A novel approach to co-expression network analysis identifies modules and genes relevant for moulting and development in the Atlantic salmon louse (*Lepeophtheirius salmonis*)

Zhaoran Zhou¹, Christiane Eichner², Frank Nilsen², Inge Jonassen¹ and Michael Dondrup¹*

**Abstract**

**Background:** The salmon louse (*Lepeophtheirius salmonis*) is an obligate ectoparasitic copepod living on Atlantic salmon and other salmonids in the marine environment. Salmon lice cause a number of environmental problems and lead to large economical losses in aquaculture every year. In order to develop novel parasite control strategies, a better understanding of the mechanisms of moulting and development of the salmon louse at the transcriptional level is required.

**Methods:** Three weighted gene co-expression networks were constructed based on the pairwise correlations of salmon louse gene expression profiles at different life stages. Network-based approaches and gene annotation information were applied to identify genes that might be important for the moulting and development of the salmon louse. RNA interference was performed for validation. Regulatory impact factors were calculated for all the transcription factor genes by examining the changes in co-expression patterns between transcription factor genes and differentially expressed genes in middle stages and moulting stages.

**Results:** Eight gene modules were predicted as important, and 10 genes from six of the eight modules have been found to show observable phenotypes in RNA interference experiments. We knocked down five hub genes from three modules and observed phenotypic consequences in all experiments. In the infection trial, no copepods with a RAB1A-like gene knocked down were found on fish, while control samples developed to chalimus-1 larvae. Also, a FOXO-like transcription factor obtained highest scores in the regulatory impact factor calculation.

**Conclusions:** We propose a gene co-expression network-based approach to identify genes playing an important role in the moulting and development of salmon louse. The RNA interference experiments confirm the effectiveness of our approach and demonstrated the indispensable role of a RAB1A-like gene in the development of the salmon louse. We propose that our approach could be generalized to identify important genes associated with a phenotype of interest in other organisms.

**Keywords:** Gene co-expression network analysis, Regularized logistic regression, Differential co-expression analysis, Regulatory impact factor, Transcriptome, Salmon louse, Moulting, RAB1A

*Correspondence: michael.dondrup@uib.no
¹Department of Informatics & Sea Lice Research Centre, University of Bergen, Thormøhlens gate 55, 5008 Bergen, Norway
Full list of author information is available at the end of the article
Background

Copepods have been suggested as the most abundant animal group, with important roles in marine ecosystems [1, 2]. The salmon louse (Lepeophtheirus salmonis) is an ectoparasitic copepod on salmonids, with a life cycle that has eight developmental stages (instars) separated by moulting, consisting of two nauplius stages, one copepodid stage, two calamus stages, two preadult stages and the adult stage [3, 4]. Salmon lice are a major challenge to cage-based aquaculture of salmonids and cause large economical losses each year [5]. The emergence of salmon lice resistances against several drugs makes the situation even worse [6, 7]. Developing novel anti-parasitic mechanisms is thus an urgent and vital issue. To achieve this, we require a thorough understanding of the molecular mechanism of life stages development of the salmon louse. Identifying key genes that influence or regulate the lifespan of the parasite is also of great importance for finding novel drug targets against salmon lice.

Moulting or ecdysis, the shedding and replacement of the exoskeleton, plays a crucial role in the survival and development of arthropods and has been extensively studied in insects. Moulting consists of different events, including detachment of the old cuticle, synthesis of new cuticle, shedding of the old cuticle, hardening of the new cuticle and absorption of the old cuticle. Steroid hormones such as 20-hydroxyecdysone (20E) play a crucial role in arthropod ecdysis by regulating a series of pathways [8, 9]. Synthesis of 20E and other steroid hormones from cholesterol via ecdysone as direct precursor to 20E is mediated via a conserved pathway of enzymatic reactions. This pathway includes 7,8-dehydrogenase, encoded by the gene neverland (nvd) in Drosophila melanogaster followed by a cascade of cytochrome P450 mono-oxygenases, encoded by the so called Halloween genes: phantom (phm), disembodied (dib), shadow (sad), shade (shd), and spook (spo) [10–15].

The binding of moulting hormones to nuclear receptors leads to a complex hormonal cascade controlling moulting [9]. The ecdysone receptor (EcR) and the mammalian retinoid X receptor (RXR) with its insect homologue, Ultraspiricle (USP), belong to the group of nuclear receptors. RXR/USP and EcR form heterodimers which can bind to regulatory elements in the promoter of key genes involved in ecdysis and ontogenesis has only recently gained traction in Copepoda. In early studies, the ultrastructure of the cuticle of moulting salmon louse larvae was visualized by electron microscopy [21] and ecdysteroid levels were measured during a single instar in Calanus pacificus [22]. Transcriptional profiling of the last juvenile instar of Calanus finmarchicus identified genes with expression profiles changing significantly over the course of the moulting cycle [23].

Within the Copepoda, moulting occurs cyclically during development until the adult stage through a sequence of instars that is thought to be evolutionarily conserved within this taxonomic group [4, 24, 25]. Based on the annotated genome sequence of the Atlantic salmon louse [26] and expressed sequence tags [27], a limited number of ecdysis-related genes have been characterized. From the 20E biosynthetic pathway, homologous sequences of the insect genes neverland (nvd) and all the Helloween genes are present in the genome but thus far, only orthologues of neverland (nvd) disembodied (dib) and shade (shd) have been partially characterized [28, 29].

In the salmon louse, orthologous genes coding for the EcR/USP pair of nuclear receptors (LsEcR/LsRXR) have been characterized. Two genes were characterized by RNA interference (RNAi) mediated gene knock-down and by measuring ontogenic and tissue-specific expression [30–32]. Unlike in other crustaceans and insects, only a combined knock-down of LsEcR/LsRXR but not each gene individually, resulted in moulting arrest. Very recently, the nuclear receptor FushiTarazu Factor-1 (FTZ-F1) has been characterized in the salmon louse [33]. Two distinct transcript isoforms, aFTZ-F1 and βFTZ-F1, are expressed. Out of these, only the ablation of the most highly expressed isofrom βFTZ-F1 resulted in altered phenotypes of moulting arrest and oocyte maturation as well as significant differential regulation of genes associated with proteolysis and chitin binding.

Recently, genes from the conserved chitin-biosynthetic pathway have also been identified in the salmon louse genome [34, 35]. Like insects, the salmon louse genome contains two homologous genes for chitin synthase, LsCHS1 and LsCHS2. Knock-down of LsCHS1 resulted in a lethal phenotype with cuticle deformation and knock-down of LsCHS2 affected the digestive system [36]. In another study, five genes of the same pathway and three additional putative chitin deacetylases were targeted, also yielding ful abrogation of infectivity when targeting LsCHS1, fructose-6-phosphate aminotransferase (LsG-FAT) and a putative chitin deacetylase (CDA5956) [35]. From the chitin catabolic pathway, three gene coding for chitinases have been identified (LsChi1, LsChi2, LsChi4). Knock-down of LsChi2 in larval stages resulted in reduced infectivity [30, 31].
Recent investigations on the impact of chitin synthesis inhibitors, compounds belonging to the benzoylurea family (for example diflubenzuron, lufenuron, teflubenzuron) demonstrate the importance of chitin metabolism for parasite survival and as a target for pest management [37,38]. Still, there is only limited and often circumstantial knowledge of the molecular mechanisms driving developmental processes in copepods.

In recent years, high-throughput technologies have enabled us to study a large number of genes in parallel and thus facilitate the study of complex biological systems [39]. Being tremendously successful, high-throughput sequencing produces large volumes of data and has enabled a new era of genome research [40]. Our group has recently performed a comprehensive transcriptome time-series analysis using RNA sequencing data from three developmental stages of salmon lice (chalimus-1, chalimus-2 and preadult-1) [24] wherein we applied a method for improved developmental staging of samples by instar-age [41]. That way, we identified genes that may regulate development in this parasite.

A research area that is particularly important for systems biology is the study of dynamic interfaces and crosstalks between different processes and components of biological systems [42]. Recently, a great deal of attention has been devoted to the area of network-based analysis. Network analysis provides a powerful framework for studying a large number of interactions among biological processes and components. Gene co-expression networks (GCNs) have been widely used to capture and mine the interactions among components of the transcriptome [42,43].

Signatures of hierarchical modularity have been suggested to be present in all cellular networks investigated so far, ranging from metabolic to protein–protein interaction and regulatory networks [44]. In gene co-expression networks, modules are defined as groups of genes with similar expression patterns and can be identified by using clustering methods [45–47]. GCN modules have facilitated a better understanding of a number of biological phenomena [45, 48, 49], and an increasing number of studies based on GCN have been conducted to identify condition-specific gene modules and predict potential genes involved in a certain phenotype [50–53].

In this study, by re-analyzing the staged time-series data produced by Eichner et al. [24], we aim at providing a framework for identifying important genes through GCN analysis and contributing to a better understanding of the molecular mechanisms of moulting in copepods. By combining GCN analysis, sample traits and annotation information from public databases we identified relevant modules and hub genes and propose novel candidates with association to moulting and development. For validation, we performed gene knock-down by RNA interference (RNAi) of five genes.

**Methods**

**Gene expression data and genome annotation**

A normalized gene expression matrix was generated from the RNA-seq data provided by Eichner et al. [24], by extracting samples from middle instar ages and old/moulting instar ages of chalimus-1, chalimus-2 and preadult-1 larvae (Fig. 1). Transcripts with low expression (not having at least 3 cpm in at least 3 samples) were excluded from the analysis. In this manuscript we are using Ensembl Metazoa stable identifiers, consisting of a 13 digit numerical suffix, with prefixes EMLSAG or EMLSAT, to unanimously refer to predicted genes and transcripts, respectively, in the _L. salmonis salmonis_ genome annotation [26]. Gene annotation data were obtained from LiceBase [54].

**Identification of moulting-associated genes and transcription factor (TF) genes**

By combining data from the published literature and LiceBase, we collected genes which are involved in the moulting of salmon lice or known to be associated with the moulting of other arthropods with high confidence. We named these genes as “moulting-associated genes”.

Gene Ontology (GO) annotation information for the salmon louse genes was obtained as previously described [24]. Any salmon louse gene that was annotated by GO terms related to transcription factor (TF) (GO:0006351, GO:001071, GO:0008134, GO:0000988, and GO:0005667) or child-terms are annotated as TF genes.

**Gene co-expression network (GCN) analysis for identifying important modules and genes associated with moulting and development of salmon louse**

In this study, we define the modules and genes that might play a role in the regulation of moulting and development of salmon louse as “important modules” and “important genes”, and we proposed a workflow to identify these important modules and genes based on GCN analysis (Fig. 2). Using gene expression profiles, sample traits and gene annotation information as input, this workflow is used to predict the important modules and genes for moulting and development of salmon louse.

**GCN construction, module identification and module eigengene calculation**

**GCN construction and power parameter estimation**

GCNs were constructed using the R package WGCNA [55]. A modified version of the biweight midcorrelation (bicor) [56] was adopted to calculate the absolute correlation between pairwise genes (transcripts) ($S_{ij}$):
Fig. 1 Grouping of sample data and photographs of representative _L. salmonis_ chalimus-1, chalimus-2 and preadult-1 larvae. Within each stage, lice were divided into groups of same instar age: directly after moulting (young), in the middle of the stage (middle) and directly before the moulting to the next stage (old/moulting). Moults are represented by a green arrow and a shedded exoskeleton. In this study, data from lice of the middle and old/moulting instar age were used.

\[ S_{ij} = \left| \text{bicor}(x_i, x_j) \right|, \]

where \( x_i \) denotes the expression profile across all samples of transcript \( i \). The function bicor is implemented in the R package WGCNA.

By transforming the correlation by power function, we obtained the adjacency between pairwise transcripts \((A_{ij})\):

\[ A_{ij} = S_{ij}^\beta, \]

where \( \beta \) is the power parameter, and \( \beta \) is determined based on whether the corresponding co-expression network exhibits scale-free characteristics and has relatively high connectivities. We chose the suitable power parameter from integers ranging from 1 to 20 by plotting the signed scale-free topology fitting index \( R^2 \) against different power parameters, and we also plotted the corresponding network mean connectivity against different power parameters. Details about how the power parameter \( \beta \) was estimated can be found in Additional file 1.

With the adjacency matrix \( A \) we can construct the co-expression network, where each node represents a gene, and the weight possessed by edges between nodes indicates the co-expression relationship between nodes. Although our data is from a transcriptome study we use the terms “gene co-expression network” and “eigen-gene” because transcript quantification was done based on gene-level counts [24].

We constructed three GCNs, based on the gene expression profiles from middle samples, old/moult samples, and all samples (samples from both middle instar ages and old/moult instar ages).

**GCN module identification and eigengene calculation**

For each GCN, hierarchical clustering was performed for the nodes based on their adjacencies and a dendrogram was obtained. Using this dendrogram as input, a top-down algorithm cutreeDynamicTree was applied to identify gene modules. Each module was assigned a unique name as color. For each gene co-expression network, nodes that could not be assigned to any modules were moved to a module called “grey”. The grey module in each of the network was not considered in further analysis.

After identifying modules from each network, a sub-adjacency matrix can be extracted for all the gene members in each module. Then the eigengene for each module was computed as the eigenvector for the largest eigenvalue of the module gene expression matrix by the function moduleEigengenes in WGCNA.

**Intramodular centrality measurements and intramodular hub identification**

In this study, we adopted three types of centrality measurements to measure the centralities of nodes within each module and identified intramodular hubs.

**Intramodular connectivity (kIM)**

The connectivity of the \( i \)th node \( (k_i) \) in the weighted network is defined as the sum of connection weights between node \( i \) and the other nodes [57]:

\[ k_i = \sum_{j \neq i} A_{ij}. \]

Suppose that there are \( Q \) modules detected in a network, and they are labeled by \( q = 1, 2, \ldots, Q \), so the connectivity
Fig. 2 The workflow of gene co-expression network (GCN) analysis for identifying important modules and genes. Methods are highlighted in light-green.
of a node $i$ within a module $q$ is defined as intramodular connectivity ($k_i^{(q)}$ or $kIM_i^{(q)}$):

$$k_i^{(q)} = kIM_i^{(q)} = \sum_{j \in M_q, j \neq i} A_{ij}^{(q)}, \quad (4)$$

where $M_q$ denotes the set of node indices that correspond to the nodes in module $q$, and $A^{(q)}$ is the adjacency matrix of module $q$. High intramodular connectivity implies that a node could be a hub within the module.

**Module membership / module eigengene-based connectivity (kME)**

The module membership (or module eigengene-based connectivity) is defined as the value of correlation between module eigengene and the expression profile of the genes (or transcripts) assigned to this module [58]:

$$kME_i^{(q)} = \text{cor} \left( x_i^{(q)}, E^{(q)} \right), \quad (5)$$

where $x_i^{(q)}$ specifies the expression profile in different samples of transcript $i$ that is assigned to the module $q$, and $E^{(q)}$ denotes the eigengene of module $q$.

Since our gene co-expression networks were constructed based on the absolute correlation values between gene expression profiles, we used the absolute value of module membership to measure the centrality of each node within a module:

$$kMEall_i^{(q)} = \left| \text{cor} \left( x_i^{(q)}, E^{(q)} \right) \right|, \quad (6)$$

In addition, the module membership of a node for module $q$ can be calculated for all nodes in the network:

$$kMEall_i^{(q)} = \left| \text{cor} \left( x_i, E^{(q)} \right) \right|, \quad (7)$$

and this definition can be used in the module preservation analysis. The details can be found in Additional file 1.

**Intramodular weighted betweenness centrality (BC)**

The betweenness centrality of a node in an unweighted network (or module) is the number of shortest paths between all other nodes in the network that pass through the node [59]. To calculate the betweenness centralities of nodes in our weighted networks, a generalization of betweenness centrality proposed by Brandes [60] was employed. The approach is implemented in the R package tnet [61].

**Definition of intramodular hubs**

We evaluated the centralities of nodes in each module, using intramodular connectivity, absolute module membership and intramodular weighted betweenness centrality. The nodes ranking among the highest ten percent in any of the three centrality measurements of all nodes within a module were defined as intramodular hubs. The node obtaining highest scores in all of the three centrality measurements was defined as “absolute hub”.

Based on the ranks of nodes in three types of centrality measurement, we can calculate the average rank of nodes within each module. Therefore, the absolute hub should have an average rank as 1.

**Module preservation analysis**

The preservation of a module between the reference network and a test network can be evaluated based on the alterations in connectivity patterns and density. A well-preservation module in two or more networks should have similar connectivity patterns and nodes in the module should remain being tightly connected. WGCNA provides a series of approaches to evaluate whether a module is preserved and reproducible in another network [62]. In this study, module preservation statistics were computed to compare the two networks constructed based on middle samples and old/moulting samples. For each module preservation statistic, permutation tests were performed to evaluate the significance of the observed value and a Z score was obtained. The Z scores for all of the module preservation statistics were integrated as a composite summary statistic $Z_{summary}$. Details about how to calculate module preservation statistics and $Z_{summary}$ can be found in Additional file 1.

The networks were unsigned, and we set the number of permutation as 200. All the correlations were calculated using the biweight midcorrelation ($bicor$). Modules with a $Z_{summary}$ smaller than 2 was regarded as non-preserved, while a $Z_{summary}$ larger than 10 indicated that a module was well preserved across different networks. Since we aimed to identify modules playing a role in the regulation of moulting, the non-preserved modules from the moulting network were of particular interest.

**Regularized logistic regression using module eigengenes as independent variables**

We made use of the eigengenes of modules in the global network to perform logistic regression with an elastic-net penalty ($\alpha = 0.5$). This task was achieved by setting the binary dependent variable as the label of middle or old/moulting (old/moulting stages were labeled as 1), and using the eigengenes of each module as independent variables.

We used the R package glmnet [63] to perform this analysis, and we adopted the $\lambda$ that gives minimum mean cross-validated error.

**Integrating information from external databases and enrichment analysis**

Data from FlyBase [64, 65] and GenomeRNAi [66] were extracted and used to identify homologous observable phenotypes and lethal phenotypes enriched modules.
To detect homologous sequences in *D. melanogaster*, we ran BLASTP with E-value cutoff as 1e-10 on the corresponding protein sequences of salmon louse transcripts against protein sequences from *Drosophila*. Only best hits were considered. After mapping the protein IDs of the homologues from *Drosophila* to gene IDs, RNAi knock-down phenotype information were mapped to data from GenomeRNAi. If a salmon louse protein had more than one *Drosophila* homologue with identical maximum bitscore, all the homologues were used to search for RNAi phenotypes. BLASTP searches of all salmon louse predicted amino-acid sequences were performed to find paralogues.

**Enrichment analysis of modules**

Based on the GO annotation file for salmon louse genes from LiceBase, GO enrichment analyses were performed for each modules identified in the middle, moulting and global network using the fisher statistic and the “elim” algorithm provided by R package topGO [67].

Furthermore, with the information from the *Drosophila* homologues-based transcript-phenotype list (Additional file 2-Table S1), we conducted two enrichment analyses for each module identified in all networks. The *p*-values of these enrichment analyses were obtained based on hyper-geometric tests, to determine whether transcripts with homologue observable phenotypes or homologue lethal phenotypes in *Drosophila* were significantly enriched within a module. Based on the suggestions from [68], we used the raw *p*-values of our enrichment analyses, and the cutoff of *p*-values was set as 0.05 for all the enrichment analyses.

**Selecting important modules for further analyses**

We were interested in identifying gene modules which are likely to play a role in the moulting and development of salmon louse, and we chose important modules based on three analyses: the module preservation analysis, the regularized logistic regression analysis and the *Drosophila* homologues-based enrichment analysis. According to the guilt-by-association (GBA) heuristic [69], nodes in the moulting-associated transcripts-containing modules are more likely to play a role in the moulting and development of salmon louse, and we conducted a focused search among modules containing at least one known moulting-associated transcript. Therefore, moulting-associated transcripts-containing modules satisfying any of the following criteria were chosen for further studies: 1) non-preserved modules in the moulting network (*Z*~summary~ < 2); 2) the eigengenes of modules from the global network obtained positive coefficients from the regularized logistic regression analysis (the module with largest coefficient value should be prioritized); 3) modules that are significantly enriched by transcripts with observable and lethal RNAi phenotypes from homologues (*p*-value < 0.05) (Fig. 2).

**Selecting important genes as Knock-Down candidates from important modules**

Since many researchers have proposed that hubs in a biological network tend to be more important [70–72], we chose RNAi knock-down candidates among the hubs of the important modules. For each selected module, we gave prime consideration to the absolute hub. If no absolute hub was detected, knock-down candidates were chosen from other intramodular hubs. Hubs with less paralogues and little annotation information were then given priority.

**Differential gene expression (DGE) analysis and regulatory impact factor (RIF) calculation**

We calculated the regulatory impact factors (RIF) for all the transcripts annotated as TF, based on the metric proposed in [73].

The first step was to perform differential gene expression (DGE) analysis to compare the middle group and the old/moulting group, and the statistics used to test the null hypotheses were calculated based on standardized rank-sum Wilcoxon test. We computed the permutation adjusted *p*-values using the step-down maxT multiple testing procedures, which provide strong control of the family-wise Type I error rate (FWER). The functions are implemented in the R package multtest [74]. Transcripts with an adjusted *p*-value smaller than 0.05 were identified as differentially expressed between middle instar ages and old/moulting instar ages.

The first RIF value (*RIF*1) for *f*th TF transcript was defined as:

\[
RIF_{1f} = \frac{1}{n_{de}} \sum_{i=1}^{n_{de}} A_i D_{Ei} D_{Wi}^{2},
\]

where *n*~de~ is the number of DE transcripts; *A*~i~ represents the average expression of *i*th DE transcript across the two groups, and *D*~Ei~ is the statistics obtained from the previous DGE analysis. *D*~Wi~ is the abbreviation for differential wiring, which means the change of correlation between *f*th TF and the *i*th DE transcript across the two groups:

\[
D_{Wi} = r_{o_i} - r_{m_i},
\]

where *r*~o~ and *r*~m~ are the co-expression correlation between *f*th TF and the *i*th DE transcript in the old/moulting samples and the middle samples, respectively.

The second RIF value (*RIF*2) was computed as followed:

\[
RIF_{2f} = \frac{1}{n_{de}} \sum_{i=1}^{n_{de}} \left( EO_i^2 r_{oi}^2 - EM_i^2 r_{mi}^2 \right),
\]
where $EO_i^2$ and $EM_i^2$ denote the square of the average expression value of the $i$th DE transcript in the old/moulting samples and the middle samples; $ro_{f,i}^2$ and $rm_{f,i}^2$ are the square of the co-expression correlation between $f$th TF and the $i$th DE transcript in the old/moulting samples and the middle samples.

**RNA interference experiments**

**Double-stranded RNA synthesis**

Double-stranded RNA (dsRNA) was produced using MEGAscript® RNAi Kit (Ambion) according to supplier’s instructions using the primers given in Table 1 with the prefixed T7 sequence. The control group was treated with the non target control fragment CPY185 [75].

**RNA interference on nauplia and fish challenge with the emerging copepods**

RNA interference was conducted in nauplia as described in [30], but with 2 μg fragment. Infection was done in single tanks with 60 copepodids per fish on three fish for each fragment and for the control as described in [76] (10°C, full salinity). The remaining copepodids were stored on RNAlater® (Invitrogen) for later measurement of transcript down regulation. Sampling was done after 16 days when lice from control group were in chalimus or preadult-1 stage. All lice were sampled from fish and photographs were taken under the binocular in a drop of seawater with a cover slide on top. Number of lice and size measurements on photographs were recorded. Genes to be knocked down were: EMLSAG00000001458, EMLSAG00000003179 and EMLSAG00000005299. A second trial with knock-down larvae of EMLSAG00000001458 with 100 copepodids on three fish each was done. The outflow water from the tanks was filtered and lice in the flow out were counted two hours after infection and 24 hours after infection. One fish from knock-down and control group each were terminated three days after infection and lice were fixed on Karnovsky’s fixative for histological investigation. The other fish were terminated eight days after infection.

**RNA interference on preadult lice**

RNA interference in preadult-2 lice was done as described in [75]. In short, the fragment was injected into preadult-2 lice, which were put on fish again until most of the control lice showed its second pair of egg strings. All lice were sampled and photographed. The lice egg strings were laid into single flow through wells for hatching observation. Lice were either stored on RNAlater® for quantitative PCR (qPCR) measurements or on Karnovski’s reagent for histological investigation. Three different experiments were conducted, one with double-stranded RNA for EMLSAG00000001458, EMLSAG00000005299 and EMLSAG00000003179 as well as a control (35 days), one with dsRNA for EMLSAG00000004347 and a control (40 days) and one with dsRNA for EMLSAG00000008959 and a control (37 days). The cod trypsin RNA was used as a non-target control fragment [75].

**RNA extraction, cDNA synthesis and qPCR measurements**

Nauplia were divided into five (or four in case of control group from second infection trial) batches of 30 to 40 nauplia. RNA from nauplia was extracted by a combination of Trizol and RNAeasy micro kit as previously described [30]. RNA was frozen at -80°C until usage. cDNA synthesis was conducted using the AffinityScript QPCR cDNA Synthesis Kit (Agilent) according to suppliers recommendations. Gene expression of the target gene was measured by quantitative real-time PCR in control and knock-down group. qPCR was carried out in duplicates using the salmon louse elongation factor 1 α (LsEF1α) [77] as well as Adenine Nucleotide Translocator 3 (LsADT3) [76] (in nauplia only) as a standard. SYBR® Green PCR master mix was used with the primers given in Table 2. Applied Biosystems 7500 Fast Real-Time PCR system was used for thermal cycling and quantification in 10μl reactions (initiation 50°C, 2 min, 95°C, 2 min, then 40 cycles of 95°C, 15 seconds, 60°C, 1 min). A melting curve 60°C to 90°C was performed. Relative gene expression was calculated using the differences in threshold cycle (CT) between gene of interest and standard genes.

### Table 1 Primer sequences used in dsRNA production for RNA interference. The T7 sequence TAATACGACTCACTATAGGGAGA was prepended to the 5'-ends of all primers

| Gene ID         | Forward primer | Reverse primer |
|-----------------|----------------|----------------|
| EMLSAG00000001458 | CAAGCTGTTATTGATTGGCGATTC | CGGCATATTTAACTGATCAGCGTA |
| EMLSAG00000003179 | GCGTAAAAGTTGCGTACAATCTGA | GTTATGGGTTGATGATGAACTCAGA |
| EMLSAG00000005299 | GTATGATGACGGACATGCTCAAGG | GGCCTGTTTATAGGCGACCAAGC |
| EMLSAG00000004347 | AAACGGCGCGAGGAGGTGAATA | GGTGGGTTTCTTTCGTGCTGTT |
| EMLSAG00000008959 | GTATGATGACGGACATGCTCAAGG | GGCCTGTTTATAGGCGACCAAGC |
| CPY185 (control) | TCACTCAACCCTACCAGTAGTTGA | AAAGGAAACAAGCTGAGGAC |
Table 2 Primers used for qPCR measurements

| Gene ID                   | Forward primer | Reverse primer |
|---------------------------|----------------|----------------|
| EMLSA00000001458          | CAAGCTTCTTGTGGAAACAAATG | GCTGCCATAGTCAAAGCTTTTC |
| EMLSA00000003179          | GGTTCCGATTTCACACACCA | CGTGGAGGATAGGCAAAACTTTC |
| EMLSA00000004299          | GTTCTCGGAGAAAATAATGTTGCG | TGTTCAGTGATTTCCAGTTCC |
| EMLSA00000005299          | CCATCAATTTCCAAGTGGAGGA | CCTTCGGAGGGATAGGCAAAACTTTC |
| LsEF1α                    | GGTCGACAGACGTACTGGTAAATCC | TGCGGCCTTGAGTGGTGGTTC |
| LsADT3                    | CTGGAGAGGGAATTCCAAGCTTG | GACCCCTGGACACCCGTCAAGTCTTC |

Histology
Lice were prepared as described in [78] for histological investigation. For plastic embedding, lice were washed twice in phosphate buffered saline, dehydrated in a graded ethanol series, pre-infiltrated with Technovit/ethanol (50/50) for four hours (Technovit 7100, Heraeus Kulzer Technique) and infiltrated with Technovit and hardener overnight. Two micrometre thick sections were cut with a microtome (Leica RM 2165) and stained with toluidine blue (1% in 2% borax) for one minute. The stained sections were mounted using Mountex (Histolab Products).

Results
Identification of moulting-associated and TF genes
Among the transcripts in our RNA-seq data, we found 40 moulting-associated transcripts and 32 of them were retained after low expression filtering. The list of moulting-associated transcripts and the relevant publications can be found in Table 3.

There were 433 transcripts annotated as TF, and 231 TF were retained after low expression filtering (Additional file 2-Table S8).

GCN construction and module identification
To detect the genes that might be involved in the regulation of moulting and development, we analyzed the RNA-seq data sampled from the middle instar ages and old/moulting instar ages of chalimus-1, chalimus-2 and preadult-1 larvae (Fig. 1). Of the 45 samples, 18 samples were from middle instar ages, and 27 samples were from old/moulting instar ages. After filtering transcripts with low expresssions, 7108 transcripts were retained for network analysis.

Three GCNs were constructed for different sample groups. The first GCN was generated using all samples labeled as “middle instar age”. Meanwhile, an “old/moulting instar age” GCN was created for all old/moulting samples. To further facilitate our analysis, a GCN based on all samples from middle and old/moulting instar ages was built. The three GCNs were thus denoted as middle network, moulting network, and global network, respectively.

We set the power parameter β = 7 to make sure that the networks satisfied scale-free topology approximately while having relatively high mean connectivities (Fig. 3). The adjacency matrices of the three networks can be found in Additional file 7.

In the module identification process, 83, 60 and 78 modules were found in the middle, moulting and global network, respectively, and the module sizes ranged from 32 to 333. There were 203, 444 and 506 genes assigned to the grey module of the middle, moulting and global network, respectively (Additional file 2-Table S2-S4). Genes in the grey modules were not considered for further analysis. Dendrogram and module identification results for the three networks were displayed in Additional file 3-Figure S1. Eigengenes for each module in all the three networks were also obtained.

The centralities and distribution of moulting-associated genes across modules
To preliminarily examine the essentiality of intramodular hubs, we calculated the centralities for the 32 moulting-associated transcripts. Based on our definition of intramodular hubs, there were 6, 12 and 8 moulting-associated transcripts identified as intramodular hubs in the middle, moulting and global network, respectively (Additional file 2-Table S5-S7). The transcript EMLSAT00000005083 (annotated to encode hexokinase type 2, Table 1) was identified as intramodular hub in all the three networks, and it was the absolute hub in the module “lightcyan1” of the global network.

We examined how the 32 moulting-associated transcripts were distributed across modules in the three networks. There were 25, 20 and 24 modules containing moulting-associated transcripts in the middle, moulting and global network, accounting for 30.1%, 33.3% and 30.8% in the three networks, respectively. The numbers of moulting-associated transcripts in these modules ranged from 1 to 4 (Additional file 2-Table S2-S4).
Table 3 ID, annotation and relevant publications for the moult-associated transcripts

| Transcript ID       | Annotation from LiceBase                                      | Publication |
|---------------------|-------------------------------------------------------------|-------------|
| EMLSAT000000012651  | Nuclear receptor subfamily I group I member 2               | [32]        |
| EMLSAT00000004055   | Phosphoacetylglucosaminemutase                             | [122]       |
| EMLSAT00000002853   | Chitin synthase 8                                           | [38]        |
| EMLSAT00000007308   | Chitin synthase 8                                           | [35]        |
| EMLSAT000000012864  | Probable glucosamine 6-phosphate N-acetyltransferase        | [123]       |
| EMLSAT00000000683   | Glutamine–fructose-6-phosphate aminotransferase [isomerizing] 2 | [124]       |
| EMLSAT00000008812   | Acidic mammalian chitinase                                  | [38]        |
| EMLSAT00000005289   | Probable chitinase                                          | [38]        |
| EMLSAT00000008235   | Acidic mammalian chitinase                                  | [38]        |
| EMLSAT00000005464   | Chitooligosaccharidolytic beta-N-acetylglucosaminidase      | [125]       |
| EMLSAT00000008487   | Beta-hexosaminidase subunit alpha                           | [126]       |
| EMLSAT00000000150   | N-acetylglucosamine-6-phosphate deacetylase                 | [37]        |
| EMLSAT00000001322   | Glucosamine-6-phosphate isomerase                           | [37]        |
| EMLSAT00000004643   | Hexokinase type 2                                           | [127]       |
| EMLSAT00000005083   | Hexokinase type 2                                           | [127]       |
| EMLSAT0000001264    | Hexokinase type 2                                           | [127]       |
| EMLSAT00000008931   | Glucose-6-phosphate isomerase                               | [128]       |
| EMLSAT00000003220   | Mannose-6-phosphate isomerase                               | [129]       |
| EMLSAT00000007185   | Phosphoglucomutase-2                                        | [129]       |
| EMLSAT00000011599   | Phosphoglucomutase-1                                        | [129]       |
| EMLSAT00000001990   | TMEM47                                                      | [130]       |
| EMLSAT00000001904   | Ecdysone-induced protein 78C                                | [131]       |
| EMLSAT0000000733    | Probable nuclear hormone receptor HR3                      | [132]       |
| EMLSAT00000008543   | Hormone receptor 4                                          | [132]       |
| EMLSAT00000008902   | Nuclear hormone receptor FTZ-F1                            | [133]       |
| EMLSAT00000008651   | Ecdysone-induced protein 74EF isoform A                     | [32]        |
| EMLSAT000000001888  | Zinc finger protein 142                                     | [134]       |
| EMLSAT00000009170   | Growth hormone secretagogue receptor type 1                 | [30]        |
| EMLSAT00000008881   | Cathepsin L                                                 | [135]       |
| EMLSAT0000001278    | Ion transport peptide                                        | [136]       |
| EMLSAT00000008115   | CHH-like protein                                             | [137]       |
| EMLSAT0000001150    | Cytochrome P450 307a1                                       | [88]        |

**Module preservation analysis**

To identify genes which may play a role in the moult of salmon lice, we detected non-preserved modules from the moult network based on module preservation analysis. Five modules from the moult network were found as non-preserved, and the module sizes ranged from 41 to 100 (Additional file 2-Table S3). Strong correlations among genes in these non-preserved modules were only observed in the moult network, and two non-preserved modules (yellowgreen and lavenderblush3) contained moult-associated transcripts (EMLSAT00000008812 and EMLSAT0000012651) (Additional file 2-Table S6). Notably, the moult-associated transcripts were also identified as intramodular hubs in these modules. The transcript EMLSAT00000008812 (annotated to encode chitinase, Table 1) was ranked eighth (based on connectivity) in the yellowgreen module, and the transcript EMLSAT0000012651 (annotated as EcR, Table 1) was ranked third (based on betweenness centrality) in the lavenderblush3 module. We thus hypothesized that transcripts in these two modules could be important for salmon louse moultting, and hubs from these modules should be considered as important.

Eight modules from the middle network were identified as non-preserved, and the module sizes ranged from 54 to 109 (Additional file 2-Table S2). Three non-
preserved modules (darkseagreen4, brown4 and lightcyan1) were found containing one moulting-associated transcript (Additional file 2-Table S5). However, none of these moulting-associated transcripts were intramodular hubs in the middle network.

**Regularized logistic regression analysis on the global co-expression network**

To compare the intramodular overall gene expression patterns between the middle sample group and old/moulting sample group, we performed elastic net regularization-based logistic regression using the eigengenes of module from the global network as independent variables. As a result, we found modules with eigengenes that were highly expressed in one sample group but lowly expressed in the other sample group. From the 78 module eigengenes, we identified 15 eigengenes with non-zero coefficient (ranging from -1.75 to 0.963), and six of the 15 corresponding modules contained one known moulting-associated transcripts (Additional file 2-Table S4). It was noteworthy that module steelblue possessed the largest positive coefficient and contained one moulting-associated transcript (EMLSAT00000001150) as intramodular hub, which was ranked second in the betweenness centrality measurement (Additional file 2-Table S7).

When checking the absolute value of regression coefficients, three modules (magenta, lightcyan, and ivory) were found with absolute coefficient larger than 1. The moulting-associated transcripts were found in two of the three modules (lightcyan and ivory). The coefficients of all the three modules were negative, indicating that genes in these modules exhibited much higher expressions in middle samples. Notably, two modules (indianred4 and lavendar blush3) with negative regression coefficients contained moulting-associated transcripts (EMLSAT0000000733 and EMLSAT0000008543) annotated as hormone receptor 3 (Hr3) and hormone receptor 4 (Hr4) (Table 1). In the module indianred4, EMLSAT0000000733 was ranked eighth in the connectivity measurement.

Differentially expressed transcripts between middle group and old/moulting group were found in all the modules with non-zero regression coefficients, and the proportions ranged from 36.9% to 98.4% (Additional file 2-Table S4).

**Integrating information from external databases**

We identified homologous genes in *D. melanogaster* for the salmon louse transcripts and then searched for RNAi phenotypes for these homologues in the GenomeRNAi database. We found homologous RNAi phenotypes for 3059 salmon louse transcripts. Additional file 2-Table S1 maps salmon louse transcripts to the RNAi phenotypes of the corresponding homologues in *Drosophila*.

**Enrichment analysis of modules**

Based on the GO annotation file for the salmon louse transcripts, we performed GO enrichment analysis for all the modules to preliminarily elucidate the biological functions of the modules. The GO term with smallest p-value in each category (Biological Process(BP), Molecular Func-
ion (MF) and Cellular Component (CC)) were recorded (Additional file 2-Table S2-S4).

To further identify modules which are more likely to contain important genes for lice development, we conducted enrichment analyses for all the modules based on the homologues-based transcript-phenotype list. The transcripts with observable RNAi phenotypes were significantly enriched in 16, 13, and 14 modules in the middle, moulting and global network, accounting for 19.3%, 21.7%, and 17.9% in total modules, respectively. Analogously, 14, 14, and 9 modules were detected as enriched by transcripts with lethal RNAi phenotypes in the middle, moulting and global network, accounting for 16.9%, 23.3% and 11.5% in total modules. We found a relatively large overlap between the two enrichment analyses: 10, 11 and 7 modules (accounting for 12.0%, 18.3% and 9.0% in total modules) were identified as being significantly enriched by both observable and lethal RNAi phenotypes in the middle, moulting and global network (Additional file 2-Table S2-S4).

**DGE analysis and Rif calculation**

All 45 samples were divided into middle and old/moulting groups to find DE transcripts. There were 1364 transcripts differentially expressed (DE) between the two groups. The list of DE transcripts facilitated calculation of the RIF scores for all transcripts with GO annotation as TF.

For the 231 TF transcripts, RIF scores were computed (Additional file 2-Table S8). It is noteworthy that EMLSAT0000003844 (annotated as forhead box protein O (FOXO)) obtained highest RIF scores from both methods. This transcript is also an intramodular hub of a moulting-associated transcripts-containing module in both of the middle and moulting network.

**Selecting important modules for further analyses**

In the module preservation analysis, two modules (yellowgreen and lavenderblush3) from the moulting network were detected based on our criteria. In the regularized logistic regression analysis, two modules (steelblue and green) from the global network passed the criteria. In the homologues-based enrichment analysis, one (mediumpurple3), two (darkolivegreen and violet) and one (turquoise) module were found satisfying the criteria from the middle, moulting and global network, respectively. In summary, one, four and three modules from the middle, moulting and global network were selected for further analyses.

**Examining intramodular hubs and selecting important genes as Knock-Down candidates from important modules**

After determining important modules, we investigated the hubs of these modules to understand their roles in moulting and development of salmon louse. For each of the eight chosen modules, we examined their hub with highest average rank (Additional file 2-Table S12). The absolute hubs (EMLSAT0000003844 and EMLSAT0000001458) of the two non-preserved modules (lavenderblush3 and yellowgreen) selected from the moulting network are annotated as epithelial cell transforming 2 (ECT2) and Ras-related protein Rab-1A (RAB1A), respectively. For the four modules selected from the enrichment analysis, EMLSAT0000000929, EMLSAT0000005299, and EMLSAT0000012769 were identified as absolute hubs, annotated as high density lipoprotein-binding protein (HDLBP), ER membrane protein complex subunit 3 (EMC3) and laminin subunit beta-1 (LanB1). EMLSAT0000010555 had the highest average rank in the module (turquoise), from the global network, annotated as stress-induced-phosphoprotein 1 (STIP1). No absolute hubs were found in the two global modules selected from the regression analysis, and the nodes with highest average rank in these modules were EMLSAT0000007421 and EMLSAT0000012693, both of them were identified as differentially expressed transcripts between the middle group and old/moulting group. EMLSAT0000007421 was annotated as cuticular protein 62Bb (Cpr62Bb), and little annotation was found for the hub EMLSAT0000012693.

To validate the importance of genes in the selected modules in moulting and development of salmon louse, within the limited accesses to RNAi experiments, we selected RNAi knock-down candidates from three important modules. Since the important modules were selected based on three analyses, we selected one module from each of the three analyses. Firstly, we choose the module yellowgreen from the two non-preserved modules (the other one is lavenderblush3) in the moulting network. The module yellowgreen had larger size than the module lavenderblush3, and the absolute hub of the module yellowgreen had higher score of the absolute module membership. From the two moulting modules (darkolivegreen and violet) selected from the enrichment analysis, we chose the module violet for further analysis, since it contained more transcripts annotated as TF, and these transcripts obtained higher scores in the regulatory impact factor analysis than those found in the module darkolivegreen. Furthermore, with regards to the proportion of transcripts annotated as TF, the module violet and yellowgreen ranked as first and third among all modules in the moulting network. Finally, we selected the module steelblue in the global network for further analysis, because the eigengene of this module obtained largest coefficient in the regularized logistic regression analysis. Details of the three selected modules can be found in Table 4 and Additional file 2.
For the module yellowgreen and violet from the moult-
ing network, we chose the absolute hub for RNAi exper-
iment. According to the criteria discussed in the method
section, we chose another one hub without paralogues
from each module to knock down.

No absolute hub was found in the module steelblue from
the global network, and the hub (EMLSAT000000012111,
EMLSAT00000008158 and EMLSAT00000012113) were
annotated to encode proteins with the chitin
binding peritrophin-A domain (PF01607); (EMLSAT000000012111,
EMLSAT00000008158 and EMLSAT00000012113) were
annotated to encode proteins with the chitin
binding peritrophin-A domain (PF01607); four
(EMLSAT00000007421, EMLSAT00000007422, EML-
SAT00000009987, and EMLSAT00000012029) were
annotated to encode cuticle proteins (PF00379); one
(EMLSAT000000004870) was annotated to encode protein
with the polypropen synthetase domain (PF00348), and the
moult-associated transcript (EMLSAT00000001150) was
annotated to encode cytochrome P450 (PF00067)
(Additional file 2-Table S10, Table 1). Among the three
hubs with few annotation information, we chose one
(EMLSAT00000004347) with least number of paralogues
to knock down. The details of all the knock-down
candidates are summarized in Table 5.

RNA interference on nauplia and infection of salmon with
the emerging copepodids

Measurement of gene expression in copepodids before
infection showed down regulation of all targeted genes
test: p-value < 0.05) with varying knock-down
efficiency. For genes EMLSAG00000001458, EML-
SAT000000005299 and EMLSAG00000003179, efficiency
was 94%, 84% and 89%, respectively. At termination
after 16 days no lice were found on the fish infected
with copepodids from EMLSAG00000001458 knock-
down (KD) group (Fig. 4, left panel). There was no
significant difference in the number of lice between
the control group and EMLSAG000000005299-KD and
EMLSAG00000003179-KD groups and the develop-
ment of lice from all groups found on fish was similar
(Fig. 4, right panel). No difference in the phenotype
could be observed under the binocular or by size
measurements.

Second trial with eMLSA00000001458 Knock-Down

Since no lice with EMLSAG00000001458-KD were found
at termination after 16 days on the fish, we were inter-
ested in finding out whether this could be due to reduced
infection success or due to problems with development
and moult. A second infection trial for qualitative
measurement was done. Knock down efficiency measured in
copepodids before infection was 95%. After two hours,
30,37 and 33 lice were found in the filtered flow through
water of tanks from fish of the control group, and 32, 57
and 35 lice were found from tanks of the knock-down
group. After 24 hours, 9, 9 and 4 lice were found in the
flow out from control fish, and 9, 8 and 4 lice were found
from knock-down fish. No lice were found in the filters
after three days. At termination of the first fish at day three
after infection there were 10 lice on the control fish and
14 on the knock-down fish. These were sampled for histo-
logical investigation. No differences were observed in the
histological sections (Additional file 3-Figure S3). Eight
days after infection, lice had developed to chalimus-1 on
control fish (13 on one fish, 39 on the other), but no lice
were found on one of the fish with knock-down samples
and two copepodids on the other fish.

Knock-Down in preadult lice

At sampling lice were in the adult stage. Down regulation
was on average 77% for EMLSAG00000001458-KD group, 47% for EMLSAG00000003179-KD
and 68% for EMLSAG00000004347-KD group.
Lice from EMLSAG00000001458-KD group and
EMLSAG00000008959-KD group had no egg string.
Length measurements for body parameters (cephalotho-
rax and genital segment length) as well as egg strings
are shown in Table 6. Egg strings of all groups with egg
strings present hatched and produced viable normal
looking offspring. Histological sections were done for
EMLSAG00000001458-KD and EMLSAG00000004347-
KD lice. Histology of different tissues is shown in
Additional file 3-Figure S2. EMLSAG00000001458-KD
lice did not develop normal looking oocytes (Additional
file 3-Figure S2 f) and the ovaries did not contain any
oogonia (Additional file 3-Figure S2 j). The cellular
structures of the subcuticular tissue of the cephalotho-
rax were changed and only loose connection between

| Module       | Network      | Size | $Z_{summary}$ | Number of moult-associated genes | Number of TF genes | p-value | RNAi phenotype | RNAi lethal phenotype |
|--------------|--------------|------|---------------|----------------------------------|--------------------|---------|----------------|-----------------------|
| yellowgreen  | moultin      | 81   | 1.92          | 7                                | 0.84               | 0.54    |                |                       |
| violet       | moultin      | 83   | 8.52          | 10                               | 0.026              | 0.0076  |                |                       |
| steelblue    | global       | 79   | -             | 1                                | 0.90               | 0.94    |                |                       |

Table 4 Information on selected modules for selection of knock-down candidates
Table 5 Information of the selected knock-down candidate genes

| TranscriptID | Module      | Network | Number of paralogues within salmon louse genome | Differentially expressed | Annotated as TF gene | Rank | kIM-score | BC-score | kME-score | Average |
|--------------|-------------|---------|------------------------------------------------|--------------------------|---------------------|------|-----------|----------|-----------|---------|
| EMLSAT00000001458 | yellowgreen | moulting | 60 | False | False | 1 | 1 | 1 | 1 | 1 |
| EMLSAT00000008959 | yellowgreen | moulting | 0 | False | False | 6 | 19 | 7 | 10.7 |  |
| EMLSAT00000005299 | violet | moulting | 0 | False | False | 1 | 1 | 1 | 1 | 1 |
| EMLSAT00000003179 | violet | moulting | 0 | False | True | 18 | 3 | 17 | 12.7 |  |
| EMLSAT00000004347 | steelblue | global | 2 | True | False | 24 | 4 | 24 | 17.3 |  |
Fig. 4 Number of *L. salmonis* per fish (left panel) and distribution of stage and instar age (right panel). Lice were collected at termination of the knock down experiment, 16 days post infection, dsRNA was introduced in naupliar larvae. pad = preadult, ch = chalimus. Sample names represent dsRNA target transcripts; KD1458 : EMLSAT00000001458, KD5299 : EMLSAT00000005299, KD3179 : EMLSAT00000003179. Error bars represent +− 1 standard deviation, N=3.

Examining the RNA interference experiments data from liceBase
From LiceBase, RNAi experiments for 188 genes were collected, and 112 genes among them appeared in our three networks. 10 genes in six of the eight selected modules were found with observable RNAi phenotypes (including the RNAi experiments results from this study). One gene from the selected module darkolivegreen had been knocked down, but no phenotype was observed. No RNAi results were found for the genes in the selected module lavenderblush3 (Additional file 2-Table S13).

Notably, one hub (EMLSAG00000009839) from one non-preserved module (skyblue) of the moulting network show reduced survival in the RNAi experiments, although this module did not contain any known moulting-associated genes. The absolute hub (EMLSAG00000005382) of the module blue2 in the middle network show shorter eggstrings in RNAi experiments, and this module contained two known moulting-associated genes.

We also found RNAi experiment records of four genes (EMLSAG00000010968, EMLSAG00000006642, EMLSAG00000007048 and EMLSAG00000004159) which obtained highest average rank in four modules. These four modules did not satisfy any of the criteria for being an important module for salmon louse moulting and devel-

Table 6 Observed phenotypes by RNAi for the selected knock-down candidates

| TranscriptID | Gene annotation | Eggstring length | Cephalothorax length | Genital segment length | Other phenotypic changes |
|--------------|-----------------|------------------|----------------------|------------------------|--------------------------|
| EMLSAT00000001458 | Ras-related protein ORAB-1 | no eggstrings | no significant difference to control lice | significantly shorter (29%) than control lice | abrogated transition copepodid to chalimus-1* |
| EMLSAT00000008959 | Digestive organ expansion factor | no eggstrings | no significant difference to control lice | no significant difference to control lice | – |
| EMLSAT00000005299 | ER membrane protein complex subunit 3 | on AVG 20% shorter | no significant difference to control lice | slightly but significantly shorter (4%) than controls | – |
| EMLSAT00000003179 | Cold shock domain-containing protein C2 | no significant difference to control lice | slightly but significantly shorter (3%) than control lice | slightly but significantly shorter (7%) than control lice | – |
| EMLSAT00000004347 | Hypothetical protein | on AVG 18% shorter | slightly but significantly longer (2%) than control lice | slightly but significantly shorter (6%) than control lice | weak cuticle of GS (easily squeezed and torn apart) |

All phenotypes were assessed in adult female lice after injection of dsRNA at the pre-adult 2 stage except for *: phenotypic change was observed at larval stages after dsRNA treatment of nauplii (see Fig. 4)
opment, and the four genes did not show any observable phenotype in RNAi experiments (Additional file 2-Table S13).

Examining the modules where the rNAi candidates were selected
In the RNAi experiments, all five selected gene candidates show observable phenotypes, and we thus examined the three modules they were from. We plotted heatmaps of scaled gene expression profiles and barplot of scaled eigengene expression for each of the three module (Fig. 5).

Gene expression profiles within a module were strongly correlated. Genes in the module yellowgreen tended to be highly expressed in the chalimus-1 old female samples, chalimus-1 moulting samples and preadult-1 old female samples. Genes in the module violet were highly expressed in the chalimus-1 male samples and chalimus-1 moulting samples. For the module steelblue from the global network, genes were highly expressed in almost all samples from the old and moulting instar stages (except two preadult-1 old male samples), especially in the chalimus-1 moulting samples. The preadult lice with the two genes of the module yellowgreen knocked down failed to develop eggstrings. Further study is necessary to understand the role of genes from the module yellowgreen in the fecundity of female lice.

The topological graph for each of three modules (Fig. 6, Additional file 4–6) shows that moulting-associated genes in the module yellowgreen and steelblue obtained relatively high average ranks, and they were tightly connected with other hubs. In the module violet, the moulting-associated gene obtained a low average rank, but the genes annotated as TF obtained high average ranks and were tightly connected with other hubs. For these modules, the proportion of differentially expressed genes was highest in the module steelblue identified from the global network.

The enriched GO terms in the module yellowgreen included GO:0008152, GO:0001071 and GO:0005667, indicating nucleic acid transcription factor activity. The enriched GO terms in the module violet included GO:0006355, GO:0003677 and GO:0044454, indicating transcriptional regulation. In the module steelblue, the most enriched GO terms were GO:0042302, GO:0008061 and GO:0006030, indicating metabolic processes of chitin and cuticle.

Discussion
RNAi has enormously facilitated rapid and straightforward analysis of gene function for parasites and other organisms [79–81], and whole-genome RNAi screens have been successfully applied to detect genes with important functions for many biological processes in Caenorhabditis elegans and mammalian cultured cells [82–84]. Although a robust RNAi method for knock-down salmon louse genes has been established [81], genome-wide RNAi screening is both labour-intensive and time-consuming due to the parasitic lifestyle of salmon louse [3, 4]. Currently, biologists choose RNAi gene candidates subjectively based on their research interests, and little work has yet been carried out to develop bioinformatic methods for objectively predicting salmon louse genes that have a crucial role in biological processes of interest and are likely to show visible phenotypes when targeted in RNAi experiments.

In this study, we systematically analyzed the RNA-seq data of salmon lice from different life stages and proposed an approach (a workflow) for identifying important genes involved in the moulting and development of salmon louse (Fig. 2). Subsequently, RNAi experiments were performed on the genes identified by the network-based approach and gene annotation information. The results of our RNAi experiments and the RNAi records from LiceBase indicate the effectiveness of our approach.

The module preservation analysis allowed us to identify two important genes (EMLSAG00000001458 and EMLSAG00000008959 annotated as RAB1A and digestive organ expansion factor (DIEXF)), and both of the genes were from a non-preserved module (yellowgreen) in the moulting network. The non-preserved modules in the moulting network may be co-regulated and play an indispensable role in moulting or development of the salmon louse. Further studies are required to clarify the biological meaning of the non-preserved modules in the middle network as well as the well-preserved modules between the middle and moulting network.

In the regularized logistic regression analysis, all module eigengenes were calculated using the same method, thus they are on the same scale and it is feasible to identify the most important module by comparing the regression coefficients of eigengenes. We found that the module (steelblue) whose eigengene obtained largest coefficient was enriched for GO categories related to cuticle and chitin metabolic process. All the annotated hubs in this module are associated with chitin binding peritrophin-A domain, cuticle proteins, and cytochrome P450, which have been reported as important proteins for the moulting of arthropods [85–88]. We knocked down a hub (EMLSAG00000004347) with little annotation and observed both reduced fecundity and fragile cuticle. Based on RNAi results and the annotations of other hubs in this module, we speculate that gene EMLSAG00000004347 may participate in building the louse exoskeleton during the moulting process to adult stage. Our approach offers an effective solution in proposing and annotating novel putative genes that play a role in the moulting process of salmon louse. Although we focused on analyzing the modules containing moulting-associated genes due to the limited access to RNAi experiments, the module preservation analy-
Fig. 5 Heatmap of scaled gene expression profiles and barplot of scaled eigengene expression values for module yellowgreen extracted from the moulting network (a), module violet extracted from the moulting network (b), and module steelblue extracted from the global network (c). The heatmap color-codes gene expression values for each gene in a module: higher expression values are represented in red, and lower expression values are represented in greens according to the color legend. The barplot below each heatmap depicts the expression levels of module eigengenes in different samples (x-axis). The barplots are colored based on the name of each module.
Fig. 6 Graph representation of the modules yellowgreen extracted from the moulting network (a), violet extracted from the moulting network (b), and steelblue extracted from the global network (c). Nodes represent genes, edges represent correlation between nodes. For readability, only edges with absolute correlation > 0.84 are drawn. Nodes are labeled based on their average rank over three centrality statistics, node label font size is proportional to average nodes ranks. TF genes are colored in red, purple nodes represent moulting-associated genes from the literature. Differentially expressed genes between non-moulting and moulting groups are displayed as diamond shapes. Node diameter is proportional to weighted betweenness centrality. Annotation information of the top 10 genes based on rank is listed for each module.

**Rank Annotation**

(a) Ras-related protein ORAB-1
2 MAP kinase-interacting serine/threonine-protein kinase 1
3 Cytochrome P450 3A24
4 Wee1-like protein kinase
5 Kinesin-like protein KIF11
6 Programmed cell death protein 4
7 Homeobox protein TGIF2
8 Embryonic polarity protein dorsal
9 Probable splicing factor, arginine/serine-rich
10 NA

(b) ER membrane protein complex subunit 3
2 Negative elongation factor B
3 DNA-directed RNA polymerases I, II, and III subunit RPABC1
4 DNA-(apurinic or apyrimidinic site) lyase
5 Serine/threonine-protein kinase MRCK beta
6 Fibrillin-3
7 NA
8 Microtubule-actin cross-linking factor 1
9 Cold shock domain-containing protein CG9705
10 Dystonin

(c) Cuticle protein 7
2 Chitin binding Peritrophin-A domain
3 Decaprenyl-diphosphate synthase subunit 1
4 NA
5 Cuticle protein 7
6 Chitin binding Peritrophin-A domain
7 Cytochrome P450 307a1
8 Cuticle protein 7
9 Chitin binding Peritrophin-A domain
10 NA
sis and regularized logistic regression analysis identify important modules without taking any prior knowledge into account. These methods are suitable to analyze the expression data from less well-annotated organisms.

Instead of focusing on the moulting process directly, the emphasis of homologue-based enrichment analysis is on detecting important modules that are enriched for genes yielding observable phenotypic changes in another species. Four modules were identified in the first step. Besides the two RNAi experiments performed in this study, RNAi records were found for genes in each of the four modules. Strong RNAi phenotypes were observed on one and three genes in the module mediumpurple3 and module turquoise, respectively. Therefore, homologue-based phenotype enrichment analysis can contribute to rational selection of important modules, especially for studying less well-annotated organisms.

For scale free protein–protein interaction (PPI) networks, many groups have argued that highly connected hub nodes are more likely to be essential than sparsely connected nodes [89–91]. Although the underlying reason is in dispute [92], the centrality-lethality rule [70] has been widely accepted. A recent study on centrality in GCNs arrived at a similar conclusion [93]. Since virtually no PPI data are available for the salmon louse, we focussed on the essentiality of hubs in GCNs instead. Taking the topological characteristics of weighted GCNs into consideration, we used three different methods to identify intramodular hubs. In many cases, the hubs identified with these three measurements were coherent and complementary, enabling us to define absolute hubs. This not only had the advantage of evaluating the intramodular centrality of nodes from different angles, but also increases robustness of our approach. 17 of the 32 moulting-associated genes were detected as intramodular hubs in the three GCNs, and a hexokinase orthologue was found as absolute intramodular hub in the global network and intramodular hub in the other two networks. For the two modules (yellowgreen and violet) from which we chose two hubs in each to knockdown, we found that both of the two absolute hubs (EMLSA00000001458 and EMLSA00000005299) show stronger phenotypic consequences than the other middle-ranked hubs (EMLSA00000008959 and EMLSA00000003179). Interestingly, the absolute hub (EMLSAT00000005382) of another module containing moulting-associated transcripts has recently been identified as a novel intestinal heme scavenger receptor (LsHSCARB) with significant phenotypic effect on reproduction and body heme levels [94].

In addition to demonstrating the biological importance of intramodular hubs, RNAi experiments also highlight the role of our selected genes in moulting and development. Ablation of the RAB1A-like gene (EMLSA00000001458) resulted in reduced survival and fecundity. Human RAB1 regulates vesicle trafficking between the endoplasmatic reticulum and the Golgi complex [95, 96]. As a member of the Ras guanosine triphosphatase (GTPase) protein family, Ras-related protein RAB1 has important roles in many biological processes, such as vesicle formation, translocation and targeting, autophagy-mediated host defense, and secretion of interleukin-8 [97–99].

Small GTPases, including Ras GTPases, are versatile molecular switches that alternate between a GTP-bound and a GDP-bound conformation. The GTP-bound state is generally considered the active state in which the GTPase interacts with various effector proteins [100]. Few studies have been conducted on Rab GTPases in crustaceans but insect Rab proteins have a role in secretion of prothoracicotropic hormone (PTTH) [101, 102], an important regulator of ecystoideroidogenesis [103]. Recently, human RABIA has been found to be involved in receptor-mediated endocytosis (RME) by regulating motility of early endocytotic vesicles [104]. While the mechanism of RME has not been studied in the salmon louse, the RABIA-like gene may have a similar role in the uptake and trafficking of nutrients via receptors, such as LsHSCARB, which is itself an absolute hub and abundantly expressed in the intestinal epithelium of the parasite [94].

According to our module preservation analysis, another top-scoring hub is annotated as epithelial cell transforming 2 (ECT2) that is associated with GTPase activity [105–107]. Human ECT2 is a guanine nucleotide exchange factor that mediates the exchange of guanosine diphosphate for guanosine triphosphate and thereby activates RhGTPases [107]. ECT2, and its Drosophila orthologue Pebble (Pbl), have an essential function in cytokinesis and potentially cell polarity [108]. Loss-of-function and gain-of-function mutations in Pbl result in embryonic lethal phenotypes with altered cuticle structure [109]. Human ECT2 has been identified as an oncogene, its expression is highly elevated in different human tumors, and it promotes tumorigenesis in vivo and in vitro [110].

To our knowledge, the only other small GTPase-like gene targeted by RNAi in the salmon louse is Ras homolog Enriched in Brain (LsRheb) which is a part of the Target of Rapamycin (TOR) signaling pathway. The TOR pathway is important for regulating growth and development in many arthropods and ablation of LsRheb expression leads to complete abrogation of oocyte development [111]. Thus, we propose that small GTPases and interacting proteins are promising targets for understanding the metabolism and development of the salmon louse and other parasites and warrant inclusion in functional studies.

The top-scoring transcript (EMLSAT00000003849) by RIF analysis is an orthologue of the highly conserved TF FOXO. The importance of FOXO in metabolism, cellular
proliferation, stress tolerance and lifespan has long been recognized in the model organisms *D. melanogaster* and *C. elegans*, and mammals [112, 113]. TFs of the FOXO family are crucial regulators of cellular homeostasis that have a conserved role in modulating organismal aging and fitness [114]. More interestingly, several recent studies have demonstrated that FOXO-like TFs control growth and molting in insects [115–117]. A homologous gene in *D. melanogaster* (dFOXO) was reported to be involved in regulation of developmental timing through interaction with molting hormone ecdysone [118]. Combining our analysis results with these published papers, we propose that it is worth investigating whether the FOXO-like TFs have a crucial role in salmon louse development.

Our findings provide support for the importance of intramodular hubs in GCNs. On the other hand, when looking at all other public RNAi experiments in LiceBase [54], we discovered four additional hub genes that had been tested previously, all of which had highest average rank distributed across four different modules. However, these modules passed none of our criteria for module selection, and negative results had been recorded. We thereby conclude that not all the intramodular hubs may be equally important, even if they have highest ranks, supporting the need for an initial step of module selection. Combining our RNAi results and public records, we argue that our rational approach is more likely to yield genes with measurable phenotypic effect under ablation of gene expression than random selection. Nonetheless, more work is needed to affirm the relationship between centrality and gene essentiality in this organism.

Neither the essential *RAB1A*-like nor *FOXO*-like genes are detected as DE, indicating that DGE analysis might not always be the best choice when it comes to identify genes that play a key role in regulating a certain phenotype. In a standard DGE analysis, only single genes are taken into account disregarding possible correlations. On the other hand, some genes linked to a phenotype or disease are not differentially expressed across samples [73, 119], because mutations or post-translational modifications may alter coding potential and function without affecting expression levels [120]. A powerful advantage of network-based analysis is that it can reveal interactions across different groups of samples, even in case of high within-group variability. Furthermore, GCN-based analysis circumvents the multiple testing problem that plagues conventional differential gene expression analysis. Nevertheless, it should be noted that the dataset we analyzed contains only three replicates per condition and may therefore provide limited power to detect DE transcripts. Furthermore, our dataset did not allow for detection of variation in transcript isoforms, for example *αFTZ-F1* and *βFTZ-F1* [33] as only gene-level quantification was performed [24] and only very few transcript variants have been identified in this organism. On the other hand, we assume our method readily lends itself to co-expression of transcript-level quantification and could equally profit from increased replication.

In summary, our results support the hypothesis that GCN-based approaches are effective in identifying genes with association to a phenotype of interest. The widely accepted view that hubs in biological networks are more likely to be essential has for the first time been successfully tested in a marine parasite. Because of the high level of modularity, it was necessary to break down our rational approach of candidate selection by GCN into a two-step process with selecting interesting modules first. In our opinion, improving prioritization of genes is in strong demand in functional genomics of parasites. This is due to the fact that slow parasite growth as well as labor- and time-intensive handling and collection procedures often render genome-wide functional assays intractable in host-parasite systems. We therefore propose that our selection method may guide gene selection towards candidates with high probability of success in functional studies of salmon lice and other parasites. Prospectively, new multifactorial gene-expression data may also allow to transfer our approach to a broader range of phenotypes.

**Abbreviations**

- 20E: 20-hydroxyecdysone
- BC: Betweenness centrality
- BLAST: Basic local alignment search tool
- BP: Biological process
- CT: Threshold cycle
- CC: Cellular component
- DE: Differentially expressed
- DGE: Differential gene expression
- diRNA: double-stranded RNA
- DW: Differential wiring
- FWER: Family-wise type I error rate
- GCN: Gene co-expression network
- GO: Gene ontology
- KD: Knock down
- kM: Intramodular connectivity
- KWE: Module eigengene-based connectivity
- MF: Molecular function
- PPI: Protein-protein interaction
- qPCR: Quantitative PCR
- RIF: Regulatory impact factor
- RME: Receptor mediated endocytosis
- RNAi: RNA interference
- RNA-seq: RNA sequencing
- TF: Transcription factor

**Supplementary Information**

The online version contains supplementary material available at https://doi.org/10.1186/s12864-021-08054-7.

**Additional file 1:** The methods used to estimate the power parameter and calculate the module preservation statistics.

**Additional file 2:** The main analysis results of this study.

Table S1: Homologues-based transcript-phenotype list

Table S2-S3: Analysis results for the modules in the middle network and molting network, including module size, enriched GO terms with smallest *p*-values, *p*-values of enrichment analyses based on homologues, module preservation *Z*~summary~, number of known molting-associated genes, proportion of TF and DE genes.

Table S4: Analysis results for the modules in the global network, including module size, enriched GO terms with smallest *p*-values, *p*-values of enrichment analyses based on homologues, regularized logistic regression coefficients, number of known molting-associated genes, proportion of TF and DE genes.

Table S5-S7: Module assignment results for the known molting-associated transcripts in the middle, molting and global network; the ranks of these transcripts based on three types of centrality measurements within modules; whether they were intramodular hubs or not.

Table S8: Two types of RIF scores for the transcripts annotated as transcription factors.
Acknowledgements
We would like to thank Heidi Kongshaug, Lars Are Hamre and Per Gunnar Espedal for technical help in the laboratory, and we would like to thank Kjell Petersen for helpful comments and suggestions on the manuscript.

Authors’ contributions
ZZ performed bioinformatics analyses and drafted the initial manuscript. MD and CE designed the study, provided and analyzed data. U and FN contributed to the design of the study and provided funding. CE performed RNAi experiments and drafted the corresponding methods and results. ZZ, MD and CE edited the manuscript. All authors read, revised and approved the final manuscript.

Funding
This project was funded by the Research Council of Norway, SFI-Sea Lice Research Centre, grant number 203513/030. This work was funded by the ELIXIR2 (270068) infrastructure grant from the Research Council of Norway to MD.

Availability of data and materials
The datasets analyzed in this study and code are available from the corresponding author on reasonable request. Supplementary data have been made available on figshare [121].

Declarations

Ethics approval and consent to participate
All experiments were performed according to Norwegian animal welfare regulations with the approval of the governmental Norwegian Animal Research Authority (ID 7704, no 2010/245410).

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
1 Department of Informatics & Sea Lice Research Centre, University of Bergen, Thomahølsgata 55, 5008 Bergen, Norway. 2 Department of Biological Sciences & Sea Lice Research Centre, University of Bergen, Thomahølsgata 55, 5008 Bergen, Norway.

Received: 3 May 2021 Accepted: 4 October 2021 Published online: 18 November 2021

References
1. Martin JH. The Possible Transport of Trace Metals Via Moulting Copepod Exoskeletons. Limnol Oceanogr. 1970;15(5):756–61.
2. Turner JT. The importance of small planktonic copepods and their roles in pelagic marine food webs. Zool. Stud. 2004;4255–66.
3. Johnson SC, Albright LJ. The developmental stages of Lepeophtheirus salmonis (Kroyer, 1837) (Copepoda: Caligidae). Can J Zool. 1991;69(4):929–50. https://doi.org/10.1139/v91-138.
4. Hamre LA, Eichner C, Caipang CMA, Dalvin ST, Bron JE, Nilsen F, Boxshall G, Skern-Maurtzen R. The Salmon Louse Lepeophtheirus salmonis (Copepoda: Caligidae) Life Cycle Has Only Two Chalimus Stages. PLoS ONE. 2013;8(9). https://doi.org/10.1371/journal.pone.0073539.
5. Abolofia J, Asche F, Wilen JE. The Cost of Lice: Quantifying the Impacts of Parasitic Sea Lice on Farmed Salmon. Mar Resour Econ. 2017;32(3):329–49. https://doi.org/10.1086/691981.
6. Denholm I, Devine G, Horsberg TE, Sev paddal S, Fallang A, Nolan DV, Powell R. Analysis and management of resistance to chemotherapeutics in salmon lice, Lepeophtheirus salmonis (Copepoda: Caligidae). Pest Manag Sci. 2002;58(6):528–36. https://doi.org/10.1002/ps.482.
7. Aaen SM, Helgesen KO, Bakke MJ, Kaur K, Horsberg TE. Drug resistance in sea lice: a threat to salmonid aquaculture. Trends Parasitol. 2015;31(2):72–81. https://doi.org/10.1016/j.pt.2014.12.006.
8. Chang ES, Mykles DL. Regulation of crustacean molting: A review and our perspectives. Gen Comp Endocrinol. 2011;172(3):323–30. https://doi.org/10.1016/j.ygcen.2011.04.003.
9. Zitnan D, Kim Y-J, Zitnava O, Roller L, Adams ME. Complex steroid-peptide-receptor cascade controls insect ecysis. Gen Comp Endocrinol. 2007;153(1-3):88–96. https://doi.org/10.1016/j.ygcen.2007.04.002.
10. Yoshiyama T, Namiki T, Mita K, Kataoka H, Niwa R. Neverland is an evolutionally conserved Rieske-domain protein that is essential for ecdysone synthesis and insect growth. Development. 2006;133(13):2565–74. https://doi.org/10.1242/dev.024248.
11. Petryk A, Warren JT, Marqués G, Jarcho MP, Gilbert LI, Kahler J, Pary JP, Li Y, Dauphin-Villemant C, O’Connor MB. Shade is the Drosophila P450 enzyme that mediates the hydroxylation of ec dysone to the steroid insect molting hormone 20-hydroxyecdysone. Proc Natl Acad Sci U S A. 2003;100(SUPPL. 2):13773–78. https://doi.org/10.1073/PNAS.2336088100.
12. Niwa R, Matsuda T, Yoshiyama T, Namiki T, Mita K, Fujimoto Y, Kataoka H. CYP306A1, a cytochrome P450 enzyme, is essential for ecdysoid biosynthesis in the prothoracic glands of Bombbyx and Drosophila. J Biol Chem. 2004;279(24):35942–49. https://doi.org/10.1074/jbc.M404514200.
13. Warren JT, Petryk A, Marqués G, Jarcho M, Pary JP, Dauphin-Villemant C, O’Connor MB, Gilbert LI. Molecular and biochemical characterization of two P450 enzymes in the ecdysonoidogenic pathway of Drosophila melanogaster. Proc Natl Acad Sci U S A. 2002;99(17):11043–48. https://doi.org/10.1073/PNAS.162575799.
14. Warren JT, Petryk A, Marqués G, Pary JP, Shinoda T, Itoyama K, Kobayashi J, Jarcho M, Li Y, O’Connor MB, Dauphin-Villemant C, Gilbert LI. Phantom encodes the 25-hydroxylase of Drosophila melanogaster and Bombbyx mori. A P450 enzyme critical in ecdysone biosynthesis. Insect Biochem Mol Biol. 2004;34(9):991–1010. https://doi.org/10.1016/j.ibmb.2004.06.009.
15. Reutitz KE, Ryczynski R, Warren JT, Gilbert LI. The Halloween genes code for cytochrome P450 enzymes mediating synthesis of the insect molting hormone. Biochem Soc Trans. 2006;34(6):1256–60. https://doi.org/10.1042/BJT2011035.
16. Thomas HE, Stunnenberg HG, Stewart AF. Heterodimerization of the Drosophila ecysone receptor with retinoid X receptor and ultraspire. Nature. 1993;362(6419):471–75. https://doi.org/10.1038/362471a0.
17. Hall BL, Thummel CS. The RXR homolog ultraspire is an essential component of the Drosophila ecysone receptor. Development. 1998;125(3):4709–17. https://doi.org/10.1242/dev.125.23.4709.
18. Barchuk AR, Figueiredo VLC, Simões ZLP. Downregulation of ultraspire gene expression delays pupal development in honeybees. J Insect Physiol. 2008;54(6):1035–40. https://doi.org/10.1016/j.jinphys.2008.04.006.
19. Xu J, Tan A, Pali SR. The function of nuclear receptors in regulation of female reproduction and embryogenesis in the red flour beetle,
28. Sandlund L, Kongshaug H, Horsberg TE, Male R, Nilsen F, Dalvin S. Chitin biosynthesis genes in the salmon louse (Lepeophtheirus salmonis) (copepoda: Caligidae). Contrib Zool. 2000;69(1-2):29–49. https://doi.org/10.2246/CZ.1837.
27. Eichner C, Frost P, Dysvik B, Jonassen I, Kristiansen B, Nilsen F. Effects of chitin synthesis inhibitor treatment on Lepeophtheirus salmonis (Copepoda, Caligidae) larvae. PLoS ONE. 2019;14(9):e0225220.
26. Schram TA. Supplementary description of the developmental stages of the salmon louse (Lepeophtheirus salmonis). J Fish Dis. 2018;41(6):1005–29. https://doi.org/10.1111/jfd.12770.
25. Eichner C, Dondrup M, Nilsen F. RNA sequencing reveals distinct gene expression patterns during the development of parasitic larval stages of the salmon louse (Lepeophtheirus salmonis). J Fish Dis. 2018;41(6):1005–29. https://doi.org/10.1111/jfd.12770.
24. Bron JE, Shinn AP, Sommerville C. Ultrastructure of the cuticle of the chalimus larva of the salmon louse (Lepeophtheirus salmonis) (Kroyer, 1837) (Copepoda: Caligidae). Contrib Zool. 2000;69(1-2):39–49. https://doi.org/10.2246/CZ.1837.
23. Tarrant AM, Baumgartner MF, Hansen BH, Altin D, Nordtug T, Olsen AJ. Transcriptional profile of reproductive development, lipid storage and molting throughout the last juvenile stage of the marine copepod Calanus finmarchicus. Front Zool. 2014;11:91. https://doi.org/10.1186/s12983-014-0091-8.
22. Johnson C. EcdysteroidsintheoceaniccopepodCalanuspacificusð. 1993;30–47. https://doi.org/10.1037/10166-015-1277-y.
21. Bron JE, Shinn AP, Sommerville C. Ultrastructure of the cuticle of the chalimus larva of the salmon louse (Lepeophtheirus salmonis) (Kroyer, 1837) (Copepoda: Caligidae). Contrib Zool. 2000;69(1-2):39–49. https://doi.org/10.2246/CZ.1837.
20. Merzendorfer H, Zimoch L. Chitinmetabolismininsects:structure,characterizationandgeneexpressionprofileduringsynthesisofanewco-factorforreproduction,molting,andeugeronickinthesalmonlouse(L. salmonis). Mol Biochem Parasitol. 2016;208(2):65–73. https://doi.org/10.1016/j.molbiopara.2016.06.007.
19. Eichner C, Hamre LA, Nilsen F. Instar growth and molt increments in Lepeophtheirus salmonis (Copepoda: Caligidae) chalimus larvae. Parasitol Int. 2015;64(1):86–96. https://doi.org/10.1016/j.parint.2014.10.006.
18. Park ST, Kim J. Trends in Next-Generation Sequencing and a New Era for Whole Genome Sequencing. Int Neutrojoul. 2016;20(Suppl 2):76–83. https://doi.org/10.5213/inj.1632742.371.
17. Park ST, Kim J. Trends in Next-Generation Sequencing and a New Era for Whole Genome Sequencing. Int Neutrojoul. 2016;20(Suppl 2):76–83. https://doi.org/10.5213/inj.1632742.371.
16. Saelens W, Cannoodt R, Saeyes Y. A comprehensive evaluation of module detection methods for gene expression data. Nat Commun. 2019;10(1):1–15. https://doi.org/10.1038/s41467-019-1184-x.
15. Oldham MC, Horvath S, Geschwind DH. Conservation and evolution of co-expression networks in human and chimpanzee brains. Proc Natl Acad Sci. 2006;103(47):17973–78. https://doi.org/10.1073/pnas.0605938103.
14. Oldham MC, Horvath S, Geschwind DH. Conservation and evolution of co-expression networks in human and chimpanzee brains. Proc Natl Acad Sci. 2006;103(47):17973–78. https://doi.org/10.1073/pnas.0605938103.
13. Oldham MC, Horvath S, Geschwind DH. Conservation and evolution of co-expression networks in human and chimpanzee brains. Proc Natl Acad Sci. 2006;103(47):17973–78. https://doi.org/10.1073/pnas.0605938103.
12. Oldham MC, Horvath S, Geschwind DH. Conservation and evolution of co-expression networks in human and chimpanzee brains. Proc Natl Acad Sci. 2006;103(47):17973–78. https://doi.org/10.1073/pnas.0605938103.
11. Oldham MC, Horvath S, Geschwind DH. Conservation and evolution of co-expression networks in human and chimpanzee brains. Proc Natl Acad Sci. 2006;103(47):17973–78. https://doi.org/10.1073/pnas.0605938103.
10. Oldham MC, Horvath S, Geschwind DH. Conservation and evolution of co-expression networks in human and chimpanzee brains. Proc Natl Acad Sci. 2006;103(47):17973–78. https://doi.org/10.1073/pnas.0605938103.
9. Oldham MC, Horvath S, Geschwind DH. Conservation and evolution of co-expression networks in human and chimpanzee brains. Proc Natl Acad Sci. 2006;103(47):17973–78. https://doi.org/10.1073/pnas.0605938103.
8. Oldham MC, Horvath S, Geschwind DH. Conservation and evolution of co-expression networks in human and chimpanzee brains. Proc Natl Acad Sci. 2006;103(47):17973–78. https://doi.org/10.1073/pnas.0605938103.
7. Oldham MC, Horvath S, Geschwind DH. Conservation and evolution of co-expression networks in human and chimpanzee brains. Proc Natl Acad Sci. 2006;103(47):17973–78. https://doi.org/10.1073/pnas.0605938103.
6. Oldham MC, Horvath S, Geschwind DH. Conservation and evolution of co-expression networks in human and chimpanzee brains. Proc Natl Acad Sci. 2006;103(47):17973–78. https://doi.org/10.1073/pnas.0605938103.
5. Oldham MC, Horvath S, Geschwind DH. Conservation and evolution of co-expression networks in human and chimpanzee brains. Proc Natl Acad Sci. 2006;103(47):17973–78. https://doi.org/10.1073/pnas.0605938103.
4. Oldham MC, Horvath S, Geschwind DH. Conservation and evolution of co-expression networks in human and chimpanzee brains. Proc Natl Acad Sci. 2006;103(47):17973–78. https://doi.org/10.1073/pnas.0605938103.
3. Oldham MC, Horvath S, Geschwind DH. Conservation and evolution of co-expression networks in human and chimpanzee brains. Proc Natl Acad Sci. 2006;103(47):17973–78. https://doi.org/10.1073/pnas.0605938103.
2. Oldham MC, Horvath S, Geschwind DH. Conservation and evolution of co-expression networks in human and chimpanzee brains. Proc Natl Acad Sci. 2006;103(47):17973–78. https://doi.org/10.1073/pnas.0605938103.
1. Oldham MC, Horvath S, Geschwind DH. Conservation and evolution of co-expression networks in human and chimpanzee brains. Proc Natl Acad Sci. 2006;103(47):17973–78. https://doi.org/10.1073/pnas.0605938103.

pituitary and brain areas that affect estrous behavior in dairy cows. J Dairy Sci. 2013;96(4):2583–9. https://doi.org/10.3168/jds.2012-5814.

54. Dondrup M. LiceBase. https://licebase.org. Accessed 03 May 2021.

55. Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9. https://doi.org/10.1186/1471-2105-9-559.

56. Langfelder P, Horvath S. Fast R Functions for Robust Correlations and Hierarchical Clustering. J Stat Softw. 2012;46(11):1.

57. Steve H. Weighted Network Analysis: Applications in Genomics and Systems Biology: Springer, 2011, pp. 4–16. https://doi.org/10.1007/978-1-4419-8819-3.

58. Horvath S, Dong J. Geometric Interpretation of Gene Coexpression Network Analysis. PLoS Comput Biol. 2008;4(8):1000117.

59. Freeman L. A Set of Measures of Centrality Based on Betweenness. Sociometry. 1977;40:35–41. https://doi.org/10.2307/3033543.

60. Opsahl T. Structure and Evolution of Weighted Networks. PhD thesis. New York: Queen Mary College University of London; 2009.

61. Brandes U. A faster algorithm for betweenness centrality. J Math Sociol. 2001;25(2):163–77. https://doi.org/10.1080/00222500199900249.

62. Langfelder P, Luo R, Oldham MC, Horvath S. WGCNA: An R package for weighted correlation network analysis. Bioinformatics. 2012;28(6):92–6. https://doi.org/10.1093/bioinformatics/btr656.

63. Eichner C, Harasimczuk E, Nielsen F, Grotmol S, Dalvin S. Molecular characterization and functional analysis of LsChi2, a chitinase found in the salmon louse (Lepeophtheirus salmonis, salmonis, Kreyer 1838). Exp Parasitol. 2015;151:152–39. https://doi.org/10.1016/j.exparpar.2015.01.011.

64. Zhou E. Essentiality in Three Eukaryotic Protein-Interaction Networks. Mol Biol Evol. 2004;21(5):803–6. https://doi.org/10.1093/molbev/msi072.

65. Aragon S, Claudinot S, Blais C, Maibeche M, Dauphin-Villemant C. Analogous to Peritrophin-Motif Proteins are Involved in the Regulation of Chitin Layer in the Silk Gland of Bombyx mori at the Molting Stage. PROTEOMICS. 2018;18(19):170389. https://doi.org/10.1002/pmic.201703839.

66. Zhou E. Essentiality in Three Eukaryotic Protein-Interaction Networks. Mol Biol Evol. 2004;21(5):803–6. https://doi.org/10.1093/molbev/msi072.

67. Said MR, Begley TJ, Oppenheim AV, Lauffenburger DA, Samson LD. Toxicity modulation in Saccharomyces cerevisiae to the parasitc Oropocetes limosus (Lepeophtheirus salmonis) salmonis. J Exp Parasitol. 2015;159:79–93. https://doi.org/10.1016/j.exparpar.2015.09.001.

68. Moyer BD, Allan BB, Balch WE. Rab1 interaction with a GM130 effector in the parasitc Oropocetes limosus (Lepeophtheirus salmonis) salmonis. J Exp Parasitol. 2015;159:79–93. https://doi.org/10.1016/j.exparpar.2015.09.001.
Allan BB, Moyer BD, Balch WE. Rab1 recruitment of p115 into a cis-SNARE complex: Programming budding COP11 vesicles for fusion. Science. 2000;289(5478):444–8. https://doi.org/10.1126/science.289.5478.444.

Goody RS, Müller MP, Wu YW. Mechanisms of action of Rab proteins, key regulators of intracellular vesicular transport. De Gruyter. 2017. https://doi.org/10.1515/hz-2016-0274. https://www.degruyter.com/document/doi/10.1515/hz-2016-0274/html.

Huang J, Birmingham CL, Shahnazari S, Shiu J, Zheng YT, Smith AC, Campellone KG, Heo WD, Gruenheid S, Meyer T, Welch MD. Kistakis NT, Kim PK, Klionsky DJ, Brumell JH. Antibacterial autophagy occurs at PtdIns(3)P-enriched domains of the endoplasmic reticulum and requires Rab1 GTase. Autophagy. 2011;7(11):17–26. https://doi.org/10.1002/aut.7.1.13840.

Dong N, Zhu Y, Lu Q, Hu L, Zheng Y, Shao F. Structurally distinct bacterial TBC-like GAPs link Arf GTase to Rab1 inactivation to counteract host defenses. Cell. 2012;150(5):1029–41. https://doi.org/10.1016/j.cell.2012.06.050.

Takai Y, Sasaki T, Matozaki T. Putative regulatory mechanism of prothoracicotropic hormone (PTTH) secretion in the American cockroach, Periplaneta americana as inferred from co-localization of Rab8, PTTH, and protein kinase C in neurosecretory cells. Cell Tissue Res. 2009;335(3):607–15. https://doi.org/10.1007/s00441-008-0747-9.

Niwa YS, Niwa R. Transcriptional regulation of insect steroid hormone biosynthesis and its role in controlling timing of molting and metamorphosis. Dev Growth Differ. 2016;58(1):94–105. https://doi.org/10.1111/dgd.12124.

Mukhopadhyay A, Quirao JA, Wolkoff AW. Rab1a regulates sorting of early endocytic vesicles. Am J Physiol Gastrointest Liver Physiol. 2014;306(5):412. https://doi.org/10.1152/ajpgi.00118.2013.

Stemmer H, Cikkonen VM. The Rab GTase family. Genome Biol. 2001;2(5):3007–1. https://doi.org/10.1186/9464640.116057.

Field MC, Horn D, Carrington M. Analysis of Small GTPase Function in Lepeophtheirus salmonis. Cell. 2013;153(2):299–308. https://doi.org/10.1016/j.cell.2013.01.039.

Hiragaki S, Uno T, Takeda M. Peptide-like regulatory mechanism of prothoracicotropic hormone (PTTH) secretion in the American cockroach, Periplaneta americana as inferred from co-localization of Rab8, PTTH, and protein kinase C in neurosecretory cells. Cell Tissue Res. 2009;335(3):607–15. https://doi.org/10.1007/s00441-008-0747-9.

Niwa YS, Niwa R. Transcriptional regulation of insect steroid hormone biosynthesis and its role in controlling timing of molting and metamorphosis. Dev Growth Differ. 2016;58(1):94–105. https://doi.org/10.1111/dgd.12124.

Schumann I, Hering L, Mayer G, Kenny N, Hui J. Halloween genes in Drosophila. Cell. 2001;104(4):777–87. https://doi.org/10.1016/S0092-8674(01)00307-1.

Stenmark H, Olkkonen VM. The Rab GTPase family. Genome Biol. 2001;2(5):3007–1. https://doi.org/10.1186/9464640.116057.

Mukhopadhyay A, Quirao JA, Wolkoff AW. Rab1a regulates sorting of early endocytic vesicles. Am J Physiol Gastrointest Liver Physiol. 2014;306(5):412. https://doi.org/10.1152/ajpgi.00118.2013.

Stemmer H, Cikkonen VM. The Rab GTase family. Genome Biol. 2001;2(5):3007–1. https://doi.org/10.1016/j.cell.2002.05.005.

Field MC, Horn D, Carrington M. Analysis of Small GTPase Function in Trypanosoma brucei, the causative agent of African trypanosomiasis. Methods Enzymology. 1st edn. Berlin: Academic Press; 2008. p. 57–76. https://doi.org/10.1016/S0076-6879(08)08051-9.

Frand AR, Russel S, Ruvkun G. Functional genomic analysis of miRNA expression and sequence conservation in Atlantic and Pacific salmon lice (Lepeophtheirus salmonis). BMC Genomics. 2016;17(1):483. https://doi.org/10.1002/ujbio.2013.06.007.

Bock M, Mengel M, Dierich A, Gubitz G, Scheller HR, Fornace AJ Jr. miR-8-5p and miR-2a-3p modulate chitin biosynthesis in response to 20-hydroxyecdysone signaling in the brown plant hopper, Nilaparvata lugens. Insect Biochem Mol Biol. 2013;43(9):839–48. https://doi.org/10.1016/j.ibmb.2013.06.002.

Surholt B. Formation of glucosamine-6-phosphate in chitin synthesis during ecdysis of the migratory locust, Locusta migratoria. Insect Biochem. 1975;5(3):585–93. https://doi.org/10.1016/0020-2657(75)90040-2.

Chen J, Yang L, Yang R, Zhang W. Conserved microRNAs mir-8-5p and mir-2a-3p modulate chitin biosynthesis in response to 20-hydroxyecdysone signaling in the brown plant hopper, Nilaparvata lugens. Insect Biochem Mol Biol. 2013;43(9):839–48. https://doi.org/10.1016/j.ibmb.2013.06.002.

Hudson NJ, Reverter A, Dalrymple BP. A Differential Wiping Analysis of Expression Data Correctly Identifies the Gene Containing the Causal Mutation. PLoS Comput Biol. 2000;6(5):1000382. https://doi.org/10.1371/journal.pcbi.1000382.

Ishii M, Numata T, Inoue M, Shimada A, Akiyama M, Kitamura T, et al. Novel miRNAs identified by next-generation sequencing: An initial Catalog of miRNAs from the silkworm, Bombyx mori. BMC Genomics. 2012;13:1587. https://doi.org/10.1186/1471-2164-13-1587.

Yoshimura T. Expression analysis of miRNA in the silkworm, Bombyx mori. J Biol Chem. 2017;292(28):11659–69. https://doi.org/10.1074/jbc.M117.777797.

Cai M, Zhao W, Liu Y, Wang J, Zhao X-F. 20-hydroxyecdysone activates Forkhead box O2 to promote proteolysis during Helicoverpa armigera molting. Development. 2016;143(6):1005–15. https://doi.org/10.1242/dev.126894.

Hossain MS, Liu Y, Zhou S, Li K, Tian L, Li S. 20-Hydroxyecdysone-induced transcriptional activity of FoxO upregulates brummer and acid lipase-1 and promotes lipolysis in Bombyx fat body. Insect Biochem Mol Biol. 2013;43(9):829–38. https://doi.org/10.1016/j.ibmb.2013.06.007.

Domingue C, Qiao M, Qian C, Jiang Y, Chen H, et al. FOXO4 and FOXO3 regulate chitin biosynthesis and insect growth. Dev Biol. 2014;393:1244–57. https://doi.org/10.1016/j.ydbio.2014.08.007.

Lavorgna G, Karim FD, Thummel CS, Wu C. Potential role for a FTZ-F1 transcription factor superfamily member in the control of Drosophila...
134. Piulachs M-D, Pagone V, Bellés X. Key roles of the Broad-Complex gene in insect embryogenesis. Insect Biochem Mol Biol. 2010;40(6):468–75. https://doi.org/10.1016/j.ibmb.2010.04.006.

135. Maldonado-Aguayo W, Chávez-Mardones J, Gonçalves AT, Gallardo-Escárate C. Cathespin Gene Family Reveals Transcriptome Patterns Related to the Infective Stages of the Salmon Louse Caligus rogercresseyi. PLoS ONE. 2015;10(4):e0123954.

136. Nagai C, Mabashi-Asazuma H, Nagasawa H, Nagata S. Identification and Characterization of Receptors for Ion Transport Peptide (ITP) and ITP-like (ITPL) in the Silkworm Bombyx mori*. J Biol Chem. 2014;289(46):32166–77. https://doi.org/10.1074/jbc.M114.590646.

137. Webster SG, Keller R, Dircksen H. The CHH-superfamily of multifunctional peptide hormones controlling crustacean metabolism, osmoregulation, moulting, and reproduction. Gen Comp Endocrinol. 2012;175(2):217–33. https://doi.org/10.1016/j.ygcen.2011.11.035.

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