

RESEARCH

Open Access



Comparative transcriptomics reveal stage-dependent parasitic adaptations in the myxozoan *Sphaerospora molnari*

Monika M. Wiśniewska^{1,2*}, Jiří Kyslík¹, Gema Alama-Bermejo^{1,3}, Alena Lövy¹, Martin Kolísko^{1,2}, Astrid S. Holzer^{1,3} and Anush Kosakyan^{1,4,5*}

Abstract

Background Parasitism as a life strategy has independently evolved multiple times within the eukaryotic tree of life. Each lineage has developed mechanisms to invade hosts, exploit resources, and ensure replication, but our knowledge of survival mechanisms in many parasitic taxa remain extremely limited. One such group is the Myxozoa, which are obligate, dixenous cnidarians. Evidence suggests that myxozoans evolved from free-living ancestors to endoparasites around 600 million years ago and are likely one of the first metazoan parasites on Earth. Some myxozoans pose significant threats to farmed and wild fish populations, negatively impacting aquaculture and fish stocks; one such example is *Sphaerospora molnari*, which forms spores in the gills of common carp (*Cyprinus carpio*), disrupting gill epithelia and causing somatic and respiratory failure. *Sphaerospora molnari* undergoes sequential development in different organs of its host, with large numbers of morphologically distinct stages occurring in the blood, liver, and gills of carp. We hypothesize that these parasite life-stages differ in regards to their host exploitation, pathogenicity, and host immune evasion strategies and mechanisms. We performed stage-specific transcriptomic profiling to identify differentially expressed key functional gene groups that relate to these functions and provide a fundamental understanding of the mechanisms *S. molnari* uses to optimize its parasitic lifestyle. We aimed to identify genes that are likely related to parasite pathogenicity and host cell exploitation mechanisms, and we hypothesize that genes unique to *S. molnari* might be indicative of evolutionary innovations and specific adaptations to host environments.

Results We used parasite isolation protocols and comparative transcriptomics to study early proliferative and spore-forming stages of *S. molnari*, unveiling variation in gene expression between each stage. We discovered several apparent innovations in the *S. molnari* transcriptome, including proteins that are likely to function in the uptake of previously unknown key nutrients, immune evasion factors that may contribute to long-term survival in hosts, and proteins that likely improve adhesion to host cells that may have arisen from horizontal gene transfer. Notably, we identified genes that are similar to known virulence factors in other parasitic organisms, particularly blood and

*Correspondence:
Monika M. Wiśniewska
monika.wisniewska@paru.cas.cz
Anush Kosakyan
anush.kosakyan@unimore.it

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

intestinal parasites like *Plasmodium*, *Trypanosoma*, and *Giardia*. Many of these genes are absent in published cnidarian and myxozoan datasets and appear to be specific to *S. molnari*; they may therefore represent potential innovations enabling *Sphaerospora* to exploit the host's blood system.

Conclusions In order to address the threat posed by myxozoans to both cultured fish species and wild stocks, it is imperative to deepen our understanding of their genetics. *Sphaerospora molnari* offers an appealing model for stage-specific transcriptomic profiling and for identifying differentially expressed key functional gene groups related to parasite development. We identified genes that are thus far unique to *S. molnari*, which reveal their evolutionary novelty and likely role as adaptations to specific host niches. In addition, we describe the pathogenicity-associated genetic toolbox of *S. molnari* and discuss the implications of our discoveries for disease control by shedding light on specific targets for potential intervention strategies.

Highlights

- We revealed *Sphaerospora molnari* developmental stage-specific expression profiles in infected host tissues.
- In gill stage parasites, genes related to development and cytoskeletal rearrangements are mostly up-regulated, while those up-regulated in the bloodstream stage are related to metabolism and host immune evasion strategies.
- We propose a list of “pathogenicity-related” gene families.
- We uncovered stage-specific up-regulation of *S. molnari* genes that have no homologues or were not found in other cnidarian lineages.

Keywords *Sphaerospora molnari*, Differential expression, Myxozoans, Pathogenicity related, Species specific genes

Background

Parasitic species are remarkably diverse, most likely representing approximately half of the life on our planet [1–3]. Parasites have evolved specific tools and strategies to expand into new host habitats [4]; yet, the genetic mechanisms that govern their adaptive strategies, including the ability to invade the host and reproduce successfully, remain largely unexplored.

The phylum Cnidaria is one of the earliest metazoan lineages [5]. Free-living cnidarians inhabit predominantly marine and freshwater environments. Evidence based on molecular clock analysis [6] suggests that about 600 million years ago, some of the free-living cnidarians began to evolve into endoparasites, leading to the emergence of Myxozoa, which presently accounts for 20% of all known cnidarian species [3, 7]. However, myxozoan diversity is likely greatly underestimated, as it has been suggested that the total number of Myxozoa is greater than that of free-living taxa [3, 8]. Parasitism has evolved several times within Cnidaria [9–11], but only myxozoans have had a massive diversification in aquatic ecosystems [6, 12], most likely due to their complex life cycle. Myxozoan life cycles involve both invertebrate (annelids or bryozoans) and vertebrate (mostly fish) hosts, in which transmissible spore stages are produced (actinospores and myxospores, respectively) [13]. Infection is initiated by the discharge of nematocysts—weaponry organelles—that facilitate the spore's attachment to the host [14–16]. Subsequently, the infectious sporoplasm released from the spore penetrates the host and proliferates within specific organs and tissues [17, 18]. After the propagative phases, the parasites enter either the terminal tissues or

organs of the host, where the final transmission stages, myxospores, are formed [14]. Mature myxospores are then released from the host to the environment to infect the definitive invertebrate host [19].

Numerous studies have been conducted on myxozoans, as they are associated with significant fish diseases that lead to substantial mortalities in both cultured and wild fish populations [20–28]. Moreover, due to increases in water temperature as a result of climate change, some myxozoan parasites have expanded massively, endangering aquacultures and wild stocks [17, 18, 21, 23–27]. However, to date, no specific methods have been developed to control myxozoan diseases. Understanding the genetic mechanisms responsible for invasion, survival, and proliferation of these parasites within their hosts is essential for designing targeted methods for predicting and controlling myxozoan diseases.

In this study, we used the cnidarian parasite *Sphaerospora molnari* (Cnidaria: Myxozoa: Myxosporaea) as a model to explore the gene expression profile and genetic mechanisms underlying feeding, propagation, and immune evasion strategies. *S. molnari* (Fig. 1) is a parasite of common carp (*Cyprinus carpio*) and is responsible for mortalities of carp stocks in central Europe [29, 30]. The parasite ultimately develops into mature myxospores in the gill filaments of the fish, causing branchial necrosis and respiratory failure [29]. Before spore formation in the gills (the so-called “gill stage” (GS)), *S. molnari* undergoes specific development within the blood [29, 31]. In the vascular system, the parasite forms multicellular extrasporogonic motile stages (the so-called “blood stage” (BS); Fig. 1) [29, 30]. It has been suggested that the

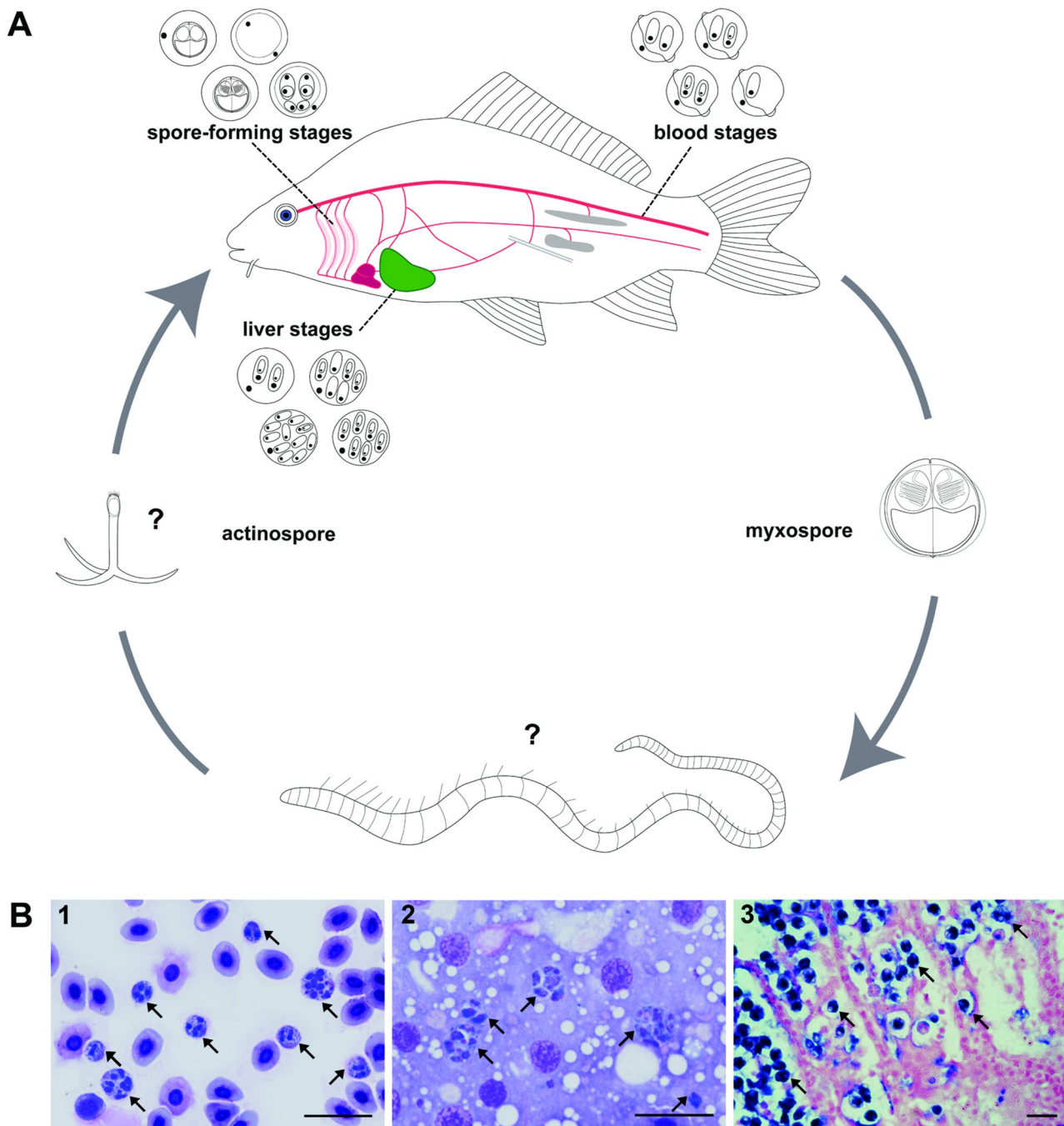


Fig. 1 The schematic life cycle of *Sphaerospora molnari*. **A** Within its known intermediate host - common carp (*Cyprinus carpio*)– the parasite undergoes extrasporogonic development in blood and liver, followed by sporogonic development in the gills, leading to the creation of myxospores, which are released into the freshwater environment. The life cycle continues with the myxospore infecting the final host - currently unknown but likely an annelid– where it develops into actinospores, which are released (not known, shown as neoactinomyxum type) to infect the intermediate host (common carp). **B** 1 & 2 Kwik-Diff stain of infected carp cells/tissues. (1) Blood smear showing multicellular parasite stages (indicated by arrowheads) surrounded by mostly red blood cells; (2) Liver tissue with large multicellular liver stages (indicated by arrowheads); (3) DNA in-situ hybridization of a gill tissue section of common carp fingerlings heavily infected with *S. molnari* spore-forming stages (indicated by arrowheads). The monosporous pseudoplasmodia develop in large numbers in the epithelium of the gill filaments. *S. molnari* is shown by dark blue coloration (arrowheads). Scale bars = 20 μ m

parasite adopted special mechanisms to avoid the host immune system and to use the bloodstream to circulate freely through host organs (for example, intracellular disguise, motility, polyclonal activation of B cell responses, skewing of host response to an anti-inflammatory phenotype, etc.) [31].

The infection kinetics of *S. molnari* have been systematically examined using quantitative RT-PCR in different organs of infected carp. An initial peak of parasite multiplication was documented at 28 days post-infection (dpi) in the blood, followed by invasion of the gills at 53 dpi. Notably, an unexpected re-emergence and the highest peak of parasite proliferation was detected at 42 dpi in the liver (studied over a total of 63 days; [31]). The study demonstrated that *S. molnari* completes the main stages of its development in the blood, liver, and gills of carp. Here, we aimed to better understand the pathogenicity-related molecular toolbox of *S. molnari*. We performed differential expression analysis between the blood and gill stages to help identify virulence factors used by *S. molnari* during its blood stage development (i.e.: genes that are specifically up-regulated in the blood stage samples). Additionally, we used comparative transcriptome analysis of cnidarians to identify genes that appear to be specific to *S. molnari* and discovered a relatively large overlap with the differentially expressed genes.

Table 1 Comparison between the *Sphaerospora molnari* genome and transcriptome assemblies, including BUSCO scores, predicted proteins, and orthogroups

	Genome Assembly	Genome-guided Transcriptome Assembly	De novo Transcriptome Assembly
Number of contigs	40	20,699	49,042
Total length (bp)	40,172,413	16,945,146	50,900,386
N50 (bp)	2,792,954	1,122	1,615
BUSCO scores against Metazoa (% complete)	46.4	41.3	45.5
BUSCO scores against Eukaryota (% complete)	64	57.3	67
Total predicted proteins	14,957	13,464	30,087
Total number of orthogroups	6,369	8,564	10,573
Number of orthogroups unique to the assembly	323	170	2,294
Number of orthogroups shared with genome assembly	-	523	408
Number of orthogroups shared with genome-guided transcriptome assembly	523	-	2,756
Number of orthogroups shared with de novo transcriptome assembly	408	2,756	-

Results
Outcome of different approaches used in *S. molnari* transcriptome assemblies

In general, when access to the genome of an organism or its closely related counterpart is feasible, it is preferable to use a genome-guided approach to assemble transcriptomic data [32, 33]. Despite the availability of a well-curated genome of *S. molnari*, the transcriptomic data presented in this study were assembled using both *de novo* and genome-guided approaches, and the results were compared. The *de novo* transcriptome assembly generated almost twice the number of contigs and protein-coding isoforms (49,042 transcripts; 30,087 protein-coding isoforms) compared to the genome-guided approach (20,699 transcripts; 13,464 protein-coding isoforms). The *de novo* assembly also demonstrated a higher N50 value, and its BUSCO score, when assessed against eukaryotic lineages, surpassed that of the genome-guided assembly (Table 1).

From the Orthofinder analysis, we retrieved 8,564 orthogroups (OGs) in the genome-guided assembly, 10,573 OGs in the *de novo* assembly, and 6,369 in the genome assembly. In total, 5,115 OGs were common to all assemblies; whereas, 170 OGs, 2,294 OGs, and 323 OGs were specific to the genome-guided, *de novo*, and genome assemblies, respectively (Fig. 2A). Comparing pairs of assemblies, 2,756 OGs were common between the genome-guided and *de novo* assemblies, 408 were shared between the genome and *de novo* assemblies, and 523 were found in both genome and genome-guided assemblies (Fig. 2A). Additionally, the completeness and quality of both transcriptome assemblies were assessed using protein sequences predicted by TransDecoder as queries for BLASTp searches against the UniProt database. As presented in Fig. 2B, the genome-guided assembly recovered longer and more continuous contigs that resulted in predictions of more complete proteins, in most cases more than 50% complete. Furthermore, when comparing transcripts from the genome-guided assembly with the genome, we observed that most of the contigs correspond to full genes from the genome. On the other hand, the *de novo* transcriptome assembly computed many truncated contigs that resulted in incomplete proteins, often less than 40% complete. Given that the *de novo* assembly generated mostly fragmented transcripts resulting in an overestimation of KEGG pathways, we opted to use the genome-guided assembly for all further analyses, including the differential expression analysis.

Exploratory data analysis
After the read filtration step, the number of reads retained in the liver samples, L1 and L2, was below the threshold value set to one million reads, as suggested in Liu et al. 2014 [34]. These samples were therefore

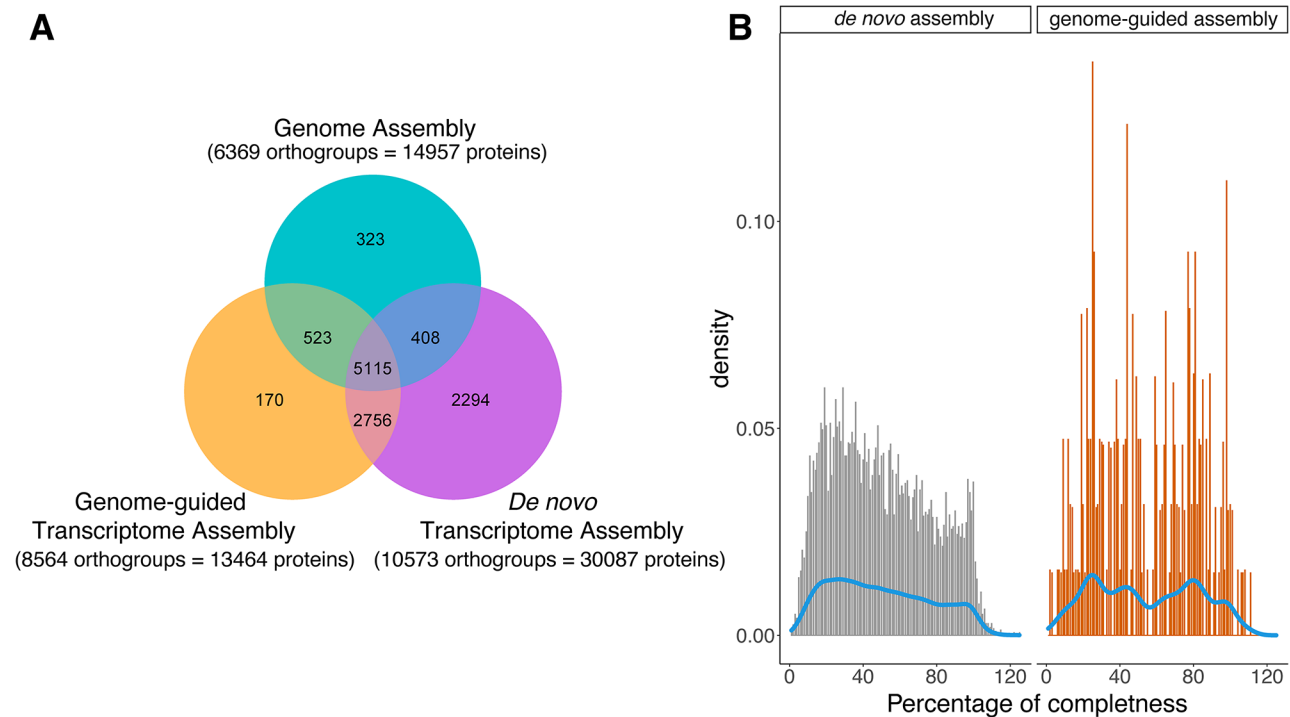


Fig. 2 Comparison of the genome and transcriptome assemblies. **A** Venn diagram depicting shared and unique Orthogroups between the assemblies. **B** Histogram showing the density of the completeness of the transcripts. The continuous line (blue) denotes the density of the completeness for each group of data (using kernel density estimation). The genome-guided assembly recovered more complete transcripts than the *de novo* assembly, the latter was characterized by < 1% of transcripts with completeness > 80%. Bars with completeness greater than 100% have query sequences longer than the subject sequence. They most likely represent signal peptides

removed from the differential expression analysis (see Additional file 1 A).

The total number of reads for two blood samples (B3 and B4) was low compared to the remaining blood samples. This may have been caused by differences in sample collection methods (the blood stage replicates B1 and B2 were parasites isolated from blood using ion exchange; the B3 and B4 samples were isolated using hematocrit capillaries and represent a mixture of parasites with white blood cells). Additionally, the samples had different origins: while B1 and B2 were collected from laboratory models, each from a single fish, the B3 and B4 samples were collected from a pond, each pooled from multiple pondinfected fish. To confirm that our results were not affected by the heterogeneity of these samples, we also performed differential expression analysis without the B1 and B2 samples. We observed similar results after removing B1 and B2, as most of the differentially expressed genes overlapped between the two methods (Additional File 2). A small decrease in the number of up-regulated genes (~16.5% fewer up-regulated genes) and a small increase in the number of down-regulated genes (9% more down-regulated genes) was observed. There was also a slight increase in the log Fold Change (lgFC) values after removing B1 and B2; however, the differences could be a result of the low number of replicates for the analysis

without B1 and B2 rather than actual biological differences or the extraction method (Additional File 2).

Out of 14,124 contigs in the assembly, 12,530 passed the filtration step (counts per million (CPM) 0.5 in at least one library) (Additional file 1 A). During exploratory data analysis, we observed changes in thousands of genes when comparing the blood and gill stages (Additional file 1 A-D). The top 2,000 genes ranked by standard deviation were divided into four groups based on their expression patterns using k-means clustering (Additional file 3). Principal component analysis (PCA) indicated that the first two components explained 88% of the variance between the blood and gill stages of the parasite (Additional file 4).

Next, we performed enrichment analysis for the four clusters (A-D) calculated by k-means clustering using clusterProfiler [35] with KEGG KOs, OGs, KEGG Pathways, and Pfam domains, as well as topGO [36] with GO terms (Additional file 3). Information on the enriched cellular components, molecular functions, and biological processes in the investigated clusters is presented in Additional file 5.

Differentially expressed genes (DEGs) and gene set enrichment

Using the DESeq2 package, we identified 2,283 up-regulated and 1,609 down-regulated DEGs (Fig. 3A) using a threshold of false discovery rate (FDR) < 0.01 and fold-change (FC) > 4. The volcano plot (Fig. 3B) illustrates a substantial number of genes that are significantly differentially expressed between the blood and gill stages of the parasite. Upon examination of the top 30 genes ranked by their absolute values of fold change (FC), distinctions were identified in various biological processes, including the lysosome, peptidase activity, transcription factors, homeostasis, and protein catabolic processes.

To further explore the functions of the DEGs, we performed enrichment analysis (Fig. 4). In the blood stage, we observed an up-regulation of genes involved in immunity, redox reactions, glycosylation, glycolysis and gluconeogenesis, calcium metabolism and protein kinase C activity, Golgi function, and locomotion. The most significantly enriched down-regulated genes related to signaling and sensing, host invasion, cysteine peptidase activity, the endomembrane system, and water homeostasis. Interestingly, in both sets of DEGs (up-regulated and down-regulated) the fluid shear stress and atherosclerosis pathway and the cathepsin L KO were significantly enriched. For more details, see Table 2.

Differentially expressed genes related to parasitic strategies and *S. molnari*-specific genes

Next, we examined genes that could be directly linked to pathogenicity, considering their homologues had been identified as virulence factors found in other pathogenic organisms (e.g., *Plasmodium*, *Giardia*, *Trypanosoma*, and different species of bacteria). We selected 159 genes as “pathogenicity-related” genes putatively involved in *S. molnari* invasion of the host, that were further divided into the following categories: “Adhesion molecules”, “Anticoagulants”, “Genes involved in heme metabolism and transport”, “Host invasion-related molecules”, “Immunity-related”, “Toxins”, and “Others”. A brief description of these categories and examples are given in Additional file 6.

We compared the *S. molnari* transcriptome assembly to all publicly available cnidarian high throughput sequencing datasets, including Myxozoa. We revealed that out of 13,464 protein-coding transcripts, 5,560 (corresponding to 763 orthogroups) are specific to *S. molnari* (Fig. 5). Among these, 45% of the transcripts (1,850) were classified as differentially expressed, 491 of which were functionally annotated (1359 could not be functionally annotated or were designated as domain of unknown function (DUF)). The expression values and categories of these genes are presented in Fig. 6.

We observed an up-regulation of genes in the blood stage that allow for proliferation, such as flagellar-associated PapD-like domain, p85-binding domain, PI3-kinase family, G-rich domain on putative tyrosine kinase,

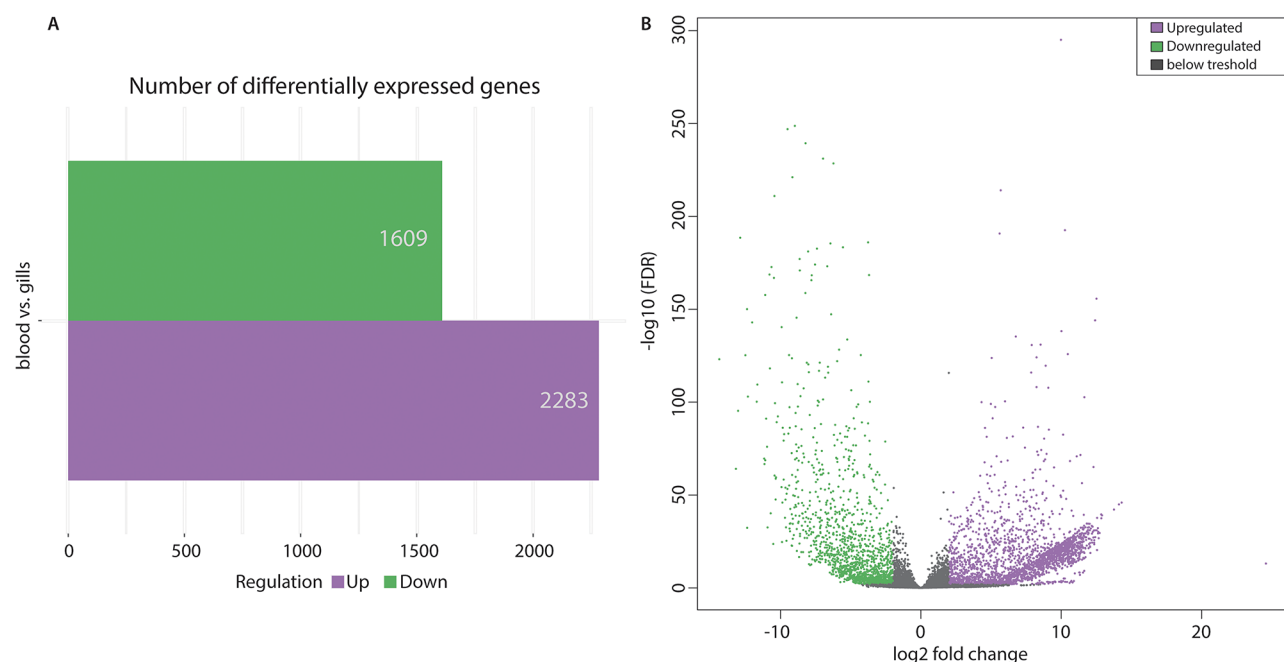


Fig. 3 DEGs calculated with the DESeq2 package with a threshold of false discovery rate (FDR) < 0.01 and fold-change (FC) > 4. **A** Bar plot showing the number of DEGs. **B** Volcano plot showing significant DEGs between blood and gill stages of the parasite

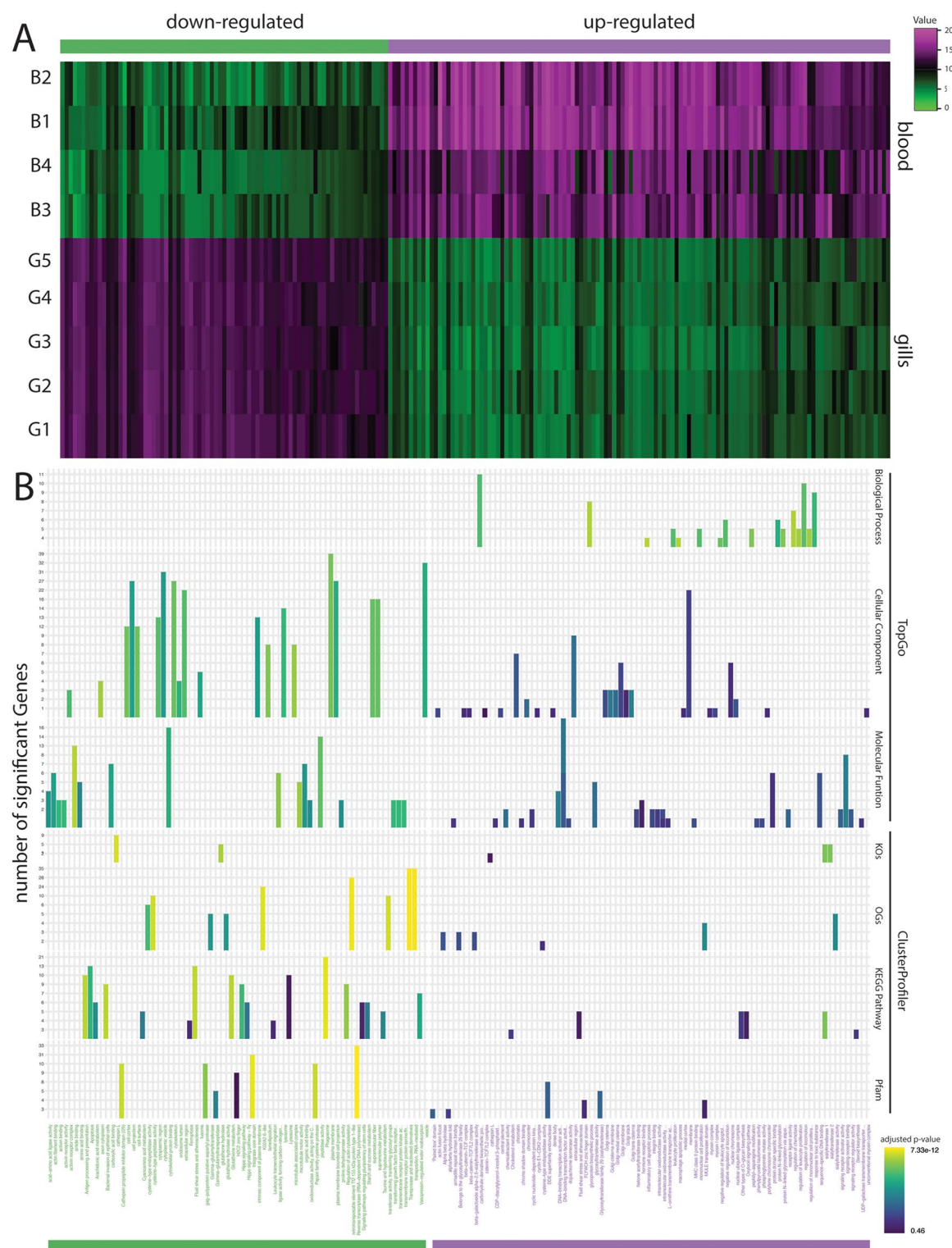


Fig. 4 Enrichment of DEGs. **A** Heatmap showing the expression patterns of the up-and down-regulated genes for the blood versus gill stage comparisons. **B** Bar plot showing enrichment using topGO (with the GO terms assigned to Biological Process, Cellular Component, and Molecular Function) and ClusterProfiler (with KEGG KOs, OGs, KEGG Pathways, and Pfam domains) for the genes in each cluster. The X-axis represents enriched groups in the up-regulated (purple bar) and down-regulated (green bar) genes. The Y-axis denotes methods used for the enrichment analysis and the significance of enriched terms (the higher and more yellow a bar is, the more significantly enriched is its term)

Table 2 List of the most significantly enriched gene categories based on the annotations (GO terms and KO identifiers)

Genes up-regulated in blood stage involved in / related to	Examples
Immunity	macrophage apoptotic process, regulation of leukocyte proliferation, leukocyte proliferation, phenylpyruvate tautomerase activity
Redox reactions	intramolecular oxidoreductase activity
Glycosylation	sialyltransferase, glycosyltransferase activity, protein N-linked glycosylation
Glycolysis and gluconeogenesis	phosphoglycerate mutase activity
Calcium metabolism and protein kinase C activity	CDP-diacylglycerol-inositol 3-phosphatidyltransferase activity
Golgi	Golgi trans cisterna, Golgi membrane
Locomotion	Unconventional myosin complex, myosin I complex, regulation of locomotion
Genes up-regulated in gill stage involved in / related to	Examples
Signaling and sensing	activin receptor complex, transmembrane receptor protein tyrosine kinase signaling pathway
Host invasion	lamellipodium, microtubule associated complex, lysosome, plasma membrane bounded cell projection, microtubule motor activity
Cysteine peptidase activity	peptidase activity, cysteine-type peptidase activity, cysteine-type endopeptidase activity, cathepsin propeptide inhibitor domain (I29), papain family cysteine protease)
Endomembrane system	endosome lumen, vesicle, cytoplasmic vesicle), and water homeostasis (vasopressin-regulated water reabsorption

GldH lipoprotein, and cell-division protein ZapC, most of which are classified within the “Cellular Processes” group (Fig. 6). Some transcription factors and translation regulators classified as “Genetic info processing” (Fig. 6), such as SAMdependent carboxyl methyltransferase, ferric uptake regulator family, and ERF superfamily (DNA binding factor) were significantly up-regulated in the blood stage. Additionally, we identified 368 “secreted proteins” in the *S. molnari* unique gene set (Fig. 7), most of which were annotated as domain of unknown function (DUF) or were not functionally annotated. Interestingly, we observed multiple copies of genes for surface proteins annotated as VSPs, ETRAMP, and Riffin/Stevor family in the *S. molnari* specific gene set that fulfilled the requirements of secretome proteins (presence of a signal peptide and no transmembrane domain after the signal peptide).

Discussion

Here, we analyzed the functional annotation of differentially expressed genes between two discrete intrapiscine life stages of the myxozoan *Sphaerospora molnari* and compared its transcriptome across closely related species

(other myxozoans and cnidarians). Within the *S. molnari* transcriptome, we identified multiple gene families that are differentially expressed between the blood and gill stages, and distinguished a set of genes that are unique to *S. molnari*. The results support our hypothesis that variation in gene expression profiles amongst parasite stages may reflect distinct mechanisms for proliferation, feeding, and the evasion of host immunity. This variability in gene expression profiles is consistent with experimental evidence demonstrating morphological and behavioral differences between blood and gill stages.

Blood stage

Bloodstream stage (BS) parasites utilize specific mechanisms to feed, evade the host’s immune system, multiply, and move within the host’s circulatory system. BS parasites are easily identifiable by their characteristic non-directional and consistent movements, often referred to as “dancing” or Membrane Fold Induced Tumbling (MFIT) [37]. The external membrane of BS parasites is used as a motility effector to rotate and move the parasite [37]. We observed several up-regulated genes related to motility, multiplication, and growth in the BS samples: for example, actin, cell-division protein ZapC, inner centromere protein, ARK binding region, flagellar-associated PapD-like, and microtubule motor activity.

The attachment process of BS parasites and their feeding on host erythrocytes have been previously documented [31], and in this study, we affirm the critical role of these processes in host colonization and the establishment of infection. We have identified several genes that are up-regulated in BS parasites– such as cytoadhesion molecules such as cadherin-like beta -sandwich domain, MAP domain, fibronectin-binding protein (FBP), ornatin, and glycoprotein GP40– which have been shown in other systems to play important roles in attachment to hosts and host cells, as well as in cell invasion [38–41]. For instance, *Trichomonas vaginalis* uses cadherin-like proteins for adherence to human mucosal epithelial cells, thereby establishing infection [38]. Similarly, GP40 is implicated in the attachment of *Cryptosporidium parvum* sporozoites to the host cell surface [39]. This suggests that *S. molnari* may use products of these genes for attaching to the host erythrocyte. Although there is no reported intracellular development of myxozoans in erythrocytes, it is possible they are used as a nutrient source. Blood itself provides nutrients, being rich in amino acids, sugars, lipids, and heme [40, 41]. It has been shown that many parasites have lost genes and pathways that would be essential for survival outside the host; they have instead gained and retained the ability to scavenge and utilize metabolites from the host, like for example sugars as a swift source of energy [42, 43]. For instance, *Thelohanellus kitauei* has adapted to inhabit

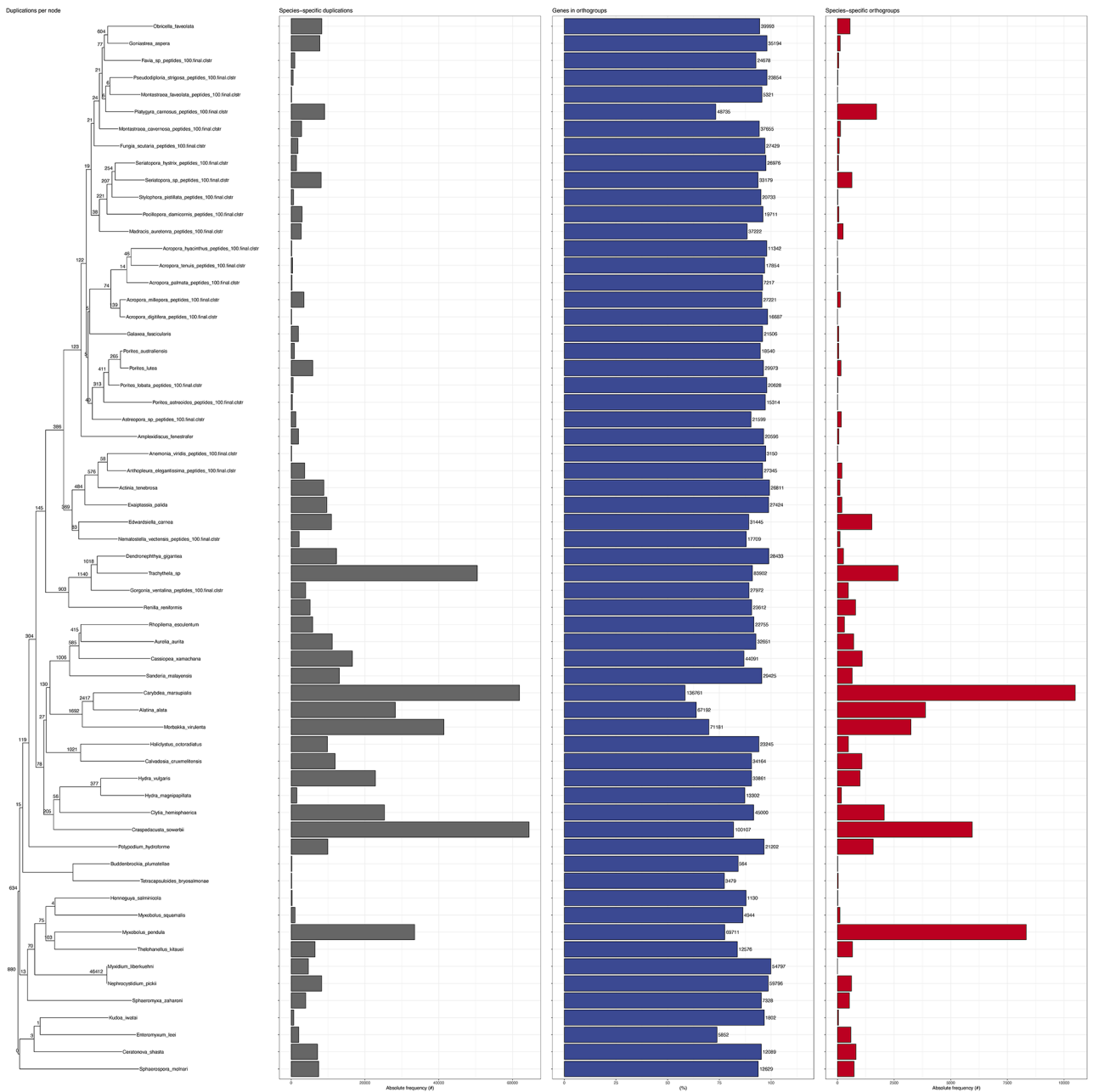


Fig. 5 Comparisons between cnidarian (including myxozoan) orthogroups, numbers above the branches correspond to duplication per node. Gene families shared and unique amongst cnidarian datasets. Orthofinder summary plot showing species-specific duplications (grey bars), number of genes in orthogroups (blue bars), and species-specific orthogroups (red bars) amongst Cnidarian datasets used in this study

the host gut submucosa, which is extremely oxygen- and nutrient-rich (especially for glucose) [43]. The *T. kitauei* genome showed an absence of classic and complimentary anaerobic pathways, gluconeogenesis, and *de novo* synthesis of fatty acids, amino acids, and nucleotides. It can be expected that the BS of *S. molnari* also scavenges glucose from host blood to serve as “input” for the glycolysis pathway. This inference is supported by our detection of the three key enzymes of glycolysis: hexokinase, phosphofructokinase, and pyruvate kinase in *S. molnari*. This

suggests that the parasite can produce glucose-6-phosphate from glucose, which is a key precursor for several essential metabolic pathways (see KEGG reference pathway map in Additional file 7). Another major group of proteins important for blood-feeding parasites are heme transporters. Various hemoparasites, such as *Plasmodium*, *Schistosoma*, *Trypanosoma*, and *Aeromonas* can lack some enzymes of the heme biosynthetic pathway. Consequently, these parasites rely on heme transporters or intermediates

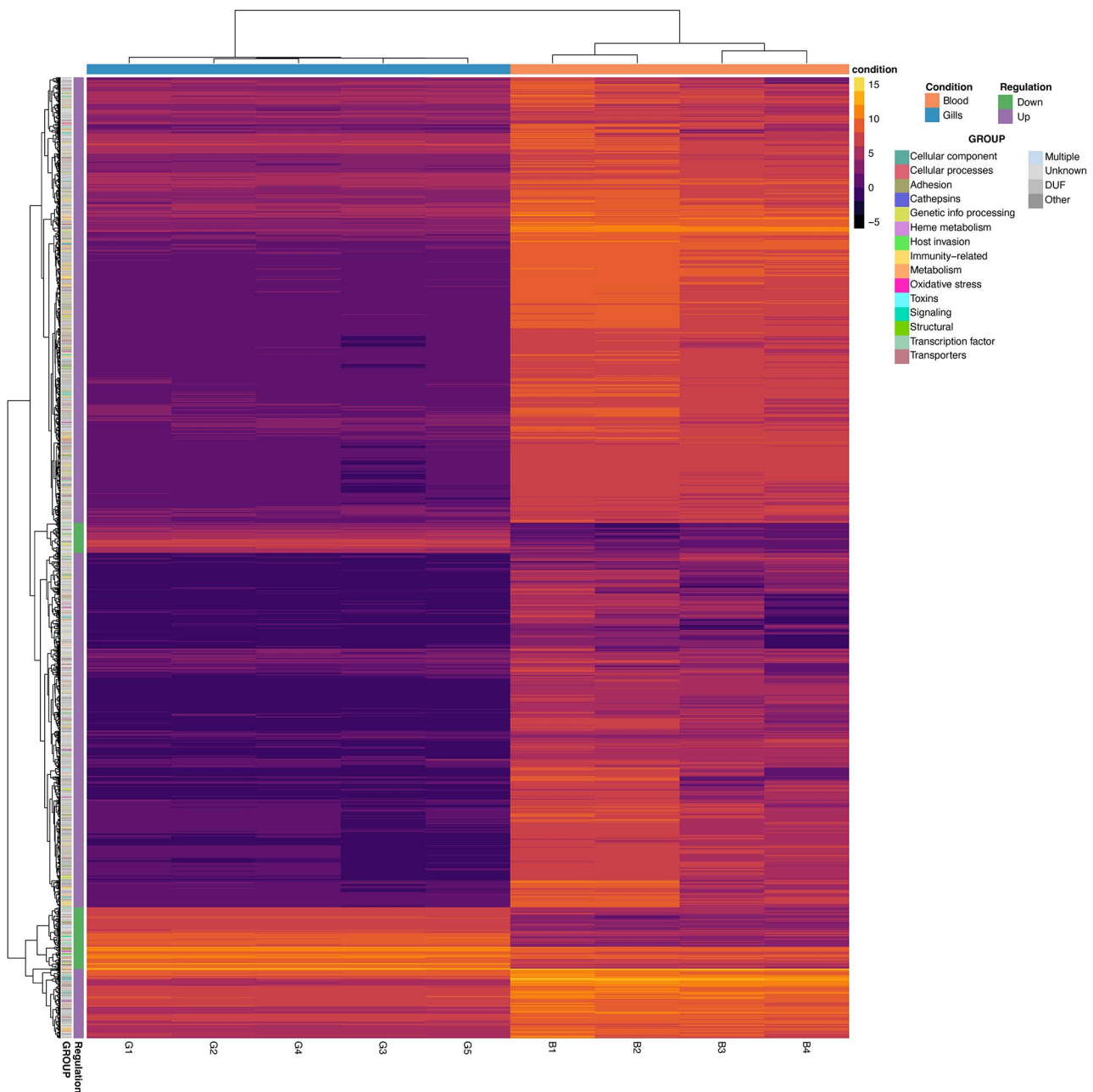


Fig. 6 Heatmap showing the *S. molnari* unique transcripts that were also differentially expressed. The up- and down-regulated genes are color-coded on the left side of the heatmap (green for down-regulated and purple for up-regulated genes), as well as the functional Gene Ontology Terms categories (column GROUP). The condition (blood or gill stage) is denoted above the heatmap (orange and blue, respectively)

for their metabolism and growth, including cell surface heme-binding proteins, heme exporter proteins, siderophore-interacting flavin adenine dinucleotide (FAD), and heme nitric oxide (NO) binding associated proteins to scavenge, transport, and assimilate heme [41, 44–46]. An almost complete pathway for heme biosynthesis is present in the *S. molnari* transcriptome; the only enzyme missing is uroporphyrinogen decarboxylase [EC: 4.2.1.75], which is responsible for the conversion of hydroxymethylbilane to uroporphyrinogen

(see porphyrin metabolism KEGG reference pathway in Additional file 7). The absence of this particular gene within our transcriptomic data may be ascribed to the incompleteness of the dataset, given that we successfully identified uroporphyrinogen decarboxylase in the *S. molnari* reference genome (SRA accession SAMN45157437 [47]). While our data do not conclusively determine the completeness of *S. molnari*'s heme biosynthesis pathway or whether the parasite necessitates scavenging heme or heme metabolites from the host, our findings

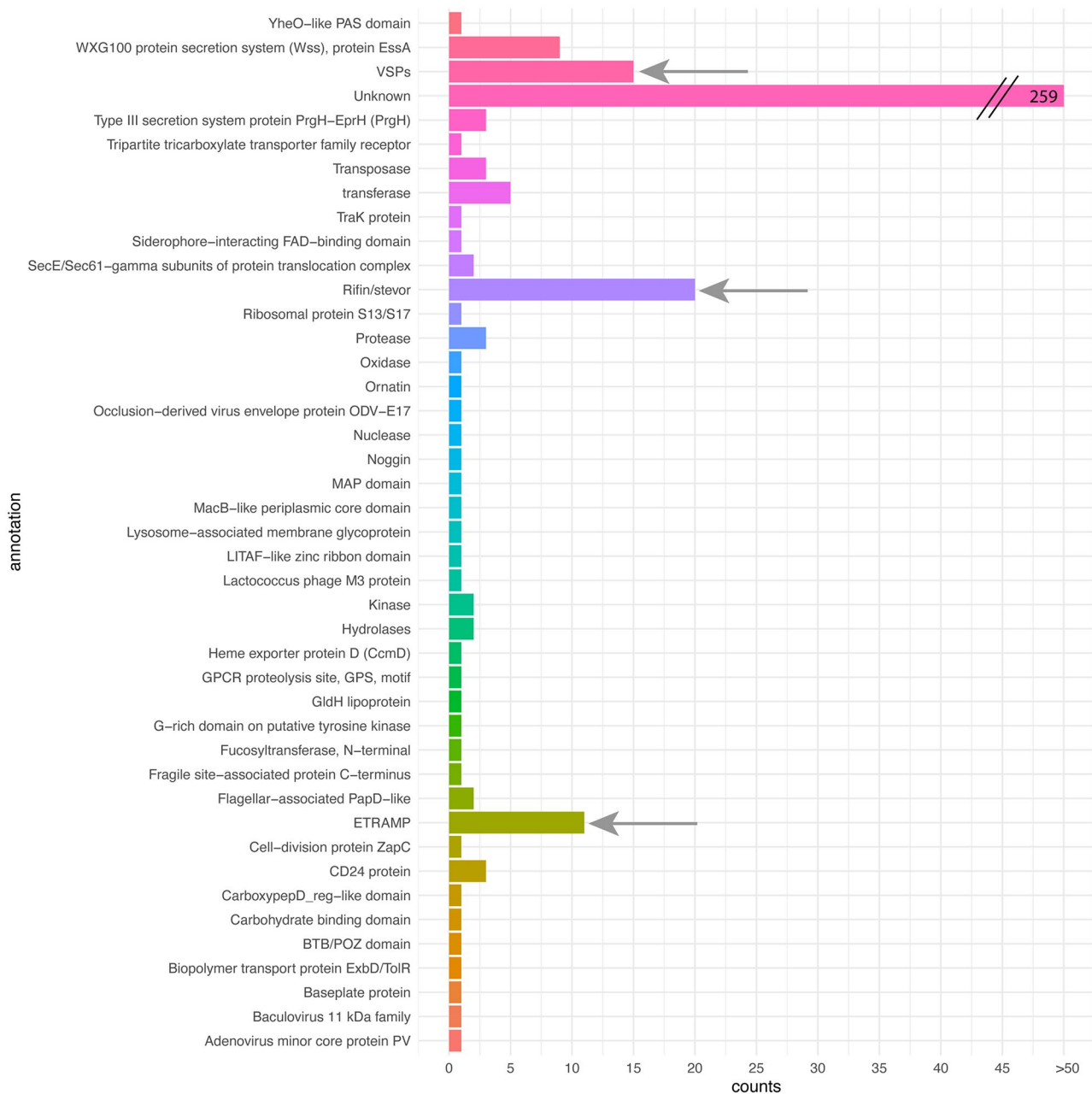


Fig. 7 Bar plot showing the *S. molnari* species-specific genes that were identified as secreted proteins. The colors of the bars denote different functional categories of secreted proteins. The size of the bars represents gene counts. Arrows indicate the protein groups analogous to known virulence factors in other parasitic organisms. The largest group, containing 259 genes, was comprised of unknown proteins

unequivocally indicate a substantial increase in both heme and glucose binding and transport across the parasite's cell membrane, as well as metabolic activity in BS parasites when compared to the GS. In particular, we noted significantly higher relative expression of specific genes involved in these processes, including cell surface heme-binding protein Shp, glycosyltransferase family 29 (sialyltransferase), ABC transporter, heme exporter protein D (CcmD), heme NO binding associated, iron transport-associated domain, and maltose transport system

permease protein MalF P2 domain, among others. The simultaneous presence of enzymes used in the heme biosynthesis pathway and the up-regulation of heme transporters that likely enable scavenging heme from host blood implies that *S. molnari* uses alternate tissue-specific strategies to obtain heme. In the blood, where heme is highly accessible, *S. molnari* may activate transporters. Conversely, in the GS, or potentially within the invertebrate host, the parasite might activate its heme

biosynthesis pathway. However, this hypothesis requires further investigation and experimental validation in the future.

Gill stage

S. molnari completes its development in the gills by forming myxospores as transmission stages to the final annelid host. This development includes early sporogonic stages (pseudoplasmodia with sporoblasts [29]) and mature myxospores. GS parasites are morphologically very different from BS parasites (Fig. 1). Mature *S. molnari* myxospores comprise highly differentiated cells, hard-shell valves, sporoplasm, and nematocysts with filaments for attachment to the host [29].

In the GS, the most significantly enriched groups of genes were classified as structural genes, toxins, host cell invasion-related genes, proteases, and genes associated with growth and development. Structural genes, including ARP2/3 complexes, cytoplasmic Fragile-X interacting family genes, chitin binding Peritrophin-A domain, and cofilin/tropomyosin-type actin-binding protein, are likely implicated in shaping spore structure, as described for other myxozoans (for example: *Tetracapsuloides bryosalmonae*, see [48], and *Ceratonova shasta*, see [49]) and free-living cnidarians [50]. The differentially expressed toxin genes exhibit similarities to bacterial toxins involved in virulence and host invasion mechanisms: for example, anthrax toxin lethal factor, which is a primary virulence factor for *Bacillus anthracis* and is capable of disrupting host signaling pathways and causing cell destruction, as well as diphtheria toxin, which is secreted by *Corynebacterium diphtheriae* to aid in host cell penetration. Another example is the LXG domain of the WXG superfamily, which is predicted to play a role in the secretion pathway for toxin export in various bacterial species [51–53]. Most probably *S. molnari* uses toxin genes in combination with cell breakdown genes (i.e., AhpC/TSA family proteins) to facilitate the disruption of host epithelium, thereby promoting the release of spores into the environment. Previous studies on toxins of free-living cnidarians and myxozoans suggested that myxozoans can produce a diverse array of toxins, which may have initially served as venoms in their free-living ancestors. However, these toxins could have been repurposed during the transition to an endoparasitic lifestyle, potentially for developmental processes and interactions with hosts rather than for prey capture or predator defence [54].

GS parasites also seem to up-regulate genes involved in modulating the cell invasion machinery. Notable examples include membrane-associated apoptosis protein, GTPase-activator protein for Ras-like GTPase, MSP1 EGF domain 1, and Rab-GTPase-TBC domain. These proteins are known to block or limit host immunity and parasite elimination mechanisms in other systems:

for instance, inhibition of host autophagy machinery in *Leishmania* parasites [55]. Additionally, a few immunity-related gene families, such as those belonging to the immunoglobulin-like superfamily (IgSF) and proteases, are significantly up-regulated in GS parasites. The proteases are likely part of the nematocyst's "venom-like compounds" that facilitate future invasion [56]. We speculate that the parasite uses IgSF for non-self-recognition, as has been shown in other cnidarians [57]. The most enriched GO terms in the GS are mainly linked to the plasma membrane, phagosome, and locomotion processes (i.e.: actin filament binding, lamellipodium, microtubule complex). These likely indicate changes in the actin cytoskeleton and facilitation of migration from blood vessels into the gill epithelia, which supports the formation of the mature spore structure, as observed in the myxozoan *Ceratonova shasta* [49]. Finally, up-regulation of lysosomal genes may play a role in digesting materials scavenged from the host or breaking down components of spore-forming cells during the maturation of myxospores.

Host immune evasion by *Sphaerospora molnari*

Parasites have evolved various mechanisms to avoid host immune systems. This avoidance can be achieved by altering the antigenic surfaces of parasites during infection, a phenomenon observed in *Plasmodium*, *Trypanosoma*, and bacteria (reviewed in [58]). Alternatively, parasites may produce molecules (e.g. adhesins and invasins) to block or modulate specific steps during the host's immune response. Another method involves injecting modulatory proteins directly into host cells using specialized secretion systems, as reported in bacteria [59].

One prominent category of enzymes related to host exploitation are proteases, involved in the various mechanisms of pathogenicity. To date, several studies have highlighted myxozoan proteases as virulence factors [43, 54, 56, 60–63]. Such virulence factors represent candidate drug targets in therapies for myxozoan diseases. Hartigan et al. [63] specifically examined how *S. molnari* proteases that act as virulence factors can provide potential vaccine candidates. Here, we identified multiple up-regulated papain family cysteine proteases and aspartyl proteases in the GS, as well as OTU-like cysteine protease.

We also identified an up-regulation of species-specific integral surface proteins facilitating host immunity evasion in BS parasites (e.g., variant-specific surface proteins (VSPs; *Giardia intestinalis*), Malarial early transcribed membrane protein (ETRAMP; *Plasmodium falciparum*), and RIFINRifin/STEVOR Stevor family proteins (*Plasmodium falciparum*); Fig. 7). These proteins have been linked with the secretome in other systems, including diplomonad parasites such as *Giardia intestinalis* and *Spirionucleus salmonicida*, which encode variant-specific

surface proteins (VSPs). These proteins are variably expressed through cell generations and are presumed to be an antigenic adaptation enabling evasion of the host immune system [64, 65]. VSPs are cysteine-rich proteins with frequent CXXC motifs followed by a transmembrane domain and a short intracellular N-terminal highly conserved motif. Remarkably, *Giardia* changes its expression from one VSP to another every sixth to fourteenth generation [64]. We have identified multiple copies of genes annotated as VSPs in our DEGs. The putative VSPs identified in *S. molnari* show high similarity to the structure of *S. salmonicida* VSPs and have not been identified in any other myxozoan lineages; thus, we believe they are lineagespecific. Comparative analysis of *S. molnari* VSPs with those of diplomonad species such as *Giardia muris*, *Giardia intestinalis*, and *S. salmonicida*, revealed similarities between the N-terminal motifs of *S. molnari* (“[MMKK]”) and *S. salmonicida* (“[KR][KR]X[KR][KR]”). The initial fasta file and sequence alignments (full and trimmed) can be found in the Figshare repository [66].

To date, ETRAMPs constitute a protein family exclusive to *Plasmodium* species [67]. We found multiple copies of genes that are similar to the ETRAMP and RIFINRifin/STEVORStevor family proteins up-regulated in the *S. molnari* blood stage. Export of proteins in the RIFIN and STEVOR families to the cell surface of red blood cells infected by *Plasmodium* [68] enables escape from host immune responses and mediates adhesion of RBCs [69–71], thus establishing long-lasting infections [68]. *S. molnari* may similarly use the products of these genes to evade the host immune system.

In summary, our findings suggest convergent evolution of immune evasion mechanisms, as *S. molnari* expresses a variety of proteins used in the immune evasion strategies of other parasites. However, the evolutionary origin of these proteins and their specific functions need further investigation to confirm convergence.

***Sphaerospora molnari* species-specific and pathogenicity-related genes**

This study unveiled a *Sphaerospora*-specific gene repertoire that is not present in other publicly available cnidarian, including other myxozoan, datasets. The observed absence of genes in other cnidarian lineages may stem from incomplete datasets, or overpredictions/mis-assembly of the *S. molnari* transcriptome. Gene duplication, divergence, and neofunctionalization during evolution appear to account for most species-specific genes [72] and are exhibited in other parasitic species [73, 74]. The genome of myxozoan, *Myxobolus honghuensis*, shows a similar pattern, as a notable proportion of its genes are absent in other myxozoan species, which may represent specific adaptations for host exploitation [75]. Comparison of the *S. molnari* dataset with other

myxozoan transcriptomic and proteomic profiles [48, 49], indicates that a set of annotated species-specific genes (491 out of 5560 *S. molnari* species-specific genes) may be related to parasitic niches, exploitation strategies, tissue tropism, and phenotypic plasticity of developmental stages and organelles. Some species-specific genes that are up-regulated in BS parasites may originate from horizontal gene transfers (HGTs) from bacteria, including Gliding motility-associated lipoprotein GldH and Cell-division protein ZapC. These genes may be involved in cell motility, as documented in bacteria [76, 77], contributing to cell crawling mobility and the movement of BS parasites out of the vascular system.

Like other parasites, *S. molnari* demonstrates stage-specific protein secretion that is likely involved in host cell invasion, immune evasion, feeding, and replication, ensuring exploitation of host resources [15, 78, 79]. A total of 160 identified transcripts, grouped into 6 functional categories (see Additional file 6), are probably associated with the pathogenicity of *S. molnari*. These genes are inferred to contribute to immune evasion, mediated by proteases and surface proteins, and exhibit similarities to bacterial toxins (as discussed above). Homologs of these genes are characterized in various pathogenic protozoans, making them promising candidates for developing control measures against a broad range of microparasites (see references in Additional file 6). Interestingly, 35 of these 160 “pathogenicity-related” genes are *S. molnari*-specific and show no homology to other cnidarian datasets. This may be the result of incomplete publicallyavailable genomic/transcriptomic datasets of parasitic cnidarians, particularly given the limited number of genomic/transcriptomic datasets from other *Sphaerospora* species [80]. Alternatively, these genes could signify an evolutionary divergence related to the distinctive features of *S. molnari* proliferating in the bloodstream of fish.

Our transcriptome analyses revealed numerous species-specific genes that lack annotation, a phenomenon observed in other organisms [81, 82], including myxozoans [20, 48]. Rapid evolution and associated fast molecular clocks [6], elevated positive gene selection [14], genome reduction, and mosaic evolution in Myxozoa [75] may contribute to gene divergence and hence challenging homology detection, as evident in some lineage-specific genes [83]. Despite the apparent lack of annotation for many of the *S. molnari* genes, we were able to identify several genes that may offer insights into the evolution of novelty related to adaptation to new host niches [84, 85] or myxozoan parasite strategies. Identification of genes that are known to play important roles in the parasitic strategies of other microbial eukaryotes, such as *Giardia* or *Plasmodium*, might provide new insights into novel *S. molnari* specific strategies.

Table 3 Sampling details for *Sphaerospora molnari* developmental stages collected from common carp, *Cyprinus carpio*

Origin of samples	Parasite developmental stage	Host tissue	No. of biological replicates
laboratory carp culture originating from Malá Outrata pond, CZE	presporogonic blood stage (BS)	column separated clean parasite	2
wild-type carp from Malá Outrata pond, CZE	presporogonic blood stage (BS)	infected blood	2
infected carp held at Szarvas facility, HUN	sporogonic stage (GS)	gills	5
laboratory carp culture originating from Malá Outrata pond, CZE	liver stage (LS)	liver	2

Conclusions

By employing highly effective protocols for isolating parasites and utilizing comparative transcriptomics, we identified differentially expressed functional gene groups involved in immune evasion, cytoskeletal rearrangement, cellular differentiation, and nutrient scavenging from host blood. Additionally, we observed differential expression of proteins with possible homologues related to pathogenic mechanisms in other blood and intestinal parasites. Finally, we compiled a list of “pathogenicity-related” gene families crucial for different developmental stages of *S. molnari*, suggesting potential candidates for disease control of this parasite and potentially other myxozoans. Our results indicate several possible evolutionary novelties in *S. molnari*, which require confirmation through further functional studies and insights from other myxozoan datasets.

Methods

Animal and sample collection

We used information on the infection kinetics of *S. molnari* from a previous study [31] (i.e., the initial peak of parasite multiplication in host blood at 28 days post-infection (dpi), followed by parasite proliferation at 42 dpi in the liver, and at 53 dpi in the gills) to sample infected fish blood, liver, and gills. Blood stages of *S. molnari* samples (B1, B2) were collected from a parasite in vivo culture line maintained in the Laboratory of Fish Protistology (Institute of Parasitology, BC CAS, Czech Republic) that has been cycled between fish individuals by intraperitoneal injection of parasites into specific parasite-free (SPF) common carp (*Cyprinus carpio*) [86], and two blood samples (B3 and B4) were collected from infected wild fish from different Czech ponds (see Table 3). Blood stage of *S. molnari* samples B1 and B2 were enriched for parasites using ion exchange chromatography according to a protocol described by

Born-Torrijos et al. [87]. *Sphaerospora molnari* blood samples B3 and B4 were concentrated from several fish and co-isolated with host white blood cells (WBCs) from total blood by centrifugation for 5 min at 3500 rpm in heparinized hematocrit tubes. Liver stages of *S. molnari* were collected and characterized from two infected wo carp individuals (RNA isolation from infected liver tissue). Gill stages ($n = 5$) were obtained from five fish that were collected from a commercial carp hatchery and housed in a recirculation system of the Research Institute for Fisheries and Aquaculture (Szarvas, Hungary). Parasite-containing gill tissues were placed into RNAlater (Sigma-Aldrich, USA), stored for the first three days at 4 °C, and then transferred to -20 °C until RNA extraction. All experimental protocols were approved by the Resort Professional Commission of the CAS for Approval of Projects of Experiments on Animals. Prior to the collection of *S. molnari* blood stages (Table 3), the fish specimens were euthanized utilizing clove oil. Fish sampling protocols and manipulations were in strict adherence to the provisions of the Czech legislation governing the welfare of animals, as set forth in the Protection of Animals Against Cruelty Act No. 246/1992. All procedures were authorized by the Czech Ministry of Agriculture. The study is reported in accordance with ARRIVE guidelines (<https://arriveguidelines.org>).

RNA isolation and sequencing

Total RNA was isolated using the NucleoSpin RNA Kit (Macherey-Nagel, Germany) following the manufacturer’s instructions. RNA concentration and purity were assessed with a NanoDrop ND-1000 Spectrophotometer (ThermoFisher, USA). All RNA samples were commercially processed for library preparation (TrueSeq Library prep.), and Illumina sequencing (NovaSeq 6000 run, 150 bp paired-end reads) was performed by Future Genomics Technologies B.V., Netherlands. The quality of sequence reads was assessed with FastQC (<https://qubeshub.org/resources/fastqc>). Adapter sequences and low-quality regions were trimmed using BBDuk (with quality trimming threshold qtrim = rl, trimq = 20, minlength = 30) (v37.62) (<https://sourceforge.net/projects/bbmap/>).

Read filtering and transcriptome assemblies

Removal of host contaminants from the read dataset was performed in two steps, following the workflow outlined in Alama-Bermejo et al. [88]. First, host reads were removed prior to assembly by read mapping to the reference genomes of both *S. molnari* (GenBank accession: SAMN45157437 [47]) and *Cyprinus carpio* (GenBank accession: ASM127010v1 [89]). All reads mapping to the *C. carpio* genome but not to the *S. molnari* genome or failing to align to either of the genomes were excluded from the analysis. Only the reads that mapped to the *S.*

molnari genome were used for assembly. Generally, when the genome of an organism or its close relative is available, a genome-guided assembly is preferred rather than a *de-novo* assembly [32]. In our case, despite the availability of a well-curated *S. molnari* genome, we used both strategies for comparative purposes: (i) a *de novo* assembly and (ii) a genome-guided assembly using the Trinity assembly suite v2.10.0 with default settings [90]. Both assemblies (i.e. *de-novo* and genome-guided) were further investigated using BLAST and in-house Python scripts to remove non-parasite contigs (the best hit for each query was selected based on the smallest e-value and all contigs with hits assigned to Bacteria, Archaea, Fungi, Metazoa-Chordata, Orthornavirae, and Viridiplantae were discarded). TransDecoder, implemented in the Trinity package [90], was used (with default settings) to find and translate open reading frames (ORFs), and translated peptides were then used for searches against NCBI, Pfam, SwissProt/UniProt, EggNog, KEGG, HMMER, and SignalP databases [90–97]. To enhance annotation, Kegg Orthology (KOs) identifiers were assigned to each predicted protein by running the BlastKOALA [98] search. The assemblies (both full and cleaned), predicted proteins for the genome-guided assembly, and annotations are available in the Figshare repository [66].

Comparison of transcriptome assemblies

To evaluate the completeness of the resulting assemblies (*de novo*, genome-guided), we performed BUSCO v4.1.4 (Benchmarking Universal Single-Copy Orthologs) [99] searches against the eukaryotic and metazoan datasets and compared the scores with the published *S. molnari* genome assembly (GenBank accession: SAMN45157437 [47]).

We used Orthofinder v2.3.11 [100] to identify orthologues shared and unique between the *de novo* and genome-guided assemblies. We also assessed the assemblies for full-length versus truncated proteins by conducting BLAST searches against the UniProt database [93] (using e-value threshold set to 1e-10 and max_target_seqs set to 25). The top hits for each query, determined by the highest bit score value, were extracted, and the percentage of their completeness was calculated using an in-house Python script. Histograms were plotted in R using the tidyverse v1.3.2 R package [101].

Differential gene expression and enrichment analyses

The RNA-seq read data for blood, gills, and liver samples were aligned to the longest isoform of each contig extracted from the genome-guided transcriptome assembly using the Bowtie v2.4.1 aligner (with default parameters) [102]. Reads with ambiguous assignments or mapping to different contigs were excluded. The resulting alignments were utilized to quantify read per

contig using featureCounts of the subread v2.0.1 R package [103]. The number of reads retained in the liver samples was below the recommended threshold value of one million reads [34], and thus data from these lifestages were excluded from analyses. The expression matrix was uploaded to iDEP.96 [104], filtered using CPM with a cut-off set to 0.5 in at least one library, and transformed using the rlog (regularized log) option [105]. The mapped data were then used for exploratory data analysis (EDA), with k-means clustering, hierarchical clustering, and PCA analyses. Gene clusters predicted by the k-means were subjected to enrichment analysis using the ClusterProfiler v4.6.2 [35] and topGo v2.50.0 [36] R packages.

Differentially expressed genes between blood and gill stages were assessed using the DESeq2 v1.38.3 [105] package. All genes with an adjusted p-value < 0.01 and fold-change > 4 were analyzed by enrichment analysis. To summarize and visualize significant DEGs, we used the tidyverse v1.3.2 [101], viridis v0.6.2 [106], and pheatmap v1.0.12 [107] R packages.

Sphaerospora-specific gene identification

We used 50 Cnidaria and 12 Myxozoa datasets (proteins derived from genomes and transcriptomes; see Additional file 8; including our *S. molnari* genome-guided transcriptome assembly) to identify *Sphaerospora*-specific genes. Transcriptome assembly and protein predictions were lacking from publicly available *Myxobolus pendula* RNA-Seq data, so we assembled three SRA datasets using SPAdes v3.15.5 [109]. TransDecoder v5.5.0 (<https://github.com/TransDecoder/TransDecoder>) was used for ORF prediction and translation of transcripts into amino acids. For the myxozoan species for which only a genome assembly was available, we performed ab initio gene predictions using Augustus v3.3.3 [108] with *Amphimedon queenslandica* as a reference species.

Next, the amino acid sequences from all the datasets were used as input for OrthoFinder v2.5.4 [100] analysis. The cogeqc v1.3.1 R package [110] was used to visualize OrthoFinder results. Orthogroups containing only *S. molnari* sequences were considered *Sphaerospora*-specific genes. Next, we examined *S. molnari*-specific genes for the presence of signal peptides and transmembrane domains to identify any secreted proteins. We used SignalP v5.0b (with the -org option set to eukaryotic) [97] to identify proteins with a D-score above 0.5, which were then used as an input for TargetP v1.1 (with option -org 'no-pl') [111] to identify all proteins with subcellular targeting. The remaining proteins underwent a sequential search using WOLFPsort (using option 'animal') [112] and TMHMM v2.0c [113]. Proteins annotated as extracellular in the WOLFPsort search and lacking a transmembrane domain after the signal peptide in the TMHMM search were classified as “secreted proteins”.

Pathogenicity-related gene identification

To identify genes that are differentially expressed between the blood and gill stages and may be involved in pathogenicity, a thorough literature search was conducted. Starting with the complete set of differentially expressed genes and their functional annotations (available in the Figshare repository [66]; see section “Read filtering and transcriptome assemblies”). Publications were retrieved from the literature based on the functional annotation and abstracts were reviewed to determine whether the study implies the annotated gene(s) involvement in host interactions, toxicity, or pathogenicity. Such genes were then considered putative pathogenicity-associated proteins. These were assembled into a table containing their expression pattern and a selection of background literature (Additional file 6).

Abbreviations

GS	Gill Stage
BS	Bloodstream Stage
dpi	Days Post-infection
OGs	Orthogroups
CPM	Counts Per Million
PCA	Principal Component Analyses
DEGs	Differentially Expressed Genes
FDR	False Discovery Rate
FC	Fold-Change
DUF	Domain of Unknown Function
MFIT	MEMBRANE Fold Induced Tumbling
FBP	Fibronectin-Binding Protein
FAD	Flavin Adenine Dinucleotide
NO	Nitric Oxide
IgSF	Immunoglobulin-like Superfamily
VSPs	Variant-specific Surface Proteins
ETRAP	Malarial Early Transcribed Membrane-like Protein
HGT	Horizontal Gene Transfer
SPF	Specific Parasite-Free
WBCs	White Blood Cells
ORFs	Open Reading Frames
KOs	Kegg Orthology identifiers
EDA	Exploratory Data Analysis

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-025-11265-x>.

Supplementary Material 1: Additional file 1: Fig. S1 Exploratory data analysis: differences in read counts in the samples used for expression analysis. A The total count of reads (in millions) that mapped to the reference transcriptome. Samples denoted by an asterisk (L1 and L2) retained fewer reads than the threshold value set to one million reads and were removed from further analyses. B Density plot showing the expression values of the data after transformation (after applying r-log transformation implemented in DESeq2). C Distribution of the transformed data for each sample. D Scatter plot of transformed expression in the first two samples (two gill samples; S3A and S3B).

Supplementary Material 2: Unique DEGs for the analyses with and without the B1 and B2 replicates. A Heatmap representing DEGs and a cartoon depiction of expression patterns with the classification of unique genes for the analyses with and without B1 and B2 samples. B Venn diagram representing the overlap between the analyses with and without B1 and B2 samples.

Supplementary Material 3: Additional file 3: Fig. S3 Exploratory data analysis: k-means clustering and enrichment analysis. A Heatmap with 4 clusters

calculated by k-means clustering. B Enrichment analysis of the 4 clusters using clusterProfiler with KEGG KOs, OGs, KEGG Pathways, and Pfam domains, and topGo with GO terms. The height of the bar depicts the number of significantly enriched genes and the shade of the bar denotes the level of significance based on the adjusted p-value. The more yellow the bar, the more significantly enriched.

Supplementary Material 4: Additional file 4: Fig. S4 Exploratory data analysis: Principal component analysis (PCA) using the first two principal components, showing the variance between blood and gill stages of the parasite.

Supplementary Material 5: Additional file 5: Supplemental Table S1 Exploratory data analysis: List of enriched cellular components, molecular functions, and biological processes in the investigated clusters.

Supplementary Material 6: Additional file 6: Supplemental Table S2 List of *S. molnari* genes related to “pathogenicity”.

Supplementary Material 7: Additional file 7: Fig. S5 and Fig. S6 KEGG reference pathway maps and tables. KEGG reference pathway map and table for: A glycolysis and gluconeogenesis, as well as p B porphyrin metabolism. Enzymes denoted by green color are present in *S. molnari*.

Supplementary Material 8: Additional file 8: Supplemental Table S3 Cnidarian datasets used to identify *S. molnari* species-specific genes.

Acknowledgements

We would like to express our sincere appreciation to all the organizations and individuals who have contributed to the publication of this manuscript. We are grateful to our colleagues at the Institute of Parasitology, Biology Centre, Czech Academy of Science for their feedback, and support. In particular, we would like to thank Rachel Kolisko, Jeffrey Donald Silberman, and Elisabeth Hehenberger for their valuable insights and suggestions.

Author contributions

MMW, AK, GAB, JK, and AL ran the analyses. MMW and AK drafted the manuscript. ASH and MK contributed to writing the manuscript. All authors have read and approved the final version of the manuscript and agree with the order of presentation of the authors.

Funding

This research was funded by the Czech Science Foundation, grant numbers, 20-30321Y (A.K.), 19-25536Y (A.L.), 19-28399X (A.H.), and the ERD fund “Centre for Research of Pathogenicity and Virulence Parasites” (no. CZ.02.1.01/0.0/0.0/16_019/0000759) (M.K.).

Data availability

Sequence data generated for this study are deposited in the NCBI SRA archive under BioProject PRJNA1067885 [114] and GenBank under SUB14832144 [47]. Data supporting the conclusions of this article are available in the Figshare repository under <https://doi.org/10.6084/m9.figshare.24639318> [66].

Declarations

Ethics approval and consent to participate

All experimental protocols were approved by the Resort Professional Commission of the CAS for Approval of Projects of Experiments on Animals. Prior to the collection of *S. molnari* blood stages, the fish specimens were subjected to euthanasia utilizing clove oil. The manipulation and sampling protocols were executed with a consistent approach and in strict adherence to the provisions of the Czech legislation governing the welfare of animals, as set forth in the Protection of Animals Against Cruelty Act No. 246/1992. All procedures were authorized by the Czech Ministry of Agriculture. The study is reported in accordance with ARRIVE guidelines (<https://arriveguidelines.org>).

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Institute of Parasitology, Biology Centre, Czech Academy of Sciences, České Budějovice, Czech Republic

²Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic

³Fish Health Division, University of Veterinary Medicine, Vienna, Austria

⁴Department of Life Sciences, University of Modena and Reggio Emilia, Modena, Italy

⁵National Biodiversity Future Center (NBFC), Palermo, Italy

Received: 28 February 2024 / Accepted: 20 January 2025

Published online: 03 February 2025

References

- Windsor DA. Controversies in parasitology, most of the species on earth are parasites. *Int J Parasitol.* 1998;28:1939–41.
- de Meeüs T, Renaud F. Parasites within the new phylogeny of eukaryotes. *Trends Parasitol.* 2002;18:247–51.
- Okamura B, Hartigan A, Naldoni J. Extensive uncharted biodiversity: the parasite dimension. *Integr Comp Biol.* 2018;58:1132–45.
- Parker GA, Chubb JC, Ball MA, Roberts GN. Evolution of complex life cycles in helminth parasites. *Nature.* 2003;425:480–4.
- Schultz DT, Haddock SHD, Bredeson JV, Green RE, Simakov O, Rokhsar DS. Ancient gene linkages support ctenophores as sister to other animals. *Nature.* 2023;618:110–7.
- Holzer AS, Bartosová-Sojková P, Born-Torrijos A, Lövy A, Hartigan A, Fiala I. The joint evolution of the Myxozoa and their alternate hosts: a cnidarian recipe for success and vast biodiversity. *Mol Ecol.* 2018;27:1651–66.
- Zhang Z-Q. Animal biodiversity: an update of classification and diversity in 2013. In: Zhang, z.-q, editor animal biodiversity: an outline of higher-level classification and survey of taxonomic richness (addenda 2013). *Zootaxa.* 2013;3703:5–11.
- Atkinson SD, Bartholomew JL, Lotan T. Myxozoans. Ancient metazoan parasites find a home in phylum Cnidaria. *Zoology.* 2018;129:66–8.
- Weinstein SB, Kuris AM. Independent origins of parasitism in Animalia. *Biol Lett.* 2016;12:20160324.
- Rodríguez E, Barbeitos MS, Brugler MR, Crowley LM, Grajales A, Gusmao L, et al. Hidden among sea anemones: the first comprehensive phylogenetic reconstruction of the order Actiniaria (Cnidaria, Anthozoa, Hexacorallia) reveals a novel group of xexacorals. *PLoS ONE.* 2014;9:e96998.
- Bentlage B, Osborn KJ, Lindsay DJ, Hopcroft RR, Raskoff KA, Collins AG. Loss of metagenesis and evolution of a parasitic life style in a group of open-ocean jellyfish. *Mol Phylogenet Evol.* 2018;124:50–9.
- Okamura B, Gruhl A. Evolution, origins and diversification of parasitic cnidarians. The evolution and fossil record of parasitism: identification and macroevolution of parasites. Springer; 2021. pp. 109–52.
- Wolf K, Markiw ME. Biology contravenes taxonomy in the Myxozoa - New discoveries show alternation of invertebrate and vertebrate hosts. *Science.* 1984;225:1449–52.
- Feist SW, Morris DJ, Alama-Bermejo G, Holzer AS. Cellular processes in myxozoans. In: Okamura B, Gruhl A, Bartholomew J, editors. *Myxozoan evolution, ecology and development.* Cham: Springer; 2015.
- Ben-David J, Atkinson SD, Pollak Y, Yossifon G, Shavit U, Bartholomew JL, et al. Myxozoan polar tubules display structural and functional variation. *Parasites Vectors.* 2016;9:549.
- Guo QX, Whipps CM, Zhai YH, Li D, Gu ZM. Quantitative insights into the contribution of nematocysts to the adaptive success of cnidarians based on proteomic analysis. *Biology.* 2022;11:91.
- El-Matbouli M, McDowell TS, Antonio DB, Andree KB, Hedrick RP. Effect of water temperature on the development, release and survival of the triacrinomyxon stage of *Myxobolus cerebralis* in its oligochaete host. *Int J Parasitol.* 1999;29:627–41.
- Lom J, Dyková I. Myxozoan genera: definition and notes on taxonomy, life-cycle terminology and pathogenic species. *Folia Parasitol.* 2006;53:1–36.
- Eszterbauer E, Atkinson S, Diamant A, Morris D, El-Matbouli M, Hartikainen H. Myxozoan life cycles: practical approaches and insights. In: Okamura B, Gruhl A, Bartholomew J, editors. *Myxozoan evolution, ecology and development.* Cham: Springer; 2015. pp. 175–98.
- Alvarez-Pellitero P, Sijtj-Bobadilla A. Pathology of Myxosporea in marine fish culture. *Dis Aquat Organ.* 1993;17:229–38.
- Kent ML, Margolis L, Corliss JO. The demise of a class of protists - taxonomic and nomenclatural revisions proposed for the protist phylum Myxozoa Grasse, 1970. *Can J Zool.* 1994. pp. 932–7.
- Nehring RB, Walker PG. Whirling disease in the wild: the new reality in the intermountain west. *Fisheries.* 1996;21:28–30.
- Pote LM, Hanson LA, Shivaji R. Small subunit ribosomal RNA sequences link the cause of proliferative gill disease in channel catfish to *Henneguya* n. Sp (Myxozoa: Myxosporea). *J Aquat Anim Health.* 2000;12:230–40.
- Moran JDW, Whitaker DJ, Kent ML. A review of the myxosporean genus *Kudoa* Meglitsch, 1947, and its impact on the international aquaculture industry and commercial fisheries. *Aquaculture.* 1999. pp. 163–96.
- Palenzuela O, Redondo MJ, Alvarez-Pellitero P. Description of *Enteromyxum Scopthalmi* gen. nov., sp. nov. (Myxozoa), an intestinal parasite of turbot (*Scophthalmus maximus* L.) using morphological and ribosomal RNA sequence data. *Parasitology.* 2002;124:369–79.
- Jones SR, Bartholomew JL, Zhang JY. Mitigating myxozoan disease impacts on wild fish populations. In: Okamura B, Gruhl A, Bartholomew J, editors. *Myxozoan evolution, ecology and development.* Cham: Springer; 2015.
- Hutchins PR, Sepulveda AJ, Hartikainen H, Staigmiller KD, Opitz ST, Yamamoto RM, et al. Exploration of the 2016 Yellowstone river fish kill and proliferative kidney disease in wild fish populations. *Ecosphere.* 2021;12:e03436.
- Holland JW, Holzer AS. Myxozoan research forum 2021-the MyxoMixer: advances, methods, and problems yet to be solved in myxozoan research. *Bull Eur Association Fish Pathologists.* 2022;41:216–24.
- Lom J, Dyková I, Pavlasková M, Grupcheva G. *Sphaerospora molnari* sp. nov. (Myxozoa, Myxosporea), an agent of gill, skin and blood sphaerosporosis of common carp in Europe. *Parasitology.* 1983;86:529–35.
- Dykova I, Lom J. Review of pathogenetic myxosporeans in intensive culture of carp (*Cyprinus carpio*) in Europe. *Folia Parasitol.* 1988;35:289–307.
- Korytář T, Wiegertjes GF, Zusková E, Tomanová A, Lisnerová M, Patra S, Sieranski V, Sima R, Born-Torrijos A, Wentzel AS, Blasco-Monleon S, Yanes-Roca C, Polcar T, Holzer AS. The kinetics of cellular and humoral immune responses of common carp to presporogonic development of the myxozoan *Sphaerospora molnari*. *Parasites Vectors.* 2019;12:208.
- Lu BX, Zeng ZB, Shi TL. Comparative study of de novo assembly and genome-guided assembly strategies for transcriptome reconstruction based on RNA-Seq. *Sci China-Life Sci.* 2013;56:143–55.
- Lischer HEL, Shimizu KK. Reference-guided *de novo* assembly approach improves genome reconstruction for related species. *BMC Bioinformatics.* 2017;18:1–12.
- Liu Y, Zhou J, White KP. RNA-seq differential expression studies: more sequence or more replication? *Bioinformatics.* 2014;30:301–4.
- Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innov.* 2021;2:100141.
- Alexa A, Rahnenfuhrer J. topGO: enrichment analysis for gene ontology. R Package Version. 2010;2:2010.
- Hartigan A, Estensoro I, Vancová M, Bílý T, Patra S, Eszterbauer E, et al. New cell motility model observed in parasitic cnidarian *Sphaerospora molnari* (Myxozoa: Myxosporea) blood stages in fish. *Sci Rep.* 2016;6:39093.
- Chen Y-P, Riestra AM, Rai AK, Johnson PJ. A novel cadherin-like protein mediates adherence to and killing of Host Cells by the parasite *trichomonas vaginalis*. *mBio.* 2019;10:e00720–19.
- O'Connor RM, Wanyiri JW, Cevallos AM, Priest JW, Ward HD. *Cryptosporidium parvum* glycoprotein gp40 localizes to the sporozoite surface by association with gp15. *Mol Biochem Parasitol.* 2007;156:80–3.
- Dalton JP, Skelly P, Halton DW. Role of the tegument and gut in nutrient uptake by parasitic plathyhelminths. *Can J Zool.* 2004;82:211–32.
- Toh SQ, Glanfield A, Gobert GN, Jones MK. Heme and blood-feeding parasites: friends or foes? *Parasites Vectors.* 2010;3:108.
- Dean P, Major P, Nakjang S, Hirt RP, Embley TM. Transport proteins of parasitic protists and their role in nutrient salvage. *Front Plant Sci.* 2014;5:153.
- Yang YL, Xiong J, Zhou ZG, Huo FM, Miao W, Ran C, et al. The genome of the Myxosporean *Thelohanellus kitauei* shows adaptations to nutrient acquisition within its fish host. *Genome Biol Evol.* 2014;6:3182–98.
- Reyes-López M, Aguirre-Armenta B, Piña-Vázquez C, de la Garza M, Serrano-Luna J. Hemoglobin uptake and utilization by human protozoan parasites: a review. *Front Cell Infect Microbiol.* 2023;13:111150054.
- Shang F, Lan J, Wang L, Liu W, Chen Y, Chen J, et al. Crystal structure of the Siderophore-interacting protein SIP from *Aeromonas hydrophila*. *Biochem Biophys Res Commun.* 2019;519:23–8.
- Lombardo ME, Araujo LS, Batlle A. 5-aminolevulinic acid synthesis in epimastigotes of *Trypanosoma cruzi*. *Int J Biochem Cell Biol.* 2003;35:1263–71.

47. Wiśniewska MM, Kyslík J, Alama-Bermejo G, Lövy A, Kolísko M, Holzer A et al. Comparative transcriptomics reveal stage-dependent parasitic adaptations in the myxozoan *Sphaerospora molnari*, GenBank. 2024. <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1193866>
48. Faber M, Shaw S, Yoon S, Alves ED, Wang B, Qi ZT, et al. Comparative transcriptomics and host-specific parasite gene expression profiles inform on drivers of proliferative kidney disease. *Sci Rep*. 2021;11:2149.
49. Brekhan V, Ofek-Lazar M, Atkinson SD, Alama-Bermejo G, Maor-Landau K, Malik A, et al. Proteomic analysis of the oasitic cnidarian *Ceratonova shasta* (Cnidaria: Myxozoa) reveals diverse roles of actin in motility and spore formation. *Front Mar Sci*. 2021;8:632700.
50. Leclère L, Horin C, Chevalier S, Lapébie P, Dru P, Peron S, et al. The genome of the jellyfish *Clytia hemisphaerica* and the evolution of the cnidarian life-cycle. *Nat Ecol Evol*. 2019;3:801–10.
51. Murphy JR. *Corynebacterium diphtheriae*: diphtheria toxin production. In: Baron S, et al. editors. *Medical Microbiology*. 4th ed. Galveston, Texas: University of Texas Medical Branch; 1996.
52. Shoop WL, Xiong Y, Wiltse J, Woods A, Guo J, Pivnichny JV, et al. Anthrax lethal factor inhibition. *Proc Natl Acad Sci USA*. 2005;102:7958–63.
53. Bowran K, Palmer T. Extreme genetic diversity in the type VII secretion system of *Listeria monocytogenes* suggests a role in bacterial antagonism. *Microbiology*. 2021;167:0001034.
54. Hartigan A, Jaimes-Becerra A, Okamura B, Doonan LB, Ward M, Marques AC, et al. Recruitment of toxin-like proteins with ancestral venom function supports endoparasitic lifestyles of Myxozoa. *PeerJ*. 2021;9:e11208.
55. Crauwels P, Bohn R, Thomas M, Gottwalt S, Jäckel F, Krämer S, et al. Apoptotic-like *Leishmania* exploit the host's autophagy machinery to reduce T-cell-mediated parasite elimination. *Autophagy*. 2015;11:285–97.
56. Americus B, Hams N, Klompen AML, Alama-Bermejo G, Lotan T, Bartholomew JL, et al. The cnidarian parasite *Ceratonova shasta* utilizes inherited and recruited venom-like compounds during infection. *PeerJ*. 2021;9:e12606.
57. Huene AL, Sanders SM, Ma Z, Nguyen AD, Koren S, Michaca MH, et al. A family of unusual immunoglobulin superfamily genes in an invertebrate histocompatibility complex. *Proc Natl Acad Sci USA*. 2022;119:e2207374119.
58. Deutsch KW, Lukehart SA, Stringer JR. Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens. *Nat Rev Microbiol*. 2009;7:493–503.
59. Schmid-Hempel P. Parasite immune evasion: a momentous molecular war. *Trends Ecol Evol*. 2008;23:318–26.
60. Piriatskiy G, Atkinson SD, Park S, Morgenstern D, Brekhan V, Yossifon G, et al. Functional and proteomic analysis of *Ceratonova shasta* (Cnidaria: Myxozoa) polar capsules reveals adaptations to parasitism. *Sci Rep*. 2017;7:9010.
61. Alama-Bermejo G, Bartošová-Sojková P, Atkinson SD, Holzer AS, Bartholomew JL. Proteases as therapeutic targets against the parasitic cnidarian *Ceratonova shasta*: characterization of molecules key to parasite virulence in salmonid hosts. *Front Cell Infect Microbiol*. 2022;11:804864.
62. Bartošová-Sojková P, Kyslík J, Alama-Bermejo G, Hartigan A, Atkinson SD, Bartholomew JL, et al. Evolutionary analysis of cystatins of early-emerging metazoans reveals a novel subtype in parasitic cnidarians. *Biology*. 2021;10:110.
63. Hartigan A, Kosakyan A, Pecková H, Eszterbauer E, Holzer AS. Transcriptome of *Sphaerospora molnari* (Cnidaria, Myxosporea) blood stages provides proteolytic arsenal as potential therapeutic targets against sphaerosporosis in common carp. *BMC Genomics*. 2020;21:404.
64. Nash TE. Surface antigenic variation in *Giardia lamblia*. *Mol Microbiol*. 2002;45:585–90.
65. Xu F, Jerlström-Hultqvist J, Einarsson E, Astvaldsson A, Svärd SG, Andersson JO. The genome of *Spironucleus salmonicida* highlights a fish pathogen adapted to fluctuating environments. *PLoS Genet*. 2014;10:e1004053.
66. Wiśniewska MM, Kyslík J, Alama-Bermejo G, Lövy A, Kolísko M, Holzer AS et al. Comparative transcriptomics reveal stage-dependent parasitic adaptations in the myxozoan *Sphaerospora molnari*. *FigShare*. 2024. <https://doi.org/10.6084/m9.figshare.24639318>
67. Spielmann T, Ferguson DJP, Beck H-P. Etramps, a new *Plasmodium falciparum* gene family coding for developmentally regulated and highly charged membrane proteins located at the parasite–host cell interface. *Mol Biol Cell*. 2003;14:1529–44.
68. Kyes S, Horrocks P, Newbold C. Antigenic variation at the infected red cell surface in malaria. *Annu Rev Microbiol*. 2001;55:673–707.
69. Niang M, Yam XY, Preiser PR. The *Plasmodium falciparum* STEVOR multigene family mediates antigenic variation of the infected erythrocyte. *PLoS Pathog*. 2009;5:e1000307.
70. Bachmann A, Scholz JAM, Janssen M, Klinkert MQ, Tannich E, Bruchhaus I, et al. A comparative study of the localization and membrane topology of members of the RIFIN, STEVOR and PfMC-2TM protein families in *Plasmodium falciparum*-infected erythrocytes. *Malar J*. 2015;14:1–18.
71. Andersson A, Kudva R, Magouloupolou A, Lejcarre EQ, Lara P, Xu PB, et al. Membrane integration and topology of RIFIN and STEVOR proteins of the *Plasmodium falciparum* parasite. *FEBS J*. 2020;287:2744–62.
72. Huminiecki L, Wolfe KH. Divergence of spatial gene expression profiles following species-specific gene duplications in human and mouse. *Genome Res*. 2004;14:1870–9.
73. Tautz D, Domazet-Lošo T. The evolutionary origin of orphan genes. *Nat Rev Genet*. 2011;12:692–702.
74. Singh U, Wurtele ES. Genetic novelty: how new genes are born. *eLife*. 2020;9:e55136.
75. Guo QX, Atkinson SD, Xiao B, Zhai YH, Bartholomew JL, Gu ZM. A myxozoan genome reveals mosaic evolution in a parasitic cnidarian. *BMC Biol*. 2022;20:51.
76. McBride MJ, Braun TF, Brust JL. *Flavobacterium johnsoniae* GldH is a lipoprotein that is required for gliding motility and chitin utilization. *J Bacteriol*. 2003;185:6648–57.
77. Hale CA, Shiomi D, Liu B, Bernhardt TG, Margolin W, Niki H, et al. Identification of *Escherichia coli* ZapC (YcbW) as a component of the division apparatus that binds and bundles FtsZ polymers. *J Bacteriol*. 2011;193:1393–404.
78. Ravindran S, Boothroyd JC. Secretion of proteins into host cells by Apicomplexan parasites. *Traffic*. 2008;9:647–56.
79. Portes J, Barrias E, Travassos R, Attias M, de Souza W. *Toxoplasma gondii* mechanisms of entry into host cells. *Front Cell Infect Microbiol*. 2020;10:294.
80. Patra S, Bartošová-Sojková P, Pecková H, Fiala I, Eszterbauer E, Holzer AS. Biodiversity and host-parasite cophylogeny of *Sphaerospora* (sensu stricto) (Cnidaria: Myxozoa). *Parasites Vectors*. 2018;11:1–24.
81. Marsden RL, Ranea JAG, Sillero A, Redfern O, Yeats C, Maibaum M, et al. Exploiting protein structure data to explore the evolution of protein function and biological complexity. *Philosophical Trans Royal Soc B-Biological Sci*. 2006;361:425–40.
82. Wilson GA, Feil EJ, Lilley AK, Field D. Large-scale comparative genomic ranking of taxonomically restricted genes (TRGs) in bacterial and archaeal genomes. *PLoS ONE*. 2007;2:e324.
83. Weisman CM, Murray AW, Eddy SR. Many, but not all, lineage-specific genes can be explained by homology detection failure. *PLoS Biol*. 2020;18:e3000862.
84. Wilson GA, Bertrand N, Patel Y, Hughes JB, Feil EJ, Field D. Orphans as taxonomically restricted and ecologically important genes. *Microbiology*. 2005;151:2499–501.
85. Ma DN, Lai ZF, Ding QS, Zhang K, Chang KZ, Li SH, et al. Identification, characterization and function of orphan genes among the current Cucurbitaceae genomes. *Front Plant Sci*. 2022;13:872137.
86. Holzer AS, Pimentel-Santos J. *Sphaerospora molnari*. Fish parasites: a handbook of protocols for their isolation, culture and transmission. In: Sitja-Bobadilla A, Bron J, Wiegertjes G, Piazzon M, editors. *European association of fish pathologists*. 5 M Books Ltd.; 2021.
87. Born-Torrijos A, Kosakyan A, Patra S, Pimentel-Santos J, Panicucci B, Chan JTH, et al. Method for isolation of myxozoan proliferative stages from fish at high yield and purity: an essential prerequisite for in vitro, in vivo and genomics-based research developments. *Cells*. 2022;11:377.
88. Alama-Bermejo G, Meyer E, Atkinson SD, Holzer AS, Wiśniewska MM, Kolísko M, et al. Transcriptome-wide comparisons and virulence gene polymorphisms of host-associated genotypes of the cnidarian parasite *Ceratonova shasta* in salmonids. *Genome Biol Evol*. 2020;12:1258–76.
89. Kolder ICRM, van der Plas-Duivesteyn SJ, Tan G, Wiegertjes GF, Forlenza M, Guler AT, et al. A full-body transcriptome and proteome resource for the European common carp. *BMC Genomics*. 2016;17:701.
90. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 2011;29:644–52.
91. Pruitt KD, Tatusova T, Maglott DR. NCBI Reference sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res*. 2005;33:D501–4.
92. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: the protein families database in 2021. *Nucleic Acids Res*. 2021;49:D412–9.
93. UniProt Consortium. The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res*. 2010;38:D142–148.

94. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol*. 2021;38:5825–9.
95. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*. 2016;44:D457–462.
96. Eddy SR. Accelerated Profile HMM searches. *PLoS Comput Biol*. 2011;7:e1002195.
97. Armenteros JJA, Tsirigos KD, Sonderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol*. 2019;37:420–3.
98. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for functional characterization of genome and metagenome sequences. *J Mol Biol*. 2016;428:726–31.
99. Manni M, Berkeley MR, Seppely M, Simao FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol*. 2021;38:4647–54.
100. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 2019;20:1–14.
101. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the Tidyverse. *J Open Source Softw*. 2019;4:1686.
102. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9.
103. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30:923–30.
104. Ge SX, Son EW, Yao RN. iDEP: an integrated web application for differential expression and pathway analysis of RNA-Seq data. *BMC Bioinformatics*. 2018;19:1–24.
105. Love MI, Huber W, Anders S. Moderated estimation of Fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:1–21.
106. Garnier S, Ross N, Rudis R, Camargo AP, Sciaini M, Scherer C. Rvision-color-blind-friendly color maps for R. R package version 06. 2021;2.
107. Kolde R, Kolde MR. Package 'pheatmap'. R Package. 2015;1:790.
108. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics*. 2008;24:637–44.
109. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012;19:455–77.
110. Almeida-Silva F, Van de Peer Y. Assessing the quality of comparative genomics data and results with the cogeqc R/Bioconductor package. *MEE*. 2023;14:2942–2952. Available: <https://github.com/almeidasilva/cogeqc>
111. Emanuelsson O, Nielsen H, Brunak S, von Heijne G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol*. 2000;300:1005–16.
112. Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, et al. WoLF PSORT: protein localization predictor. *Nucleic Acids Res*. 2007;35:W585–7.
113. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. Predicting trans-membrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol*. 2001;305:567–80.
114. Wiśniewska MM, Kyslik J, Alama-Bermejo G, Lövy A, Kolisko M, Holzer AS et al. Comparative transcriptomics reveal stage-dependent parasitic adaptations in the myxozoan *Sphaerospora Molnari*, GenBank. 2024. <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1067885>

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.