and show that, like other
721 myxozoan species, *Myxobolus sp.* has lost several well-conserved eukaryotic genes, and at the
722 same time, has also experienced expansion in several gene families. Our analysis demonstrated
723 the value of long read sequencing platforms for all myxozoan parasites. We also propose that
724 *Myxobolus sp.* presents a promising potential model to study host-parasite interactions in
725 myxozoans.
726

727 # 6. Acknowledgements

728 We are grateful to Molly Tilley for help collecting *Myxobolus sp.* samples. We acknowledge the
729 high-performance computing resources by the Faculty of Veterinary Medicine and Research
730 Computing at the University of Calgary.
731

732 # 7. Funding

733 This work was supported by: the Office of the Chief Scientist, Alberta Environment and
734 Protected Area to JDW; NSERC Discovery Grant (No. 04589-2020) to JDW; a University of
735 Calgary Eyes High Postdoctoral award to VRM. The funders had no role in study design, data
736 collection and analysis, decision to publish, or preparation of the manuscript.
737
738

References

Alama-Bermejo, G., & Holzer, A. S. (2021). Advances and Discoveries in Myxozoan Genomics. *Trends in Parasitology*, *37*(6), 552–568. https://doi.org/10.1016/j.pt.2021.01.010

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J., Staden, R., & Young, I. G. (1981). Sequence and organization of the human mitochondrial genome. *Nature*, *290*(5806), 457–465. https://doi.org/10.1038/290457a0

Andrews, S. & others. (2010). *FastQC: a quality control tool for high throughput sequence data*. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.

Bartholomew, J. L., Atkinson, S. D., Hallett, S. L., Lowenstine, L. J., Garner, M. M., Gardiner, C. H., Rideout, B. A., Keel, M. K., & Brown, J. D. (2008). Myxozoan parasitism in waterfowl. *International Journal for Parasitology*, *38*(10), 1199–1207. https://doi.org/10.1016/j.ijpara.2008.01.008

Bartholomew, J. L., & Reno, P. W. (2002). The history and dissemination of whirling disease. *American Fisheries Society Symposium*, 3–24.

Bayih, A. G., Daifalla, N. S., & Gedamu, L. (2014). DNA-protein immunization using Leishmania peroxidoxin-1 induces a strong CD4+ T cell response and partially protects mice from cutaneous leishmaniasis: Role of fusion murine granulocyte-macrophage colony-stimulating factor DNA adjuvant. *PLoS Neglected Tropical Diseases*, *8*(12), e3391. https://doi.org/10.1371/journal.pntd.0003391

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., & Stadler, P. F. (2013). MITOS: Improved de novo metazoan mitochondrial genome annotation. *Mitogenomics and Metazoan Evolution*, *69*(2), 313–319. https://doi.org/10.1016/j.ympev.2012.08.023

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M., & Borodovsky, M. (2021). BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics and Bioinformatics*, *3*(1), lqaa108. https://doi.org/10.1093/nargab/lqaa108

Buchfink, B., Reuter, K., & Drost, H.-G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods*, *18*(4), 366–368. https://doi.org/10.1038/s41592-021-01101-x

Bushmanova, E., Antipov, D., Lapidus, A., & Prjibelski, A. D. (2019). rnaSPAdes: A de novo transcriptome assembler and its application to RNA-Seq data. *GigaScience*, *8*(9), giz100. https://doi.org/10.1093/gigascience/giz100

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*(1), 421. https://doi.org/10.1186/1471-2105-10-421

Canadian Food Inspection Agency. (2019). https://inspection.canada.ca/animal-health/aquatic-animals/diseases/reportable-diseases/whirling-disease/alberta/eng/1520358140712/1520358141169

Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution*, *38*(12), 5825–5829. https://doi.org/10.1093/molbev/msab293

Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics (Oxford, England)*, *25*(15), 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Chang, E. S., Neuhof, M., Rubinstein, N. D., Diamant, A., Philippe, H., Huchon, D., & Cartwright, P. (2015). Genomic insights into the evolutionary origin of Myxozoa within Cnidaria. *Proceedings*

of the National Academy of Sciences, *112*(48), 14912–14917. https://doi.org/10.1073/pnas.1511468112

COSEWIC assessment and status report. (2017). https://www.canada.ca/en/environment-climate-change/services/species-risk-public-registry/cosewic-assessments-status-reports/western-silvery-minnow-2017.html

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Dong, S., Raffaele, S., & Kamoun, S. (2015). The two-speed genomes of filamentous pathogens: Waltz with plants. *Genomes and Evolution*, *35*, 57–65. https://doi.org/10.1016/j.gde.2015.09.001

Dupeyron, M., Singh, K. S., Bass, C., & Hayward, A. (2019). Evolution of Mutator transposable elements across eukaryotic diversity. *Mobile DNA*, *10*(1), 12. https://doi.org/10.1186/s13100-019-0153-8

Eiras, J. C. (2005). An overview on the myxosporean parasites in amphibians and reptiles. *Acta Parasitol*, *50*(4), 267–275.

El-Matbouli, M., McDowell, T. S., Antonio, D. B., Andree, K. B., & Hedrick, R. P. (1999). Effect of water temperature on the development, release and survival of the triactinomyxon stage of Myxobolus cerebralis in its oligochaete host. *International Journal for Parasitology*, *29*(4), 627–641. https://doi.org/10.1016/s0020-7519(99)00009-0

Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*(1), 238. https://doi.org/10.1186/s13059-019-1832-y

Eszterbauer, E., Atkinson, S., Diamant, A., Morris, D., El-Matbouli, M., & Hartikainen, H. (2015). Myxozoan Life Cycles: Practical Approaches and Insights. In B. Okamura, A. Gruhl, & J. L. Bartholomew (Eds.), *Myxozoan Evolution, Ecology and Development* (pp. 175–198). Springer International Publishing. https://doi.org/10.1007/978-3-319-14753-6_10

Eszterbauer, E., Szegő, D., Ursu, K., Sipos, D., & Gellért, Á. (2021). Serine protease inhibitors of the whirling disease parasite Myxobolus cerebralis (Cnidaria, Myxozoa): Expression profiling and functional predictions. *PloS One*, *16*(3), e0249266. https://doi.org/10.1371/journal.pone.0249266

Figueroa-Martínez, F., Vázquez-Acevedo, M., Cortés-Hernández, P., García-Trejo, J. J., Davidson, E., King, M. P., & González-Halphen, D. (2011). What limits the allotopic expression of nucleus-encoded mitochondrial genes? The case of the chimeric Cox3 and Atp6 genes. *Mitochondrion*, *11*(1), 147–154. https://doi.org/10.1016/j.mito.2010.09.003

Fletcher, C. & Pereira da Conceicoa, Lyndell. (n.d.). The genome sequence of the starlet sea anemone, Nematostella vectensis (Stephenson, 1935) [version 1; peer review: 1 approved, 1 approved with reservations]. *Wellcome Open Research 2023.*

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences*, *117*(17), 9451–9457. https://doi.org/10.1073/pnas.1921046117

Fontes, I., Hallett, S. L., & Mo, T. A. (2015). Comparative Epidemiology of Myxozoan Diseases. In B. Okamura, A. Gruhl, & J. L. Bartholomew (Eds.), *Myxozoan Evolution, Ecology and Development* (pp. 317–341). Springer International Publishing. https://doi.org/10.1007/978-3-319-14753-6_17

Friedrich, C., Ingolic, E., Freitag, B., Kastberger, G., Hohmann, V., SKOFITSCH, G., NEUMEISTER, U., & KEPKA, O. (2000). A myxozoan-like parasite causing xenomas in the brain of the mole, Talpa europaea L., 1758 (Vertebrata, Mammalia). *Parasitology*, *121*(5), 483–492. Cambridge Core. https://doi.org/10.1017/S0031182099006769

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, *28*(23), 3150–3152. https://doi.org/10.1093/bioinformatics/bts565

Funes, S., Davidson, E., Claros, M. G., van Lis, R., Pérez-Martínez, X., Vázquez-Acevedo, M., King, M. P., & González-Halphen, D. (2002). The typically mitochondrial DNA-encoded ATP6 subunit of

the F1F0-ATPase is encoded by a nuclear gene in Chlamydomonas reinhardtii. *The Journal of Biological Chemistry*, *277*(8), 6051–6058. https://doi.org/10.1074/jbc.M109993200

Grant, J. R., Enns, E., Marinier, E., Mandal, A., Herman, E. K., Chen, C., Graham, M., Van Domselaar, G., & Stothard, P. (2023). Proksee: In-depth characterization and visualization of bacterial genomes. *Nucleic Acids Research*, *51*(W1), W484–W492. https://doi.org/10.1093/nar/gkad326

Gretes, M. C., Poole, L. B., & Karplus, P. A. (2012). Peroxiredoxins in parasites. *Antioxidants & Redox Signaling*, *17*(4), 608–633. https://doi.org/10.1089/ars.2011.4404

Guo, qingxiang. (2021). *A myxozoan genome reveals mosaic evolution in a parasitic cnidarian* (Version V2) [dataset]. Harvard Dataverse. https://doi.org/10.7910/DVN/INLEPM

Guo, Q., Atkinson, S. D., Xiao, B., Zhai, Y., Bartholomew, J. L., & Gu, Z. (2022). A myxozoan genome reveals mosaic evolution in a parasitic cnidarian. *BMC Biology*, *20*(1), 51. https://doi.org/10.1186/s12915-022-01249-8

Gupta, A., & Kaur, H. (2017). A new pathogen, Myxobolus holzerae (Myxosporea: Myxozoa) causing severe gill disease in an Indian major carp Labeo rohita in a cold water wetland, Punjab (India). *Microbial Pathogenesis*, *111*, 244–251. https://doi.org/10.1016/j.micpath.2017.08.044

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution*, *35*(2), 518–522. https://doi.org/10.1093/molbev/msx281

Holzer, A. S., Bartošová-Sojková, P., Born-Torrijos, A., Lövy, A., Hartigan, A., & Fiala, I. (2018). The joint evolution of the Myxozoa and their alternate hosts: A cnidarian recipe for success and vast biodiversity. *Molecular Ecology*, *27*(7), 1651–1666. https://doi.org/10.1111/mec.14558

Hoskins, R. A., Carlson, J. W., Wan, K. H., Park, S., Mendez, I., Galle, S. E., Booth, B. W., Pfeiffer, B. D., George, R. A., Svirskas, R., Krzywinski, M., Schein, J., Accardo, M. C., Damia, E., Messina, G., Méndez-Lago, M., de Pablos, B., Demakova, O. V., Andreyeva, E. N., … Celniker, S. E. (2015). The Release 6 reference sequence of the Drosophila melanogaster genome. *Genome Research*, *25*(3), 445–458. https://doi.org/10.1101/gr.185579.114

Hutchins, P. R., Sepulveda, A. J., Hartikainen, H., Staigmiller, K. D., Opitz, S. T., Yamamoto, R. M., Huttinger, A., Cordes, R. J., Weiss, T., Hopper, L. R., Purcell, M. K., & Okamura, B. (2021). Exploration of the 2016 Yellowstone River fish kill and proliferative kidney disease in wild fish populations. *Ecosphere*, *12*(3), e03436. https://doi.org/10.1002/ecs2.3436

James, C. T., Veillard, M. F., Martens, A. M., Pila, E. A., Turnbull, A., Hanington, P., Luek, A., Alexander, J., & Nehring, R. B. (2021). Whirling disease in the Crowsnest River: An emerging threat to wild salmonids in Alberta. *Canadian Journal of Fisheries and Aquatic Sciences*, *78*(12), 1855–1868. https://doi.org/10.1139/cjfas-2020-0484

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, *30*(9), 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Kallert, D. M., Eszterbauer, E., Erséus, C., El-Matbouli, M., & Haas, W. (2005). Life cycle studies of Myxobolus parviformis sp. N. (Myxozoa: Myxobolidae) from bream. *Diseases of Aquatic Organisms*, *66*(3), 233–243. https://doi.org/10.3354/dao066233

Kallert, D. M., Ponader, S., Eszterbauer, E., El-Matbouli, M., & Haas, W. (2007). Myxozoan transmission via actinospores: New insights into mechanisms and adaptations for host invasion. *Parasitology*, *134*(Pt 12), 1741–1750. https://doi.org/10.1017/S0031182007003290

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589. https://doi.org/10.1038/nmeth.4285

Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, *30*(14), 3059–3066. https://doi.org/10.1093/nar/gkf436

Koel, T. M., Mahony, D. L., Kinnan, K. L., Rasmussen, C., Hudson, C. J., Murcia, S., & Kerans, B. L. (2006). Myxobolus cerebralis in Native Cutthroat Trout of the Yellowstone Lake Ecosystem. *Journal of Aquatic Animal Health*, *18*(3), 157–175. https://doi.org/10.1577/H05-031.1

Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology*, *37*(5), 540–546. https://doi.org/10.1038/s41587-019-0072-8

Kriventseva, E. V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F. A., & Zdobnov, E. M. (2019). OrthoDB v10: Sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Research*, *47*(D1), D807–D811. https://doi.org/10.1093/nar/gky1053

Laetsch, D., & Blaxter, M. (2017). BlobTools: Interrogation of genome assemblies [version 1; peer review: 2 approved with reservations]. *F1000Research*, *6*(1287). https://doi.org/10.12688/f1000research.12232.1

Lavrov, D. V., & Pett, W. (2016). Animal Mitochondrial DNA as We Do Not Know It: Mt-Genome Organization and Evolution in Nonbilaterian Lineages. *Genome Biology and Evolution*, *8*(9), 2896–2913. https://doi.org/10.1093/gbe/evw195

Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, *49*(W1), W293–W296. https://doi.org/10.1093/nar/gkab301

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Liu, Y., Lövy, A., Gu, Z., & Fiala, I. (2019). Phylogeny of Myxobolidae (Myxozoa) and the evolution of myxospore appendages in the Myxobolus clade. *International Journal for Parasitology*, *49*(7), 523–530. https://doi.org/10.1016/j.ijpara.2019.02.009

Liu, Y., Whipps, C. M., Gu, Z. M., Zeng, C., & Huang, M. J. (2012). Myxobolus honghuensis n. Sp. (Myxosporea: Bivalvulida) parasitizing the pharynx of allogynogenetic gibel carp Carassius auratus gibelio (Bloch) from Honghu Lake, China. *Parasitology Research*, *110*(4), 1331–1336. https://doi.org/10.1007/s00436-011-2629-4

Maizels, R. M., Gomez-Escobar, N., Gregory, W. F., Murray, J., & Zang, X. (2001). Immune evasion genes from filarial nematodes. *International Journal for Parasitology*, *31*(9), 889–898. https://doi.org/10.1016/s0020-7519(01)00213-2

Manfredi, G., Fu, J., Ojaimi, J., Sadlock, J. E., Kwong, J. Q., Guy, J., & Schon, E. A. (2002). Rescue of a deficiency in ATP synthesis by transfer of MTATP6, a mitochondrial DNA-encoded gene, to the nucleus. *Nature Genetics*, *30*(4), 394–399. https://doi.org/10.1038/ng851

Marie Veillard & James Clayton. (2020). *Status of Whirling Disease in the Crowsnest River 2019: Technical Report*. Government of Alberta, Ministry of Environment and Parks. https://open.alberta.ca/publications/9781460148402

Martinson, J. W., Bencic, D. C., Toth, G. P., Kostich, M. S., Flick, R. W., See, M. J., Lattier, D., Biales, A. D., & Huang, W. (2022). De Novo Assembly of the Nearly Complete Fathead Minnow Reference Genome Reveals a Repetitive but Compact Genome. *Environmental Toxicology and Chemistry*, *41*(2), 448–461. https://doi.org/10.1002/etc.5266

Marton, S., & Eszterbauer, E. (2012). The susceptibility of diverse species of cultured oligochaetes to the fish parasite Myxobolus pseudodispar Gorbunova (Myxozoa). *Journal of Fish Diseases*, *35*(4), 303–314. https://doi.org/10.1111/j.1365-2761.2012.01347.x

Meisinger, C., Sickmann, A., & Pfanner, N. (2008). The mitochondrial proteome: From inventory to function. *Cell*, *134*(1), 22–24.

Mendes, F. K., Vanderpool, D., Fulton, B., & Hahn, M. W. (2021). CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics*, *36*(22–23), 5516–5518. https://doi.org/10.1093/bioinformatics/btaa1022

Minh, B. Q., Nguyen, M. A. T., & von Haeseler, A. (2013). Ultrafast Approximation for Phylogenetic Bootstrap. *Molecular Biology and Evolution*, *30*(5), 1188–1195. https://doi.org/10.1093/molbev/mst024

Mukherjee, K., & Moroz, L. L. (2023). Transposon-derived transcription factors across metazoans. *Frontiers in Cell and Developmental Biology*, *11*. https://www.frontiersin.org/articles/10.3389/fcell.2023.1113046

Muthye, V., & Lavrov, D. V. (2018). Characterization of mitochondrial proteomes of nonbilaterian animals. *IUBMB Life*, *70*(12), 1289–1301. https://doi.org/10.1002/iub.1961

Neverov, A. M., Panchin, A. Y., Mikhailov, K. V., Batueva, M. D., Aleoshin, V. V., & Panchin, Y. V. (2023). Apoptotic gene loss in Cnidaria is associated with transition to parasitism. *Scientific Reports*, *13*(1), 8015. https://doi.org/10.1038/s41598-023-34248-y

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. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

Novosolov, M., Yahalomi, D., Chang, E. S., Fiala, I., Cartwright, P., & Huchon, D. (2022). The Phylogenetic Position of the Enigmatic, Polypodium hydriforme (Cnidaria, Polypodiozoa): Insights from Mitochondrial Genomes. *Genome Biology and Evolution*, *14*(8). https://doi.org/10.1093/gbe/evac112

OKAMURA, B., HARTIKAINEN, H., SCHMIDT-POSTHAUS, H., & WAHLI, T. (2011). Life cycle complexity, environmental change and the emerging status of salmonid proliferative kidney disease. *Freshwater Biology*, *56*(4), 735–753. https://doi.org/10.1111/j.1365-2427.2010.02465.x

Park Eun-Ji, Kim Bo-A, & Won Yong-Jin. (2010). The Complete Mitochondrial Genome of Dendronephthya gigantea (Anthozoa: Octocorallia: Nephtheidae). *Animal Systematics, Evolution and Diversity*, *26*(3), 197–201. https://doi.org/10.5635/KJSZ.2010.26.3.197

Pett, W., Ryan, J. F., Pang, K., Mullikin, J. C., Martindale, M. Q., Baxevanis, A. D., & Lavrov, D. V. (2011). Extreme mitochondrial evolution in the ctenophore Mnemiopsis leidyi: Insight from mtDNA and the nuclear genome. *Mitochondrial DNA*, *22*(4), 130–142. https://doi.org/10.3109/19401736.2011.624611

Raffaele, S., & Kamoun, S. (2012). Genome evolution in filamentous plant pathogens: Why bigger can be better. *Nature Reviews Microbiology*, *10*(6), 417–430. https://doi.org/10.1038/nrmicro2790

Rath, S., Sharma, R., Gupta, R., Ast, T., Chan, C., Durham, T. J., Goodman, R. P., Grabarek, Z., Haas, M. E., Hung, W. H. W., Joshi, P. R., Jourdain, A. A., Kim, S. H., Kotrys, A. V., Lam, S. S., McCoy, J. G., Meisel, J. D., Miranda, M., Panda, A., … Mootha, V. K. (2021). MitoCarta3.0: An updated mitochondrial proteome now with sub-organelle localization and pathway annotations. *Nucleic Acids Research*, *49*(D1), D1541–D1547. https://doi.org/10.1093/nar/gkaa1011

Rawlings, N. D., Barrett, A. J., & Bateman, A. (2010). MEROPS: the peptidase database. *Nucleic Acids Research*, *38*(suppl_1), D227–D233. https://doi.org/10.1093/nar/gkp971

Ruan, J., & Li, H. (2020). Fast and accurate long-read assembly with wtdbg2. *Nature Methods*, *17*(2), 155–158. https://doi.org/10.1038/s41592-019-0669-3

Sarker S, Kallert DM, Hedrick RP, & El-Matbouli M. (2015). Whirling disease revisited: Pathogenesis, parasite biology and disease intervention. *Diseases of Aquatic Organisms*, *114*(2), 155–175.

Shafin, K., Pesout, T., Lorig-Roach, R., Haukness, M., Olsen, H. E., Bosworth, C., Armstrong, J., Tigyi, K., Maurer, N., Koren, S., Sedlazeck, F. J., Marschall, T., Mayes, S., Costa, V., Zook, J. M., Liu, K. J., Kilburn, D., Sorensen, M., Munson, K. M., … Paten, B. (2020). Nanopore sequencing and the Shasta toolkit enable efficient de novo assembly of eleven human genomes. *Nature Biotechnology*, *38*(9), 1044–1053. https://doi.org/10.1038/s41587-020-0503-6

Shpirer, E., Chang, E. S., Diamant, A., Rubinstein, N., Cartwright, P., & Huchon, D. (2014). Diversity and evolution of myxozoan minicollagens and nematogalectins. *BMC Evolutionary Biology*, *14*(1), 205. https://doi.org/10.1186/s12862-014-0205-0

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. https://doi.org/10.1093/bioinformatics/btv351

Székely, C., Molnár, K., Eszterbauer, E., & Baska, F. (1999). Experimental detection of the actinospores of Myxobolus pseudodispar (Myxosporea: Myxobolidae) in oligochaete alternate hosts. *Diseases of Aquatic Organisms*, *38*(3), 219–224. https://doi.org/10.3354/dao038219

Székely, C., Molnár, K., & Rácz, O. (2001). Complete developmental cycle of Myxobolus pseudodispar (Gorbunova)(Myxosporea: Myxobolidae). *Journal of Fish Diseases*, *24*(8), 461–468.

Takeuchi, F., Sekizuka, T., Ogasawara, Y., Yokoyama, H., Kamikawa, R., Inagaki, Y., Nozaki, T., Sugita-Konishi, Y., Ohnishi, T., & Kuroda, M. (2015). The Mitochondrial Genomes of a Myxozoan Genus Kudoa Are Extremely Divergent in Metazoa. *PLOS ONE*, *10*(7), 1–18. https://doi.org/10.1371/journal.pone.0132030

The UniProt Consortium. (2019). UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Research*, *47*(D1), D506–D515. https://doi.org/10.1093/nar/gky1049

Tilley, Molly. (2021). *A new and emerging myxozoan parasite of fathead minnows: Species description, life-cycle, and effects to the host* [Masters, University of Lethbridge]. https://hdl.handle.net/10133/6122

Tørresen, O. K., Star, B., Mier, P., Andrade-Navarro, M. A., Bateman, A., Jarnot, P., Gruca, A., Grynberg, M., Kajava, A. V., Promponas, V. J., Anisimova, M., Jakobsen, K. S., & Linke, D. (2019). Tandem repeats lead to sequence assembly errors and impose multi-level challenges for genome and protein databases. *Nucleic Acids Research*, *47*(21), 10994–11006. https://doi.org/10.1093/nar/gkz841

Valdivieso, E., Perteguer, M. J., Hurtado, C., Campioli, P., Rodríguez, E., Saborido, A., Martínez-Sernández, V., Gómez-Puertas, P., Ubeira, F. M., & Gárate, T. (2015). ANISERP: a new serpin from the parasite Anisakis simplex. *Parasites & Vectors*, *8*, 399. https://doi.org/10.1186/s13071-015-1006-z

Wolf, K., & Markiw, M. E. (1984). Biology Contravenes Taxonomy in the Myxozoa: New Discoveries Show Alternation of Invertebrate and Vertebrate Hosts. *Science*, *225*(4669), 1449–1452. https://doi.org/10.1126/science.225.4669.1449

Wolstenholme, D. R., Okimoto, R., & Macfarlane, J. L. (1994). Nucleotide correlations that suggest tertiary interactions in the TV-replacement loop-containing mitochondrial tRNAs of the nematodes, Caenorhabditis elegans and Ascaris suum. *Nucleic Acids Research*, *22*(20), 4300–4306. https://doi.org/10.1093/nar/22.20.4300

Wolters, J. F., Chiu, K., & Fiumera, H. L. (2015). Population structure of mitochondrial genomes in Saccharomyces cerevisiae. *BMC Genomics*, *16*(1), 451. https://doi.org/10.1186/s12864-015-1664-4

Yahalomi, D., Atkinson, S. D., Neuhof, M., Chang, E. S., Philippe, H., Cartwright, P., Bartholomew, J. L., & Huchon, D. (2020). A cnidarian parasite of salmon (Myxozoa: Henneguya) lacks a mitochondrial genome. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(10), 5358–5363. https://doi.org/10.1073/pnas.1909907117

Yahalomi, D., Haddas-Sasson, M., Rubinstein, N. D., Feldstein, T., Diamant, A., & Huchon, D. (2017). The Multipartite Mitochondrial Genome of Enteromyxum leei (Myxozoa): Eight Fast-Evolving Megacircles. *Molecular Biology and Evolution*, *34*(7), 1551–1556. https://doi.org/10.1093/molbev/msx072

Yang, Y., Xiong, J., Zhou, Z., Huo, F., Miao, W., Ran, C., Liu, Y., Zhang, J., Feng, J., Wang, M., Wang, M., Wang, L., & Yao, B. (2014). The genome of the myxosporean Thelohanellus kitauei shows adaptations to nutrient acquisition within its fish host. *Genome Biology and Evolution*, *6*(12), 3182–3198. https://doi.org/10.1093/gbe/evu247