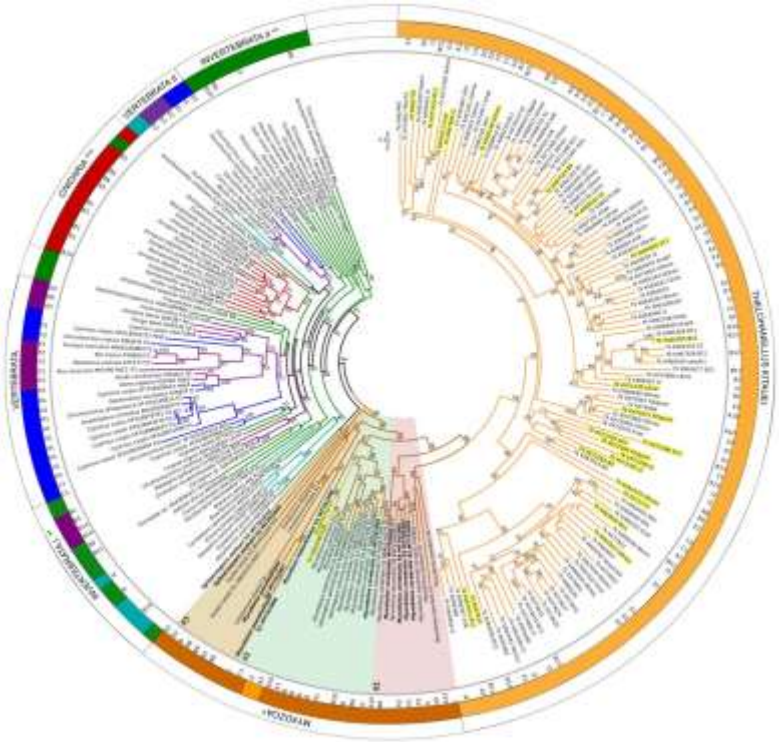
Studies on the potential therapeutic target genes of fish pathogen myxozoan parasites

Final report



Budapest, 29th December 2021

1. Summary

Myxozoans (Cnidaria, Myxozoa) are an economically relevant, widespread and special group of fish parasites. Some species cause extreme decline in natural habitats and aquaculture, like *Myxobolus cerebralis*, the causative agent of whirling disease in trout fry. Despite of the economic losses, there is no effective protection of fish developed against myxozoans until now. In the cooperation project with researchers from the Laboratory of Fish Protistology, Institute of Parasitology (Czechia), our research focused on expanding knowledge about myxozoan protease inhibitors possessing therapeutic potential. We identified, characterized and compared serine protease inhibitors (serpins) of low and high virulence myxozoan species using state-of-art genomic, transcriptomic and proteomic approaches. For the first time, we detected genetic distinction of myxozoan serpins that further strengthens the idea of an independent origin of Myxozoa, and may indicate novel protein functions potentially related to parasitism in this animal group. The gene expression profiling revealed that *M. cerebralis* serpins Mc-S3 and Mc-S5 were upregulated during early-stage fish invasion, what is the best time to interrupt the parasite's development for the prevention of disease. Furthermore, the functional analyses suggest that these caspase-like protease inhibitors are promising therapeutic targets for the development of an anti-myxozoan treatment, and they are worthy of further investigations.

2. Összefoglalás

A nyálkaspórások (Cnidaria, Myxozoa) a halparaziták gazdaságilag fontos, széles körben elterjedt és különleges csoportja. Egyes fajok komoly elhullást okoznak természetes élőhelyeken és halgazdaságokban is, mint például a pisztrángfélék kergekórját okozó *Myxobolus cerebralis*. A gazdasági károk ellenére a mai napig nincs hatékony védekezés a nyálkaspórások ellen. A cseh parazitológiai intézet halprotisztológiai laboratóriumának (Csehország) kutatóival folytatott együttműködési projektben kutatásunk a terápiás potenciállal rendelkező nyálkaspórás proteáz-gátló fehérjékkel kapcsolatos ismeretek bővítésére összpontosított. A legmodernebb genomikai, transzkriptomikai és proteomikai módszerek segítségével azonosítottuk, jellemeztük és összehasonlítottuk az alacsony és a magas virulenciájú nyálkaspórás fajok szerin proteáz inhibitorait (szerpinek). Első alkalommal mutattunk ki jelentős genetikai különbséget a nyálkaspórások szerpinjei között, ami tovább erősíti a nyálkaspórások független eredetének elképzelését, és új, potenciálisan a parazitizmussal összefüggő fehérjefunkciókat feltételez ebben az állatcsoportban. A génexpressziós profilok vizsgálata kimutatta, hogy a *M. cerebralis* Mc-S3 és Mc-S5 szerpinek felülszabályozottak voltak a halgazdában való megtelepedés korai stádiumában, ami a legjobb időpont a parazita fejlődésének megszakítására a betegség megelőzése érdekében. Továbbá a funkcionális elemzések azt mutatják, hogy ezek a kaszpáz-szerű proteáz-inhibitorok ígéretes terápiás célpontok a nyálkaspórások elleni kezelés kifejlesztéséhez, és érdemesek a további vizsgálatokra.

3. Project-related publication activity

The outcome of the project was presented on international and national scientific conferences, and published in prestigious, peer-reviewed journals. Up to now, six presentations and five papers were published with the cumulative impact factor of 16.35. The manuscript presenting the results of *Myxobolus pseudodispar de novo* genomics and transcriptomics is under preparation, and it is expected to be published soon.

4. Results

The originally three-year-long project was prolonged with another year in 2020, due to unexpected delays caused by COVID pandemic. Following the research workplan, the project composed of five work packages (WP): the laboratory maintenance of myxozoan parasites' life cycle (i.e. *Myxobolus cerebralis* and *M. pseudodispar*) (WP1) until Year 3; the *de novo* whole genome sequencing of *M. pseudodispar* (WP2), the expression analyses of parasite serine-protease inhibitor (serpin) genes (WP3) in Year 1 and 2; the functional characterization of the candidate serpin genes (WP4) in Year 3; and the supplementary transcriptome sequencing of *M. pseudodispar* and the final data analysis in Year 4.

4.1. WP1: Life cycle of myxozoans

The life cycles of the model organisms, *Myxobolus cerebralis* and *M. pseudodispar* were maintained in our laboratory, whereas *Sphaerospora molnari* was available in the laboratory of the Czech partners until Year 3, to supply sufficient material for experimental exposures and for *de novo* genome and transcriptome sequencing. Furthermore, the laboratory experiments unshed the light on relevant details regarding the early intrapiscine development of myxozoans, which helped the fine-tuning the sampling regime for subsequent transcriptome analysis and gene expression profiling.

New data on the early-stage development of *Myxobolus cerebralis* in fish (Eszterbauer et al. 2019d)

During the optimization of fish exposure protocol for *M. cerebralis*, we observed certain site-preference by the parasite in the early stage of development in fish host. To ascertain the nature of the site-preference, we experimentally studied the invasion of *M. cerebralis* in three fish species with various susceptibility levels: the type host brown trout, the highly susceptible rainbow trout, and the non-susceptible gibel carp, in which the parasite do not develop spores. We investigated the first two hours of fish invasion, and measured the site preference of triactinomyxons (TAMs) during attachment and penetration of fish in three body parts (gills, fins, skin). Infection prevalence and intensity were estimated using a species-specific nested PCR, optimized in the present study. The highest infection prevalence was detected in the most susceptible fish species, rainbow trout. Interestingly, higher prevalence was observed in gibel carp than in the type host, brown trout (95.2% vs. 85.7%). Considering body locations, remarkable differences were detected in infection intensities. The highest intensity was observed in fins, whereas skin was the least infected body part in every fish species examined. Infection prevalence and intensity did not differ significantly among fish species. Thus, we confirmed that *M. cerebralis* TAMs cannot discern fish species. Furthermore, we proved experimentally that fish fin is significantly more attractive to fish-invading parasite TAMs than gills or skin. These results were published in Acta Veterinaria Hungarica (Eszterbauer et al. 2019; IF 0.991).

4.2. WP2: Genomics

De novo whole genome and transcriptome sequencing of *Myxobolus pseudodispar* (Eszterbauer et al. 2019b, 2019c)

The whole genome of low virulence myxozoan species has not been available up to now. To be able to perform virulence-related genomic comparisons, we aimed to obtain the *de novo* whole

genome assembly of the common, low-virulence, myxozoan parasite of cyprinids, *M. pseudodispar*. After the optimization of sample collection protocol, genomic DNA extraction and single-molecule real-time sequencing on Pacific BioSciences (PacBio) Sequel new generation sequencing system was performed in Year 1, and the assembly was done with softwares Canu1.6, HGAP4 and FALCON. PacBio sequencing provided data over 7.5 Gb (gigabase). The average length of 1.6 million reads was 9,250 bp. The genome assembly predicted a genome size of about 50-70 Mb, and revealed an extremely high AT content (~70%) similarly to other myxozoans. The most accurate assembly done with Falcon supplied however only a partial genome, in 3,619 contigs. As realized that the gained 80x coverage was insufficient for the complete assembly, Illumina paired-end sequencing was performed additionally, using Truseq Nano DNA Prep Kit (insert size 550 bp) on Illumina Novaseq 6000 instrument. Illumina sequencing provided ~24.3 Gb data (~300x overall coverage), 94.58% of which with high quality. Even with our strong effort over several months of work, we could not define the entire genome, most likely because of the high AT-content and the high number of unique repeats in the genome. Therefore, with the help of bioinformatics expert in the assembly of organisms with special genome organization, we re-started the work from raw data, and the generated a hybrid assembly of PacBio and Illumina data. The descriptive statistics suggested reliable and good quality assembly (e.g. No. of contigs: 4677, N50: 37218; largest contig: 368003 bp). The estimated genome size is approximately 143 Mb. It is the second largest genome size among myxozoan parasite (the largest one detected so far is 170 Mb, whereas the smallest is around 22 Mb). The GC content is rather low (~30%), similar to other members of this parasite group. The genome annotation brought further unexpected difficulties, as a small fraction of genes could be annotated only. Therefore to aid gene annotation, we decided to complete the analysis with transcriptomic data, which was not planned originally. The transcriptomes of three parasite developmental stages were sequenced: the parasite's early stages in fish, sporogonic stages in fish, and sporogonic stages in annelid hosts. Samples were obtained from experimentally exposed hosts. The high coverage (over 100x), paired-end Illumina RNA-seq data enabled the annotation of 70% of genes. We identified potential gene orthologs of cnidarian origin, including protease inhibitors other than the already identified serpins. The outcome of the *de novo* genome sequencing was presented in national and international conferences (Akadémiai Beszámolók 2019, and EAFP conference 2019). The manuscript, summarizing the results of *de novo* whole genome characterization of *Myxobolus pseudodispar* and the complementary transcriptome analysis is under preparation, and expected to be submitted for publication soon.

Analysis of the genetic diversity and phylogenetic relationships among serine protease inhibitors of various origins (Sipos et al. 2020; Eszterbauer et al. 2019a, 2020)

To gain information about the diversity of serpins in myxozoans (in particular for our model organisms, *M. cerebralis*, *M. pseudodispar* and *S. molnari*), we identified and analyzed serpins in the genome and transcriptome of various myxozoan species, and compare them with those of other taxa. Serine protease inhibitor (serpin) sequences were mined in the genome/transcriptome of wide variety of taxa from protozoans to vertebrates. We recognized high genetic variability among serpins of myxozoans. Serpins were reported as important factors for host invasion and immune evasion, and as promising targets for the development of anti-parasitic therapies. For the first time, we identified and aligned serpin sequences from high throughput sequencing datasets of ten myxozoan species, and analyzed 146 serpins from this parasite group together with those of other taxa phylogenetically, to explore their relationship

and origins. High intra- and interspecific variability was detected among the examined serpins. The average sequence identity was 25–30% only. The conserved domains (i.e. motif and signature) showed taxon-level differences. Serpins clustered according to taxonomy rather than to serpin types, and myxozoan serpins seemed to be highly divergent from that of other taxa. None of them clustered with their closest relative free-living cnidarians. The genetic distinction of myxozoan serpins further strengthens the idea of an independent origin of Myxozoa, and may indicate novel protein functions potentially related to parasitism in this animal group. These results were presented in a national and an international conference (Akadémiai Beszámolók 2019, and EAFP conference 2019), and published in the journal Microorganisms (Eszterbauer et al. 2020; Microorganisms; IF 4.152).

4.3 WP3: Transcriptomics

Identification of suitable reference genes for gene expression profiling (Kosakyan et al. 2019)

Quantitative RT-PCR (qPCR) is the most extensively used approach for gene expression studies. However, the accuracy of the results depends on the normalization of the data to reference genes. With the cooperation partners in Czechia, we studied the expression of eight commonly used reference genes, adenosylhomocysteinase, beta actin, eukaryotic translation elongation factor 2 (EF2), glyceraldehyde-3-phosphate dehydrogenase (GAPDH), hypoxanthine-guanine phosphoribosyl transferase 1, DNA-directed RNA polymerase II, 18S and 28S ribosomal RNA across different developmental stages of three myxozoan species, *Sphaerospora molnari*, *M. cerebralis* and *Ceratonova shasta*, representing the three major myxozoan linages from the largest class Myxosporea. The stable reference genes were identified using four algorithms: geNorm, NormFinder, Bestkeeper and ΔCq method. Additionally, we analyzed transcriptomic data from *S. molnari* proliferative and spore-forming stages to compare the relative amount of expressed transcripts with the most stable reference genes suggested by qPCR. Our results revealed that GAPDH and EF2 are the most uniformly expressed genes across the different developmental stages of the studied myxozoan species, with slight differences among myxozoan species. These findings were published in Scientific Reports (Kosakyan et al. 2019; IF 3.998).

De novo transcriptome sequencing of carp infecting *Sphaerospora molnari* (Bartošová-Sojková 2019; Hartigan et al. 2020)

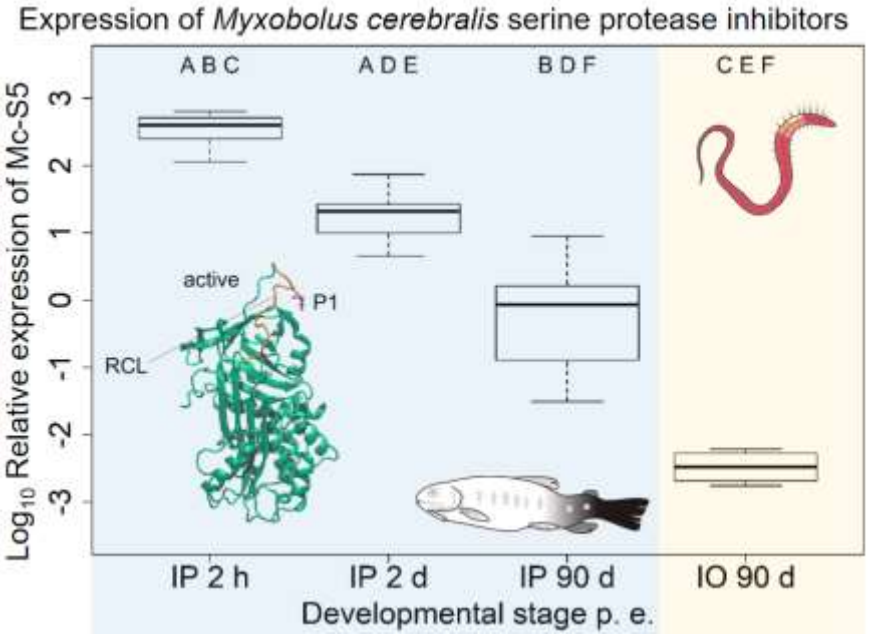
In the course of the transcriptome analysis of myxozoans, a *de novo* approach was utilized to assemble the transcriptome of the carp myxozoan, *S. molnari*, in collaboration with the Czech research partners. In the first transcriptome of the proliferative myxozoan stage of *S. molnari*, we identified proteases that are upregulated during the first stages of infection when the parasite multiplies massively in host blood. Furthermore, a subset of orthologs (including serine and cysteine proteases) was used to characterize 3D structures and putative therapeutic targets. An assembled and host filtered transcriptome containing 9436 proteins, mapping to 29,560 contigs was mined for protease virulence factors and revealed that cysteine proteases were most common (38%), at a higher percentage than other myxozoans or cnidarians (25–30%). Two cathepsin Ls that were found upregulated in spore-forming stages with a presenilin like aspartic protease and a dipeptidyl peptidase. We also identified downregulated proteases in the spore-forming development when compared with proliferative stages including an astacin metallopeptidase and lipases (qPCR). In total, 235 transcripts were identified as putative

proteases using a MEROPS database. *In silico* analysis of highly transcribed cathepsins revealed potential drug targets within this data set that should be prioritized for development. The findings regarding protease inhibitors was presented in an international conference (EAFP 2019). The *de novo* transcriptome of *S. molnari* and its comparative analysis were published in BMC Genomics (Hartigan et al. 2020; BMC Genomics; IF 3.969).

4.4. WP4: Proteomics

Expression profiling and functional analyses of *M. cerebralis* serpins (Szegő et al. 2021; Eszterbauer et al. 2021)

The existing knowledge on the gene expression patterns and on the function of myxozoan protease inhibitors is rather scarce. Our goal was, therefore, to study the serpins of the highly pathogen *Myxobolus cerebralis* species. Of the seven serpins distinguished based on the phylogenetic analysis, the structural and functional characteristics of four newly discovered inhibitors were elucidated. The relative expression of serpins were determined at different developmental stages both in fish and in annelid hosts using serpin-specific qPCR assays developed and optimized in the present study. The expression of serpin Mc-S1 was similar throughout the life cycle, whereas a significant decrease was detected in the relative expression of Mc-S3 and Mc-S5 during the development in fish, and then in the sporogonic stage in the worm host (see the graphical abstract). A decreasing tendency could also be observed in the expression of Mc-S4 in fish, which was, however, upregulated in the worm host.



(Eszterbauer et al. 2021, PLOS ONE 16(3): e0249266; graphical abstract)

For the first time, we predicted the function of *M. cerebralis* serpins by the use of several bioinformatics-based applications. The functional analyses suggested serine-type endopeptidase inhibitor activity for all studied serpins. Mc-S1 is putatively a chymotrypsin-like inhibitor that locates extracellularly and is capable of heparin-binding. The other three serpins are caspase-like inhibitors, and they are probably involved in protease and cell degradation processes during the early stage of fish invasion. The inactivation of the parasite at an early stage would interrupt its life cycle in fish, and thereby it would prevent the development of

whirling disease. As serpins Mc-S3 & Mc-S5 are most likely extracellular proteins and highly upregulated in the early stage of fish invasion (i.e. 2 hours post exposure), these caspase-like protease inhibitors seem to be promising therapeutic targets. Nevertheless, Mc-S1 as a putative antichymotrypsin should not be excluded from further functional analysis either, taking into account the anti-inflammatory potential of this putatively extracellular protease inhibitor. Our findings on gene expression profiling were presented in a conference (Szegő et al. 2021) and the results completed with functional analyses were published in PlosOne (Eszterbauer et al. 2021, IF 3.24).

5. Project-related presentations and publications

Bartošová-Sojková, P; Kosakyan, A; Pudhuvai, B; Eszterbauer, E; Holzer, A S: Expressional profiling and 3D structure modeling of cysteine protease inhibitors of *Sphaerospora molnari* (Cnidaria: Myxozoa). 19th International Conference on Diseases of Fish and Shellfish, Porto, Portugália, Paper: No.188-O, 2019

Eszterbauer, E; Sipos, D; Kaján, Gy; Herczeg, D; Holzer, A S; Bartošová-Sojková, P: Genetic diversity of serine protease inhibitors in fish-parasitic Myxozoa (Cnidaria). 19th International Conference on Diseases of Fish and Shellfish, Porto, Portugália, Paper: No.285-P, 2019a

Eszterbauer E, Sipos D, Orbán L, S Singh: Halélősködő nyálkaspórások *de novo* genom szekvenálásásnak eddigi eredményei. Akadémiai Beszámolók, Budapest, 2019. 01. 23, 2019b

Eszterbauer, E; Sipos, D; Orbán, L; Singh, S: *De novo* genome sequencing project of the fish-parasitic *Myxobolus pseudodispar* (Myxozoa): preliminary results. 19th International Conference on Diseases of Fish and Shellfish, Porto, Portugália, Paper: No.284-P, 2019c

Eszterbauer, E; Sipos, D; Szakály, Á; Herczeg, D: Distinctive site preference of the fish parasite *Myxobolus cerebralis* (Cnidaria, Myxozoa) during host invasion. ACTA VETERINARIA HUNGARICA 67:212-223., 2019d (IF:0.991)

Kosakyan, A; Alama-Bermejo, G; Bartošová-Sojková, P; Born-Torrijos, A; Šíma, R ; Nenarokova, A; Eszterbauer, E; Bartholomew, J; Holzer, A S: Selection of suitable reference genes for gene expression studies in myxosporean (Myxozoa, Cnidaria) parasites. SCIENTIFIC REPORTS 9:15073, 2019 (IF: 3.998)

Eszterbauer, E; Sipos, D; Kaján, Gy L.; Szegő, D; Fiala, I; Holzer, A S.; Bartošová-Sojková, P: Genetic diversity of serine protease inhibitors in myxozoan (Cnidaria, Myxozoa) fish parasites. MICROORGANISMS 8:10 Paper: 1502, 2020 (IF: 4.152)

Hartigan, A; Kosakyan, A; Pecková, H; Eszterbauer, Edit; Holzer, A S.: Transcriptome of *Sphaerospora molnari* (Cnidaria, Myxosporea) blood stages provides proteolytic arsenal as potential therapeutic targets against sphaerosporosis in common. BMC GENOMICS 21:1 Paper: 404, 2020 (IF: 3.969)

Sipos D, Kaján Gy, Szegő D, A S. Holzer, P Bartošová-Sojková, Eszterbauer Edit: Halélősködő nyálkaspórások (Myxozoa, Cnidaria) szerin proteáz inhibitorainak genetikai diverzitása. Akadémiai Beszámolók, Budapest, 2020. január 22., 2020

Szegő D, Sipos D, Ursu K, Eszterbauer E: Halélősködő *Myxobolus cerebralis* (Cnidaria, Myxozoa) szerin proteáz inhibitor génjeinek expressziós profilja az életciklus különböző stádiumaiban. Akadémiai Beszámolók, Budapest, 2021. január 27., 2021

Eszterbauer E, Szegő D, Ursu K, Sipos D, Gellért Á: Serine protease inhibitors of the whirling disease parasite *Myxobolus cerebralis* (Cnidaria, Myxozoa): Expression profiling and functional predictions. PLOS ONE 16(3): e0249266, 2021 (IF: 3.240)