

Identification of transcripts relating to host-parasite interactions and feeding-associated activities through de novo transcriptome sequencing of branchiuran fish lice, *Argulus foliaceus* (Linnaeus, 1758) and *Argulus coregoni* Thorell, 1865

Aisha AmbuAli (✉ hala@squ.edu.om)

University of Stirling

John B. Taggart

University of Stirling

Michael Bekaert

University of Stirling

Stefanie Wehner

University of Stirling

Sean J. Monaghan

University of Stirling

James E. Bron

University of Stirling

Research Article

Keywords: RNAseq, host-parasite interaction, immunomodulation, Branchiura, parasite, fish lice

Posted Date: March 17th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-285509/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

The genus *Argulus*, whose members are widely known as “fish lice”, comprises a group of obligate ectoparasites that cause argulosis. This parasitic disease affects both wild and cultured fish populations and can lead to significant mortalities. Our previous work confirmed the presence of immunomodulatory or other active products in the secretions of *Argulus foliaceus* (Linnaeus, 1758) using proteomic analysis conducted on secretory / excretory products (SEPs). Some of the compounds identified have also been described for other host-associated Crustacea, particularly the parasitic copepod *Lepeophtheirus salmonis* (Krøyer, 1837). The identification and characterisation of such components can contribute to the understanding of the host-parasite relationship for argulids and may also assist in the development of novel control methods, however, due to limitations in existing genomic resources, only 27 proteins could be confirmed in SEPs. In this study, high-throughput sequencing, supported by the use of a range of bioinformatics tools and methodologies were employed on host-associated stages of *A. foliaceus* and *A. coregoni* Thorell, 1865 to explore elements of the transcriptome associated with host-parasite interactions and feeding-associated activities.

Results

The transcriptome data produced in the current study comprises the largest current genomic data set for the genus *Argulus*, with 84,256,934 reads, assembled into 66,940 contigs for *Argulus foliaceus*, and 10,840,092 reads, assembled to give 40,954 contigs, representing the first transcript sequences for *Argulus coregoni*. A pre-existing transcriptome dataset for *A. siamensis* was also analysed. An overview of the functional roles of the transcripts and their respective proteins for the three *Argulus* species was performed using GO analysis. A phylogenetic tree was also constructed, using the top 100 shared genes for the four transcriptome datasets available for *Argulus*. From this analysis, *A. coregoni* was revealed to be more closely related to *A. foliaceus* than to *A. siamensis*. The application of OrthoVenn software to the *Argulus* spp. transcriptome datasets identified 6,674 shared gene clusters. The transcriptomic analysis in this study, supported by earlier proteome work, has contributed to identifying, for the first time, a range of proteins / genes in *Argulus*, and more widely for the *Branchiura*, that have been previously characterised as being important immune mediators for other aquatic and terrestrial blood-feeding arthropods. Data are available *via* EBI with the project number PRJEB34947.

Conclusions

In addition to confirming the presence of a range of key immune mediators in *Argulus* spp. for the first time, this study has provided important genomic tools for researchers seeking to better understand the biology and host interactions of *Argulus* spp. The created resources unlock the possibility of characterising protein targets relating to host-parasite interactions and feeding-associated activities and can thereby assist the future development of vaccines, veterinary drugs, functional feeds and other tools that might contribute to improved integrated pest management of these globally important parasites.

Background

Argulus is one of the most common crustacean fish ectoparasite genera encountered worldwide [1–3], being frequently referred to as “fish lice”. *Argulus* is the largest and most diverse genus of the branchiuran family *Argulidae* [4]. Members of this genus are responsible for the condition argulosis, a parasitic disease affecting both wild and cultured fish populations [5,6].

Argulus spp., such as *Argulus japonicus* Thiele, 1900 and *Argulus coregoni* Thorell, 1864, have been described as causing a skin haemorrhagic response in carp (*Cyprinus carpio* Linnaeus, 1758) and rainbow trout (*Oncorhynchus mykiss* Walbaum, 1792) respectively [7, 8, 9] and it has been suggested that *Argulus* spp. might secrete substances during feeding to modulate host immunity or to facilitate blood feeding [7]. A general lack of knowledge in this area led von Reumont *et al.* (2014) to call for studies looking into the composition of the expected active secretions originating from branchiuran exocrine glands associated with feeding [10]. Due to limited genomic resources, we recently identified only 27 proteins in the secretory / excretory products (SEPs) of *Argulus foliaceus* [11]. Some of the compounds identified have also been described from other host-associated Crustacea, particularly the parasitic copepod *Lepeophtheirus salmonis* (Krøyer, 1837) [12].

Although argulids affect aquaculture and fisheries globally, causing major economic impacts for freshwater aquacultured fish species worldwide, few studies have examined the mechanisms underlying physiological and behavioural aspects of the parasite [11]. In the UK, *Argulus* spp. have economic impacts for both aquaculture and sports fishing industries. To date, control methods employed have targeted few of the biological aspects of the parasite and have had little impact in preventing argulosis, with those controls that do exist, all having critical limitations in their use [14]. Hence new strategies for control are still urgently needed. One avenue for control is the potential for developing vaccines or other immunoprophylactic approaches such as use of immunostimulants / functional feeds. To achieve this, better knowledge of the host response to the parasite and the parasite’s mechanisms for combating host immunity are required.

A range of newer techniques, including high-throughput sequencing which allows examination of the parasite’s transcriptome, can provide a greater understanding of the biology of the parasite and of key genes involved in particular activities. Such genes and the products they code for can provide potential targets for use in control of argulosis [12, 13]. The first transcriptome analysis for argulids was conducted by Sahoo *et al.* (2013) for *Argulus siamensis* Wilson C.B., 1926, which has a severe impact on the aquaculture industry in India. An Illumina-based sequence analysis, led to the identification of 19,290 coding sequences (CDS) including 184 novel CDS and 59,019 open reading frames (ORFs). Using this same transcriptome shotgun assembly (TSA), Christie (2014) was subsequently able to identify 27 transcripts encoding putative neuropeptide precursors, these peptides being known to play a role in physiological/behavioural control across species. This was the first finding of neuropeptide encoding sequences in any member of the subclass *Branchiura* [16].

More recently, Pinnow *et al.* provided *Argulus foliaceus* (Linnaeus, 1758) transcriptome data (NCBI sequence read archive (SRA) under the Bio-Project number PRJNA293150. 52,725,850). The assembly resulted in 8,424 contigs. These transcriptome data were only used to identify and characterise hemocyanin subunits in *A. foliaceus* and *A. siamensis* and provided evidence that hemocyanin was the principal respiratory protein in the stem lineage of the *Pancrustacea* [15].

Despite the above, genomic resources for *Argulus* spp. and for the *Branchiura* more widely, remain extremely impoverished, providing a major constraint to further research in this area. The work reported in this study sought to provide new transcriptomic resources for *A. foliaceus* and *A. coregoni*, the two native *Argulus* species in the UK, and to analyse the new sequences with respect to the previous two transcriptome libraries established for *A. siamensis* [12] and *A. foliaceus* [15]. Understanding key biological processes in *Argulus* spp. will help in searching for candidate immunomodulatory or drug targets, which may assist researchers to develop new drugs or vaccines to provide supplementary control methods. This study also highlights the fact that *Argulus* spp. might prove productive as a tractable ectoparasitic arthropod model. By filling gaps in the genomic knowledge for this genus, this study provides a useful addition to crustacean genomic resources, particularly for *Branchiura* and details the first list of potential host immunomodulators for these three *Argulus* species.

Methods

Transcriptome sequencing

Transcriptome sequencing was undertaken for two species, *Argulus foliaceus* (Linnaeus, 1758) and *Argulus coregoni* Thorell, 1865, at different times. Feeding adult *A. coregoni* were collected from common carp (*Cyprinus carpio* L.) by Environment Agency staff in Brampton in England, while feeding *A. foliaceus* were collected from the skin of rainbow trout (*Oncorhynchus mykiss* W.), Isle of Bute, Scotland, UK. In both cases the samples were preserved in RNA/later™ (Thermo Fisher Scientific, UK) according to manufacturer's guidelines (24 hr at 4°C, RNA/later drained, tissue stored at -70°C until required). Total RNA was extracted using TRI Reagent® (Sigma-Aldrich, UK) according to the manufacturer's protocol, followed by a column clean up (RNAeasy Mini Kit, Qiagen, UK). Total RNAs were re-suspended in nuclease-free water, with preliminary QC performed by UV spectrometry (NanoDrop ND-1000, Thermo Scientific, USA) and 1 % agarose gel electrophoresis.

For *A. coregoni* total RNA was extracted from a whole male and female adult, with equal quantities used in library preparation. For *A. foliaceus* four different sample types were extracted; a whole adult male, a whole adult female, a pool of mixed juvenile stages and a pool of six dissected pieces from adult parasites. The latter comprised tissue taken from the expected position of the pre-oral spine gland and proboscis glands as established from histological sections [18]. Equal amounts of RNA from the four sources were used.

Library preparation from supplied total RNA and transcriptome sequencing, using Illumina based protocols, were outsourced to commercial suppliers. Different companies were used for two species,

dictated by service availability at the time. For *A. coregoni*, the Edinburgh Genomics (University of Edinburgh, UK) provided 150 base paired-end reads generated on the MiSeq platform, while for *A. foliaceus* the Theragen BioInstitute, Republic of Korea supplied 100 base paired-end reads generated on a HiSeq 2500 platform).

To assist assembly, two existing *Argulus* transcriptome datasets were used. These comprised an *A. siamensis* dataset [12] and an *A. foliaceus* dataset [15] lodged in the NCBI Bioprojects database under accession numbers PRJNA293150 and PRJNA167720, respectively.

Sequence data processing and transcriptome assembly for *A. foliaceus* and *A. coregoni*

The transcriptome reads obtained for *A. coregoni* and *A. foliaceus* were assembled using the following bioinformatics pipeline, illustrated in **Figure 1**.

Quality Control Processing

Quality control of the raw data was conducted using FastQC v0.11.4 [16]. The data were then trimmed using the Trimmomatic v0.35 program [17]. This program was run using the parameters option "ILLUMINACLIP". Deconseq v0.4.3 program [18] was used to remove any sequence contamination. For this, reads were mapped against different databases.

De novo assembly and annotation

De novo assembly was performed on the cleaned RNA-seq raw data using Trinity v2.1.1 [19]; the K-25 default parameter of the Trinity software was used for the assembly. The selection of the Trinity tool for final assembly was decided upon following a trial including other assembler software i.e. Velvet and ABySS. Trinity was found to give higher numbers of more consistent reads. All the sample reads were merged into a single dataset for each species for final assembly. Then Transdecoder v2.0.1 [23] was used to find the coding region within identified transcripts. Identification of the coding region transcripts gave ORFs and amino acid sequences, to prepare the assembled dataset for annotation. Annotation was achieved by 1) BLAST v2.2.31 software [21], BLAST+ [22] using the uniprot/trembl-invertebrate database as query [23] and 2) Annotation HMMER v3.1b2 [24] using the Pfam A v29.0 as query [25].

Gene Ontology annotations

Gene Ontology (GO) annotations were retrieved from the UniProt GO annotation project [<http://www.ebi.ac.uk/GOA>]. All proteins described in the UniProt (Swiss-Prot and TrEMBL) KnowledgeBase, including the uniprot/trembl (invertebrate) database, and used during the annotation stage, were associated with GO controlled vocabulary. The set of all GO annotations for the proteins in the UniProt KnowledgeBase (UniProtKB) was downloaded [accessed 2016-11-09] and for each, the uniprot/trembl (invertebrate) GO annotation was recovered and added to the transcriptome annotations. GO annotation was conducted for both *A. foliaceus* and *A. coregoni* transcriptomes as well as for the pre-

existing *Argulus* transcriptome libraries (*A. siamensis* with Accession: PRJNA167720 and *A. foliaceus* with Accession: PRJNA293150) curated by the National Centre for Biotechnology Information (NCBI).

Identifying shared transcript expression among the three *Argulus* species

OrthoVenn is a web-based genomics tool used for comparative and visualisation analyses of orthologous protein clusters from different species. Use of cluster information helps in identifying protein function and evolution across multiple species [29]. OrthoVenn was used to check for shared expressed genes / proteins. The tool was used to identify gene clusters enriched in the four transcriptome libraries available for *A. foliaceus*, *A. coregoni* and *A. siamensis*. To achieve this, a list of proteins was first prepared for each *Argulus* species and then the ClusterVenn tool used to generate Venn diagrams by clustering proteins of similar sequence and function, *i.e.* orthologous and paralogous genes, among the species. The tool compares between clusters only and ignores singletons where there are no orthologues between species and no paralogues within species. The default OrthoVenn parameters were used *i.e.* e-value cutoff of 1e-5 for all-to-all protein similarity comparisons and an inflation value (-I) of 1.5 for the generation of orthologous clusters.

In addition, OrthoVenn allows searching for specific proteins using the BLAST feature. In this study, it was used to search for some known immunomodulatory candidates to check if any existed within the annotated shared clusters and related functions. The selected targets were: trypsin, serpin (serine protease inhibitor), cathepsin-L, aspartic protease, ferritin, cysteine protease, enolase, phospholipase, adenosine deaminase, apyrase, metalloprotease, thrombin inhibitor and venom serine protease.

Phylogenetic reconstruction

The longest coding sequence (CDS) for selected proteins of *A. foliaceus*, *A. coregoni* and *A. siamensis* were translated into amino acid sequences and a BLASTP was performed against all oligostracan sequences present in the NCBI NR database (accessed 2016-11-09). A total of 10 species were represented in the top 100 shared proteins (**Additional file 1**). Each shared protein sequence was retrieved and aligned individually using ClustalO v1.2.3. Each alignment was concatenated before running RaxML v8.0.0 using partitioned models (one model per protein, based on a GAMMA model of rate heterogeneity with an estimate of proportion of invariable sites and using the BLOSUM62 substitution rates and maximum-likelihood estimate of the base frequencies) with 1000 bootstrap [-# 1000 -m PROTGAMMAIBLOSUM62X].

Results

De novo transcriptome sequencing of *Argulus coregoni* and *Argulus foliaceus*

Quality Control Processing

All samples shipped passed the commercial companies external initial QC (Bioanalyser), with RIN values exceeding 0.8. Transcriptome libraries were prepared by the firms using Illumina based kits and

platforms. One library sequenced by Edinburgh Genomics (MiSeq platform; 150 base paired-end reads) the other by Thetagen (HiSeq2500 platform, 100 base paired-end reads).

De novo assembly and annotation

The MiSeq transcriptome sequencing for *A. coregoni* generated 11,265,959 150 base paired-end raw sequence reads (**Table 1**). Of these, 10,840,092 reads passed quality control and filtering, and were then merged into a single dataset, assembled using Trinity v2.1.1, and blasted to generate a *de novo* transcriptome assembly comprising 40,954 transcripts (contigs). The maximum transcript contig length was 9,791 bp, the mean contig length was 1,787 and the N50 value was 2,339 bp.

The *A. foliaceus* sequencing (HiSeq2500 platform) produced 88,255,979 150 base paired-end raw sequence reads (**Table 1**) with 84,256,934 passing quality control and filtering. The high-quality reads were merged into a single dataset, assembled with Trinity v2.1.1, and blasted to get total assembled contiguous sequences of 66,940 reads. The maximum transcript contig length was 17,078 bp, the mean transcript contig length was 1,842 bp and the N50 value was 2,573 bp.

The pre-existing TSA for *A. siamensis* (Accession PRJNA167720) and *A. foliaceus* (Accession: PRJNA293150) were also examined and the total number of reads, number of assembled contigs and N50 are summarised in (**Table 1**).

Table 1 Summary data generation and statistics for *de novo* transcriptome assembly of *A. coregoni*, *A. foliaceus* and *A. siamensis*, first two (bolded) columns comprise sequencing results for the current study.

Data generation	<i>Argulus coregoni</i>	<i>Argulus foliaceus</i>	<i>Argulus siamensis</i>	<i>Argulus. foliaceus</i>
Raw reads	11,265,959	88,255,979	77,759,443	52,725,850
High quality reads	10,840,092	84,256,934		
Total contigs	73,164,334	123,272,467	50,396,610	16,894,535
Number of contigs	40,954	66,940	46,352	8,424
Largest contig (bp)	9,791	17,078	26,436	16,889
Mean contig (bp)	1,787	1,842	1,211	2,006
N50 value in bp	2,339	2,573	2,302	1,499
GC%	41.3	40.74	38.29	42.05
Illumina platform	MiSeq v2	HiSeq2500	HiSeq 2000	HiSeq 2500
Assembly software	Trinity v2.1.1	Trinity v2.1.1	Velvet/Oases	CLC-Genomics Workbench 7.5.1

NCBI Reference genomic resources. *A. siamensis* (Accession PRJNA167720) and *A. foliaceus* (Accession: PRJNA293150)

Gene ontology annotation

Gene ontology (GO) terms were retrieved for the annotated transcripts, with the longest sequence for each transcript selected. Following the use of BLASTp against the UniProt_Trembl (invertebrate) database, the maximum number of GO terms were identified. The GO analysis was applied both to the current datasets (*A. foliaceus* and *A. coregoni*) and to the pre-existing reference datasets from NCBI (*A. siamensis* and *A. foliaceus*). **Table 2** summarises the GO distribution for the three *Argulus* species from the four transcriptome datasets. The GO terms were associated with transcripts for molecular function (MF), biological process (BP) and cellular component (CC). The GO category terms for *A. coregoni* were the same as those obtained for *A. foliaceus*. In the MF category, binding activity and catalytic activity were the most heavily represented. For BP the most represented terms were “transport” and “metabolic

process” and for CC the most represented terms were “membrane” and “integral component of membrane”

Table 2 Summary of GO distribution for the three *Argulus* species.

Species	<i>A. foliaceus</i>	<i>A. coregoni</i>	<i>A. siamensis</i>	<i>A. foliaceus</i>
Total GO terms	251,456	183,125	179,710	42,361
Number of assigned transcripts	2,076	1,904	1,931	1,596
Number of cellular component terms	48,731 (19%)	36,103 (20%)	28,251 (16%)	8,263 (20%)
Number of molecular function terms	125,676 (50%)	91,041 (50%)	93,791 (52%)	20,952 (49%)
Number of biological process terms	77,049 (31%)	55,981 (30%)	57,668 (32%)	13,146 (31%)

NCBI Reference genomic resources. *A. siamensis* (Accession PRJNA167720) and *A. foliaceus* (Accession: PRJNA293150)

Overall, the GO distributions for the four transcriptomes of the three species (*A. coregoni*, *A. foliaceus* and *A. siamensis*) were highly similar (**Table 2**), and in the GO distribution charts of the first 10 clustered annotated proteins for the *Argulus* species (**Fig. 2 & 3**)

Identifying shared genes expressed in *Argulus* species

Venn diagrams (**Fig. 4A and B**) show the shared orthologous gene clusters among three *Argulus* species. The transcriptome results for *A. foliaceus* (135,679 transcripts, 57,928 unique CDS and 7,932 clusters) of this study and *A. foliaceus* with Accession: PRJNA293150 (8,424 transcripts, 8,424 unique CDS and 6,567 clusters) [15] were combined to check for shared proteins/genes. Out of 7955 clusters over both datasets, 6522 orthologous clusters and 6425 single-copy gene clusters were found (**Fig. 4A**). OrthoVenn analysis of the three *Argulus* species showed that 13,324 orthologous clusters were formed based on the protein sequences from the three species. The diagram shows that 6,674 gene clusters were shared by all three species (**Fig. 4B**).

Blasting specifically selected targets (trypsin, serpin, serine protease, cathepsin-L and aspartic protease, ferritin, cysteine protease, enolase, phospholipase, adenosine deaminase, apyrase, metalloprotease, thrombin inhibitor, venom serine protease) against the resulting shared clusters (6,674) gave a number of hits. For trypsin, 25 annotated clusters were identified and classified into six groups according to their functions. Serpin resulted in four clusters described as alaserpin (Swiss-Prot Hit database), all having the same GO annotation. Eleven clusters were found for serine protease and grouped into eight GO annotations. Only a single cluster each was identified for cathepsin (cathepsin-L), apyrase, aspartic protease (Aspartic protease 6), cysteine protease, enolase and thrombin inhibitor. Two types of ferritin

were found with one being yolk ferritin. Two cluster proteins comprised phospholipase A2, enzymes that hydrolyse phospholipids into fatty acids and other lipophilic substances, combining 11 GO annotation functions. There were 17 clusters of adenosine deaminase as determined using Swiss-Prot Hit; each one having a different GO annotation function. By checking for matches to metalloprotease, 5 clusters distributed between three GO annotation functions were identified. Finally, searching for matches to venom serine protease resulted in 6 gene (protein) clusters with 3 different GO annotations (**Table 3**).

Table 3 Identification of targeted immunomodulator/infection candidates from *Argulus* spp.

Protein Class	Protein	# Proteins	Function(s)	References
Serine protease	Trypsin	25	Digestion and anti-haemostatic	[30,31]
Protease inhibitor	Alaserpin	4	Anti-coagulant and anti-complement activation	[32,33]
	Serine protease inhibitor	2	Anti-coagulation, vasodilator Anti-inflammation	[10]
Protease	Aspartic protease	1	Haemoglobin proteolysis	[34]
	Cysteine protease	1	Anti-inflammatory Haemoglobin digestion	[35,36]
	Venom serine protease	8	Anti-coagulant, vasodilator, anti-inflammatory	[10][10][10][10][10][10][10][10][10][10][10][10][37]
	Cathepsin-L	1	Anti-coagulant	[38]
	Metalloprotease	4	Anti-haemostatic	[39]
Glycoprotein	Ferritin	2	Iron storage and transport - involved in homeostasis of iron during feeding	[40]
Metalloenzyme	Enolase	1	Degrades plasminogen, aids host penetration	[41]
Phospholipase	Phospholipase A2	2	Hydrolyses phospholipids (deactivates platelet-activating factor)	[42]
Purine metabolism enzyme	Adenosine deaminase	2	Vasodilator Antiplatelet	[32,43]
Diphosphohydrolase	Apyrase	1	Anti-pain Anti-inflammatory Antihaemostatic Platelet aggregation inhibitor	[32,43,44]
Serine protease inhibitor (serpin)	Thrombin inhibitor	1	Anticoagulant	[32]

3.1.5 Phylogenetic reconstruction

The findings of the phylogenetic analysis (**Fig. 5**) reflected those of Regier *et al.* [45] and confirms the position of the sequenced *Argulus* species with respect to the superclass Oligostraca.

Discussion

In this study we have generated sequence resources for the poorly characterised genus *Argulus* (*A. foliaceus*, *A. coregoni* and *A. siamensis*) and for the Branchiura more generally. While there were two existing genomic datasets for *Argulus* species (*A. siamensis* and *A. foliaceus*), these were limited in terms of their coverage. The new transcriptomes comprise larger data sets for *Argulus* with 84,256,934 reads assembled into 66,940 contigs for *A. foliaceus* and a first transcriptome for *A. coregoni*, 10,840,092 reads, assembled to give 40,954 contigs.

As there is no reference genome for *Argulus*, the transcriptomes of both *A. foliaceus* and *A. coregoni* were assembled *de novo* using Trinity v2.1.1 software. The assembled transcriptomes are important for both genomic and proteomic studies [11], especially in terms of providing the complete ORF information for genes coding for proteins that may play important roles in biological processes and might therefore provide potentially suitable vaccine candidates [12] or targets for anti-parasiticides.

It was reported by Clark & Greenwood [31] that 50 % of crustacean transcripts from transcriptomic studies were orthologous genes to a model organism and most of these genes belong to biological processes involving metabolism, development, or regulation of a biological process. This concurs with the *Argulus* GO outcomes, as the predominant genes from the biological process analysis were involved in response to metabolic processes. This was a predictable result as shown from other organisms, due to the enormous number of GO terms that cover basic processes required to maintain a living organism [47]. One of the most important processes, and one that has some shared properties between invertebrates and vertebrates is the innate immune system [48]. Knowledge of crustacean immune and stress responses are key to understanding crustacean physiology [46]. Although, the annotation method applied in this study is not sufficient to provide detailed information about the immune pathways, the GO molecular functions associated with the *Argulus* transcripts highlighted specific processes that can also be involved in host-parasite interactions such as hydrolase, transferase and metalloproteinase activities.

Hydrolases have been identified from different parasite venoms which are known to include proteases, peptidases and glycosidase enzymes [34, 35]. As we have not only identified hydrolase transcripts here in *Argulus* tissues, but also recently in *Argulus* SEPs through proteomic analysis [11], the presence of hydrolases might reflect their contribution to host-parasite interactions in *Argulus* species. The identified metalloproteinases from different arthropods are assumed to be related to different biological processes such as digestion, cuticle degradation and host immune suppression [47]. In general, the GO annotations from the three-species identified a high representation of genes involved in metabolism, protein binding and nucleotide binding. These are linked with high basal metabolic activity [51], and although the precise reason is not known, various studies on the effects of parasitism on infected mammals have reported

that parasites increase the basal metabolic rate of the host [52]. The patterns of the GO analysis from three *Argulus* species were similar. Further characterisation of the transcripts involved in immune pathway functions and their potential contribution to successful parasitism can provide a valuable addition to understanding of *Argulus* biology and physiology.

The construction of a limited phylogenetic tree based on the top 100 shared genes for the four transcriptome datasets from *Argulus*, available from both this study and from previous data for other arthropods, confirms the position of the sequenced *Argulus* species with respect to the superclass Oligostraca, as previously suggested by Regier *et al.* [45]. These results suggest that *A. coregoni* is more closely related to *A. foliaceus* than to *A. siamensis*. This probably reflects the fact that *A. foliaceus* and *A. coregoni* are indigenous *Argulus* species in northern Europe and UK with closer and more recent evolutionary links than *A. siamensis*, an invasive species from Asia. Long association in similar environments is also reflected by similar host specificities, or lack of them, allowing them to infect such disparate species as salmonids *e.g.* rainbow trout (*O. mykiss*) and cyprinids *e.g.* common carp (*C. carpio*). This may also mean that they share more similar mechanisms for successful parasite-host interaction.

Comparative functional annotation of overlapping clusters among the *Argulus* species identified 6,674 shared gene clusters, suggesting continued conservation after speciation. Identification of conserved genes can help in the search for more universal control targets in these species, for other *Argulus* species and perhaps for Branchiuran parasites more widely.

The saliva of almost all studied blood feeding arthropods has been shown to contain a range of immunomodulating proteins. Specific components have active functions in altering the host immune response to help blood sucking [53] and exploiting aspects of these powerful substances can provide an effective way to control disease distribution [54]. A range of proteins / genes that have been previously characterised as being important immune mediators for blood-feeding insects and other haematophagous parasites, and identified recently by our group in SEPs of the parasitic crustaceans, *L. salmonis* and *A. foliaceus* [11,12], were here discovered in *Argulus* spp. transcriptome for the first time and indeed for the Branchiura more widely. Proteins identified as trypsin, serpin, serine protease, cathepsin-L, aspartic protease, ferritin, cysteine protease, enolase, phospholipase, adenosine deaminase, apyrase, metalloprotease, thrombin inhibitor, and venom serine proteases in shared transcript clusters for the three *Argulus* species provide potential candidates for investigation of *Argulus* controls in the future.

Transcripts of a trypsin-like protease were detected in *Argulus*. Skin mucus of Atlantic salmon (*Salmo salar* L.) infected by sea lice *L. salmonis* contains trypsin-like proteases derived from the parasite, which suggests that the secretion of this protease during feeding is a mechanism to impair the host immune response [28, 29, 54–58]. Seven types of *L. salmonis* trypsin-like proteases have been inferred to play roles in digestion and other host-parasite interactions [59]. Valenzuela-Munoz *et al.* similarly identified the up-regulation of trypsin-1 gene in the sea louse *Caligus rogercresseyi* (Boxshall & Bravo, 2000) post-treatment with an organophosphate parasiticide, suggesting the gene to be associated with resistance /

susceptibility to organophosphates [56], while Braden *et al* [60] found significant upregulation of trypsin-1 in *L. salmonis* after 2 days post infection of lice feeding on Atlantic salmon (*S. salar*). Serine protease inhibitors (serpins), identified for the first time here in *Argulus*, are recognised as the largest protein group in tick saliva and were suggested to play an important role during tick feeding by modulating host immunity [33,61–65]. Serpins characterised from *Ixodes scapularis* saliva were found to inhibit thrombin and trypsin activity and lower platelet aggregation [66]. Serpins are also recognised from insects, for example salivary serpin from *Aedes aegypti* mosquitoes, serves as an active anti-coagulant due to its ability to inhibit Factor Xa, which is the activated form of the coagulation factor thrombo-kinase (Stark & James 1998). Serpins have also been noted to be differentially expressed in salivary glands of the mosquito *Anopheles culicifacies*, suggesting involvement in the blood feeding process [40]. We have previously identified serpins in *A. foliaceus* SEPs suggesting their role in facilitating parasitism and modulating host immune responses in argulids, and the presence of trypsins suggests their role in parasite-host interactions and digestion [11].

Transcripts for cathepsin-L identified from *Argulus* in the present study may signify a role for this protease in digestion and / or modulation of the host immune system. In *Hysterothylacium aduncum* (Rudolphi, 1802), a parasitic anisakid nematode of fish, cathepsin L was proposed to be a vital protein involved in digestion processes [68]. Cathepsin L was also identified as a potential anticoagulant in SEPs of adult *Haemonchus contortus*, a parasitic nematode of sheep and other ruminants [38]. Franta *et al.* suggested that this could provide a novel target for control of ticks and tick-borne pathogens [69]. Transcripts for cathepsin-L were also reported to be expressed by the copepodid life stages of sea lice, *C. rogercresseyi* [34,70], and were upregulated in feeding adult *L. salmonis* [60] and showed activity in SEPs of pre-adult and adult *L. salmonis* [12,71], suggesting a likely role in modulating the host immune system to facilitate attachment and feeding [71].

Transcripts for aspartic protease 6 were also identified from *Argulus* in this study, this being a protein common to the three species examined. Aspartic proteases have been proposed to have a role during *C. rogercresseyi* feeding through haemoglobin proteolysis, where aspartic protease D was highly expressed in the chalimus and adult stages [34]. The salivary glands of blood feeding arthropods frequently contain anticoagulation inhibitors [32] and in this study we found evidence for the existence of the thrombin inhibitor rhodniin in all three *Argulus* species studied. It has previously been reported to prevent blood clotting by inhibiting thrombin [72] to facilitate blood acquisition. Apyrase is another immunomodulatory molecule, which we found in *Argulus*. This enzyme is believed to have anti-pain; anti-inflammatory and anti-haemostatic activity, thus potentially playing a role during feeding to prevent blood clotting and host behavioural responses [32]. Ribeiro *et al.* described the apyrase from *R. prolixus* to act as a platelet aggregation inhibitor [73]. A number of metalloproteases were also identified in *Argulus*, these being a common enzyme group in other blood-feeding arthropods as well, although it is important to note that metalloproteases also perform a wide range of other roles in arthropods more generally. Transcriptome analysis in conjunction with KEGG pathway analysis conducted by Sahoo *et al.* on *A. siamensis* resulted in the identification of proteases including serine proteases and metalloproteases which have been described to have antigenic properties in some ectoparasites [14]. Apyrase and metalloprotease found in

tick salivary secretions were suggested to affect antiplatelet aggregation and coagulation activity [44] and may serve a similar function for processing blood meals in *Argulus*. Adenosine deaminase, which can act as a vasodilator and antiplatelet aggregation factor for arthropod feeding [32,43] was also found in *Argulus* in the present study. By searching for adenosine deaminase using OrthoVenn BLAST feature, two contigs of adenosine deaminase transcripts resulted from *Argulus* species, adenosine deaminase and RNA-specific adenosine deaminase-1. In the sand fly *Lutzomyia longipalpis* [54] and the mosquito *Aedes aegypti* [74] adenosine deaminase can assist feeding and modulate host immune responses by metabolize the purine 2- deoxyadenosine and adenosine to 2- deoxyinosine and inosine, correspondingly. The mechanism of immune evasion of salivary adenosine deaminase of *Lu. longipalpis* and *Ae. Aegypti*, is regulated by the removal of adenosine, a molecule associated with pain perception and mast cell degranulation. Then, the resulted product, inosine, functions as immunomodulator by inhibiting the production of pro inflammatory cytokines [53].

The exact role of the cysteine protease transcript found here in all three *Argulus* species is not known but it was suggested to have roles in nutrition and / or host immune evasion in the nematode *H. contortus* (Rudolphi, 1803) Cobb, 1898 [38]. It was also suggested to have a critical function in the infection process of the trophont parasitic stages of the ciliate, *Ichthyophthirius multifiliis* and blood feeding parasitic copepod *Phrioxcephalus cincinnatus* (C. B. Wilson, 1908) (Jousson *et al.*, 2007; Perkins *et al.*, 1997). Cysteine proteases have similarly been shown to play important roles in trematode parasites. Their roles include catalysing turnover of parasite peptides and growth factors and regulation of proteins, host tissue penetration and invasion, digestion and modulation of host defence systems [75,76]. Phospholipases were also represented in the shared *Argulus* clusters, with 14 different clusters recognised in Swiss-Prot homology blasting. One of the most notable of these is phospholipase A2 (PLA2), a platelet-activating factor acetyl hydrolase, which is known to deactivate platelet-activating factor by *Leishmania* parasites [42] and known to facilitate blood feeding [77]. Transcripts for two types of ferritin were also found in *Argulus* in the present study, Ferritin-1 heavy chain and yolk ferritin, possibly involved in processing and storing iron after blood meals. Recently, ferritin was reported up-regulated in the salivary gland of the mosquito *A. culicifaciens*, after blood feeding [40], which prevents iron overload by assisting in iron homeostasis through iron transport and storage [78]. Hajdusek *et al.*, working on *Ixodes ricinus* and *Rhipicephalus annulatus* (Say, 1821), illustrated the success of using the secreted form of ferritin 2 protein (FER2) as an antigen in an anti-tick vaccine [79]. Transcripts for enolase B, another protein suggested to be involved in blood digestion / feeding and defensive activity in ticks and other blood feeding parasites [80], were also found in *Argulus* spp. here as well as in sea lice, *L. salmonis* [70]. Verónica Díaz-Martín *et al* [81], reported that, enolase protein collected from the argasid tick *Ornithodoros moubata* saliva, known to bind with plasminogen, and hence may play a role in tick feeding by maintain blood fluidity. Investigations for both the cestode *Echinococcus granulosus*, (Batsch, 1786) [82] and the nematode *Ascaris suum*, (Goeze, 1782) have suggested enolase to be a promising vaccine candidate [41,82].

Future work is needed to determine whether the identified compounds are secreted / expressed from *Argulus* feeding glands (i.e. pre-oral spine and proboscis glands), which we have recently characterised

using histology and transmission electron microscopy[18]. This study substantially extends genomic resources for *Argulus* and is the first study to explore shared genes expressed across species, providing a preliminary list of potential feeding enhancers / immunomodulators / defensins, thus potential broad scope vaccine / drug targets.

Conclusion

In the absence of a fully annotated *Argulus* genome, and in view of the previously limited genomic resources available for argulids and branchiurans more widely, this study has made a key contribution to argulid genomics and sheds further light on crustacean fish parasites more widely. High-throughput sequencing technologies have allowed the discovery of a number of conserved proteins across the species studied and have shed light on possible immunomodulator targets. The data produced from this study for *A. foliaceus* and *A. coregoni* adds considerably to that available from previous transcriptome studies.

The current study has provided evidence, for the first time to our knowledge, of potential effectors of host immunomodulation, potentially providing future studies with targets for *Argulus* drug or vaccine development. There are 145 described species of *Argulus* worldwide [83] and it is therefore useful to find conserved secretory proteins that may be used for control across different *Argulus* species. Thus, a universal vaccine or drug treatment strategy could be applied for different *Argulus* species, as has been previously suggested with respect to the development of a universal vaccine for ticks [84]. The first step for developing any new control treatment is to search for potential targets [85] and the outcomes of the research described in this study provide important initial steps in this direction.

Declarations

Acknowledgments

The authors would like to thank Jimmy Poole at Loch Fad fisheries, Isle of Bute, Scotland and Chris Williams and Amy Reading at the Environmental Agency in Brampton, England for providing *Argulus* samples. We should also like to thank Edinburgh Genomics, the core genomics and bioinformatics facility of University of Edinburgh UK, and Theragen BioInstitute, Republic of Korea, for generating RNA-Seq libraries and sequencing of *Argulus coregoni* and *Argulus foliaceus* respectively. This work was generously supported by Sultan Qaboos University, Oman and by the Schlumberger Foundation Faculty for the Future programme.

Authors' contributions

AA, JB and JT conceived and designed the study. SM was involved in discussions leading to the selected gene candidates. AA and JT performed the experiments. MB and SW performed bioinformatics analyses and assisted in interpretation of results. AA wrote the manuscript with input from all authors. All authors discussed and commented on the manuscript and approved the final manuscript.

Funding

This work was generously supported by Sultan Qaboos University, Oman and by the Schlumberger Foundation Faculty for the Future programme.

Availability of data and materials

The sequence data for the transcriptome were submitted to the EBI (<https://www.ebi.ac.uk/ena/browser/view/PRJEB34947>) with a project number PRJEB34947 under the title “De novo transcriptome sequencing of branchiuran fish lice, *Argulus foliaceus* (Linnaeus, 1758) and *Argulus coregoni* Thorell, 1865”).

Ethics approval and consent to participate

The project was ethically approved by the University of Stirling Ethics Committee (Animal Welfare and Ethics Review Board; AWERB).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Fryer G. The parasitic Crustacea of African freshwater fishes; their biology and distribution. *Journal of Zoology*. 1968 Sep;156(1):45-95.
2. Kabata Z. Diseases of fishes. Book I. Crustacea as enemies of fishes. THF Publications, Neptune City, NJ; 1970.
3. Byrnes T. Two New *Argulus* Species (Branchiura: Argulidae) Found on Australian Bream (*Acanthopagurus* spp.). *Aust Zool*. 1985;21:579–86.
4. Møller O. Branchiura (Crustacea)—survey of historical literature and taxonomy. *Arthropod Syst Phylogeny*. 2009;67:41–56.
5. Rahman MM. Some aspects of the biology of a freshwater fish parasite *Argulus foliaceus* (L.) (Argulidae, Branchiura, Crustacea). *J Bangladesh J Zool*. 1995;23:77–86.
6. Sahoo PK, Mohanty J, Garnayak SK, Mohanty BR, Kar B, Prasanth H, Jena JK. Estimation of loss due to argulosis in carp culture ponds in India. *Indian J Fish*. 2013 Apr 1;60(2):99-102.
7. Walker P, Russon I, Haond C, Van Der Velde G, Wendelaar-Bonga S. Feeding in Adult *Argulus Japonicus* Thiele, 1900 (maxillopoda, Branchiura), an Ectoparasite on Fish. *Crustaceana*. 2011;84:307–18.

8. Forlenza M, Walker PD, de Vries BJ, Wendelaar Bonga SE, Wiegertjes GF. Transcriptional analysis of the common carp (*Cyprinus carpio* L.) immune response to the fish louse *Argulus japonicus* Thiele (Crustacea: Branchiura). *Fish Shellfish Immunol.* 2008;25:76–83.
9. Shimura S, Inoue K. toxic effects of extract from the mouthparts of the *Argulus coregoni* Thorell. (Crustacea:Branchiura) *Bull japSocScientFish.* 1984;50:729.
10. von Reumont B, Campbell L, Jenner R. Quo Vadis Venomics? A Roadmap to Neglected Venomous Invertebrates. *Toxins (Basel).* 2014;6:3488–551.
11. AmbuAli A, Monaghan SJ, McLean K, Inglis NF, Bekaert M, Wehner S, Bron JE. Identification of proteins from the secretory/excretory products (SEPs) of the branchiuran ectoparasite *Argulus foliaceus* (Linnaeus, 1758) reveals unique secreted proteins amongst haematophagous ecdysozoa. *Parasites & vectors.* 2020 Dec;13(1):1-3.
12. Hamilton S, McLean K, Monaghan SJ, McNair C, Inglis NF, McDonald H, Adams S, Richards R, Roy W, Smith P, Bron JE. Characterisation of proteins in excretory/secretory products collected from salmon lice, *Lepeophtheirus salmonis*. *Parasites & vectors.* 2018 Dec;11(1):1-9.
13. Singhal RN, Jeet S, Davies RW. The effects of argulosis-saprolehniasis on the growth and production of *Cyprinus carpoi*. *Hydrobiologia.* 1990 Aug;202(1):27-31.
14. Sahoo PK, Kar B, Mohapatra A, Mohanty J. De novo whole transcriptome analysis of the fish louse, *Argulus siamensis*: first molecular insights into characterization of Toll downstream signalling molecules of crustaceans. *Exp Parasitol.* Elsevier Inc.; 2013;135:629–41.
15. Bron JE, Frisch D, Goetze E, Johnson SC, Lee CE, Wyngaard GA. Observing copepods through a genomic lens. *Front Zool.* BioMed Central Ltd; 2011;8:22.
16. Christie AE. Peptide discovery in the ectoparasitic crustacean *Argulus siamensis*: Identification of the first neuropeptides from a member of the Branchiura. *Gen Comp Endocrinol.* Elsevier Inc.; 2014;204:114–25.
17. Pinnow P, Fabrizius A, Pick C, Burmester T. Identification and characterisation of hemocyanin of the fish louse *Argulus* (Crustacea: Branchiura). *J Comp Physiol B.* Springer Berlin Heidelberg; 2016;186:161–8.
18. AmbuAli A, Monaghan SJ, Al-Adawi K, Al-Kindi M, Bron JE. Histological and histochemical characterisation of glands associated with the feeding appendages of *Argulus foliaceus* (Linnaeus, 1758). *Parasitology international.* 2019 Apr 1;69:82-92.
19. Andrews S. FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>. 2010.
20. Bolger, A. M., Lohse, M., & Usadel B. Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics.* 2014; 1;30(15):2114-20.
21. Schmieder R and Edwards R. Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLoS One.* 2011;6.
22. 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, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C,

- Lindblad-Toh K, Friedman N RA. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat Biotechnol.* 2015;29:644–52.
23. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols.* 2013 Aug;8(8):1494-512.
 24. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215:403–10.
 25. Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K. & MTL. BLAST+: architecture and applications. *BMC Bioinformatics.* 2008;10.
 26. The UniProt Consortium UniProt. The UniProt Consortium UniProt. a hub protein Inf *Nucleic Acids Res.* 2015;43:D204–12.
 27. Eddy SR. Profile Hidden Markov Models. *Bioinformatics.* 1998;14:755–63.
 28. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA. The Pfam protein families database: towards a more sustainable future. *Nucleic acids research.* 2016 Jan 4;44(D1):D279-85.
 29. Wang Y, Coleman-Derr D, Chen G, Gu YQ. OrthoVenn: A web server for genome wide comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res.* 2015;43:W78–84.
 30. Fast MD, Burka JF, Johnson SC, Ross NW. Enzymes released from *Lepeophtheirus salmonis* in response to mucus from different salmonids. *J Parasitol.* 2003;89:7–13.
 31. Firth KJ, Johnson SC, Ross NW. Characterization of proteases in the skin mucus of Atlantic salmon (*Salmo salar*) infected with the salmon louse (*Lepeophtheirus salmonis*) and in whole-body louse homogenate. *J Parasitol.* 2000;86:1199–205.
 32. Ribeiro JMC, Francischetti IMB. Role of arthropod saliva in blood feeding: sialome and post-sialome perspectives. *Annu Rev Entomol.* 2003;48:73–88.
 33. Mulenga A, Kim T, Ibelli AMG. *Amblyomma americanum* tick saliva serine protease inhibitor 6 is a cross-class inhibitor of serine proteases and papain-like cysteine proteases that delays plasma clotting and inhibits platelet aggregation. *Insect Mol Biol.* 2013;22:306–19.
 34. Maldonado-Aguayo W, Chávez-Mardones J, Gonçalves AT, Gallardo-Escárate C. Cathepsin gene family reveals transcriptome patterns related to the infective stages of the salmon louse *Caligus rogercresseyi*. *PLoS One.* 2015;10:1–21.
 35. Kotsyfakis M, Sá-Nunes A, Francischetti IMB, Mather TN, Andersen JF, Ribeiro JMC. Antiinflammatory and immunosuppressive activity of sialostatin L, a salivary cystatin from the tick *Ixodes scapularis*. *J Biol Chem.* 2006;281:26298–307.
 36. Yamaji K, Tsuji N, Miyoshi T, Islam MK, Hatta T, Alim MA, et al. Hemoglobinase activity of a cysteine protease from the ixodid tick *Haemaphysalis longicornis*. *Parasitol Int. Elsevier Ireland Ltd;* 2009;58:232–7.

37. von Reumont BM, Blanke A, Richter S, Alvarez F, Bleidorn C, Jenner RA. The first venomous crustacean revealed by transcriptomics and functional morphology: remipede venom glands express a unique toxin cocktail dominated by enzymes and a neurotoxin. *Molecular biology and evolution*. 2014 Jan 1;31(1):48-58.
38. Rhoads ML, Fetterer RH. Developmentally regulated secretion of cathepsin L-like cysteine proteases by *Haemonchus contortus*. *J Parasitol*. 1995;81:505–12.
39. Francischetti IMB, Mather TN, Ribeiro JMC. Cloning of a salivary gland metalloprotease and characterization of gelatinase and fibrin(ogen)lytic activities in the saliva of the Lyme disease tick vector *Ixodes scapularis*. *Biochem Biophys Res Commun*. 2003;305:869–75.
40. Rawal R, Vijay S, Kadian K, Singh J, Pande V, Sharma A. Towards a Proteomic Catalogue and Differential Annotation of Salivary Gland Proteins in Blood Fed Malaria Vector *Anopheles culicifacies* by Mass Spectrometry. *PLoS One*. 2016;11:e0161870.
41. Chen N, Yuan ZG, Xu MJ, Zhou DH, Zhang XX, Zhang YZ, Wang XW, Yan C, Lin RQ, Zhu XQ. *Ascaris suum* enolase is a potential vaccine candidate against ascariasis. *Vaccine*. 2012 May 14;30(23):3478-82.
42. Pawlowic MC, Zhang K. *Leishmania* parasites possess a platelet-activating factor acetylhydrolase important for virulence. *Mol Biochem Parasitol* . Elsevier B.V.; 2012;186:11–20.
43. Champagne DE. Antihemostatic strategies of blood-feeding arthropods. *Curr Drug Targets Cardiovasc Haematol Disord*. 2004;4:375–96.
44. Chmelar J, Calvo E, Pedra JHF, Francischetti IMB, Kotsyfakis M. Tick salivary secretion as a source of antihemostatics. *J Proteomics* . Elsevier B.V.; 2012;75:3842–54.
45. Regier JC, Shultz JW, Zwick A, Hussey A, Ball B, Wetzer R, Martin JW, Cunningham CW. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature*. 2010 Feb;463(7284):1079-83.
46. Clark KF, Greenwood SJ. Next-Generation Sequencing and the Crustacean Immune System: The Need for Alternatives in Immune Gene Annotation. *Integr Comp Biol* . 2016;icw023.
47. Laurino S, Grossi G, Pucci P, Flagiello A, Bufo SA, Bianco G, Salvia R, Vinson SB, Vogel H, Falabella P. Identification of major *Toxoneuron nigriceps* venom proteins using an integrated transcriptomic/proteomic approach. *Insect biochemistry and molecular biology*. 2016 Sep 1;76:49-61.
48. Cooper MD, Alder MN. The evolution of adaptive immune systems. *Cell*. 2006;124:815–22.
49. Asgari S, Rivers DB. Venom Proteins from Endoparasitoid Wasps and Their Role in Host-Parasite Interactions. *Annu Rev Entomol*. 2011;56:313–35.
50. Moreau SJM, Guillot S. Advances and prospects on biosynthesis, structures and functions of venom proteins from parasitic wasps. *Insect Biochem Mol Biol*. 2005;35:1209–23.
51. Bai X, Rivera-Vega L, Mamidala P, Bonello P, Herms DA, Mittapalli O. Transcriptomic Signatures of Ash (*Fraxinus* spp.) Phloem. *Khodursky AB, editor. PLoS One* . 2011;6:e16368.

52. Morand S, Harvey PH. Mammalian metabolism, longevity and parasite species richness. *Proc R Soc London B*. 2000;267:1999–2003.
53. Valenzuela JG. High-throughput approaches to study salivary proteins and genes from vectors of disease. *Insect Biochem Mol Biol*. 2002;32:1199–209.
54. Charlab R, Valenzuela JG, Rowton ED, Ribeiro JM. Toward an understanding of the biochemical and pharmacological complexity of the saliva of a hematophagous sand fly *Lutzomyia longipalpis*. *Proc Natl Acad Sci U S A*. 1999;96:15155–60.
55. Valenzuela-Miranda D, Gallardo-Escárate C. *Caligus rogercresseyi* serine proteases: Transcriptomic analysis in response to delousing drugs treatments. *Aquaculture*. Elsevier; 2016 [cited 2016 Sep 22];465:65–77.
56. Valenzuela-Munoz V, Chavez-Mardones J, Gallardo-Escrate C. RNA-seq analysis evidences multiple gene responses in *Caligus rogercresseyi* exposed to the anti-salmon lice drug azamethiphos. *Aquaculture*. Elsevier B.V.; 2015;446:156–66.
57. Poley JD, Sutherland BJJ, Jones SRM, Koop BF, Fast MD. Sex-biased gene expression and sequence conservation in Atlantic and Pacific salmon lice (*Lepeophtheirus salmonis*). *BMC Genomics*. 2016;17:483.
58. Kvamme BO, Frost P, Nilsen F. The cloning and characterisation of full-length trypsins from the salmon louse *Lepeophtheirus salmonis*. *Molecular and biochemical parasitology*. 2004 Aug 1;136(2):303-8.
59. Johnson S, Ewart K, Osborne J, Delage D, Ross N, Murray H. Molecular cloning of trypsin cDNAs and trypsin gene expression in the salmon louse *Lepeophtheirus salmonis* (Copepoda: Caligidae). *Parasitol Res*. 2002;88:789–96.
60. Braden LM, Sutherland BJJ, Koop BF, Jones SRM. Enhanced transcriptomic responses in the Pacific salmon louse *Lepeophtheirus salmonis oncorhynchi* to the non-native Atlantic Salmon *Salmo salar* suggests increased parasite fitness. *BMC Genomics*. *BMC Genomics*; 2017;18:110.
61. Tirloni L, Kim TK, Coutinho ML, Ali A, Seixas A, Termignoni C, Mulenga A, da Silva Vaz Jr I. The putative role of *Rhipicephalus microplus* salivary serpins in the tick-host relationship. *Insect biochemistry and molecular biology*. 2016 Apr 1;71:12-28.
62. Rodriguez-Valle M, Xu T, Kurscheid S, Lew-Tabor AE. *Rhipicephalus microplus* serine protease inhibitor family: annotation, expression and functional characterisation assessment. *Parasit Vectors*. 2015;8:7.
63. Kim TK, Tirloni L, Radulovic Z, Lewis L, Bakshi M, Hill C, da Silva Vaz Jr I, Logullo C, Termignoni C, Mulenga A. Conserved *Amblyomma americanum* tick Serpin19, an inhibitor of blood clotting factors Xa and XIa, trypsin and plasmin, has anti-haemostatic functions. *International journal for parasitology*. 2015 Aug 1;45(9-10):613-27.
64. Mulenga A, Khumthong R, Blandon MA. Molecular and expression analysis of a family of the *Amblyomma americanum* tick Lospins. *J Exp Biol*. 2007;210:3188–98.

65. Muleng A, Sugino M, Nakajim M, Sugimoto C, Onuma M. Tick-Encoded serine proteinase inhibitors (serpins); potential target antigens for tick vaccine development. *J Vet Med Sci*. 2001;63:1063–9.
66. Ibelli AM, Kim TK, Hill CC, Lewis LA, Bakshi M, Miller S, Porter L, Mulenga A. A blood meal-induced *Ixodes scapularis* tick saliva serpin inhibits trypsin and thrombin, and interferes with platelet aggregation and blood clotting. *International journal for parasitology*. 2014 May 1;44(6):369-79.
67. Stark KR, James AA. Isolation and characterization of the gene encoding a novel factor Xa-directed anticoagulant from the yellow fever mosquito, *Aedes aegypti*. *J Biol Chem* . 1998;273:20802–9.
68. Malagón D, Díaz-López M, Benítez R, Adroher FJ. Cathepsin B- and L-like cysteine protease activities during the in vitro development of *Hysterothylacium aduncum* (Nematoda: Anisakidae), a worldwide fish parasite. *Parasitol Int* . Elsevier Ireland Ltd; 2010;59:89–92.
69. Franta Z, Sojka D, Frantova H, Dvorak J, Horn M, Srba J, Talacko P, Mares M, Schneider E, Craik CS, McKerrow JH. IrCL1–The haemoglobinolytic cathepsin L of the hard tick, *Ixodes ricinus*. *International journal for parasitology*. 2011 Oct 1;41(12):1253-62.
70. Eichner C, Frost P, Dysvik B, Jonassen I, Kristiansen B, Nilsen F. Salmon louse (*Lepeophtheirus salmonis*) transcriptomes during post molting maturation and egg production, revealed using EST-sequencing and microarray analysis. *BMC Genomics* . 2008;9:126.
71. McCarthy E, Cunningham E, Copley L, Jackson D, Johnston D, Dalton JP, Mulcahy G. Cathepsin L proteases of the parasitic copepod, *Lepeophtheirus salmonis*. *Aquaculture*. 2012 Aug 1;356:264-71.
72. Friedrich T, Kröger B, Bialojan S, Lemaire HG, Höffken HW, Reuschenbach P, Otte M, Dodt J. A Kazal-type inhibitor with thrombin specificity from *Rhodnius prolixus*. *Journal of Biological Chemistry*. 1993 Aug 5;268(22):16216-22.
73. Ribeiro JM, Marinotti O, Gonzales R. A salivary vasodilator in the blood-sucking bug, *Rhodnius prolixus*. *Br J Pharmacol*. 1990;101:932–6.
74. Ribeiro JM, Charlab R, Valenzuela JG. The salivary adenosine deaminase activity of the mosquitoes *Culex quinquefasciatus* and *Aedes aegypti*. *J Exp Biol* . 2001;204:2001–10.
75. McKerrow JH, Caffrey C, Kelly B, Loke P, Sajid M. Proteases in Parasitic Diseases. *Annu Rev Pathol Mech Dis* . 2006;1:497–536.
76. Na BK, Kim SH, Lee EG, Kim TS, Bae YA, Kang I, Yu JR, Sohn WM, Cho SY, Kong Y. Critical roles for excretory–secretory cysteine proteases during tissue invasion of *Paragonimus westermani* newly excysted metacercariae. *Cellular microbiology*. 2006 Jun;8(6):1034-46.
77. Francischetti I, Sa-Nunes A, Mans B, Santos I, Ribeiro J. The role of saliva in tick feeding. *Front Biosci*. 2009;14:2051–88.
78. Zhou G, Kohlhepp P, Geiser D, Frasquillo M del C, Vazquez-Moreno L, Winzerling JJ. Fate of blood meal iron in mosquitoes. *J Insect Physiol*. 2007;53:1169–78.
79. Hajdusek O, Almazán C, Loosova G, Villar M, Canales M, Grubhoffer L, Kopacek P, De la Fuente J. Characterization of ferritin 2 for the control of tick infestations. *Vaccine*. 2010 Apr 9;28(17):2993-8.

80. Xu XL, Cheng TY, Yang H, Liao ZH. De novo assembly and analysis of midgut transcriptome of *Haemaphysalis flava* and identification of genes involved in blood digestion, feeding and defending from pathogens. *Infect Genet Evol* . Elsevier B.V.; 2016;38:62–72.
81. Díaz-Martín V, Manzano-Román R, Oleaga A, Encinas-Grandes A, Pérez-Sánchez R. Cloning and characterization of a plasminogen-binding enolase from the saliva of the argasid tick *Ornithodoros moubata*. *Vet Parasitol*. Elsevier B.V.; 2013;191:301–14.
82. Gan W, Zhao G, Xu H, Wu W, Du W, Huang J, Yu X, Hu X. Reverse vaccinology approach identify an *Echinococcus granulosus* tegumental membrane protein enolase as vaccine candidate. *Parasitology research*. 2010 Mar;106(4):873-82.
83. Walter TC, Boxshall GA. World of Copepods Database. World Register of Marine Species.
84. Parizi LF, Githaka NW, Logullo C, Konnai S, Masuda A, Ohashi K, da Silva Vaz Jr I. The quest for a universal vaccine against ticks: cross-immunity insights. *The Veterinary Journal*. 2012 Nov 1;194(2):158-65.
85. Yang Y, Jun Wen Y, Cai YN, Vallée I, Boireau P, Liu MY, Cheng SP. Serine proteases of parasitic helminths. *The Korean journal of parasitology*. 2015 Feb;53(1):1.

Figures

The summary of *de novo* whole transcriptome analysis workflow of *A. coregoni* and *A. foliaceus*

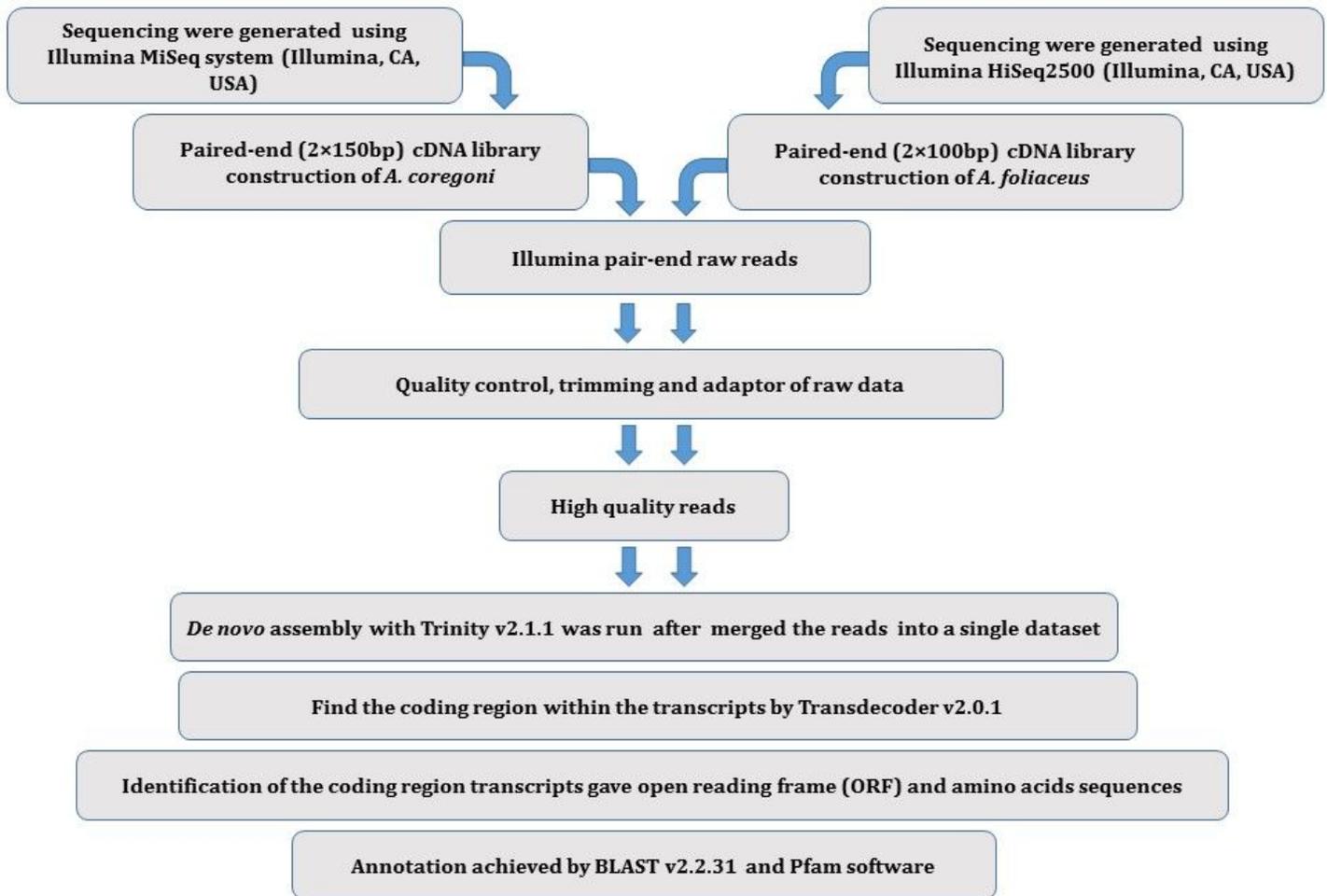


Figure 1

Workflow diagram for *de novo* transcriptome analysis of *A. coregoni* and *A. foliaceus*

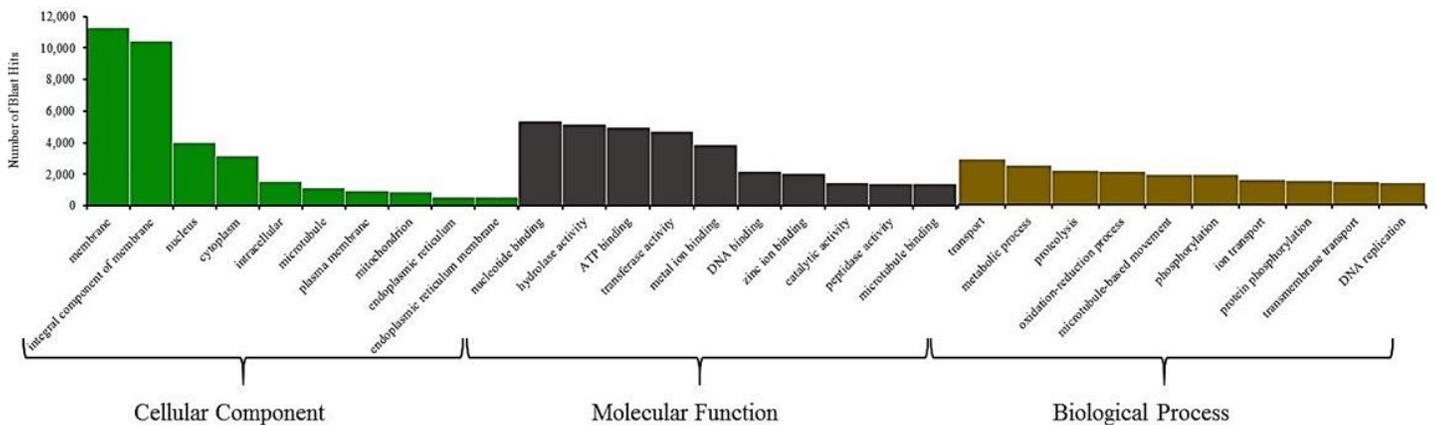


Figure 2

Distribution of GO terms from BLASTp hits with the transcriptome of *A. foliaceus* from this study.

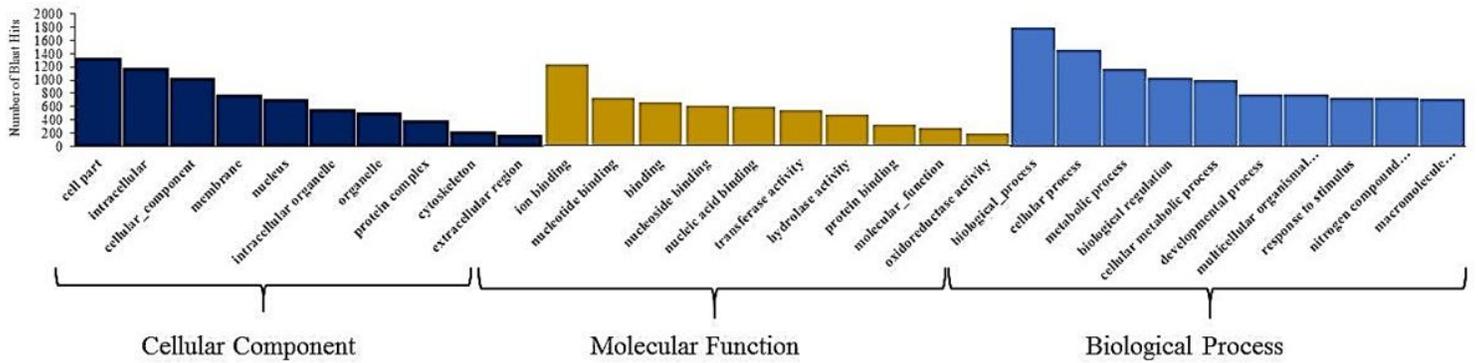


Figure 3

Distribution of GO terms of the first 10 clustered annotated proteins from the four transcriptome results of *Argulus* species.

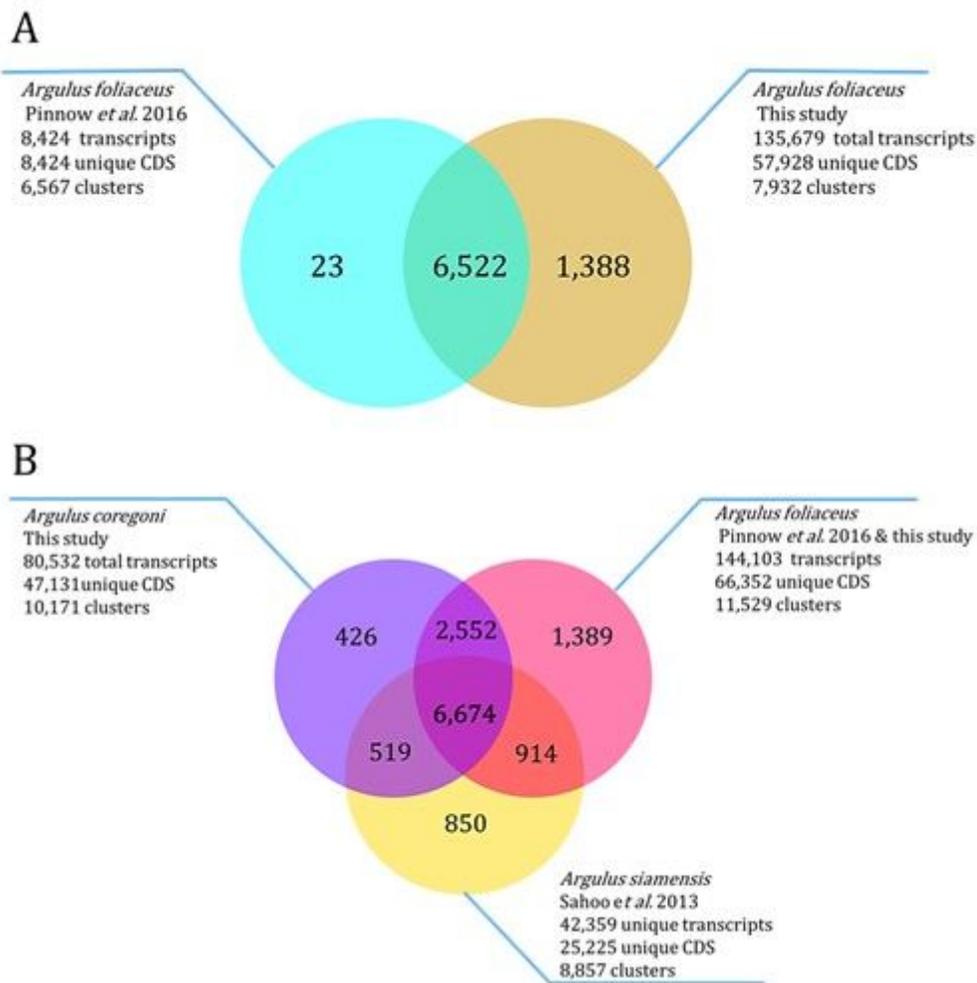


Figure 4

Venn diagrams showing number of orthologous clusters. (A) Two *A. foliaceus* transcriptomes, which shared 6,522 orthologous clusters and (B) transcriptomes from all three targeted species (*A. foliaceus*, *A.*

coregoni and *A. siamensis*). The three species display 13,324 clusters with 6,674 orthologous clusters shared between the species.

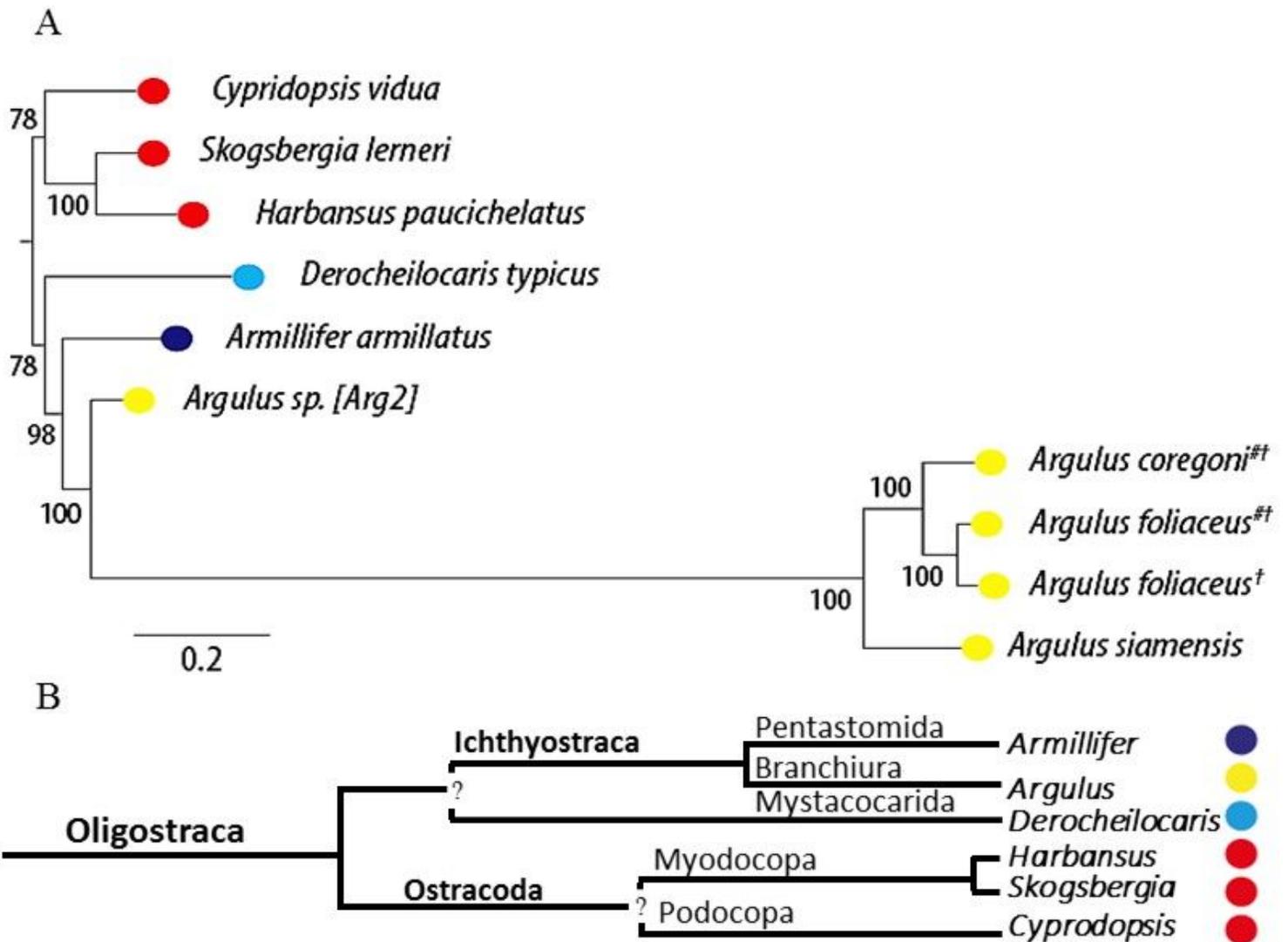


Figure 5

Phylogenetic tree showing the position of *Argulus* species from this study transcriptome (*A. foliaceus* and *A. coregoni*) and the previous transcriptome results (*A. foliaceus* and *A. siamensis*) relative to other members of the superclass Oligostraca. (A) Phylogenetic reconstruction based on the top 100 shared genes, including those generated from this study. (B) Oligostracan subtree, reproduced from Regier et al. [45]. Circle colour reflects subphylum. #† Species from this study. † from transcriptome reference [17].

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1TableS1.docx](#)