Know your enemy – transcriptome of myxozoan *Tetracapsuloides bryosalmonae* reveals potential drug targets against proliferative kidney disease in salmonids

Freed Ahmad1, Paul V. Debes2, Lilian Pukk3, Siim Kahar3, Hanna Hartikainen4, Riho Gross3 and Anti Vasemägi3,5

1Department of Biology, University of Turku, FI-20014, Finland; 2Department of Aquaculture and Fish Biology, Hōlar University, Saudárkrókur, Iceland; 3Chair of Aquaculture, Institute of Veterinary Medicine and Animal Sciences, Estonian University of Life Sciences, 51014 Tartu, Estonia; 4School of Life Sciences, The University of Nottingham, University Park, NG7 5RD, Nottingham, UK and 5Department of Aquatic Resources, Swedish University of Agricultural Sciences, 17893 Drottningholm, Sweden

**Abstract**

The myxozoan *Tetracapsuloides bryosalmonae* is a widely spread endoparasite that causes proliferative kidney disease (PKD) in salmon fish. We developed an *in silico* pipeline to separate transcripts of *T. bryosalmonae* from the kidney tissue of its natural vertebrate host, brown trout (*Salmo trutta*). After stringent filtering, we constructed a partial transcriptome assembly of *T. bryosalmonae*, comprising 3427 transcripts. Based on homology-restricted searches of the assembled parasite transcriptome and Atlantic salmon (*Salmo salar*) proteome, we identified four protein targets (Endoglycoceramidase, Legumain-like protease, Carbonic anhydrase 2, Pancreatic lipase-related protein 2) for the development of anti-parasitic drugs against *T. bryosalmonae*. Earlier work of these proteins on parasitic protists and helminths suggests that the identified anti-parasitic drug targets represent promising chemotherapeutic candidates also against *T. bryosalmonae*, and strengthen the view that the known inhibitors can be effective in evolutionarily distant organisms. In addition, we identified differentially expressed *T. bryosalmonae* genes between moderately and severely infected fish, indicating an increased abundance of *T. bryosalmonae* sporogenic stages in fish with low parasite load. In conclusion, this study paves the way for future genomic research in *T. bryosalmonae* and represents an important step towards the development of effective drugs against PKD.

**Introduction**

*Tetracapsuloides bryosalmonae* is an obligate endoparasite that causes proliferative kidney disease (PKD) in salmonid fish. PKD is considered as one of the most serious emerging temperature-dependent parasitic freshwater diseases of salmonids in the Northern Hemisphere, causing significant losses on farms and in natural populations (Okamura et al., 2001, 2011). *Tetracapsuloides bryosalmonae* belongs to the Class Myxozoa within the Phylum Cnidaria and belongs to the relative species poor Subclass Malacosporea (Giribet and Edgecombe, 2020). The parasite has a two-host life-cycle, alternating between an invertebrate host (several species of freshwater bryozoans) and a fish host (salmonids) (Okamura et al., 2011). Fish host invasion occurs via epidermal mucous cells (e.g. gills, skin; Longshaw et al., 2002), followed by dissemination to internal organs, primarily the kidney and spleen via the vascular system. Sporogonic development takes place in the kidney of permissive fish hosts, and spores infective to bryozoans are released in fish urine (Hedrick et al., 2004). The development of PKD symptoms is temperature-dependent. Exposure to water temperatures ≥15°C elicits a strong immune response, kidney hypertrophy and anaemia, which may lead to death (Hedrick et al., 1993; Betige et al., 2009). As a consequence, the decline of native salmonid fish populations in several countries has been linked with a combination of increasing temperatures and emergence of PKD (Wahl et al., 2002; Okamura et al., 2011; Sudhagar et al., 2020).

Despite the high economic and ecological significance of PKD, the genomic resources for *T. bryosalmonae* are limited to a few hundred expressed sequence tags and some other targeted loci (e.g. Saulnier et al., 1996; Holland et al., 2011; Hartikainen et al., 2014, 2015; Carraro et al., 2018). Only very recently, an intersect transcriptome of *T. bryosalmonae* consisting of *de novo* assembled transcripts present in both rainbow trout (*Oncorhynchus mykiss*) and bryozoan host (*Federicella sultana*) has been reported (Faber et al., 2021). However, the *T. bryosalmonae* transcriptome recovered from the dead-end host *O. mykiss* might lack genes essential for the parasite to reach sporogony in natural fish host. Knowledge of the parasite genomic composition and gene expression is therefore needed to better understand the underlying mechanisms of parasite life stages, host–parasite interactions, resistance, tolerance and virulence. Moreover, genomic information may help to predict potential new drug targets.
and drugs/drug-like compounds against the parasite (Cowell and Winzeler, 2019). The climate-change driven emergence of PKD underlies the urgency for the identification of effective chemotherapeutic targets against *T. bryosalmonae* that can be used to protect stocks and improve the welfare of farmed salmonids.

In this study, we report the de novo assembly and annotation of the partial *T. bryosalmonae* transcriptome based on dual RNA sequencing (Westermann et al., 2012) of both parasite and host utilizing infected brown (*Salmo trutta*) trout kidney tissue. We also identify *T. bryosalmonae* genes that are differentially expressed in relation to parasite load (PL) to obtain a better understanding of host–parasite interactions. Based on searches against known drug target proteins in the ChEMBL database, we identify four promising protein drug targets (consisting of six genes) whose suppression might obstruct the parasite growth and mitigate infections. Thus, these findings prime future experimental approaches to develop and test new control strategies against PKD. Furthermore, the assembled *T. bryosalmonae* transcriptome will be useful in understanding the parasite biology and serve as a valuable reference for characterizing the parasite development during infection.

**Materials and methods**

**Sampling and measurement of kidney hyperplasia**

Sampling and phenotyping methods of juvenile (0+) brown trout used in this study were reported in Debes et al. (2017). DNA was extracted from kidney tissue as in Aljanabi and Martinez (1997). Fourteen fish kidneys corresponding to eight infected and six uninfected fish from three wild populations were selected. All the infected fish originated from River Mustoja (Estonia) and were collected in September, 2014, whereas the uninfected fish originated from two parasite-free rivers (Priedi and Vodja, Estonia; three per river) and were collected in September, 2015. Sampling from multiple populations was necessary as *T. bryosalmonae* prevalence in the study area is often 100% (Dash and Vasemägi, 2014; Debes et al., 2017), prohibiting the sampling of both infected and uninfected individuals from the same location. We also quantified a typical disease trait associated with PKD, kidney swelling (renal hyperplasia), estimated as a body-length-standardised kidney depth, as described by Debes et al. (2017).

**PL measurement**

PL (a proxy for the inverse of resistance) was determined from infected fish kidney tissue by qPCR using previously described protocols (Bruneaux et al., 2016; Debes et al., 2017). For each sample, we quantified the occurrence of two target DNA sequences: a 166 bp fragment of *T. bryosalmonae* 18S rDNA sequence (GenBank Accession U70623) and 74 bp fragment of the *S. solar* nuclear DNA sequence (GenBank Accession BT049744.1, Prefoldin subunit 6). Simultaneous quantification of both DNA fragments enabled us to standardize the amount of parasite DNA relative to host DNA. Three technical replicates were run for both targets in each sample. Based on PL, the infected samples were divided into categories of moderate (n = 4) and severe (n = 4) infection. On average the moderately infected fish exhibited lower kidney swelling measurements, compared to severely infected fish (~twice the kidney depth; Supplementary Fig. 1).

**RNA isolation, library preparation and RNA sequencing**

We preserved half of the kidney cross-section for RNA isolation by freezing them in liquid N2 and storing them at −80°C. The other half of the kidney cross-section was stored in the same way but used for estimation of PL using qPCR. Total RNA was extracted from the kidney cross-section of the 14 individuals using NucleoSpin® RNA kit. RNA quality was tested using the Advanced Analytical Fragment Analyzer. Quantity of each total RNA sample was measured with a Nanodrop ND-2000 spectrophotometer. Library preparation was done according to Illumina TruSeq® Stranded mRNA Sample Preparation Guide (part # 15031047) starting from 100 ng of total RNA. The steps involved in cdNA library constructions were as follows: purification of poly-A containing mRNA, fragmentation of mRNA using divalent cations, first-strand (using reverse transcriptase and random primers) and second strand (using DNA Polymerase I and RNase H) synthesis of cdNA, purification and enrichment with PCR. A unique Illumina TruSeq indexing adapter was ligated to each sample during the adapter ligation step to be able to pool several samples for sequencing. The quality of the prepared libraries was evaluated with Advanced Analytical Fragment Analyzer. The concentration of each library was quantified with Qubit® Fluorometric Quantitation, Life Technologies. Due to adapter peaks, libraries were purified twice. The samples were normalized and pooled for the automated cluster preparation that was carried out using Illumina cBot station. The 14 libraries were pooled into a single pool. The samples were sequenced with Illumina HiSeq 2500 instrument using TruSeq v3 sequencing chemistry in three lanes. Paired-end sequencing with 100 bp read length was used, followed by 6 bp index run. The base calling was performed using Illumina’s standard bcl2fastq2 software.

**Data processing and removal of host RNA sequence**

The quality assessment of sequenced reads was performed with FastQC v 0.10.1 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) both before and after pre-processing. Trimmomatic (v. 0.35) was used for adapter trimming and read pre-processing (parameters: ILLUMINACLIP = TruSeq3-S.E..fa; HEADCROP = 10; SLIDINGWINDOW = 4:20; LEADING = 5; TRAILING = 5; MINLEN = 40). In order to assemble a host-free parasite transcriptome, the filtered reads were first mapped to the Atlantic salmon genome (GCF_000233375.1) and the unmapped reads were subsequently mapped to the northern pike (*Esox lucius*, Esociformes) genome (GCF_000721915.2) using Hisat2 (v. 2.0.1; score-min = 1.0,= 0.4) (Kim et al., 2015). We used chromosome level assembly of Atlantic salmon genome (Lien et al., 2016), as it is evolutionarily close to brown trout (Ahmad et al., 2018; Lemopoulos et al., 2018), and at the time of analysis, genome assembly of *S. trutta* was not available. Esociformes represents the closest order to Salmoniformes (Rondeau et al., 2014). The paired-end reads from the infected fish samples that were not aligned to either the salmon or pike genome were further taxonomically classified by using kraken2 (Wood et al., 2019) with a custom database. The custom kraken2 database consisted of two fish genomes (Atlantic salmon and Northern pike), four available myxozoan genomes (*Enteromyxum leei*, *Sphaeromyxa zaharoni*, *Kudoa iwaii* and *Thelohanellus kitauei*), both host (*Salmo trutta*, 3177 sequences, mean GC content = 43.2%) and parasite (*T. bryosalmonae*, 741 sequences, mean GC content = 33.6%) nucleotide sequences available in NCBI and ENA (Supplementary Table 1), as well as, unpublished contigs for *T. bryosalmonae*, derived from its myxozoan host *F. sultana* (mean GC content = 27.1%; H. Hartikainen unpublished), along with Kraken2-2 provided archaea, bacteria, plasmid, viral, human, fungi, protozoa (proteins) and UnilVec_Core sequences. The reads that were classified within the NCBI
bony fish division were excluded before the de novo assembly of the parasite transcriptome.

**De novo transcriptome assembly of T. bryosalmonae**

The putative parasite transcriptome was assembled using Trinity (v. 2.5.1) assembler (Grabherr et al., 2011) with the default parameters. CD-HIT-EST (Fu et al., 2012) was used to cluster 95% identical transcripts and isoforms. The paired-end reads from both the uninfected samples that were not mapped to any fish genome and the non-bony fish reads from the infected samples were aligned on the clustered transcripts using Bowtie2 (Langmead and Salzberg, 2012). The read count matrix for all samples was generated using the summarize Overlaps function implemented in R for further evaluation and differential expression analyses. In order to filter out transcripts originating from the host or from potential contamination, the longest isoforms of the assembled transcripts were subjected to Kraken2 classification and similarity searching using BLASTN and BLASTX (Altschul et al., 1990) against NR databases (only the top hits, dust = no, soft masking = false, E-value cut-off = $1 \times 10^{-4}$).

**Filtering and completeness assessment of the parasite transcriptome**

The putative parasite transcripts were filtered using the following criteria: (i) at least 40 reads mapped, (ii) 90% or more of the mapped reads originate from infected fish and (iii) retained only those transcripts that matched to GenBank’s invertebrate division or no hit was found using Kraken2, BLASTN and BLASTX. In order to assess the completeness of the transcriptome, these transcripts were subjected to Benchmarking Universal Single-Copy Orthologs (BUSCO v. 3.0) (Simão et al., 2015) analysis against 978 Metazoan single-copy orthologs.

Recently, Faber et al. (2021) reported an intersect transcriptome of T. bryosalmonae, common to both rainbow trout (O. mykiss) and bryozoan host (F. sultana) which consisted of 7362 contigs showing 50.1% completeness (excluding fragmented hits) against the BUSCO eukaryotic dataset (303 genes). Underestimation of completeness for the myxozoan genomes/criptotmes using BUSCO metazoan set has already been reported (Hartigan et al., 2020; Yahalomi et al., 2020). Therefore, in order to identify BUSCO metazoan genes that are either entirely missing or present in all myxozoans, we analysed genome annotations-based coding sequences (CDS) and transcriptome assemblies for seven species: Thelohanellus kitauei, Kudoa iwatai, Hemagogus salminicola, Myxobolus squamalis, Myxobolus cerebralis, Sphaerospora mohiari and Ceratonia nova shasta (Supplementary Table 2). The T. kitauei transcriptome assembly was not available and therefore only CDS extracted from the genome using available genome feature file (GFF) were used. For K. iwatai, using gencode, BLASTN and BLASTX (Altschul et al., 1990) against NR databases (only the top hits, dust = no, soft masking = false, E-value cut-off = $1 \times 10^{-4}$), the reduced value of minimum ORF length was used to consider the shorter transcripts (200–300 bp, n = 526) and fragmented genes for the subsequent annotation. Because myxozoans are underrepresented in the public databases and difficult to annotate, ORFs > 150 bp length (50 aa) were retained even when no homology was found. The resulting protein sequences were then used for BLASTP searches and for signal peptide prediction in Trinotate annotation suite. For Trinotate annotation (report E-value < $1 \times 10^{-5}$), we used the results from BLASTP and BLASTX homology searches against Swissprot, transmembrane helices predictions using TMHMM Server (v. 2.0), pfam domain search using hmmscan, signal peptides prediction using SignalP 5.0, and BLASTP homology search against Refseq proteins.

For the genes with multiple isoforms and protein translations, only the hits with the maximum number of GO terms were reported. Finally, the ribosomal RNAs were predicted using RNAmmer 1.2 Server (Lagesen et al., 2007).

**Overlap between two T. bryosalmonae transcriptomes**

In order to find the overlap between our T. bryosalmonae transcriptome and the recently reported T. bryosalmonae transcriptome (from the dead-end vertebrate host, rainbow trout [O. mykiss], and the bryozoan F. sultana; Faber et al., 2021), we carried out reciprocal BLASTN. A transcript (longest isoform in case of multiple isoforms) was considered as present in Faber et al. (2021) transcriptome if its top hit (95% minimum identity) matched the best hit in the reciprocal BLASTN. As the scaffolds in Faber et al. (2021) transcriptome were relatively long (N50 = 2479) compared to ours (N50 = 657), for the reciprocal BLASTN, we allowed up to 3 hits per scaffold when Faber et al. (2021) transcriptome was queried against our transcripts. All isoforms of the same gene were considered as present if the longest isoform was present in Faber et al. (2021) transcriptome and vice versa. The GO terms between the genes found in Faber et al. (2021) transcriptome vs the genes unique to our study were compared using WEGO 2 (Ye et al., 2018).

**Differential gene expression analysis of the parasite transcripts**

In order to identify potential links between differences in gene expression and parasite development, gene-level differential expression (DE) analysis of T. bryosalmonae between the moderately and severely infected fish was carried out using the DEseq function implemented in the R package DESeq2 (Love et al., 2014). For this purpose, the read counts of the different isoforms of the same genes were summed up to obtain gene-level read counts. An explorative default significance level (FDR < 0.1) was
used to identify differentially expressed genes since we were mostly interested in detecting candidates that need further validation using RT-qPCR.

**Identification of drug targets**

In order to screen for potential drug targets, *T. bryosalmonae* transcriptome was searched against ChEMBL protein sequences using TBLASTN (soft masking = false, E-value cut-off = 1 × 10−5). To identify parasite-specific targets and avoid potential drug toxicity to the host, only those transcripts lacking BLASTX hits against the *Salmo salar* proteome (soft masking = false, E-value cut-off = 1 × 10−5) were retained. To increase the sensitivity of the homology search, we focused on targets with an alignment length >100 amino acids based on TBLASTN search. The protein sequences of the putative parasite transcripts corresponding to the top matched drug targets (based on E-value) were subjected to homology-based structure modelling in Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2) using the intensive mode (Kelley et al., 2015). Only those sequences were subject to homology modelling which had a conserved alignment at active or binding sites of the drug targets. The best-supported structures were subsequently superimposed with the template structures using Chimera (v. 1.14; Pettersen et al., 2004) for visualization.

**Results**

After the quality control and filtering, 493.1 Million (401.6 M paired-end and 91.5 M single-end) reads from the host and parasite were retained (range 22.9–47.2 M reads per sample). Among these, 266 M (220.7 M paired-end and 45.3 M single-end) reads originated from infected fish. After mapping on fish reference genomes, a total of 7.6 M out of 220.7 M (3.4% of the total) paired-end reads from infected fish, that were not assigned to bony fishes by Kraken2, were assembled into 16499 transcripts (12676 genes) using Trinity. Individual transcripts with 95% or more similarity were clustered into 14741 merged transcripts (12621 genes) using CD-HIT-EST with default parameters (Table 1).

The overall average alignment rates of infected and uninfected samples against merged transcripts were 83.73 and 33.24%, respectively. Because parasite reads should originate only from infected fish, we used this criterion as an additional step to filter out host transcripts. More specifically, if >10% of the aligned reads of the particular transcript originated from the uninfected fish, the transcript was not assigned as a putative parasite sequence.

A large proportion of merged transcripts were not assigned to any taxonomic group (Fig. 1A, Kraken2 = 6508, BLASTN = 10283 and BLASTX = 6440) reflective of the highly diverged nature of myxozoans from other cnidarians (Chang et al., 2015) and a small number of available myxozoan genomes. Among the sequences that were classified using BLASTN, the majority of hits were from vertebrate and salmonid fishes (Fig. 1B). However, relatively more hits to cnidarians (*Hydra vulgaris* and *Nematostella vectensis* and *Exaiptasia pallida*) were identified using BLASTX (Fig. 1C). Only 5993 (40%), 130 (0.9%) and 22 (0.2%) transcripts were assigned to Myxozoa by Kraken2 (Fig. 1D), BLASTX and BLASTN, respectively. Among these, 11 transcripts (Fig. 1A) were designated as myxozoan by all three searches and consisted of following *T. bryosalmonae* genes: antigen PKX101 (seven transcripts including four isoforms of the same gene), minicollagen (Ncol-1, DN1177_c0_g1_i1 and DN14479_c0_g1_i1), DNA repair protein RAD51-like protein (DN15328_c0_g1_i1, excluded because of low coverage) and 60S ribosomal protein L18 (RP-L18, DN3431_c0_g1_i1). BLASTN also assigned some additional transcripts to known genes of *T. bryosalmonae*, including 28S large subunit ribosomal RNA gene (DN11757_c1_g2_i1 and DN11757_c1_g4_i1), S-adenosylmethionine synthetase (DN7631_c0_g1_i1) and ribosomal protein rpl23a-like gene (DN1241_c0_g1_i1), respectively.

**Table 1. Summary statistics of the de novo transcriptome assembly of *T. bryosalmonae***

| Characteristics | All assembled transcripts | Clustered transcripts | Putative parasite transcripts |
|----------------|--------------------------|-----------------------|-----------------------------|
| Total transcripts (genes) | 16 499 (12676) | 14 741 (12 621) | 3427 (2905) |
| GC content (%) | 36.77 | 36.38 | 31.50 |
| Contig N50 | 441 | 426 | 657 |
| Median contig length | 304 | 300 | 483 |
| Average contig length | 411.11 | 402.56 | 587.07 |
| Total assembled bases | 6 782 897 | 5 934 156 | 2 011 902 |
| Median contig length (longest isoform) | 298 | 298 | 477 |
| Average contig length (longest isoform) | 395.78 | 396.1 | 580.75 |
| Total assembled bases (longest isoform) | 5 016 954 | 4 999 154 | 1 687 084 |

**Transcriptome characterization and completeness**

The overall distribution of GC content of 14741 merged transcripts was bimodal (Fig. 2A), suggesting that the vertebrate host and myxozoan parasite genomes have different GC content. Previous studies have identified myxozoan transcriptomes and genomes to be predominantly AT-rich (Yang et al., 2014; Fox et al., 2015; Yahalomi et al., 2020). After filtering, the putative parasite transcripts (Fig. 2A; n = 3427; mean GC = 31.5%), from here on referred to as *T. bryosalmonae* transcriptome, and transcripts belonging to bony fishes were well separated (Fig. 2A; n = 1027; mean GC = 48.56%). The GC content of the *T. bryosalmonae* transcriptome was similar to other myxozoan genomes (GC content 23.6–37.5%) (Yahalomi et al., 2020). The GC content of brown trout (43.9%) (Malachowicz et al., 2017) also broadly matched with the GC content of the transcripts belonging to bony fishes. The *T. bryosalmonae* transcriptome consisted of more than 2 million bases with an average transcript length of 587 bp (range: 201–3388 bp). Among these transcripts, 929 represented isoform reconstituted of 407 genes, while the remaining 2498 transcripts corresponded to genes without isoforms. The
read counts for the large majority of the putative parasite transcripts were positively correlated with the independent estimate of relative PL obtained via qPCR assay (Fig. 2B; 92.3% of transcript r ≥ 0.5) suggesting that our bioinformatic pipeline (Supplementary Fig. 2) successfully retained T. bryosalmonae reads. Moreover, the positive correlation estimates suggested that the coverage of the majority of assembled transcripts was sufficient to accurately reflect the relative PL.

The assessment of T. bryosalmonae transcriptome completeness using BUSCO analysis indicated that only 26% of the core single copy metazoan genes were completely (14.5%) or partially (11.6%) present in our assembly (Supplementary Table 2 and 3). The recently published T. bryosalmonae transcriptome (Faber et al., 2021) on the other hand exhibited two times higher completeness as 48.3% of the metazoan single copy genes were completely (38.1%) or partially (10.1%) present. A total of 58 BUSCO genes (32 complete single; 2 complete duplicated and 24 fragmented) were identified in our T. bryosalmonae transcriptome assembly which were missing in the published transcriptome Faber et al. (2021). A similar pattern of relatively low BUSCO completeness can be observed in other published myxozoan transcriptome and genome annotations (Hartigan et al., 2020; Yahalomi et al., 2020). We subsequently evaluated the presence and absence of BUSCO core single copy metazoan genes in the published myxozoan parasite datasets (Fig. 3A–B, Supplementary Table 3). Despite variation in coverage and completeness of individual transcriptomes and genomes, we identified a set of BUSCO metazoan genes found in all myxozoans (n = 231, Fig. 3A). Furthermore, we observed a set of BUSCO metazoan genes (n = 232, 23.7%) that were missing in all seven myxozoan transcriptome and genome annotations. This suggests that these genes are either missing or highly divergent in myxozoans (Fig. 3B).

**KEGG ortholog assignment and ribosomal RNA**

Out of 3427 putative parasite transcripts, only 1742 transcripts (1509 genes) were assigned to 1172 unique KEGG orthologs
characterization of T. bryosalmonae transcripts: (A) The distributions of GC content of all assembled transcripts (grey; n = 14744; mean GC = 35.51%), putative parasitic sequences (light grey; n = 3427; mean GC = 31.67%) and transcripts belonging to bony fishes in BLAST results (black; n = 1027; mean GC = 48.56%). (B) The distribution of Pearson’s correlation coefficients between the raw read counts of putative parasite transcripts and the qPCR-inferred relative parasite load.
Comparison between the two T. bryosalmonae transcriptomes

Using an adjusted reciprocal BLASTN, 81.5% (2793 transcripts, 2377 genes) of the transcripts were assigned to 1615 contigs of the recently reported intersect transcriptome (from bryozoan and rainbow trout, Faber et al., 2021). A total of 634 transcripts (528 genes), did not have a hit within the intersect transcriptome and are hence unique to our assembly (Supplementary Fig. 6A).

In comparison of the transcripts that matched to the Faber et al., 2021 transcriptome, the unique transcripts had higher relative proportion of genes related to Ribosome (5.13%), Ribosome biogenesis (3%), Ubiquitin system (2%) and Peptidases and inhibitors (1.6%) based on BRITE hierarchies (Supplementary Fig. 6B). GO terms comparison using WEGO 2 showed that the unique T. bryosalmonae transcripts have higher (%<0.05) proportion of genes related to the cellular component terms: Endoplasmic reticulum membrane (GO:0005789) and Mitochondrial protein complex (GO:0098798), and following molecular function terms: Structural molecule activity (GO:0005198) and Structural component of ribosome (GO:0003735) (Supplementary Fig. 6C and D).

Differentially expressed T. bryosalmonae genes linked to PL

A total of 55 T. bryosalmonae genes were differentially expressed at FDR <0.1 between severely and moderately infected fish (Fig. 4; Supplementary Table 9). Similar to the whole transcriptome annotation, only a proportion of differentially expressed genes (n = 31) retrieved functional annotation using KAAS and Trinotate. In the severely infected samples, the polar capsule related minicollagens (TBNCOL-1, TBNCOL-2 and TBNCOL-4) were downregulated, indicating that severely and moderately infected hosts may differ in the extent of polar capsule formation of T. bryosalmonae. Furthermore, in severely infected fish several lysosome and lipid metabolism-related genes were downregulated. These include cathepsin L (CTSL; 2 genes), Niemann-Pick C2 protein (NPC2), hydroxymethylglutaryl-CoA synthase (HMGCOS), hydroxymethylglutaryl-CoA reductase (NADPH) (HMGCOR) and PI-PLC X domain-containing protein 2 (PLCX2). Interestingly, differential expression of CTSLs has recently been found between the proliferative blood- and sporogonic gill stages of the myxozoan S. molnari (Hartigan et al., 2020). Therefore, it is possible that cathepsin L is also linked to sporogony in T. bryosalmonae. In
severely infected fish downregulation was also inferred for the expression of glutathione peroxidase (gpx; the highest mean expression level among differentially expressed genes; AvgFPKM in the Supplementary Table 9), which is an antioxidant that protects the parasite from the damage caused by free radicals released from the host immune cells (Changklungmoa et al., 2018).

Among the upregulated genes in the severely infected fish, four were retrotransposon genes (annotated as Retrovirus-related Pol polyprotein from type-1 retrotransposable element R2; pol). The upregulated gene with the highest fold change corresponded to low-density lipoprotein receptor class A-like protein 3 (MT070967.1, 100% BLASTN identity) gene. Other upregulated genes included Talin (TLN), TRAF2 and NCK interacting kinase (TNIK), Tetraspanin-9 (TSN9), Vacular-sorting protein (snf7), importin-5 (IPOS), calcium-dependent secretion activator (CADPS), Rho GTPase-activating protein 39 (ARHGAP39), synaptosomal-associated protein 25 (SNAP25), AGPAT8 lysocardioliipin and lysophospholipid acyltransferase (LCLAT1) and large subunit ribosomal protein L3e (RP-L3e).

Drug targets and homology modelling

After stringent screening of 3427 putative parasite transcripts against known drug target proteins in the ChEMBL database, four potential chemotherapeutic targets, Endoglycoceramidase (two genes; EGCase1 and EGCase2), Legumain-like protease, Carbonic anhydrase 2 and Pancreatic lipase-related protein 2 (two genes) were identified (Table 2). The two EGCase transcripts encode 283 and 382 amino acid long partial protein sequences (two genes) were identified (Table 2). The two EGCase transcripts encode 283 and 382 amino acid long partial protein sequences with 43.8% similarity. We subsequently individually modelled the 3D structures of the two EGCase and CA2 protein sequences by using Phyre2. For both EGCase1 and EGCase2, >90% of residues were modelled at >90% confidence during homology modelling. Both structures showed a high overlap with the template structure (Fig. 5, Supplementary Fig. 7 and 8). The structure of the EGCase2 (longer alignment length with the template) is shown in Fig. 5 in which the low confidence residues (92–114) were unreliably modelled into a loop because they were missing in the template sequence (Supplementary Fig. 8). The template EGCase chosen by Phyre2 (B chain of endoglycoceramidase ii from Rhodococcus sp. (PDB ID 2OYL) for the homology modelling was bound to an inhibitor molecule (cellobiose-like imidazole) using polar interactions with Lys68, His135, Asp177, Asn232, Glu233 and Glu253 (Caines et al., 2007). Apart from Lys68, which was missing from the partial template sequence, the corresponding His65, Asp67, Asn222, Glu173 and Glu277 in the T. bryosalmonae EGCase2 (DN1345_c0_g1_i1, Fig. 5c) acquired the same positions in the binding site (Fig. 5C). The T. bryosalmonae EGCase1 (DN11176_c0_g2_i1) exhibits three amino acids occupying the similar 3D confirmation as the template structure’s binding site amino acids (Supplementary Fig 7C). These conserved binding site residues of T. bryosalmonae Endoglycoceramidase II represent a promising target for the chemotherapeutic drugs with cellobiose-like imidazole inhibitor-like structures.

More than 90% of residues of the T. bryosalmonae CA2 were modelled at >90% confidence using Phyre2. The predicted structure (Fig. 5D and E, green) was superimposed on chain A of the parasite Schistosoma mansoni carbonic anhydrase (SmCA, PDB ID 6QQM, pink). In SmCA, the active site residues are His68, Gln115, Glu122, and Thr131 and the Zinc binding Histidines are His117, His119, and His142 (D’Ara et al., 2019). These important residues are also conserved in T. bryosalmonae CA2 (Active site: His76, Gln107, Glu122 and Zinc binding Histidines: His109, His111 and His136) while only Threonine is missing. These residues also acquired similar orientation as in the SmCA (Fig. 5F) thus the inhibitors used for the S. mansoni carbonic anhydrase might also be effective for T. bryosalmonae CA2. Finally, we did not predict 3D structures of LGMN and PLRP2s as they are partial sequences and lacked important active site or binding site residues based on alignments with ChEMBL hits (Supplementary Fig. 9 and 10).

Discussion

This study reports the first annotated transcriptome of T. bryosalmonae derived from wild, infected brown trout using dual RNA sequencing. By employing multi-step filtering of parasite reads followed by de novo assembly, our work provides new insights into the differentially expressed T. bryosalmonae genes in the fish and identifies six proteins that could act as potential targets for the development of antiparasitic drugs against PKD. As such, this work demonstrates the utility of the assembled T.
bryosalmonae transcriptome and promotes experimental approaches to develop and test new control strategies against PKD.

A major challenge for the de novo transcriptome assembly using dual-RNA sequencing data is to correctly separate host and parasite RNA reads (Schulze et al., 2016; Videvall et al., 2017). In earlier studies targeting myxozoan transcriptomes within fish hosts, RNA was isolated from the appendages visible by eye, e.g. plasmodia from the K. iwatai infected gilthead seabream (Chang et al., 2015) or cysts from the T. kitaei infected common carp (Yang et al., 2014), and after careful washing steps that were followed by in silico filtering of reads to avoid inclusion of host sequences during the parasite transcriptome or genome assembly. Yet, more recent works have shown that host tissue contamination is still an important issue in next generation sequencing of myxozoans (Fox et al., 2015; Alama-Bermejo et al., 2020; Hartigan et al., 2020). Furthermore, isolation of microscopic T. bryosalmonae specimens from salmonid kidney tissue is challenging and our initial RNAseq data was dominated (more than 97% of total sequences) by brown trout RNA reads. Therefore, in order to produce a host-free parasite transcriptome, we employed a rigorous multi-step strategy, comprising host genome mapping, pre-assembly metagenomic classification and post-assembly filtering. Recently, Alama-Bermejo et al. (2020) tackled the problem of host contamination removal from the myxozoan C. shasta transcriptome by employing a two-step filtering approach. In the first step, host reads were removed by aligning reads to host and parasite reference genomes. During the second step (transcript-level filtering), Alama-Bermejo et al. (2020) used BLASTN against both host and parasite genomes to further exclude the host transcripts, while contaminant sequences (of bacterial, fungal or viral origins) were excluded based on BLASTX (Alama-Bermejo et al., 2020). In contrast, T. bryosalmonae transcriptome was not published at the time of the analysis. In addition, only a good quality genome of Salmo salar, representing a genetically close sister species to brown trout, was available for us at the time of the analysis. Therefore, to overcome these limitations, we included multiple myxozoan genomes in the Kraken2 database to classify the unmapped reads and assembled parasite transcriptome using reads that were not classified to ‘bony fishes’. Furthermore, reads from the uninfected fish that mapped on our assembled transcripts provided a useful control step which enabled to effectively remove both known and unknown transcripts originating from the fish host. Some homology between host and parasite short reads is expected and therefore we allowed up to 10% of the total reads per transcript to map to the uninfected fish. Similar to the second step transcript-level filtering in C. shasta (Alama-Bermejo et al., 2020), we also excluded transcripts that showed significant (e.g. bacterial, viral and vertebrate) BLASTX hits during the post-assembly filtering, thus any potential bacterial and human contamination during the sample collection and library preparation was also filtered out. Finally, the majority of transcripts with positive correlation between raw read counts and PL measured by an independent approach (qPCR; Bruneaux et al., 2016; Debes et al., 2017) indicated that our pipeline successfully removed the host sequences and captured true RNA reads from T. bryosalmonae.

Myxozoans are highly reduced metazoans with extremely small diverged genome sizes (22.5–188.5 Mb), which makes it challenging to evaluate the completeness of assembled transcriptomes and genomes (Chang et al., 2015; Hartigan et al., 2020; Yahalomi et al., 2020). By using a large set of published data, we identified a set of 231 BUSCO genes that were detected in all myxozoans. Based on this smaller set of BUSCO genes, the estimated T. bryosalmonae transcriptome assembly completeness is close to 56.3% (130 of 231). Thus, we suggest that this core
A set of 231 BUSCO genes can be used as an alternative to estimate the completeness of the assemblies in future myxozoan studies. Conversely, we also observed that 23.7% of metazoan single-copy gene genes were not identified in any published myxozoan genome and transcriptome assemblies. As the myxozoans have gone through an extreme reduction in body plan and genome size (Chang et al., 2015), it is possible that at least some of these genes have been truly lost during myxozoan evolution. However, we cannot exclude, at this point, the alternative explanation that unidentified BUSCO genes are still present in myxozoans, but because of highly divergent nature, were not detected using BUSCO.

Currently, there are no licenced drugs available for efficient PKD treatment. Earlier studies demonstrated an efficiency of malachite green (triarylmethane dye; Clifton-Hadley and Alderman, 1987) and fumagillin (antibiotic of the fungus Aspergillus fumigatus; Le Gouvello et al., 1999) for treatment against PKD. Malachite green has been used for many fish species to control fungal attacks, protist infections and other diseases, such as caused by helminths (Srivastava et al., 2004). Fumagillin has been shown to prevent clinical outbreaks of whirling disease caused by the myxosporean parasite Myxobolus cerebralis in rainbow trout (El-Matbouli and Hoffmann, 1991). In addition, fumagillin is an effective antibiotic against the microsporidian pathogen Nosema apis in honey bees (Burnham, 2019). However, because of the accumulation of malachite green in fish and its potential teratogenic and carcinogenic properties, its use is, unsurprisingly, restricted for food fish. Similarly, it has been shown that fumagillin treatment leads to a dramatic loss in haematopoietic tissue in rainbow trout (Wishkovsky et al., 1990). Therefore, given the high economic losses incurred by PKD on the freshwater salmonid aquaculture industry of Europe and North America (Feist, 2004; Burkhardt-Holm et al., 2005; Borsuk et al., 2006; Okamura et al., 2011), novel treatment targets and compounds are urgently needed.

The assembly of the T. bryosalmonae transcriptome enabled us to identify potential drug targets while avoiding potential toxicity to the salmonid host due to drug target homology. The top drug candidate (based on the lowest BLAST E-value) endoglycoceramidase (EGCase) is an enzyme that hydrolyses acidic and neutral glycosphingolipids to oligosaccharides and ceramides. Our homology modelling analysis predicted that the T. bryosalmonae EGCase is expected to have the similar binding site as the original template and hence the same inhibitors, such as iminocyclitols, may affect this enzyme activity in T. bryosalmonae. Among the free-living Cnidaria, EGCase has been previously reported to occur in Cyanea nozakii (Jellyfish) (Horibata et al., 2000) and Hydra vulgaris (Hydrozoa) (Horibata et al., 2004). Interestingly, EGCase is

Fig. 5. Prediction of the EGCase2 (DN1345_c0_g1_i1) and CA2 (DN419_c0_g2_i1) three-dimensional structures in T. bryosalmonae. (pink). (A) The parasite EGCase2 (green) structure is superimposed on the chain B of the template endoglycoceramidase ii from rhodococcus sp. In EGCase2, the inhibitor (yellow) is shown in ball and stick representation. Similar binding cavities between template and target can be seen around the inhibitor molecule. (B) The same structure shown in ribbons to illustrate the consistency between the template and target structures. (C) The interactions between the EGCase ii amino acids and inhibitor at the binding site are shown with black lines as reported in Caines et al. (2007). (D) The parasite CA2 (green) structure is superimposed on the chain A of the Schistosoma mansoni carbonic anhydrase (SmCA, pink). The surface is made transparent to show the active site. (E) The same structure shown in ribbons to illustrate the consistency between SmCA and the predicted T. bryosalmonae structures. (F) The active site (His 76, Glu 117, and Thr missing) and Zinc binding Histidines (His 89, His 111 and His 113) and residues of T. bryosalmonae CA2 also acquired similar orientation as the SmCA (Da'dara et al., 2019).
involved in a unique dietary glycosphingolipids catabolism pathway in *Hydra* (Horibata et al., 2004). However, until this study, endoglycoceramidases have not been reported in Myxozoa. The presence of two copies of this gene in a reduced cnidarian genome makes it possible that the glycosphingolipids catabolism pathway also plays a role in the parasite lipid metabolism.

The second identified drug target legumain-like protease, an asparaginyl endopeptidase, occurs in the gut of the hard tick (*Ixodes ricinus*) where it participates in the digestion of blood meal haemoglobin (Sojka et al., 2007). A similar legumain enzyme is also present in the parasitic blood fluke *Schistosoma mansoni* (Davis et al., 1987). The proteolytic activity of legumain can be reduced by Aza-peptidyl Michael acceptors, which are inhibitors for the legumains (Götz et al., 2008; Ovat et al., 2009; Rendón-Gandarilla et al., 2013). For example, Aza-peptidyl Michael acceptors inhibit legumain activity in the living cells of protist *Trichomonas vaginalis* and reduce the parasitic cytodehiscence to the host cells (Rendón-Gandarilla et al., 2013). Similarly, Aza-peptidyl Michael acceptors (Götz et al., 2008) and epoxide inhibitors (Ovat et al., 2009) have been shown to inhibit legumain in other evolutionary distant parasite species, such as *S. mansoni* and *I. ricinus*.

The third drug target, Carbonic anhydrase 2 (CA2), belongs to the plasma membrane anchored alpha carbonic anhydrases, which catalyse the reversible hydration of carbon dioxide and aid in intracellular pH regulation. Carbonic anhydrases are considered as potent anti-parasitic drug targets against protists, such as *Trypanosoma, Leishmania* (Vermeilho et al., 2017) and, recently, against blood flukes, schistosomes (Da’dara et al., 2019; Angeli et al., 2020). In the predicted three-dimensional model, the active site- and Zn binding residues of *T. bryosalmonae* CA2 retained similar positions as in SmCA, suggesting that similar therapeutic inhibitors may be used to target this protein in *T. bryosalmonae*. The schistosome alpha carbonic anhydrase plays a role in parasite virulence as its suppression using RNAi significantly reduced the establishment of parasite within host (Da’dara et al., 2019). Several sulphonamide, thioli and hydroxamate substances have been shown to inhibit CA2 in *Trypanosoma* and *Leishmania* schistosomes (Vermeilho et al., 2017; Angeli et al., 2020).

The fourth drug target, pancreatic lipase-related protein 2 (PLRP2), is a secretory lipase enzyme that in vertebrates hydrolyses both phospholipids and galactolipids and helps in milk fat digestion (Berton et al., 2009). This gene is also present in genomes of several cnidarians, including *Exaiptasia polida* (NCBI Gene IDs: 110254178 & 110231612), *Stylophora pistillata* (NCBI Gene IDs: 111321164 & 110253264) and *Nematostella vectensis* (NCBI Gene IDs: 5496642). Presence of PLRP2 in invertebrates and cnidarians suggest a potential role in nutrient acquisition and lipid metabolism. In humans, this enzyme has been employed as a target for the lipase inhibitors, such as Orlistat, for obesity treatment (Ballinger and Peikin, 2002). Interestingly, Orlistat has also been shown to inhibit in vitro growth of a protist parasite *Giardia duodenalis* (Hahn et al., 2013). Altogether, earlier work on endoglycoceramidases, legumain-like proteases, carbonic anhydrases and lipase inhibitors suggests that the identified potential anti-parasitic drug targets represent promising chemotherapeutic candidates also against myxozoan *T. bryosalmonae*.

The pathways and enzymes assigned by the KEGG orthology provided important functional information about the genetic make-up of *T. bryosalmonae*. Among identified pathways, the endocytosis pathway is expected to have an important role in *T. bryosalmonae*. Specifically, endocytosis is involved in acquiring nutrients in other Myxozoa (Yang et al., 2014). In addition, endocytosis of the cellular content of the host phagocyte (that has already engulfed the parasite cell) has been observed in *T. bryosalmonae* (Morris and Adams, 2008). Endocytosis has also been observed during the transfer of material from the primary cell to the enclosed secondary and tertiary cells in *T. bryosalmonae* (Morris and Adams, 2008), suggesting that endocytosis may have wider functions than acquiring nutrients. Furthermore, it has been shown that the internal cells (both secondary and tertiary) of *T. bryosalmonae* are formed via the engulfment of other cells by primary cell (Morris, 2010) and thus, the same molecular machinery i.e. endocytosis pathway genes are likely to be involved in these processes.

Peptidases (proteases) were the second most abundant enzymes found *T. bryosalmonae*. These are important catabolic enzymes that hydrolyze proteins into smaller polypeptides or single amino acids. A key role of the proteases in the nutrient digestion has already been proposed for the myxozoan *T. kitauei* (Yang et al., 2014). Furthermore, proteases are known virulence factors in many parasites and considered as important targets for the antiphrastic drug development (McKerrow et al., 2006; Pina-Vazquez et al., 2012; Siqueira-Neto et al., 2018). More recently, proteases have also been suggested as potential drug targets in myxozoan *S. molnari* (Hartigan et al., 2020). In the current *T. bryosalmonae* assembly, the distribution of KEGG assigned proteases showed higher abundance of cysteine (32%) and metallo (27%) proteases, while a smaller proportion of serine (16%), threonine (16%) and aspartic proteases (8%) were identified (Fig. 3C). In contrast, recently published *T. bryosalmonae* assembly (Faber et al., 2021) reported a slightly higher proportion of metallo proteases in relation to cysteine and serine proteases. This difference may reflect the host-specificity of proteases as the former transcriptome was common to both rainbow trout and bryozoan host, while the latter was based on a native salmonid host, brown trout. Furthermore, despite low completeness of current transcriptome assembly, the protease composition of *T. bryosalmonae* was very similar to *S. molnari, C. shasta*, *K. iwatai* and *M. cerebralis*. Notably, *T. kitauei* exhibited a clearly different protease composition compared to other myxozoans (Fig. 3C).

Between severely and moderately infected fish, altogether, 55 *T. bryosalmonae* genes were differentially expressed. Most interestingly, a higher expression of minicollagens, which are expected to be only expressed during myxozoan sporogenesis, in moderately infected fish suggests that more parasites had reached the sporogenic stage in moderately infected relative to the severely infected fish. Similarly, Niemann-Pick C2 protein and ammonium transporter Rh, whose homologos are found in the polar capsule of the myxozoan *C. shasta* (ecdysteroid-regulated 16 kDa-like and ammonium transporter Rh type C 2-like respectively) (Piriatiensky et al., 2017) were upregulated in moderately infected fish. Thus, it is possible that fish with low PL shows increased abundance of sporogenic stages of *T. bryosalmonae*. However, further experimental time-series analysis is needed to confirm this hypothesis. Among the genes upregulated in severely infected fish was Talin, which is related to the cytoskeletal adhesion to the extracellular matrix (Klapholz and Brown, 2017). Recently, the upregulation of the Talin gene in *C. shasta* has been shown to associate with the transformation from a localized (intestine) to a proliferated infection of other organs in rainbow trout (Alama-Bermejo et al., 2019). Thus, Talin, as a putative virulence factor, may also play role in proliferation and adhesion in *T. bryosalmonae*.

Based on the 231 metazoan genes detected in all published myxozoan transcriptomes/genomes using BUSCO, we estimated that the assembled transcriptome recovered approximately ~56% of *T. bryosalmonae* genes (~92% for intersect transcriptome of Faber et al., 2021). The removal of transcripts with low expression levels (less than 40 reads), is one likely reason behind the low BUSCO score and genes with low expression levels are
most likely underrepresented in the current transcriptome assembly. Secondly, mapping of the reads against multiple fish genomes is expected to efficiently exclude teleost sequences but it also removes parasite reads homologous to the host (Schulze et al., 2016). This, however, reduces potential toxicity due to nucleotide level homology between the host and parasite and facilitates drug target discovery process. Thirdly, RNA sequencing has known bias against extremes of GC content (Swindell et al., 2014; Sheng et al., 2016) and the AT-richness of the T. bryosalmonae transcripts may have resulted in an underrepresentation of the parasite reads in the raw data. This might be one potential reason why only ~10% protein sequences in the current T. bryosalmonae transcriptome assembly are complete (i.e. having start and stop codons at the opposite ends). In addition, our filtering pipeline most likely removed a proportion of the true parasite reads from the further assembly, reflecting the trade-off between false positive and negative identification of parasite reads in dual-RNA sequencing experiments. Finally, despite applying two different strategies for the transcriptome annotation, one fourth (24.6%) of the transcripts still remained uncharacterized. This demonstrates the urgent need for expanding myxozoan genomic and proteomic resources. Despite the limitations, this study illustrates the value and utility of our T. bryosalmonae transcriptome assembly by enabling the identification of novel drug targets providing new biological insights into both parasite biology and disease progression.

In conclusion, we developed an in silico pipeline to efficiently separate transcripts of T. bryosalmonae from the kidney tissue of a native fish host, followed by de novo transcriptome assembly. As such, this study represents an important milestone towards developing new effective control strategies against PKD. More generally, we expect that the transcriptome assembly of T. bryosalmonae represents a valuable resource for future studies on genomics and architecture and evolution of the myxozoans and will facilitate research on molecular mechanisms underpinning PKD.

**Supplementary material.** The supplementary material for this article can be found at https://doi.org/10.1017/S003118202100010X.

**Acknowledgements.** The authors wish to acknowledge CSC-IT Center for Science, Finland, for generous computational resources. The authors are also thankful to Dr Jason W. Holland, Dr Ashlie Hartigan and Dr Gema Alama-Bermejo for providing transcriptome sequences for C. shasta, T. bryosalmonae and S. molnari and C. shasta, respectively.

**Financial support.** The work was supported by the University of Turku Foundation (F.A. grant for PhD studies), the Ella and Georg Ehrnrooth Foundation (F.A. grant for PhD studies), Estonian Ministry of Education and Research (R.G., institutional research funding project IUT8-2), Estonian Research Council (R.G., grant number PRG852) the Academy of Finland (A.V., grant number 266321), German Research Foundation research fellowship (P.V.D., grant number DE 2405/1-1) and the Swiss National Foundation (H.H., Sinergia Program grant CRSII2_147649).

**Conflicts of interest.** The authors declare there are no conflicts of interest.

**Ethical standards.** Collection of the studied fish was carried based on permits 46/2014 and 63/2015 issued by Estonian Ministry of Environment. The fish were euthanized in accordance to the principles described in ‘DIRECTIVE 2010/63/EU OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 22 September 2010 on the protection of animals used for scientific purpose’. The authors have followed the principles of the 3Rs (Replacement, Reduction and Refinement) and have involved the minimum number of animals to produce statistically reproducible results.

**References**

Ahmad F, Debes PV, Palomar G and Vasemägi A (2018) Association mapping reveals candidate loci for resistance and anaemic response to an emerging temperature-driven parasitic disease in a wild salmonid fish. Molecular Ecology 27, 1385–1401.

Alama-Bermejo G, Holzer AS and Bartholomew JL (2019) Myxozoan adhesion and virulence: Ceratonia Shasta on the move. Microorganisms 7, 397.

Alama-Bermejo G, Meyer E, Atkinson SD, Holzer AS, Wisniewska MM, Kolisko M and Bartholomew JL (2020) Transcriptome-wide comparisons and virulence gene polymorphisms of host-associated genotypes of the cnidianarian parasite Ceratonia Shasta in salmonids. Genome Biology and Evolution 12, 1258–1276.

Aljanabi SM and Martinez I (1997) Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. Nucleic Acids Research 25, 4692–4693.

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990) Basic local alignment search tool. Journal of Molecular Biology 215, 403–410.

Ahmad F, Pintecu M, Maier S, Simionescu BC, Da’dara AA, Skelly PJ and Supuran CT (2020) Sulfonylamine inhibition studies of an α-carbonic anhydrase from Schistosoma mansoni, a platyhelminth parasite responsible for schistosomiasis. International Journal of Molecular Sciences 21, 1842.

Ballinger A and Petkin SR (2002) Orlistat: its current status as an anti-obesity drug. European journal of pharmacology 440, 109–117.

Berton A, Sebben-Kreuzer C, Rouvellas S, Lopez C and Cremonn I (2009) Individual and combined action of pancreatic lipase and pancreatic lipase-related proteins 1 and 2 on native versus Homogenized milk fat globules. Molecular Nutrition & Food Research 53, 1592–1602.

Bettge K, Wahl T, Segner H and Schmidt-Posthaus H (2009) Proliferative kidney disease in rainbow trout: time- and temperature-related renal pathology and parasite distribution. Diseases of Aquatic Organisms 83, 67–76.

Borsuk ME, Reichert P, Peter A, Schager E and Burkhardt-Holm P (2006) Assessing the decline of brown trout (Salmo Trutta) in Swiss rivers using a Bayesian probability network. Ecological Modelling 192, 224–244.

Brouneaux M, Visse M, Gross R, Pukk L, Saks L and Vasemägi A (2016) Parasite infection and decreased thermal tolerance: impact of proliferative kidney disease on a wild salmonid fish in the context of climate change. Functional Ecology 31, 216–226.

Bryant DM, Johnson K, DiTommaso T, Tickle T, Couger MB, Pasyzin-Dogra D, Lee TJ, Leigh ND, Kuo TH, Davis FG, Bateman J, Bryant S, Guzikowski AR, Tsai SL, Coyne S, Ye WW, Freeman RM, Peskhn I, Tabin CJ, Regev A, Haas J and Whited JL (2017) A tissue-mapped axolotl De Novo transcriptome enables identification of limb regeneration factors. Cell Reports 18, 762–776.

Burkhardt-Holm P, Giger W, Güttinger H, Ochsenbein U, Peter A, Scheurer K, Segner H, Staub E and Suter MJ (2005) Where have all the fish gone? Environmental Science & Technology 39, 441A–447A.

Burnham AJ (2019) Scientific advances in controlling Nosema Ceranae (Microsporida) infections in honey bees (Apis mellifera). Frontiers in Veterinary Science 6, 79.

Caines MEC, Hancock SM, Tarling CA, Wrodnigg TM, Stick RV, Stütz AE, Vaseela A, Withers SG and Smyrnaić NJC (2007) The structural basis of glycosidase inhibition by five-membered iminocycloptane: the clan A glycoside hydrolase endoglucocerebrosidase as a model system. Angewandte Chemie International Edition 46, 4474–4476.

Carraro L, Hartikainen H, Jokela J, Bertuzzo E and Rinaldo A (2018) Estimating species distribution and abundance in river networks using environmental DNA. Proceedings of the National Academy of Sciences of the United States of America 115, 11724–11729.

Chang ES, Neuhof M, Rubinstein ND, Diamant A, Philippe H, Huchon D and Cartwright P (2015) Genomic insights into the evolutionary origin of Myxozoa within Cnidaria. Proceedings of the National Academy of Sciences of the United States of America 112, 14912–14917.

Changklungmoa N, Chaiithirayanon K, Cheukamud W, Chaiwichien A, Changklungmoa N, Chaithirayanon K, Cheukamud W, Chaiwichien A, and Sadaa AA, Angeli A, Ferraroni M, Supuran CT and Skelly PJ (2019) Crystal structure and chemical inhibition of essential schistosome host-interactive virulence factor carbonic anhydrase SmCA. Communications Biology 2, 1–11.

Darda AA, Angeli A, Ferraroni M, Supuran CT and Skelly PJ (2019) Crystal structure and chemical inhibition of essential schistosome host-interactive virulence factor carbonic anhydrase SmCA. Communications Biology 2, 1–11.
Holland JW, Okamura B, Hartikainen H and Secombes CJ (2011) A novel minicollagen gene links cnidarians and myxozoans. Proceedings of the Royal Society B: Biological Sciences 278, 546–553.
Horibata Y, Okino N, Ichinose S, Omori A and Ito M (2000) Purification, characterization, and cDNA cloning of a novel acidic endoglycosidase from the jellyfish, Cyanea nozakii. Journal of Biological Chemistry 275, 297–310.
Horibata Y, Sakaguchi K, Okino N, Iida H, Inagaki M, Fujisawa T, Hama Y and Ito M (2004) Unique catalytic pathway of glycopospholipids in a hydroid, Hydra Magnapapillata, involving endoglycosidase. Journal of Biological Chemistry 279, 33579–33589.
Ishida J, Hinou H, Naruchi K and Nishimura SI (2014) Synthesis of neoglycopospholipid from methoxysan-funcationalized ceramide. Bioorganic and Medicinal Chemistry Letters 24, 1197–1200.
Kelley LA, Merulis S, Yates CM, Was M and Sternberg MJE (2015) The Phezy2 web portal for protein modeling, prediction and analysis. Nature Protocols 10, 845–858.
Kim D, Langmead B and Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. Nature methods 12, 357–360.
Klapolz B and Brown NH (2017) Talin – The master of integrin adhesions. Journal of Cell Science 130, 2435–2446.
Lagesen K, Hallin P, Rodland EA, Starfeldt H-H, Rognes T and Ussery DW (2007) 2007: NMNNamer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Research 35, 3106–3108.
Langmead B and Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nature Methods 9, 357–359.
Le Guvello R, Pobel T, Richards RH and Gould C (1999) Field efficacy of a 10-day treatment of fumagillin against proliferative kidney disease in rainbow trout Oncorhynchus mykiss. Aquaculture 171, 27–40.
Lemopolou A, Uusi-Heikkilä S, Huusko A, Vasemägi A and Vainikka A (2018) Comparison of migratory and resident populations of brown trout reveals candidate genes for migration tendency. Genome Biology and Evolution 10, 1493–1503.
Lien S, Koop BF, Sandve SR, Miller JR, Kent PM, Nøde Westen, T and Hvidsten TR, Le Gouvello R, Pobel T, Richards RH and Gould C (2007) ASTALAVISTA: dynamic and flexible analysis of alternative splicing events in custom gene datasets. Nucleic Acids Research 35, 2816–2822.
Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Ginrke A, Rhind N, Di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N and Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 29, 644–652.
Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Cooper MB, Eccles D, Li B, Lieber M, Macmanes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, Leduc RD, Friedman N and Regev A (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protocols 8, 1494–1512.
Hahn J, Seebier F, Kołodziej H, Ignatius R, Laue M, Aebischer T and Klotz C (2013) High sensitivity of Gdaruada Duodeni to tetrahydrolipstatin (orlistat) in vitro. PLoS ONE 8, 1–10.
Hartigán A, Kosakyan A, Peckóvá H, Eszterbauer E and Holzer AS (2020) Transcription of Sphaerospora Molnarii (Cnidaria, Myxosporida) blood stages provides proteolytic arsenal as potential therapeutic targets against sphaerosporiasis in common carp. BMC Genomics 21, 1–404.
Hartikainen H, Grulh A and Okamura B (2014) Diversification and repeated morphological transitions in endoparasitic cnidarians (Myxozoa: Malacosporaea). Molecular Phylogenetics and Evolution 76, 261–269.
Hartikainen H, Filippenko D, Okamura B and Vasemägi A (2015) First microsatellite loci of the myxozoan parasite Tetracapsuloides Broyalsalmonae, the causative agent of proliferative kidney disease. Diseases of Aquatic Organisms 113, 85–88.
Hedrick RP, MacConnell E and de Kinkelin P (1993) Proliferative kidney disease of salmonid fish. Annual Review of Fish Diseases 3, 277–290.
Hedrick RP, Baxa DV, De Kinkelin P and Okamura B (2004) Malacosporean-like spores in urine of rainbow trout react with antibody and DNA probes to Tetracapsuloides broyalsalmonae. Parasitology Research 92, 81–88.
peptidyl Michael acceptor and epoxide inhibitors – Potent and selective inhibi-
tors of Schistosoma mansoni and Isodes ricinus legumins (asparaginyl endopeptidases). Journal of Medicinal Chemistry 52, 7192–7210.

Pertea G and Pertea M (2020) GFF Utilities: GffRead and GffCompare [ver-
zon 2; peer review: 3 approved], F1000Research, 9, 304. Available at https://
doi.org/10.12688/F1000research.22397.2.

Pertea M, Pertea GM, Antonecchi CM, Chang T, Mendell JT and Saltzberg
SL (2015) Stringtie enables improved reconstruction of a transcriptome from RNA-seq reads. Nature Biotechnology 33, 290–295.

Petterse EN, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng
EC and Ferrin TE (2004) UCSF Chimera? A visualization system for
exploratory research and analysis. Journal of Computational Biology and
Analysis, 16, 1605–1612.

Pina-Vazquez C, Reyes-Lopez M, Ortiz-Estrada G, de la Garza M and
Serrano-Luna J (2012) Host-parasite interaction: parasite-derived and
-induced proteases that degrade human extracellular matrix. Journal of
Parasitology Research 2012, 748206–24.

Piriatinskiy G, Atkinson SD, Park S, Morgenstern D, Brekhman V,
Parkkila S and Brekhman V (2012) Analysis of the discriminative inhibition of mammalian digestive
endopeptidases).<ref>

PLoS Pathogens 7, e1002089.

Point V, Pavan Kumar KVP, Marc S, Delorme V, Parsiega G, Amara S,
Carrière F, Buono G, Fofiau F, Canaan S, Leclaire J and Cavalier JF
(2012) Analysis of the discriminative inhibition of malarial digestive
lipases by 3-phenyl substituted 13,4-oxadiazol-2(3H)-ones. European
Journal of Medicinal Chemistry 58, 452–463.

Rendón-Gandarrila FJ, Ramón-Luque I, de los A, Ortega-López J, Rosa de
Andrade I, Benchimol M and Arroyo R (2013) The TvlLEGU-1, a
legumain-like cysteine proteinase, plays a key role in Trichomonas vaginalis
Cytoskeleton. BioMed Research International 7, 561979.

Rondeau EB, Minkley DR, Leong JS, Messmer AM, Jantzen JR, von
Schalburg KR, Lemon C, Bird NH and Koop BF (2014) The genome and
linkage map of the Northern pike (Esox Lucius): conserved synteny revealed between the salmonid sister group and the Neoteleostei. PLoS ONE 9, e102089.

Saulnier D, Brémond M and De Kinkelin P (1996) Cloning, sequencing and
expression of a cDNA encoding an antigen from the Myxosporean parasite
causing the proliferative kidney disease of salmonid fish. Molecular and
Biochemical Parasitology 83, 153–161.

Schulze S, Schleicher J, Guthke R and Linde J (2016) How to predict molecu-
arial interactions between species? Frontiers in Microbiology 7, 442.

Sheng Q, Vickers K, Zhao S, Wang J, Samuels DC, Kouses O, Shyr Y and
Guo Y (2016) Multi-perspective quality control strategy of Illumina RNA sequen-
cing data analysis. Briefings in Functional Genomics 15, 196–204.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV and Zdobnov EM
(2015) BUSCO: assessing genome assembly and annotation completeness
with single-copy orthologs. Bioinformatics 31, 3210–3212.

Siqueira-Neto JL, Debnath A, McCall I, Bernatchez JA, Ndao M, Reed SL
and Rosenthal PJ (2018) Cysteine proteases in protozoan parasites. PLoS
Neglected Tropical Diseases 12, e006512.

Sojka D, Hajdušek O, Dvořák J, Sajid M, Franta Z, Schneider EL, Craik CS,
Vancová M, Buresová V, Bogyo M, Sexton KB, McKerrow JH, Caffrey
CR and Kopsáček P (2007) Irae – An asparaginyl endopeptidase (legumain)
in the gut of the hard tick Isodes ricinus. International Journal for
Parasitology 37, 713–724.

Srivastava S, Sinha R and Roy D (2004) Toxicological effects of malachite
green. Aquatic Toxicology 66, 319–329.

Sudhagar A, Kumar G and El-Mathouli M (2020) The malacosporean myx-
osporean parasite Tetracapsuloides Breyssmalonica: a threat to wild salmonids.
Pathogens 9, 16.

Sullivan JC and Finnerty JR (2007) A surprising abundance of human disease
genes in a simple basal animal, the starlet sea anemone (Nematostella vec-
tensis). Genome 50, 689–692.

Swindell WR, Xing X, Voorhees JJ, Elder JT, Johnston A and Gudjonsson
JE (2014) Integrative RNA-seq and microarray data analysis reveals GC con-
tent and gene length biases in the psoriasis transcriptome. Physiological
Genomics 46, 533–546.

Syriänen L, Tolvanen MEE, Hilvo M, Vulló D, Carta F, Supuran CT and
Parkkila S (2013) Characterization, bioinformatic analysis and diithio-
carbamate inhibition studies of two new α-carbonic anhydrases, CAH1
and CAH2, from the crustacean Drosophila melanogaster. Bioorganic and
Medicinal Chemistry 21, 1516–1521.

Vermelho AB, Capacì GR, Rodrigues IA, Cardoso VS, Mazotto AM and
Supuran CT (2017) Carbonic anhydrases from Trypanosoma and
Leishmania as anti-protozoan drug targets. Bioorganic and Medicinal
Chemistry 25, 1543–1555.

Videvall E, Cornwallis CK, Ahrén D, Palinauskas V, Valkiūnas G and
Helligren O (2017) The transcriptome of the avian malaria parasite
Plasmodium Ashfordi displays host-specific gene expression. Molecular
Ecology 26, 2939–2958.

Wahlì T, Knuesel R, Bernet D, Segner H, Pugovkin D, Burkhardt-Holm P,
Escher M and Schmidt-Posthauås H (2002) Proliferative kidney disease
in Switzerland: current state of knowledge. Journal of Fish Diseases 25,
491–500.

Westermann AJ, Gorski SA and Vogel J (2012) Dual RNA-seq of pathogen
and host. Nature Reviews Microbiology 10, 618–630.

Wishkovsky A, Groff MJ, Lauren DJ, Toth RJ and Hedrick RP (1990)
Efficacy of fumagillin against proliferative kidney disease and its toxic
side effects in rainbow trout (Oncorhynchus Mykiss). Journal of Fish Diseases
23, 141–147.

Wood DE, Lu J and Langmead B (2019) Improved metagenomic analysis
with Kraken 2. Genome Biology 20, 257.

Yahalomi D, Atkinson SD, Neuhof M, Chang ES, Philippe H, Cartwright P,
Bartholomew JL and Huchon D (2013) The Tvlegu-1, a
legumain-like cysteine proteinase, plays a key role in Trichomonas vaginalis
Cytoskeleton. BioMed Research International 7, 561979.