Comparative Biochemistry and Physiology, Part D 10 (2014) 9–21

Contents lists available at ScienceDirect

Comparative Biochemistry and Physiology, Part D

journal homepage: www.elsevier.com/locate/cbpd

High-throughput SNP discovery and transcriptome expression profiles from the salmon louse *Caligus rogercresseyi* (Copepoda: Caligidae)

Gustavo Nuñez-Acuña, Valentina Valenzuela-Muñoz, Cristian Gallardo-Escárate *

Laboratory of Biotechnology and Aquatic Genomics, Interdisciplinary Center for Aquaculture Research (INCAR), University of Concepción, P.O. Box 160-C, Concepción, Chile

ARTICLE INFO

Article history:
Received 13 November 2013
Received in revised form 29 January 2014
Accepted 31 January 2014
Available online 7 February 2014

Keywords:
Caligus rogercresseyi
SNPs
RNA-seq
MiSeq
Sea lice

ABSTRACT

The salmon louse *Caligus rogercresseyi* is the dominant ectoparasite species affecting the salmon aquaculture industry in the Southern hemisphere, and it is currently the main cause for economic losses in Chilean aquaculture. However, despite the great concern over *Caligus* infestations, genomic information on this louse is still scarce, even while the need to develop high-resolution molecular markers is growing. This study provides the first deep transcriptome survey to identify thousands of SNP markers from *C. rogercresseyi*, with a total of 69,466 SNPs identified using the MiSeq platform (Illumina®), 30,605 (52%) of which were found in contigs successfully annotated against known protein databases. Furthermore, *in silico* gene expression profiles associated with SNP variants were evaluated, and the results evidenced a wide array of genes that were down- and upregulated throughout the developmental stages of *C. rogercresseyi*. Interestingly, putative KEGG pathways involved in resistance to antiparasitic agents were also identified, where ten pathways were associated with the nervous system and one was related to ABC transporters. Taken together, this information could be highly useful for investigating the molecular underpinnings involved in the susceptibility or resistance of salmon lice to chemical treatments.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Infestations by the salmon lice *Caligus rogercresseyi* represent one of the main causes for economic losses in the Chilean salmon industry (Costello, 2009; Niklitschek et al., 2013). To mitigate these losses, various antiparasitic treatments have been used, including chemicals such as emamectin, deltametrine, diflubenzuron and cipermetrin, among others. Classically, the most used chemical in the Chilean industry has been emamectin, however, the efficacy of this chemical treatment has decreased over time, and other antiparasites have emerged (Bravo et al., 2008). One of the possible reasons for the loss of efficacy could be related to resistance mechanisms of the parasite against the chemical product, which may have genetic and genomic implications (Ffrench-Constant et al., 2004). Efficient and sensitive methods for diagnosing resistance are crucial for the management of drug resistance. Early detection of reduced sensitivity to a chemical can enable the application of effective countermeasures at the most probable point for successful treatment. For this, novel molecular markers directly linked to resistance against salmon louse treatments are required. However, genomic knowledge for *C. rogercresseyi* that would be applicable to this concern is currently unavailable.

The nucleotide and expressed sequence tag (EST) databases of the NCBI GenBank archive for *C. rogercresseyi* currently provide 2961 and 46,858 sequences, respectively. These data represent the effort to increase the genomic knowledge for this species, but it is still a low number of available sequences in comparison with other species. Moreover, the present study is the first to identify single nucleotide polymorphism (SNP) markers for this species of salmon louse. The main previous study to characterize EST sequences for this species obtained 32,135 ESTs through the construction of clone libraries and sequencing with the Sanger method (Yasuike et al., 2012). The main aim of this study was to compare different species of sea lice on the level of the mitochondrial genome, but the study also established the first genomic basis for the future construction of a 38K microarray of sea lice, which could be used in conjunction with the 44K salmon microarray to study host–parasite interactions at a transcriptomic level. Meanwhile, other sea lice species have more available genomic information than *C. rogercresseyi*, such as *Lepeophtheirus salmonis* with 56,720 nucleotide and 129,328 EST sequences registered. In this latter species, a set of SNPs and microsatellites, with 114 total molecular markers, were used to assess population structure along the Pacific coasts of Canada (Messmer et al., 2011). Until now, however, no reports have been made on sea lice SNPs that are applicable to the assessment of genotypes of interest to the salmon industry, such as the resistance or immune response of these copepod ectoparasites to antiparasitic chemicals.

SNP markers have increasingly become the most popular molecular markers due to the possibility of being linked with different functional phenotypes or genotypes of interest (Liao and Lee, 2010). Moreover, these markers have the advantage of being identifiable in non-model organisms through the use of high-throughput sequencing, a next-

* Corresponding author.
E-mail address: crisgallardo@udec.cl (C. Gallardo-Escárate).

1744-117X/$ – see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.cbd.2014.01.003

generation sequencing technique (Cahais et al., 2012). Genome wide association studies (GWAS) have made it possible to link SNP variants to traits of interest for different purposes, as evidenced in studies related to the immune system and growth, among others (Korte and Farlow, 2013). In sea lice, SNPs have not been linked with particular traits, but studies in arthropods have associated individual or a wide array of polymorphisms with phenotypes of interest. For example, an array of 1536 SNPs from the African malaria vector mosquito *Anopheles gambiae* was used to genotype around 1500 phenotypes that presented resistance to chemical disease control agents (Weetman et al., 2010). Furthermore, in the acetylcholinesterase gene of the mosquito *Culex pipiens*, 22 SNP differentiated individuals resistant to insecticide treatment were found (Mao et al., 2013). Interestingly, another study in the western corn rootworm *Diabrotica virgifera virgifera* associated a non-synonymous SNP in the γ-aminobutryic acid *(GABA) receptor* gene with phenotypes resistant to cyclodiene insecticides (Wang et al., 2013).

Given this background, the main objective of this study was to characterize thousands of SNP markers associated with the transcriptome of the salmon louse *C. rogercresseyi*. A further objective was to then classify the characterized SNPs according to biological function, thus establishing putative relationships between polymorphisms and potential traits of interest such as immune response, resistance to antiparasites, growth, and development, among others.

2. Materials and methods

2.1. Salmon lice cultures

Ovigerous specimens of *C. rogercresseyi* were collected from recently harvested fish in Puerto Montt, Chile. Individuals were transported to the laboratory on ice, and their egg strings were then removed and placed in culture buckets supplied with seawater flow at 12 °C and with gentle aeration. Eggs were allowed to hatch and develop until the infectious copepodid stage, at which point they were harvested for RNA extraction and cDNA library construction. The culture procedure was carried out according to Bravo (2010). The specific protocol to suit with the management requirements for salmon louse culture is available in the handbook in resistance management (exclusively

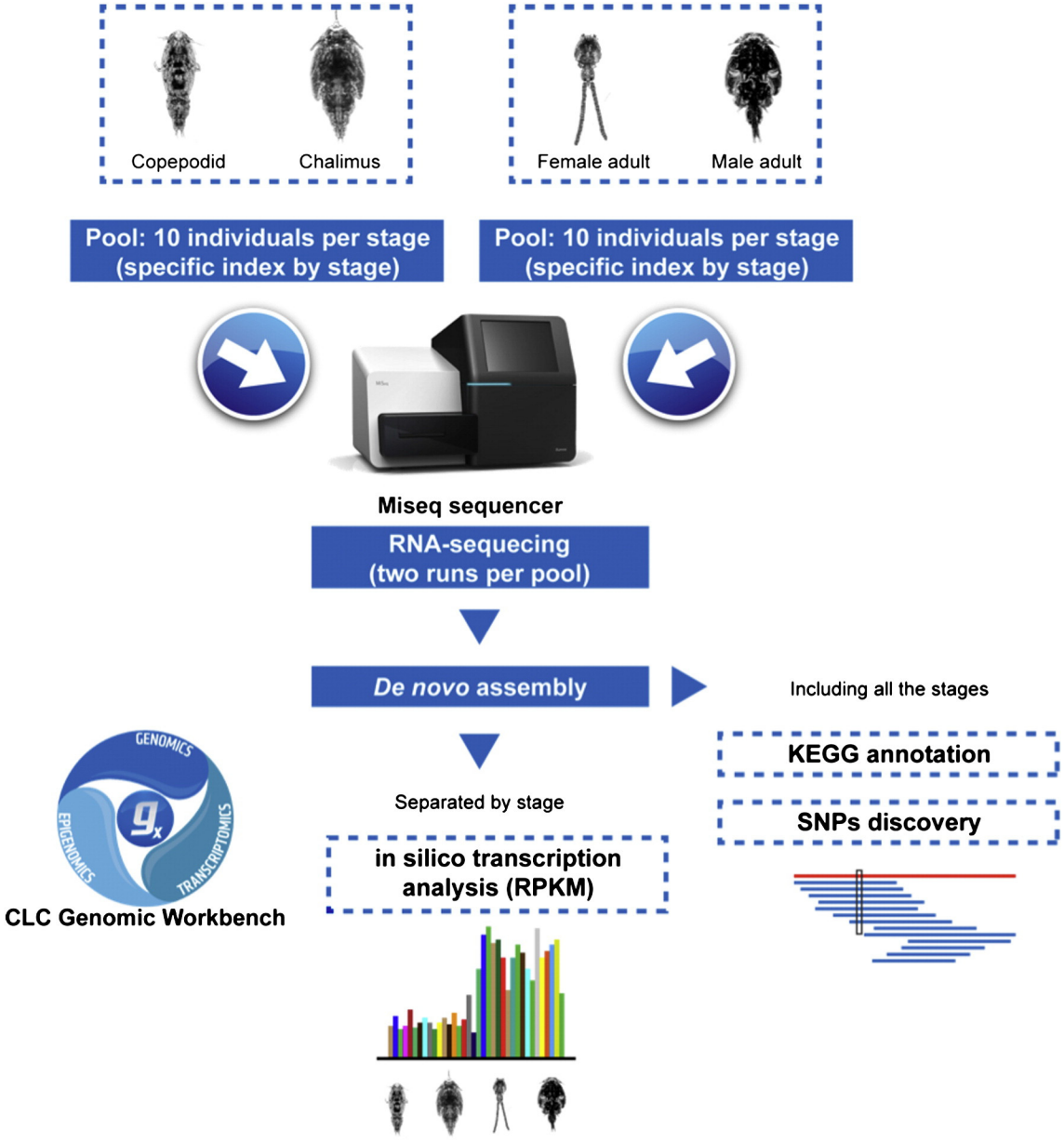


Fig. 1. Experimental set-up designed for this study.

available at http://www.iacr.bbsrc.ac.uk/pie/search-EU/). A veterinarian expert supervision was necessary to suit with ethical and management requirements. See the complete experimental set-up for this methodology at Fig. 1.

2.2. Transcriptome sequencing of C. rogercresseyi

Total RNA was extracted from pools containing ten individuals from each larval stage (copepodid and chalimus) and adult (female and male) using the Ribopure™ kit (Ambion®, Life Technologies™, USA) according to the manufacturer's instructions. Quantity, purity, and quality of isolated RNA were measured in TapeStation 2200 (Agilent Technologies Inc., Santa Clara, CA, USA) using the R6K reagent kit according to the manufacturer's instructions. Subsequently, double-stranded cDNA libraries were constructed using the TruSeq RNA Sample Preparation kit v2 (Illumina®, San Diego, CA, USA). Two biological replicates for each sample pool (n = 8) were sequenced by the MiSeq (Illumina®) platform using sequenced runs of 2 × 250 paired-end reads at the Laboratory of Biotechnology and Aquatic Genomics, Interdisciplinary Center for Aquaculture Research (INCAR), University of Concepción, Chile.

2.3. De novo assembly

Bioinformatic analyses were carried out using the CLC Genomics Workbench software (Version 6.0.1, CLC Bio, Denmark). *De novo* assembly was applied to mine SNP variants from the transcriptome of each pool of *C. rogercresseyi*. The assembly parameters used were the following: mismatch cost = 2, deletion cost = 3, insert cost = 3, minimum contig length = 200 bp, and trimming quality score = 0.05. Furthermore, to exclude paralogous sequence variants (PSVs) an overlap criterion of 70% and a similarity of 0.9 were applied.

Transcriptomic data were deposited in Sequence Read Archive (SRA) at GenBank database (NCBI, National Center for Biotechnology Information). SRA accession code is SRR1106551. These data correspond to trimmed unassembled sequences generated by MiSeq sequencing. The *de novo* assembly generated with raw data of the different stages studied is available upon author request.

2.4. SNP identification and annotation

Candidate SNPs were called with the following settings: window length = 11, maximum gap and mismatch count = 2, minimum average quality of surrounding bases = 15, minimum quality of central base = 20, maximum coverage = 100, minimum coverage = 20, minimum variant frequency (%) = 25.0, and maximum expected variations (ploidy) = 2. Frequencies of SNPs found in all the contigs were calculated using the same software and were tabulated in excel spreadsheets.

Transcriptome annotation was based on a tBLASTx analysis against UniProt and GenBank EST databases. Subsequently, Gene Ontology analysis was conducted using the Blast2GO plugin incorporated in the CLC Genomic Workbench software (CLC Bio). In addition, the consensus

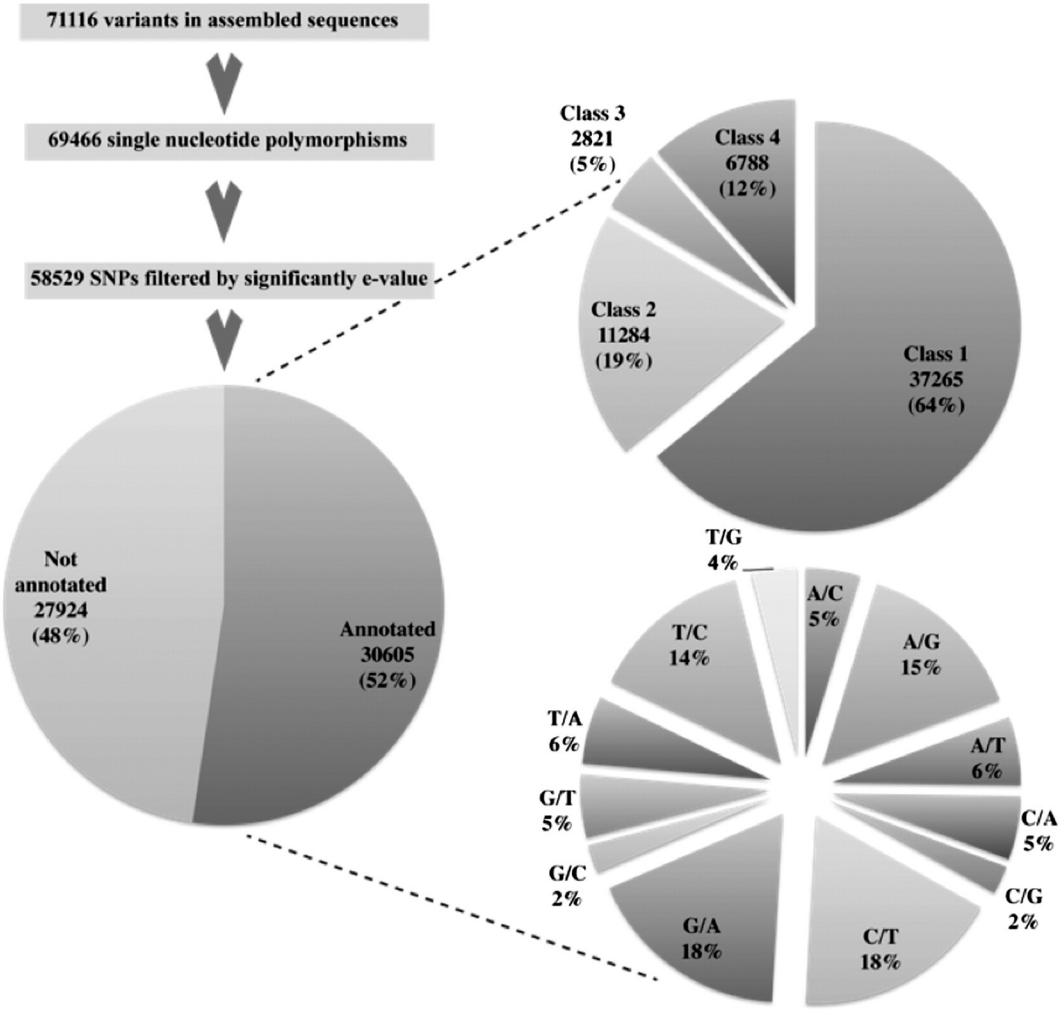


Fig. 2. General description of SNPs identified in the transcriptome of *Caligus rogercresseyi*. Proportion of annotated and not annotated contigs and also their distribution by classes and polymorphism type are displayed.

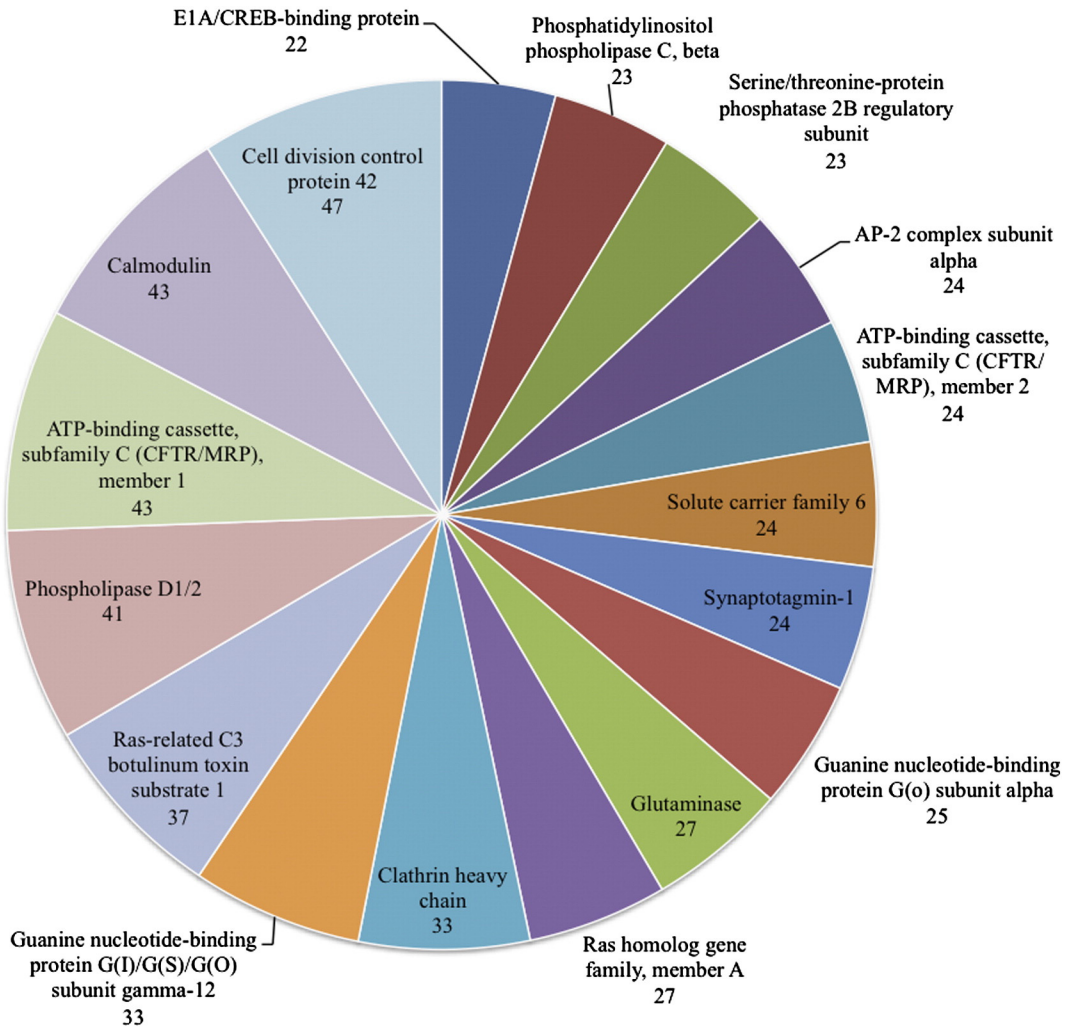


Fig. 3. Most polymorphic contigs identified in the KEGG pathways related to nervous system and ABC transporters selected for this work. The numbers represent the count of SNPs found in the respective transcript.

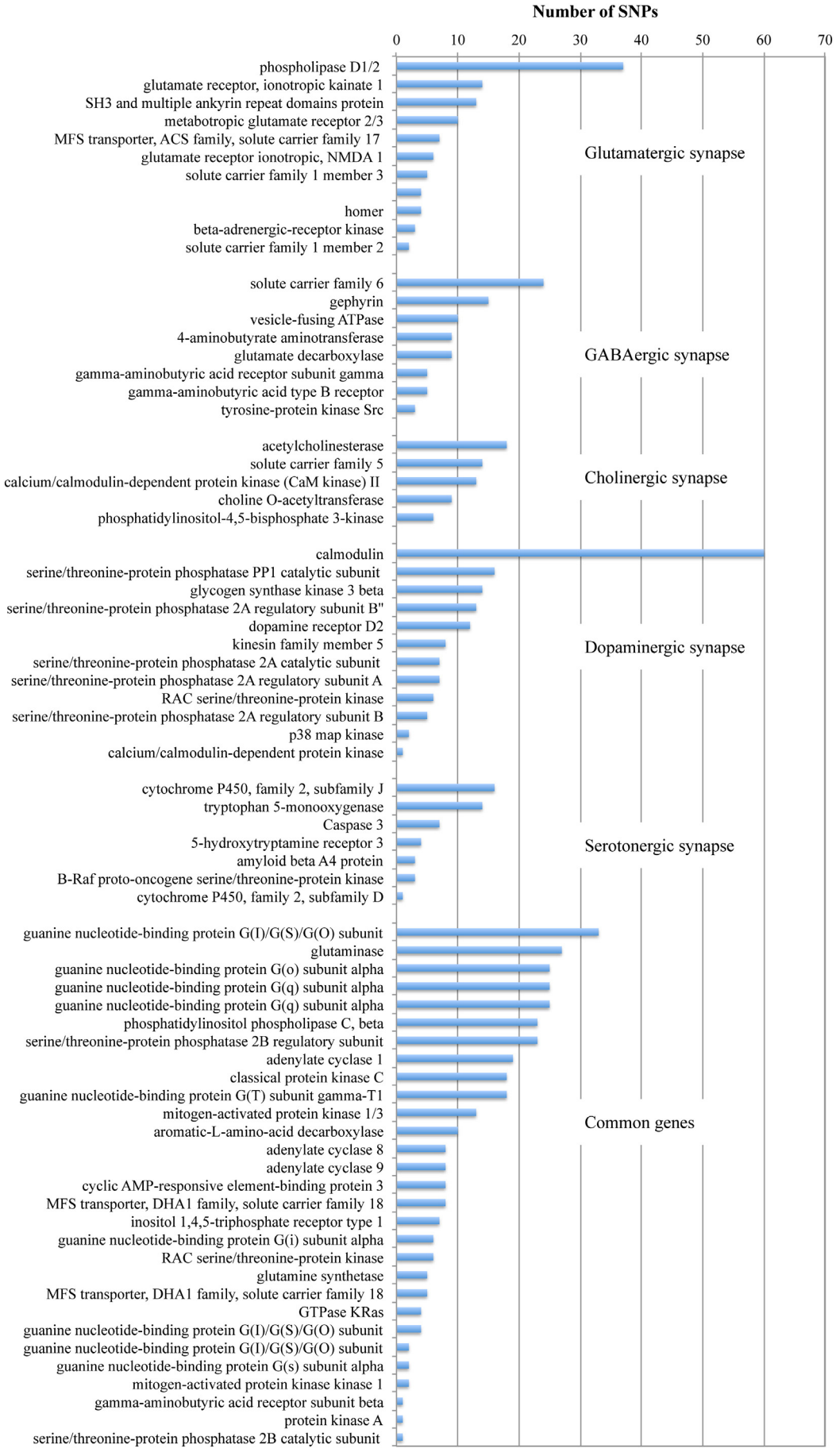
sequences of contigs were annotated by KEGG ontology using the KEGG Automatic Annotation Server (KAAS) (Moriya et al., 2007). From this latter annotation, KEGG pathways related to the nervous system and ABC transporters were selected to identify the contigs that might be associated with tolerance to chemical treatments. As the model species available in KEGG database are not genetically close to *C. rogercresseyi*, we searched for all the genes presented in the selected pathways in GenBank databases in species more related to copepods (as other crustacean species). Then, these transcripts were blasted against the transcriptomic data generated for *C. rogercresseyi* to extract the transcripts that were not found through KAAS procedure. Contigs selected from these pathways were downloaded, and SNPs were searched for within the sequences according to the SNP list generated.

2.5. In silico gene expression of sequences containing SNPs

Contigs containing SNPs were extracted from the total assembly sequences using the plugin *extract annotations* of the CLC Genomic Workbench software (CLC Bio). RNA-seq analyses were performed using the contigs containing SNPs from the four salmon louse stages in order to identify the transcripts, which were overexpressed or suppressed. To perform this analysis, a mapping of the polymorphic sequences was done, allowing the reads to map only in the most suitable region of the reference used (ignoring non-specific matches). The reference used was the *de novo* assembly generated in Section 2.3. Maximum gap and mismatch cost was set as 2.

Gene expression was calculated measuring the read per kilobases per million mapped reads (RPKM) values, which were normalized by total reads. To obtain the significantly expressed transcripts, proportion comparisons were calculated using Kal's test with corrected Bonferroni p-values. Multi-paired comparisons were performed to generate three volcano plots for copepodid/chalimus, chalimus/adult, and copepodid/adult stages. These three volcano plots were used to extract the significantly expressed contigs, which were those that had at least a p-value > 10 on a $\log_{10}$ scale and a fold change higher than 4 or lower than −4 on a $\log_2$ scale. Fold change values were calculated in CLC Genomic Workbench as the division between RPKM values of the query stage and the RPKM values of the reference stage (reference and query stages are selected arbitrary according to each required result, see Supplementary Tables S2 to S7 for the results of specific pathways). With these contigs, a sub-experiment was carried out in order to generate a hierarchical clustering of features with differential

Fig. 4. Distributions of SNPs in the KEGG pathway related to synapsis function in *Caligus rogercresseyi*. Selected pathways are: GABAergic synapse, cholinergic synapse, dopaminergic synapse and serotonergic synapse. The histogram shows the sum of SNPs (number of SNPs) identified in all the contigs annotated for each respective transcript shown in Y-axis. Common genes correspond to transcripts presented in all the pathways mentioned in the figure.


Number of SNPs
0
10
20
30
40
50
60
70
phospholipase D1/2
glutamate receptor, ionotropic kainate 1
SH3 and multiple ankyrin repeat domains protein
metabotropic glutamate receptor 2/3
MFS transporter, ACS family, solute carrier family 17
glutamate receptor ionotropic, NMDA 1
solute carrier family 1 member 3
homer
beta-adrenergic-receptor kinase
solute carrier family 1 member 2
Glutamatergic synapse
solute carrier family 6
gephyrin
vesicle-fusing ATPase
4-aminobutyrate aminotransferase
glutamate decarboxylase
gamma-aminobutyric acid receptor subunit gamma
gamma-aminobutyric acid type B receptor
tyrosine-protein kinase Src
GABAergic synapse
acetylcholinesterase
solute carrier family 5
calcium/calmodulin-dependent protein kinase (CaM kinase) II
choline O-acetyltransferase
phosphatidylinositol-4,5-bisphosphate 3-kinase
Cholinergic synapse
calmodulin
serine/threonine-protein phosphatase PP1 catalytic subunit
glycogen synthase kinase 3 beta
serine/threonine-protein phosphatase 2A regulatory subunit B"
dopamine receptor D2
kinesin family member 5
serine/threonine-protein phosphatase 2A catalytic subunit
serine/threonine-protein phosphatase 2A regulatory subunit A
RAC serine/threonine-protein kinase
serine/threonine-protein phosphatase 2A regulatory subunit B
p38 map kinase
calcium/calmodulin-dependent protein kinase
Dopaminergic synapse
cytochrome P450, family 2, subfamily J
tryptophan 5-monooxygenase
Caspase 3
5-hydroxytryptamine receptor 3
amyloid beta A4 protein
B-Raf proto-oncogene serine/threonine-protein kinase
cytochrome P450, family 2, subfamily D
Serotonergic synapse
guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit
glutaminase
guanine nucleotide-binding protein G(o) subunit alpha
guanine nucleotide-binding protein G(q) subunit alpha
guanine nucleotide-binding protein G(q) subunit alpha
phosphatidylinositol phospholipase C, beta
serine/threonine-protein phosphatase 2B regulatory subunit
adenylate cyclase 1
classical protein kinase C
guanine nucleotide-binding protein G(T) subunit gamma-T1
mitogen-activated protein kinase 1/3
aromatic-L-amino-acid decarboxylase
adenylate cyclase 8
adenylate cyclase 9
cyclic AMP-responsive element-binding protein 3
MFS transporter, DHA1 family, solute carrier family 18
inositol 1,4,5-triphosphate receptor type 1
guanine nucleotide-binding protein G(i) subunit alpha
RAC serine/threonine-protein kinase
glutamine synthetase
MFS transporter, DHA1 family, solute carrier family 18
GTPase KRas
guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit
guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit
guanine nucleotide-binding protein G(s) subunit alpha
mitogen-activated protein kinase kinase 1
gamma-aminobutyric acid receptor subunit beta
protein kinase A
serine/threonine-protein phosphatase 2B catalytic subunit
Common genes

expression levels. In addition, six clusters with similar expression patterns based on Manhattan distances between transcripts were identified. These clusters were extracted and annotated according to the previously obtained BLAST hits.

Furthermore, a comparison between the total of contigs presenting SNPs and the total of contigs without any polymorphism was conducted through *in silico* gene transcription analysis. For this purpose, two datasets were generated containing consensus sequences of contigs, one with the contigs containing SNPs and another with the contigs that did not evidenced polymorphisms. These datasets were used as reference for mappings performed with the same conditions that the previous analyses mentioned above. Mapped reads corresponded to two datasets of all the obtained sequences; one corresponded to the larval stages (copepodid and chalimus) and another with the adult stages (male and female). Paired analyses were conducted on mappings for each reference based on Kal's test, and differentially expressed genes were visualized in volcano plots. From the plots, contigs presenting lower fold change values than −4 or higher than 4, and at the same time a p value lower than 0.01, were selected to generate a subexperiment, which were plotted in heatmap forms though hierarchical clustering of features.

3. Results

3.1. SNP discovery and annotation

Transcriptome sequencing for different developmental stages of the salmon louse *C. rogercresseyi* yielded 108,637,988 high-quality reads which were successfully assembled into 84,023 contigs with a mean length of 635 bp. Of these contigs, 15,422 presented at least one SNP within its sequences, although the total number of polymorphisms

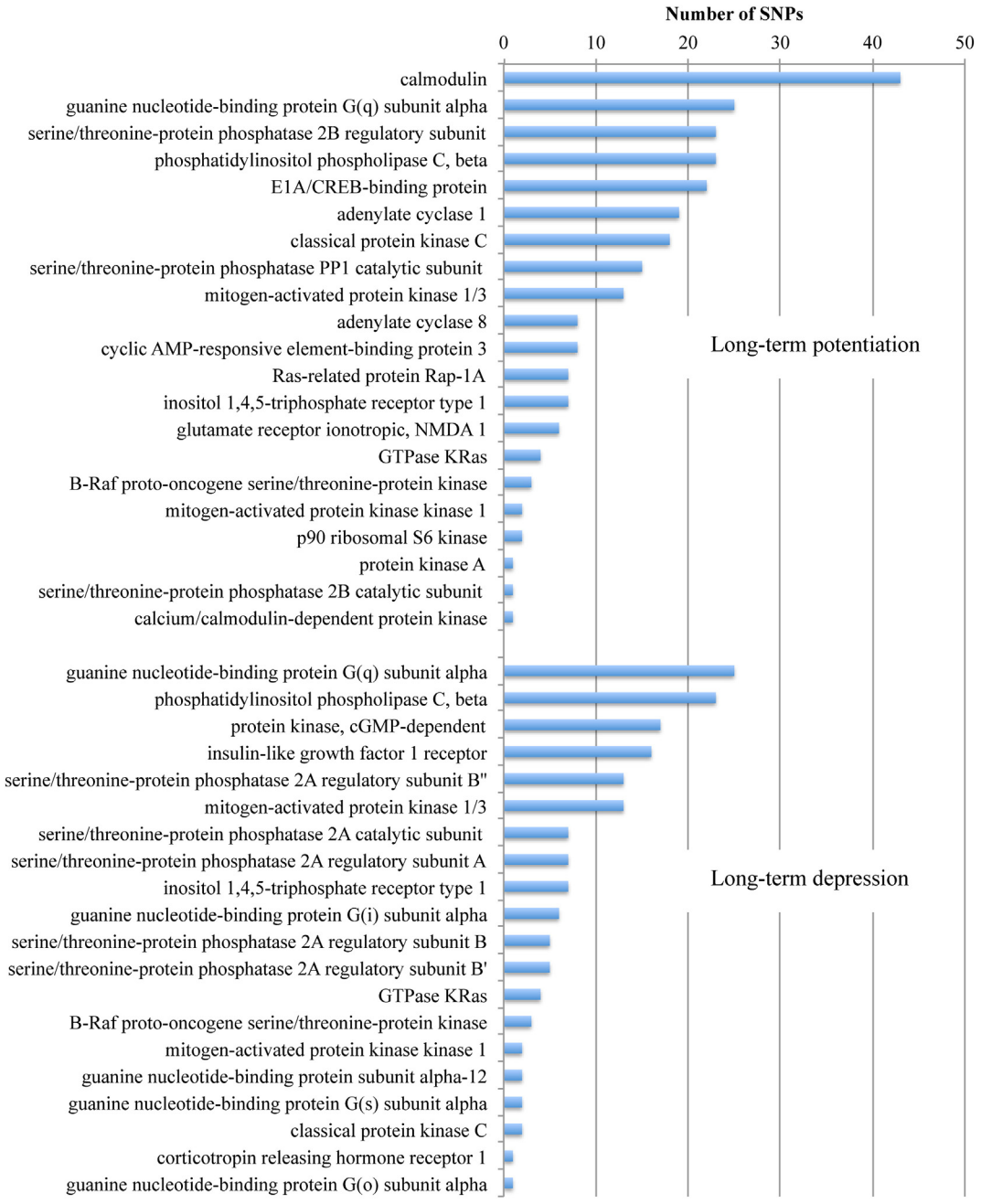


Fig. 5. Distributions of SNPs in the KEGG pathways related to long-term potentiation and long-term depression. The histogram shows the sum of SNPs (number of SNPs) identified in all the contigs annotated for each respective transcript shown in Y-axis.

was 71,116; 69,466 of which were SNPs, and the remaining variants corresponded to insertions/deletions (Indels). The 15,422 contigs were annotated, and SNPs were selected according to an e-value cutoff of 1E−05, resulting in 58,529 SNPs (see a full list of SNPs and respective BLAST hits in Table S1). Fifty-two percent of these had BLAST hits with known genes or EST sequences from public databases, whereas the remaining 48% of SNPs were present in contigs that did not have BLAST hits with available reported sequences (Fig. 2).

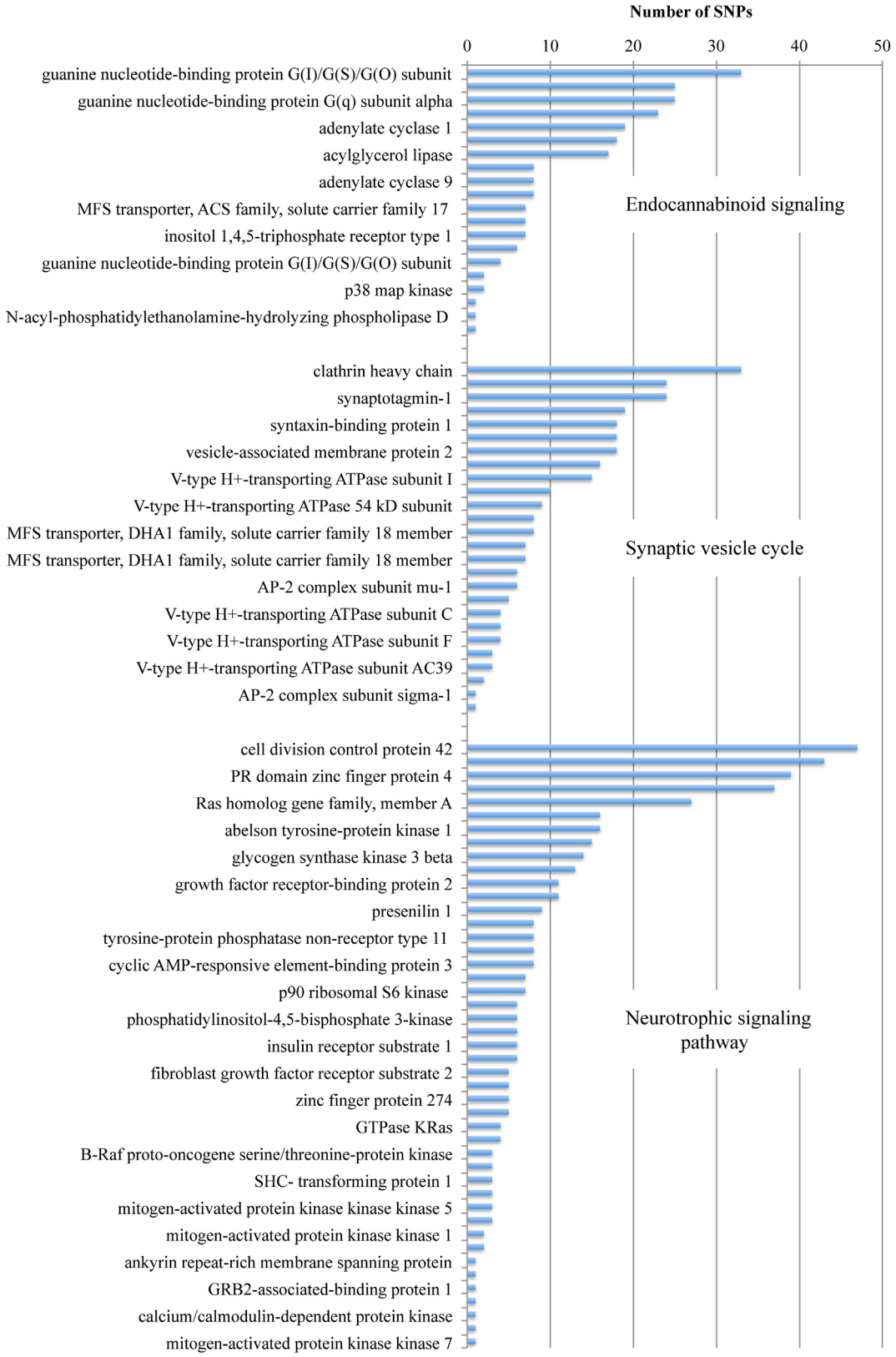


Fig. 6. Distributions of SNPs in the KEGG pathways related to retrograde endocannabinoid signaling, synaptic vesicle cycle and neurotrophic signaling pathway. The histogram shows the sum of SNPs (number of SNPs) identified in all the contigs annotated for each respective transcript shown in Y-axis.

SNPs were then categorized according to the classes described by Liew et al. (2004), which are related to the differential melting temperatures that a specific sequence could have depending on the nucleotides involved in the polymorphism. From this, 37,265 SNPs, or 64%, were classified as Class 1 (C/T, T/C, G/A, and A/G polymorphisms); 11,284, or 19%, were classified as Class 2 (C/A, A/C, G/T, and T/G polymorphisms); 2821, or 5%, were classified as Class 3 (C/G and G/C polymorphisms), representing the class with the lowest number of SNPs; and 6788, or 12%, were classified as Class 4 (A/T and T/A) (Fig. 2).

3.2. SNPs identified in KEGG pathways potentially involved in resistance to antiparasites

KAAS annotation of putative KEGG pathways involved in the resistance to antiparasites was performed to identify SNPs in the successfully annotated contigs, where ten pathways associated with the nervous systems and one related to ABC transporters were identified. This annotation was then enriched by downloading and searching, from GenBank database, transcripts presumably absent in *C. rogercresseyi* but present in other crustaceans. From this annotation, 17 candidate genes were selected on a basis of presenting more than 20 polymorphisms within its contigs. The cell division control protein 42 gene, a component of the neurotrophic signaling pathway, was the most polymorphic and presented 47 SNPs. Following this, 43 SNPs were present in the ATP-binding cassette subfamily C member 2 gene, which corresponded to an ABC transporter, and the calmodulin gene, which was present in the dopaminergic synapse, long-term potentiation, and the neurotrophic signaling pathway. Finally, the phospholipase D1/2 gene, a component of the glutamatergic synapse, presented 41 SNPs (Fig. 3). The following were also analyzed: glutamatergic synapse, GABAergic synapse, cholinergic synapse, dopaminergic synapse, and serotonergic synapse (Fig. 4), long-term potentiation and long-term depression (Fig. 5), retrograde endocannabinoid signaling, synaptic vesicle cycle and neurotrophic signaling (Fig. 6), and ABC transporters (Fig. 7). The distribution of SNPs among all of the studied pathways is available in Supplemental material (Table S1). Notably, among these pathways the highest number of polymorphic genes was found in neurotrophic signaling with 45 polymorphic transcripts and the lowest number of polymorphic transcripts was found in the ABC transporters with 9 polymorphic genes.

3.3. In silico gene expression analysis of polymorphic contigs

RPKM values from the contigs containing SNPs during the different stages of *C. rogercresseyi* were calculated and then visualized through a heat map based on a hierarchical clustering of features. This analysis showed that various contigs had differential expression values according to the developmental stage (Fig. 8). Furthermore, six clusters were identified according to the calculated Manhattan distances (Suppl. Tables S2–S7). In particular, Cluster 1 showed a group of contigs with an expression pattern mainly associated with the copepodid stage, and it was less expressed in the two remaining stages. Cluster 2 had contigs mainly associated with the adult stage, and it showed an overexpression or downregulation of some contigs in this stage. Cluster 3 was significantly more related to the copepodid stage, but in contrast, Cluster 4 showed a set of genes most expressed in both the chalimus and adult stages, in addition to showing lower expression levels in the

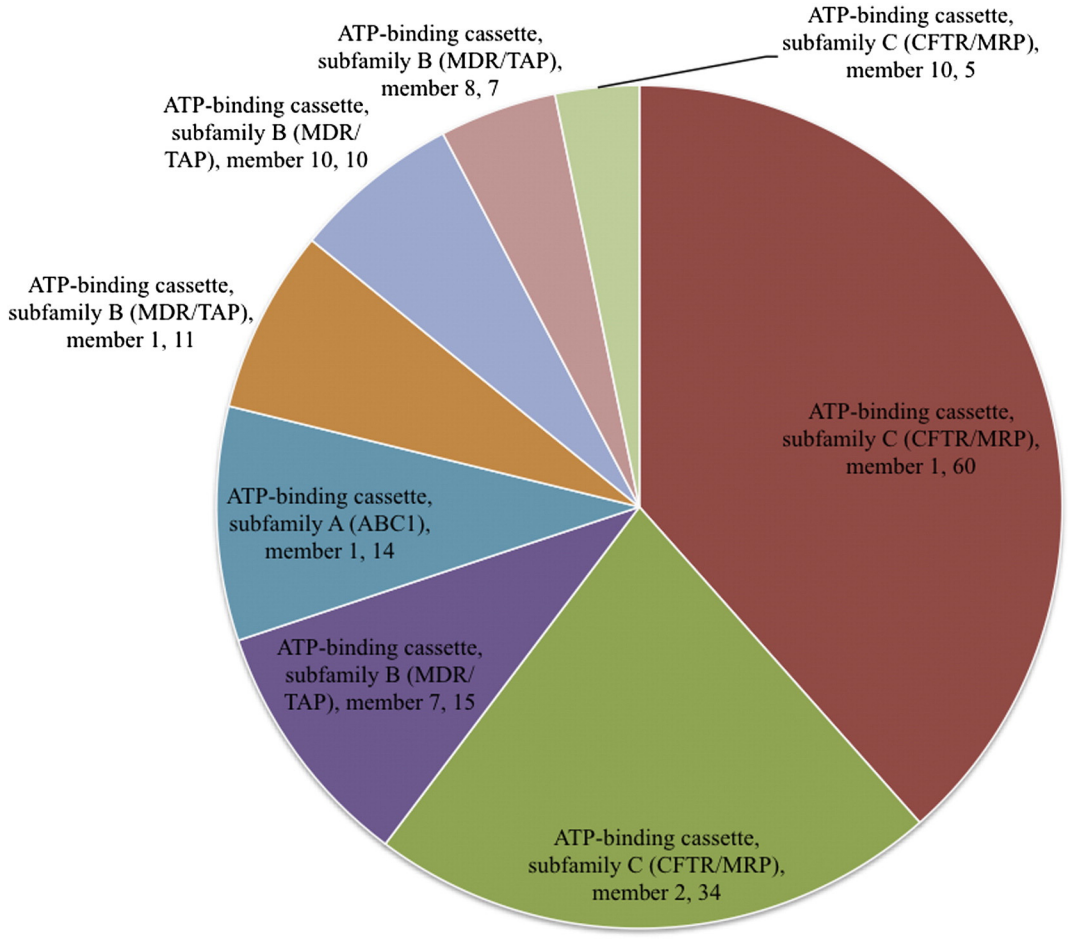


Fig. 7. Distributions of SNPs in the KEGG pathway related to ABC transporters. The pie chart shows the sum of SNPs (number of SNPs) identified in all the contigs annotated for each respective transcript shown in Y-axis.

chalimus stage. Cluster 5 exhibited a similar pattern as Cluster 4. Finally, Cluster 6 was highly expressed in both chalimus and adult stages, but it was less expressed in copepodids (Fig. 8, Tables S2–S7). Among the transcripts with expression patterns associated with the copepodid

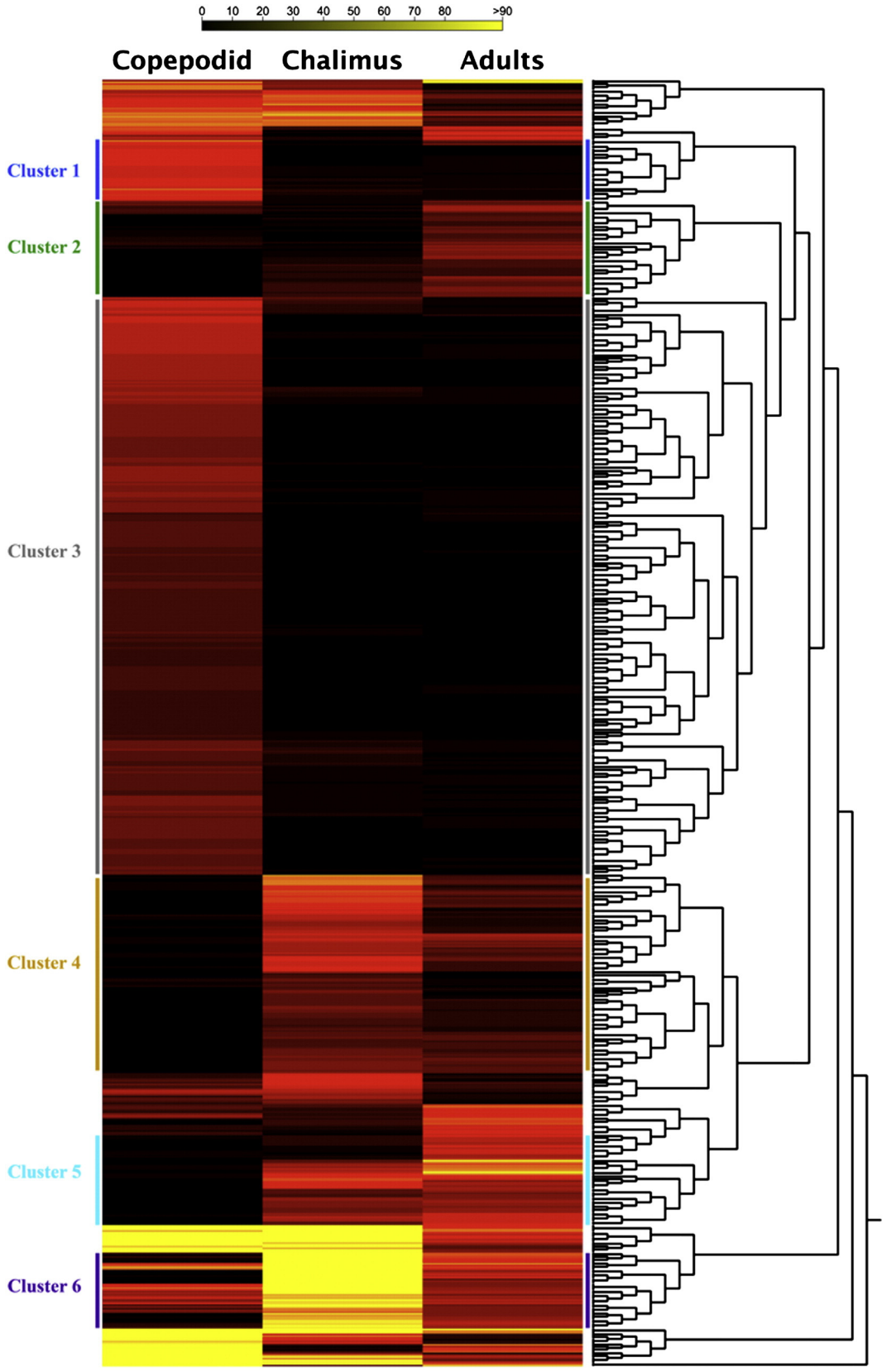


Fig. 8. Heat-map representing the hierarchical clustering of features of RPKM transformed values (log_2) calculated among the three stages studied. Yellow values correspond to highly expressed transcripts, while red and black represent lower expression levels, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

stage (Clusters 1 and 3), genes were found related to mitochondrial metabolism, such as Acyl carrier protein, cytochrome c oxidase II, and NADH dehydrogenase 5; structural proteins, such as collagen alpha chain, ankyrin, and chondroadherin; cell adhesion, such as E-selectin and L-selectin; and, interestingly, diapause, the couch potato protein (Tables S2 and S4). On the other hand, Cluster 2 had transcripts mainly associated with the adult stage and different biological process; nonetheless it should be noted that at this stages, overexpression started in genes related to detoxification, such as ascorbate peroxidase precursor, prophenoloxidase activating factor, and quinone oxidoreductase, and two different isoforms of serine proteases were also found (Table S3). Meanwhile, cluster-grouping transcripts associated with both chalimus and adult stages were related to a wide array of biological process, such as mitochondrial metabolism, transcription factors, and binding, among others (Tables S5, S6, and S7).

Regarding the most differentially expressed contigs that contained SNPs, thirty transcripts with the highest fold change values throughout the three sequenced stages were selected (Table 1). Among these contigs, transcripts related to different biological functions were found. For instance, the contig with the highest calculated fold change value corresponded to a *C-type lectin* involved in the immune response, and it was strongly overexpressed in the copepodid stage. Furthermore, other transcripts with SNPs related to the innate immune signaling pathways were overexpressed in this stage, such as *Toll*, *Toll-7*, and *serine/threonine-protein kinase 16*. Nonetheless, higher levels of serine proteases were expressed in the chalimus and adult stages. Genes related to other functions were also found, such as the sex-related *vitellogenin* gene, which was expressed mainly in the chalimus and adult stages, and the growth and development-related growth arrest-specific protein 1, Cuticle protein 18.6, and other isoform B genes, which were almost exclusively expressed in copepodids.

On the other hand, the comparison between contigs with SNPs and without SNPs showed different expression profiles in larval and adult stages. Although the number of significantly differentially expressed transcripts was similar between both groups (182 contigs with SNPs and 205 non-polymorphic SNPs), their expression patterns were not equally distributed. Non-polymorphic contigs expressed with higher uniformity than contigs with SNPs, but at the same time fold change levels were more contrasting between larvae and adults (Fig. 9).

4. Discussion

In the present study, 64,699 novel SNPs were identified in the transcriptomic regions of the salmon louse *C. rogercresseyi*. This work is the first effort towards developing SNP markers in this ectoparasite species. Furthermore, this study constitutes the most comprehensive SNP mining of all salmon lice species. Two previous reports, which characterized SNP markers, of a salmon louse from the Northern hemisphere, *L. salmonis*, were found. First, Messmer et al. (2011) characterized 87 SNP markers from an EST library of 76,642 sequences in order to evaluate the population structure of *L. salmonis* along the Canadian coasts, but the molecular markers, which also included microsatellites, could not resolve the population structures in these locations, suggesting a widely homogeneous population as a product of the dispersion of this salmon louse. The second study carried out a RAD-sequencing project in this species and obtained more than 85 million high-quality Illumina reads (100 pb) that produced 281,838 unique RAD-tags, from which 31,555 were polymorphic (Carmichael et al., 2013). From these RAD-tags, the authors identified 24,538 bi-allelic markers, from which one marker found in the *prohibin-2* gene was fully associated with sex determination in the species. Another notable study regarding other copepod species was carried out by Ning et al. (2013), which identified 284,154 SNP markers from two 454 runs in the transcriptome of the copepod *Calanus sinicus*. This is a higher number of markers than reported in the present study, even though the sequencing throughput was lower. This could be caused by the parameters used for the SNP calling procedure,

Table 1
Top differentially expressed contigs containing SNPs among the different stages of the *Caligus rogercresseyi* lifecycle.

Contig number	Best BLAST hit	e-Value	Fold change	RPKM normalized values		
				Copepodid	Chalimus	Adults
contig_15081	C-type lectin – galactose binding [*Papilio polytes*]	2.29E−43	1950.56	351.25	1.00	3.43
contig_1128	ADP-ribosylation factor-directed GTPase activating protein isoform B	0	1779.84	1.76	1.00	5.54
contig_51746	Cuticle protein 18.6, isoform B [*Caligus clemensi*]	9.87E−28	1773.21	404.09	1.00	4.29
contig_9738	Vitellogenin-like protein [*Lepeophtheirus salmonis*]	7.31E−71	1476.32	1.00	1180.28	862.29
contig_27713	Elongation of very long chain fatty acids protein, putative	1.03E−93	1216.78	28.46	1.00	5.73
contig_5086	E1a binding protein P400 [*Aedes aegypti*]	6.77E−57	1002.77	1.35	1.91	1.00
contig_6817	Cytoplasmic aconitate hydratase-like [*Aplysia californica*]	1.52E−53	927.38	1.37	1.00	8.63
contig_4896	Protein lin-7 homolog C-like [*Bombyx mori*]	1.88E−96	592.32	3.90	1.00	1.84
contig_15051	RT10125p [*Drosophila melanogaster*]	6.75E−81	493.35	23.23	1.50	0.99
contig_6594	Vacuolar protein sorting-associated protein 8 homolog isoform X2	0	411.97	1.43	1.33	1.00
contig_23664	Mucin-22-like [*Ceratitis capitata*]	7.60E−14	317.79	13.25	1.84	0.99
contig_15801	C-type lectin – galactose binding [*Papilio polytes*]	2.29E−43	295.45	8.53	2.77	0.99
contig_22948	Lebercilin, partial	2.48E−19	295.39	23.43	0.99	1.22
contig_64573	Hypothetical protein [*Lepeophtheirus salmonis*]	1.32E−22	279.01	1.02	425.25	379.06
contig_30224	Myotubularin-related protein 9	0	254.17	8.83	1.15	0.99
contig_1695	Carbonic anhydrase 3 [*Caligus rogercresseyi*]	1.30E−60	239.96	1.03	715.42	688.15
contig_15610	Mediator complex subunit 12-like	0	235.21	2.92	1.00	2.04
contig_7561	Toll [*Culex quinquefasciatus*]	0	231.97	10.05	0.99	1.00
contig_1165	Probable histone-lysine N-methyltransferase NSD2-like	0	229.08	3.71	4.10	0.99
contig_3620	Serine protease K12H4.7 precursor [*Caligus clemensi*]	7.55E−28	227.50	1.02	22.26	38.33
contig_13111	Serine protease [*Tigriopus japonicus*]	1.86E−18	224.64	1.01	10.33	6.35
contig_15611	Serine/threonine-protein kinase 16 [*Caligus rogercresseyi*]	0	220.60	6.62	3.45	0.99
contig_10783	KN motif and ankyrin repeat domain-containing protein 1-like, partial	6.48E−101	212.66	1.01	3.92	2.11
contig_6235	Trifunctional nucleotide phosphoesterase protein YfkN-like	0	211.24	5.84	2.94	0.99
contig_19388	Zinc finger and SCAN domain-containing protein 2-like	1.92E−10	209.53	1.01	4.90	4.22
contig_6506	Microprocessor complex subunit DGCR8-like	1.84E−137	196.50	2.82	0.99	1.84
contig_24062	Vitellogenin 2 [*Lepeophtheirus salmonis*]	4.74E−90	180.11	1.05	4574.61	5281.31
contig_10503	Growth arrest-specific protein 1 precursor [*Lepeophtheirus salmonis*]	1.37E−52	163.82	51.35	0.98	8.41
contig_9634	Tyrosine-protein kinase RYK [*Crassostrea gigas*]	3.31E−75	144.70	3.14	0.99	1.31
contig_11937	Toll-7 [*Nilaparvata lugens*]	2.60E−124	143.74	24.25	0.98	3.37
contig_7303	Inositol hexakisphosphate and diphosphoinositol-pentakisphosphate kinase	0	132.88	3.31	1.21	0.99
contig_21519	Steryl-sulfatase [*Crassostrea gigas*]	1.68E−30	128.17	1.03	40.20	12.94

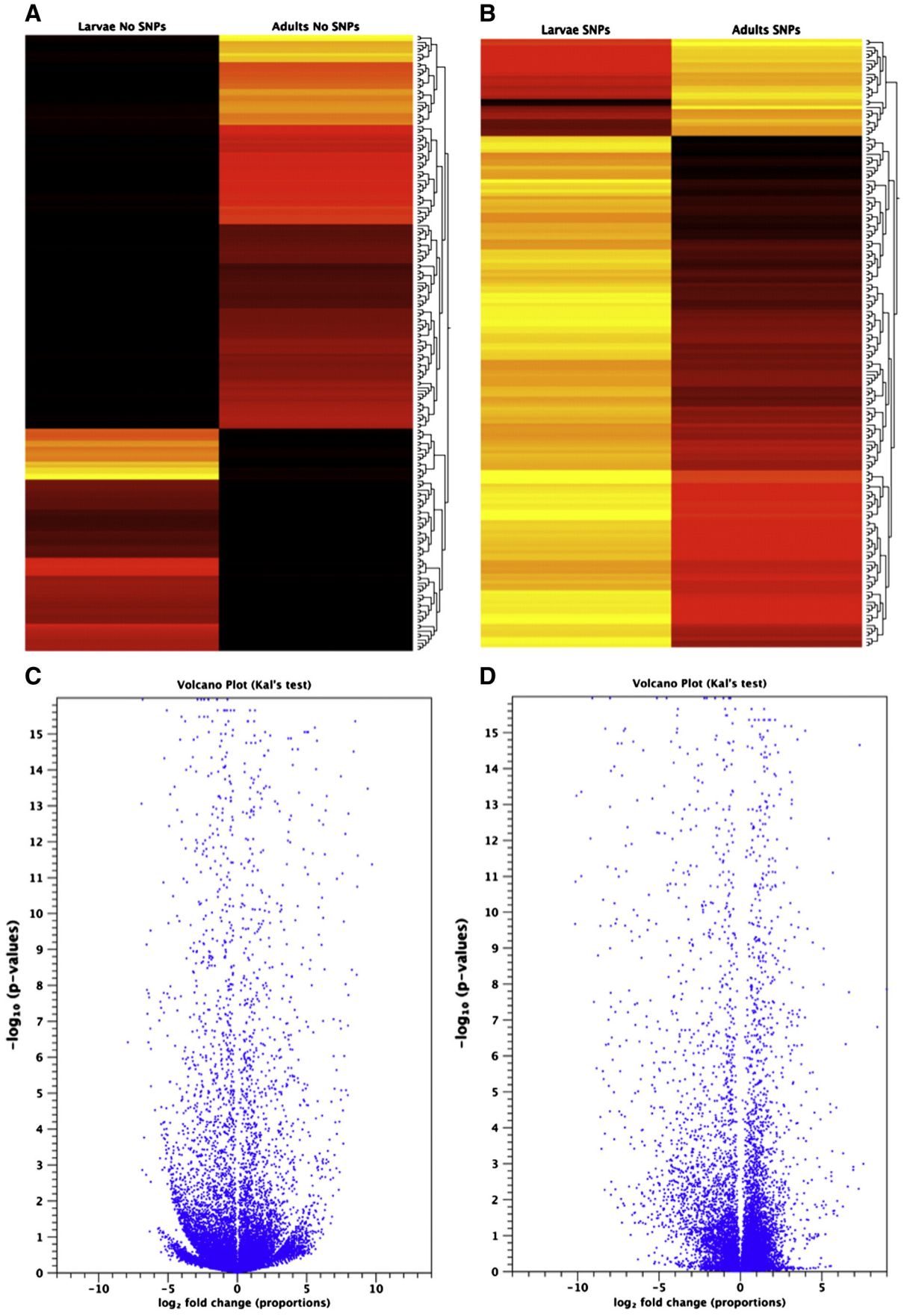


Fig. 9. RNA-seq analysis comparison of all the contigs that presented at least one SNP (SNP contigs) and all the contigs absent of polymorphisms (no-SNP contigs). A–B: Heatmap constructed based on hierarchical clustering constructed in CLC Genomic Workbench of most differentially expressed contigs for each state. C–D: Kal-test represented as volcano plot to show all the contigs with SNPs or without polymorphisms for each stage of the sea louse lifecycle.

specifically that Ning et al. (2013) considered SNPs presented in at least four reads, while the present study opted for a higher coverage of at least 20 reads.

The KEGG annotation performed in this study was useful to identify SNPs in the pathways related to the nervous system and ABC transporters, which might be related with the resistance in *C. rogercresseyi* to the antiparasitic chemical treatments that the salmon industry currently applies. Among the most polymorphic pathways, 431 SNPs were identified in the neurotrophic-signaling pathway. In superior animals, this pathway is comprised by a gene family of neurotrophins and Trk tyrosine kinase receptors, and its main function is to initiate the early signaling that leads to cell survival and the differentiation of neurons (Gunn-Moore and Tavaré, 1998). The dopaminergic synapse was another polymorphic pathway that had 402 SNPs. The role of this pathway in higher vertebrates is related with motricity and the modulation of emotions and cognitions through the regulation of the dopamine neurotransmitter (Seutin, 2005), but there is still debate in invertebrates about the specific role of this pathway (Magoski and Bulloch, 1997; Weislogel et al., 2013). Following these pathways in quantity of identified SNPs was the glutamatergic synapse with 379 SNPs. This pathway in both vertebrates and invertebrates is involved in chemosensations, among other functions (Horie et al., 2008). In summary, these three polymorphic pathways have some biological functions that could be interesting for evaluating antiparasitic treatments in *C. rogercresseyi*.

Among all of the KEGG pathways analyzed, the most polymorphic transcripts were related to genes with different biological functions, but all of these are pivotal components of the pathways involved in nervous system development. The most polymorphic transcript corresponded to *cell division control protein 42*, which contained 47 SNPs. This gene is one of the main components of different pathways, including neurotrophic signaling, that promote proliferation and apoptosis signals in the cell (Melendez et al., 2011). Following this transcript, the second most polymorphic transcripts were subfamily C (CFTR/MRP) member 1 and the calmodulin gene. *ATP-binding cassette* corresponds to the ABC (ATP-binding cassette) transporters, which are known as *multidrug resistance protein 1* and are implicated in drug tolerances found in mammalians and other taxa such as marine vertebrates (Zaja et al., 2008; He et al., 2011). Based on this study, this gene is suggested as a potential marker for evaluating the resistance to antiparasitic treatments due to its polymorphisms and function. In addition, it should be noted that previous studies have characterized another ABC transporter, the P-glycoprotein gene, as a marker for evaluating resistance to the antiparasite emamectin benzoate (commercially known as SLICE®) in the sea louse *L. salmonis* (Igboeli et al., 2012). Regarding *calmodulin*, it is a widely studied gene that has implications on different pathways, including three pathways evaluated in the present study, neurotrophin signaling, dopaminergic synapse, and long-term potentiation. This gene is mainly involved in calcium ion binding, thus regulating some nervous impulses. Furthermore, the calmodulin protein was detected as one of the main components in salmonid mucus during sea lice infection (Easy and Ross, 2009). Given this information, this is suggested as another gene of interest for the purposes mentioned above.

Since RNA-seq analysis was applied in this study, it was possible to measure the relative abundance of transcripts that SNP markers contained within their sequences. Using this approach, groups of transcripts with expression patterns associated with a particular stage or stages of the *C. rogercresseyi* life cycle could be evaluated. This transcriptomic information possibly has great relevance since it could be used to combine expression levels with the respective polymorphisms found. Specifically, a panel of genes was characterized with the greatest differences in expression levels between the developmental stages. In this set of genes, *vitellogenins* showed a strong association with the chalimus and adult stages. Since it is known that this gene is implicated in female reproduction and sexual differentiation, it is suggested that the female differentiation process starts in the chalimus stage in this species (Baumann et al., 2013). Another gene, which shared this expression pattern, was *carbonic anhydrase 3*, which is available in the GenBank dataset for *C. rogercresseyi*. Moreover, other marine organisms have shown the functioning of this gene in molting stages, supporting the basis that carbonic anhydrase 3 is another gene relevant for association studies (Jasmani et al., 2008, 2010). On the other hand, other genes showed the opposite expression pattern, with greater abundance in the copepodid stage. This was the case for the immune-related *C-type lectin* gene, which is known to be related with the innate immune response in a wide array of organisms, but it has also been proposed as a modulator in some part of the molting stages of crustaceans, specifically contributing to the hardening process of the cuticle (Kuballa et al., 2011). The *cuticle protein*, as previously characterized in *Caligus clemensi* (GenBank accession number: BT080137), was also related to the same process and developmental stage.

5. Conclusion

The SNP markers identified in this study could be useful for genetic studies in the salmon louse *C. rogercresseyi*. A special focus for future studies should be on the genes identified in pathways involved in the nervous system of this ectoparasite as one of the main unsolved problems for the salmon industry in the Southern hemisphere is the resistance of *C. rogercresseyi* to antiparasitic chemicals.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.cbd.2014.01.003.

Acknowledgments

This study was funded by a grant awarded by CONICYT-Chile (FONDAP project 15110027).

References

Baumann, L., Holbech, H., Keiter, S., Kinnberg, K.L., Knörr, S., Nagel, T., Braunbeck, T., 2013. The maturity index as a tool to facilitate the interpretation of changes in vitellogenin production and sex ratio in the Fish Sexual Development Test. Aquat. Toxicol. 128–129, 34–42.

Bravo, S., 2010. The reproductive output of sea lice *Caligus rogercresseyi* under controlled conditions. Exp. Parasitol. 125, 51–54.

Bravo, S., Sevatdal, S., Horsberg, T.E., 2008. Sensitivity assessment of *Caligus rogercresseyi* to emamectin benzoate in Chile. Aquaculture 282, 7–12.

Cahais, V., Gayral, P., Tsagkogeorga, G., Melo-Ferreira, J., Ballenghien, M., Weinert, L., Chiari, Y., Belkhir, K., Ranwez, V., Galtier, N., 2012. Reference-free transcriptome assembly in non-model animals from next-generation sequencing data. Mol. Ecol. Resour. 12, 834–845.

Carmichael, S.N., Bekaert, M., Taggart, J.B., Christie, H.R.L., Bassett, D.I., Bron, J.E., Skuce, P.J., Gharbi, K., Skern-Mauritzen, R., Sturm, A., 2013. Identification of a sex-linked SNP marker in the salmon louse (*Lepeophtheirus salmonis*) using RAD sequencing. PLoS One 8, e77832.

Costello, M.J., 2009. The global economic cost of sea lice to the salmonid farming industry. J. Fish Dis. 32, 115–118.

Easy, R.H., Ross, N.W., 2009. Changes in Atlantic salmon (*Salmo salar*) epidermal mucus protein composition profiles following infection with sea lice (*Lepeophtheirus salmonis*). Comp. Biochem. Physiol. D 4, 159–167.

Ffrench-Constant, R.H., Daborn, P.J., Le Goff, G., 2004. The genetics and genomics of insecticide resistance. Trends Genet. 20, 163–170.

Gunn-Moore, F.J., Tavaré, J.M., 1998. Progress toward understanding the molecular mechanisms of neurotrophic factor signalling. Cell. Signal. 10, 151–157.

He, S.-M., Li, R., Kanwar, J.R., Zhou, S.F., 2011. Structural and functional properties of human multidrug resistance protein 1 (MRP1/ABCC1). Curr. Med. Chem. 18, 439–481.

Horie, T., Kusakabe, T., Tsuda, M., 2008. Glutamatergic networks in the *Ciona intestinalis* larva. J. Comp. Neurol. 508, 249–263.

Igboeli, O.O., Fast, M.D., Heumann, J., Burka, J.F., 2012. Role of P-glycoprotein in emamectin benzoate (SLICE®) resistance in sea lice, *Lepeophtheirus salmonis*. Aquaculture 344–349, 40–47.

Jasmani, S., Jayasankar, V., Wilder, M.N., 2008. Carbonic anhydrase and Na/K-ATPase activities at different molting stages of the giant freshwater prawn *Macrobrachium rosenbergii*. Fish. Sci. 74, 488–493.

Jasmani, S., Jayasankar, V., Shinji, J., Wilder, M.N., 2010. Carbonic anhydrase and Na/K-ATPase activities during the molt cycle of low salinity-reared white shrimp *Litopenaeus vannamei*. Fish. Sci. 76, 219–225.

Korte, A., Farlow, A., 2013. The advantages and limitations of trait analysis with GWAS: a review. Plant Methods 9, 29.

Kuballa, A.V., Holton, T.a, Paterson, B., Elizur, A., 2011. Moult cycle specific differential gene expression profiling of the crab *Portunus pelagicus*. BMC Genomics 12, 147.

Liao, P.-Y., Lee, K.H., 2010. From SNPs to functional polymorphism: the insight into biotechnology applications. Biochem. Eng. J. 49, 149–158.
Liew, M., Pryor, R., Palais, R., Meadows, C., Erali, M., Lyon, E., Wittwer, C., 2004. Genotyping of single-nucleotide polymorphisms by high-resolution melting of small amplicons. Clin. Chem. 50, 1156–1164.
Magoski, N.S., Bulloch, a.G., 1997. Localization, physiology, and modulation of a molluskan dopaminergic synapse. J. Neurobiol. 33, 247–264.
Mao, Y., Tan, F., Yan, S.-G., Wu, G.-X., Qiao, C.-L., Zhang, W.-X., Cui, F., 2013. High-throughput genotyping of single-nucleotide polymorphisms in ace-1 gene of mosquitoes using MALDI-TOF mass spectrometry. Insect Sci. 20, 167–174.
Melendez, J., Grogg, M., Zheng, Y., 2011. Signaling role of Cdc42 in regulating mammalian physiology. J. Biol. Chem. 286, 2375–2381.
Messmer, A.M., Rondeau, E.B., Jantzen, S.G., Lubieniecki, K.P., Davidson, W.S., Koop, B.F., 2011. Assessment of population structure in Pacific *Lepeophtheirus salmonis* (Krøyer) using single nucleotide polymorphism and microsatellite genetic markers. Aquaculture 320, 183–192.
Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C., Kanehisa, M., 2007. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35, W182–W185.
Niklitschek, E.J., Soto, D., Lafon, A., Molinet, C., Toledo, P., 2013. Southward expansion of the Chilean salmon industry in the Patagonian Fjords: main environmental challenges. Aquaculture 5, 172–195.
Ning, J., Wang, M., Li, C., Sun, S., 2013. Transcriptome sequencing and de novo analysis of the copepod *Calanus sinicus* using 454 GS FLX. PLoS One 8, e63741.
Seutin, V., 2005. Dopaminergic neurones: much more than dopamine? Br. J. Pharmacol. 146, 167–169.
Wang, H., Coates, B.S., Chen, H., Sappington, T.W., Guillemaud, T., Siegfried, B.D., 2013. Role of a gamma-aminobutryic acid (GABA) receptor mutation in the evolution and spread of *Diabrotica virgifera virgifera* resistance to cyclodiene insecticides. Insect Mol. Biol. 22, 473–484.
Weetman, D., Wilding, C.S., Steen, K., Morgan, J.C., Simard, F., Donnelly, M.J., 2010. Association mapping of insecticide resistance in wild *Anopheles gambiae* populations: major variants identified in a low-linkage disequilibrium genome. PLoS ONE 5, e13140.
Weislogel, J.-M., Bengtson, C.P., Muller, M.K., Hortzsch, J.N., Bujard, M., Schuster, C.M., Bading, H., 2013. Requirement for nuclear calcium signaling in *Drosophila* long-term memory. Sci. Signal. 6, ra33.
Yasuike, M., Leong, J., Jantzen, S.G., von Schalburg, K.R., Nilsen, F., Jones, S.R.M., Koop, B.F., 2012. Genomic resources for sea lice: analysis of ESTs and mitochondrial genomes. Mar. Biotechnol. 155–166 (XX).
Zaja, R., Munić, V., Klobucar, R.S., Ambriović-Ristov, A., Smital, T., 2008. Cloning and molecular characterization of apical efflux transporters (ABCB1, ABCB11 and ABCC2) in rainbow trout (*Oncorhynchus mykiss*) hepatocytes. Aquat. Toxicol. 90, 322–332.