Manuscript 6

Identification of transcripts regulated during sexual maturation and egg production in adult female salmon lice *Lepeophtheirus salmonis* (*Crustacea*, Copopoda), using EST-sequencing and microarray analysis

Christiane Eichner[1,6], Petter Frost[1,5,6], Bjarte Dysvik[2], Bjørn Kristiansen[3], Inge Jonassen[2,4] and Frank Nilsen[1]
[1] Institute of Marine Research, Bergen, Norway
[2] Department of Informatics, University of Bergen, Bergen, Norway
3 The Norwegian microarray consortium, The Norwegian Radium Hospital, Oslo, Norway
4 Computational Biology Unit, Bergen Center for Computational Science, University of Bergen, Bergen, Norway
[5] Corresponding author
[6] These two authors contributed equally to this work.

ABSTRACT

Even though copepods are very abundant arthropods in the aquatic environment relatively limited biological information is available for copepods. The salmon louse (*L.salmonis*) is a marine ectoparasitic copepod feeding on skin, mucous and blood from salmonid hosts. The life cycle consists of 10 developmental stages separated by ecdysis. After final molting, adult female lice continuously produce eggs for life. Sexual maturation of copepods is poorly investigated and no information regarding the reproductive system on the molecular or biochemical level is available. During the course of the present study 7,021 expressed sequence tags (ESTs) from five cDNA libraries (adult and preadult stages) were produced and sequenced. A large proportion (43%) of the sequences had no significant hit in GenBank. Assembling sequences by ContigExpress resulted in 556 contigs and 1614 singletons. To link the unknown genes to biological functions and to characterize female salmon louse transcriptomes during the post molting growth, sexual maturation and egg production, microarray analysis was performed using a 7k salmon louse array produced together with the Norwegian Microarray Consortium. Gene expression from the preadultII stage and 6 different adult time points (after the last molting) of the female *L.salmonis* were analyzed. The microarray data analysis revealed two major expression patterns, one with a substantial increase in gene transcription during the first steps of adult development (group1) and one with a significant decrease (group2). In addition a third minor group was found containing ESTs up regulated in preadultII stage only. The number of novel genes was high in the major groups, 49% and 39% respectively. The majority of annotated genes in group 2 and 3 gave hits with cuticula proteins or other genes involved in growth related processes while in group 1 two large contigs encode two vitellogenins and are clearly involved in reproduction. Only 1 of the 76 contigs in group 2 contained more than 10 ESTs while in group 1, 10 of the 35 contigs, consist of more than 10 ESTs indicating a generally higher transcription level in group 1 genes compared to group 2 genes. The two vitellogenin contigs contains approximately 100 ESTs which are in compliance with the high transcription level typically seen for vitellogenins that are to be deposited in eggs. One novel gene contig in group 1 contained 106 ESTs indicating an expression level comparable to the vitellogenins. The microarray transcription profiles were confirmed by Northern blot analysis.

INTRODUCTION

Copepods are very abundant Arthropods in the aquatic environment. In the marine environment the free-living species are extremely abundant, e.g. *C.firmarcicus* the most numerous animal on the planet. In spite of this, relatively limited biological information is available for copepods and no model organism exist. In addition to being numerous a wide range of copepods are parasitic for instance on fish.

The salmon louse (*L.salmonis*) is a marine ectoparasitic copepod feeding on skin, mucous and blood from salmonid hosts. The life cycle consists of 10 developmental stages separated by ecdysis. The first free-living naupli larvae hatch directly from egg-strings attached to adult females and the free-living larvae stages (naupli I, naupli II and the infectious copepodid) can be transported by the ocean currents over large distances depending on hydrographical conditions [1]. After host settlement the infectious copepodids stage molt into chalimus, which are physically attached to the host by a frontal filament. All four chalimus stages, separated by moulting/growth, are anchored to the host by the frontal filament, which restricts the feeding area. However, in the preceding pre-adult and adult stages the salmon louse can move unrestricted on the host surface resulting in an significant increase in virulence [2].

Sexual maturation is one of the major physiological and behavioural changes in animal life cycle. Germ cells are typically established early in development but arrested in development until the onset of sexual maturation. The generation of gametes is most conserved in males where as a large variation is seen between different species for the development of female gametes. The ovum (i.e. the mature unfertilized egg) is a highly complex cell that is energetically expensive to produce. In order to produce high quality ova the females must undergo physiological adaptations that initiate further gamete development and maturation.

Since the reproduction strategy is highly variable between different species and different life history strategies (e.g. free living or parasitic lifestyle), there is considerable variation in the processes of sexual maturation and the production of eggs between different species. However, there are some common hallmarks that are expected in most animals. After fertilisation, the egg must be capable to initiate development and contain sufficient energy to ensure development until external energy can be taken up and utilised. The ovarium is the site for initial development for female gametes during the sexual maturation and reproduction. At ovulation oogonia are released from the ovarium into the oviduct where a significant growth and maturation take place before the ovum can be fertilised. This process is typically divided into pre-vitellogenic and vitellogenic development. During vitellogenesis yolk proteins are incorporated into the oocytes. A molecular hallmark for this process is the transcription of genes encoding egg yolk proteins like vitellogenins (Vgs). Depending on animal group, the transcription of Vgs takes place in different tissues/cell types (liver (in vertebrates), fat body (insects) and hepatopacrease (e.g. decapods)) and subsequently transported by the blood or hemolymph to the maturing oocytes, where they are taken up by receptor-mediated endocytosis.

Production of vitellogenin is controlled by steroid hormones, which induce transcription of the target gene through binding of a steroid steroid-receptor complex to the gene promoter. In arthropods, ecdyson (i.e. E20) has been shown to be the regulator hormone that binds to a heterodimeric receptor, the ultraspiracle (UsP) and ecdysteroid receptor (EcR) that is capable of binding to an ecdysteroid response (ERE) element in the promoter region of vitellogenins. Ecdysteroids is also a key regulatory component in the molting process in Arthropods [3]. At present, few articles are available that describes the gross features in the salmon louse reproduction biology.

The salmon louse reproductive systems have been described at the anatomical and histological levels [4] but there is no information regarding the timing of the different events during the sexual maturation. It has been proposed that fertilisation is the signal triggering the egg-production in the salmon louse [5], but according to our observations using unfertilised laboratory kept animals this is not the case. Adult female salmon louse produce eggs and egg-strings whether they are fertilised or not. Unlike crustaceans like shrimp, that produce eggs in seasons and grow/molt their entire life, adult female salmon louse stop have a final molting, stop growing when egg production has started and they continuously produce eggs for life. It has been shown that *L. salmonis* can produce up to 11 sets of egg-strings from a single fertilization [6].

However, following the last molting and prior to egg production the size of the gonad complex and abdomen increases where as the cephalotorax appears unchanged in size, indicating a process of sexual maturation. Initial production and analysis of EST sequences from pre adult and adult stages of *L. salmonis* revealed a large proportion of transcripts with no significant hit in public databases. In order to be able to link these unknown genes or transcripts to biological functions we have in the present study combined EST sequencing and microarray analysis (7k salmon louse array) to characterize female salmon louse transcriptomes during the post molting growth/sexual maturation and egg production.

RESULTS

EST analyses

During the course of the present study 7,021 ESTs from five cDNA libraries (adult and preadult stages) were sequenced (Table 1). The initial annotation of the individual ESTs revealed that a large proportion (43%) of the sequences had no significant hit in GenBank (Figure 1). Five percent of the ESTs showed significant hit with proteins with unknown function, 9% with proteases and 8% with ribosomal proteins. Approximately 6% of the ESTs encoded mitochondrial transcripts (Figure 1) of which 81.6 % were16S rRNA. All non-mitochondrial sequences with a length over 100 bp (4,563 ESTs) were assembled using ContigExpress, resulting in 556 contigs and 1614 singletons (Table 2). The majority of the 556 contigs (2,948 ESTs) contained 2-4 ESTs. Two clusters contained more than 100 EST. Contig 63 contains 147 ESTs encoding trypsins (LsTryp1 and LsTryp2) like transcripts [7, 8] while Contig 72 (124 ESTs) has no significant hits in GenBank and no detectable conserved domains present in the putative protein (172aa) encoded by the ORF.

Library	N clones	Percent of total	N EST	Percent of total
Female total	4,320	61.5	3,262	63.8
Female intestine	864	12.3	464	9.1
Adult male	768	10.9	646	12.6
Pre-adult II female	768	10.9	571	11.2
Female male subtracted	301	4.3	166	3.2
Total	**7,021**	**100.0**	**5,109**	**100.0**

Table 1. An overview of isolated clones from the various cDNA libraries and the number of ESTs. The number of ESTs is sequences with a remaining length over 100 after vector trimming and quality assessment (see above).

	ContigExpress analysis
Sequences analysed:	4563
Number of ESTs in contigs:	2948
Number of contigs:	556
Number of singletons:	1614
Number of contigs containing:	
2-4 ESTs	418
5-10 ESTs	92
11-20 ESTs	25
>20 ESTs	16

Table 2. Summary of data obtained after EST assembly using ContigExpress module in Vector NTI.

About 1,900 clones were not included in the contig assembling, primarily due to lack of high quality sequence data but also due to empty clones or clones contain vector *E. coli* sequences (approximately 1%) and insert sizes less than 100 bp (less than 1%). Based on these considerations, the proportion of singletons (35%) in the clustering results and the proportion of mitochondrial sequences (6%), the number of additional salmon louse transcripts among the 1,900 clones without sequence data was estimated to be approximately 500. This indicates that the 7021 clones, from which cDNA probes were printed on the microarray, represents approximately 2,600 different *L. salmonis* transcripts.

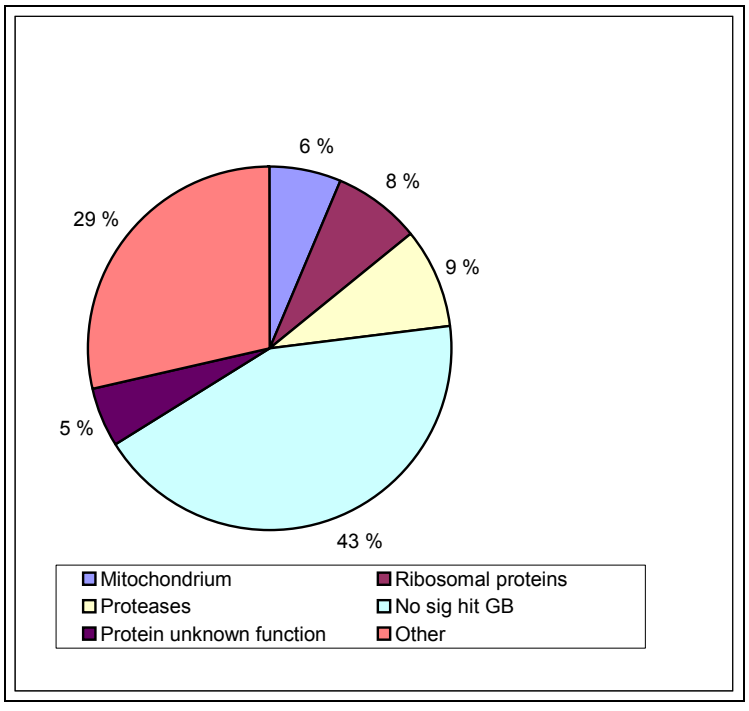

Figure 1: Distribution of individual ESTs in major functional groups based on primary Blast annotation.

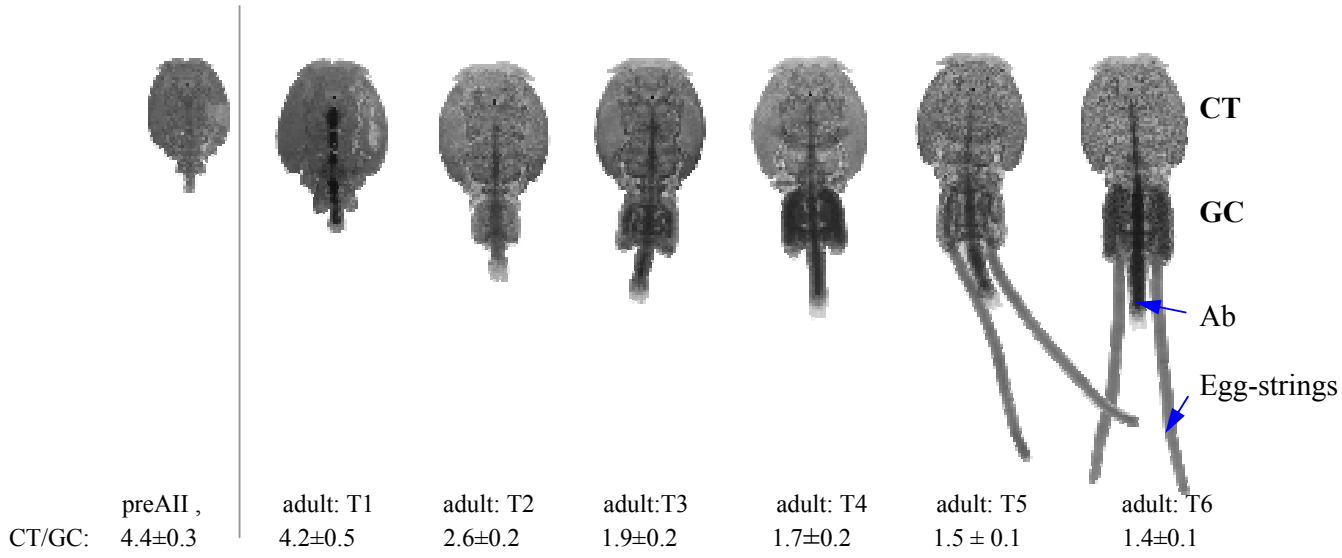

Figure 2: Development of adult female *L. salmonis* after the last molting. The size of the cephalothorax (CT) is stable during the devekopment, while the gonade complex (GC) and the Abdomen (Ab) grows as indicated by the decreasing CT/GC ratio. The time frame from T1-T6 is approximately 13-18 days at 10°C. External egg strings may be observed approximately from day 10 post molting

Microarray analysis

Microarray analysis was performed on a total of 34 female *L. salmonis*, representing the preadultII stage and 6 different adult time points after the last molting (Figure 2). From each stage we included at least 5 biological replicates. We used a self-organizing map [9, 10] (25 cluster (5x5) neuron map (Figure 5)), to group ESTs according to expression profiles. This revealed two clusters with a substantial increase in gene transcription during the first steps of adult development (clusters 1, 2, 6) and two (cluster 24 and 25) with a significant decrease. All other clusters showed very little within cluster variance.

Although the grouped samples are not clear-bordered developmental stages but floating transitions, correspondence analysis of the complete dataset revealed strong similarities among the biological parallels. With the exception of a few outliers, samples were located in adjacency with other samples from the same time point or in the vicinity of neighbouring stages (Figure 6). The developmental group membership, and the few outliers, of individual samples were also evident in a 10x10 neuron heat SOM (Figure 7). Most samples within a developmental group show a similar heat map pattern, while outlier samples resemble no other patterns or patterns from neighbouring groups (in terms of development stage).

Correspondence analysis [11] (CA) using trimmed mean values for all lice from each time point (means were trimmed to reduce the influence of outliers), shows the main direction genes are connected with the respective developmental stages (Figure 3). The two first principal axes explain 89% percent of all variation in the data set. As expected, the vectors representing the most different developmental stages are on opposite sides in the plot. The vectors corresponding to time points fall into an anti clockwise pattern for the sequential transitions between stages.

The single spots in the CA plot represent different ESTs. The further the spots are from the centre, the more strongly they are connected to the stage in which direction they lay [11]. Most genes are not differentially transcribed between stages and these ESTs are found in the centre of the correspondence analysis plot. ESTs that are clearly connected with special developmental stages are located far from the centre and in direction of one of the development stage vectors. Two major and one minor cloud of ESTs, strongly separated from the centre, are evident in Figure 3. The two main patterns (one for ESTs with expression increasing through T1-T6, and one for ESTs with decreasing expression level) accounts for 14% of all ESTS and are almost completely correlated to the first principal component. The third minor "cloud" of ESTs (n=49) strongly correlates with the PreAII group and consists of ESTs that are transcribed in the PreAII stage but down regulated after molting into the adult stage (T1).

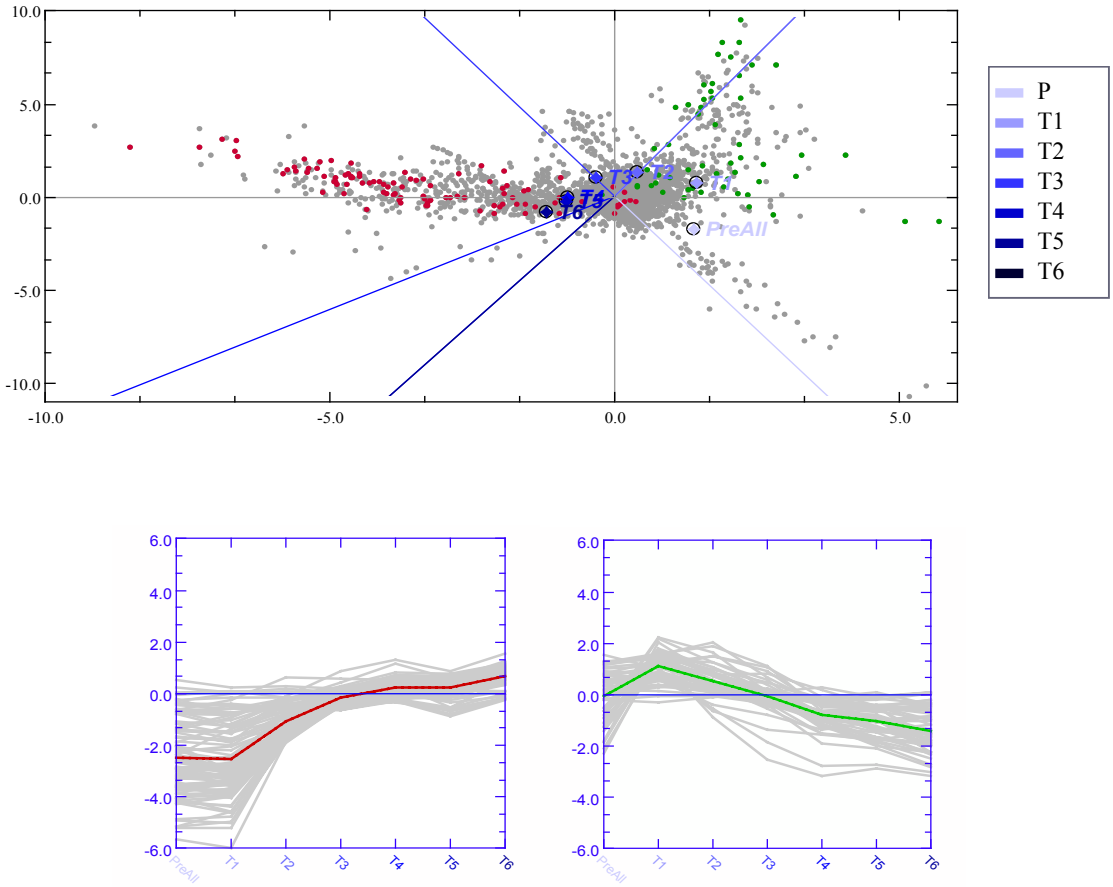

Figure 3: Correspondence analysis plot (top) and expression profiles of selected gene sets below. Spots represents individual EST and vectors represents sample group trimmed means. Two groups of annotated genes are shown in red and blue color respectively, representing the vitellogenin genes known to be involved in reproduction and genes were annotation indicates a growth related function (including cuticula proteins). The average expression profiles for these groups are shown separately, indicating the form of the two most significant components in the CA plot. Any gene located near the plot origin is poorly correlated with any of the principal axes and sample groups. Correlation to both the principal axes and sample groups increases as genes are located further from the origin in the direction of one of the axes or sample group vectors.

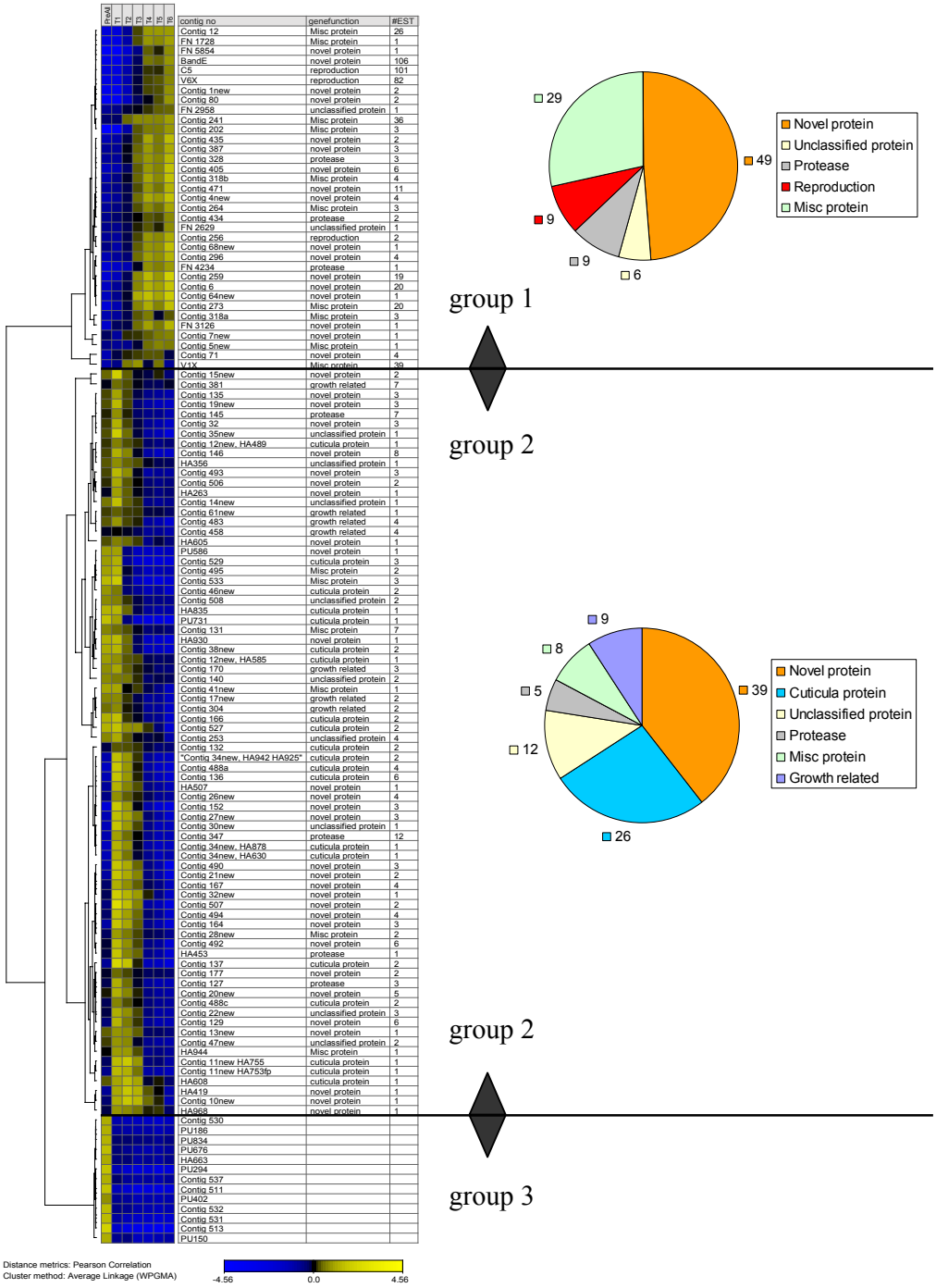

Figure 4: Hierarchical clustering of the merged contigs. The first 7 columns of the heat map represents mean group, mean EST log (sample vs. reference – mean over biological replicates and probes from the contig) signal. 3 distinct patterns can be seen in this figure; the upper sub-tree shows genes up-regulated during most of the maturation phase. The middle sub-tree shows genes turned on in T1 (while low expressed in the pre-adult phase) and continuously down-regulated trough the rest of the phases. The lower sub-tree shows genes down-regulated throughout all adult phases.

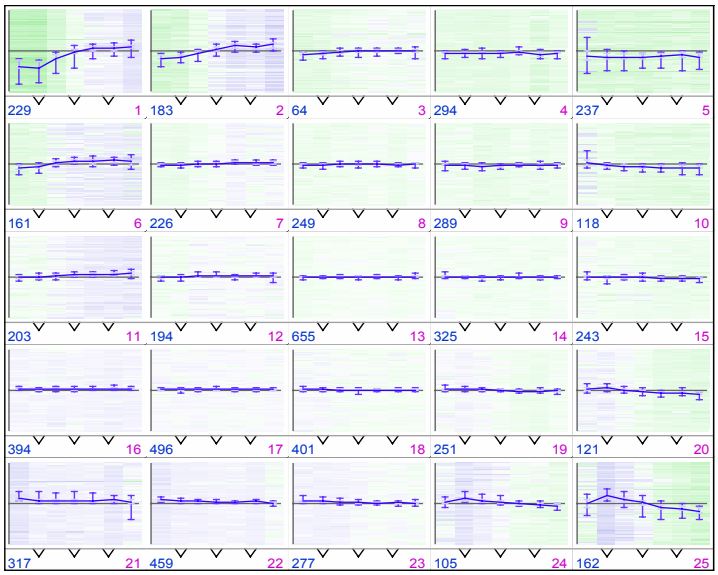

Figure 5: Grouping of expression profiles using a 5x5 self-organizing map (SOM). Interesting profiles were found in cluster 1, 2, 6, 24 and 25. The first 3 clusters show a relatively strong up-regulation while cluster 24 and 25 are down regulated during adult development (although up-regulated in regards to the PreAII sample). Dark blue colors corresponds to the mean expression profile for each cluster, dark blue bars corresponds to cluster standard deviation, light blue color bars represents cluster max-min values, the green number represents cluster size and the red number represents cluster ID.

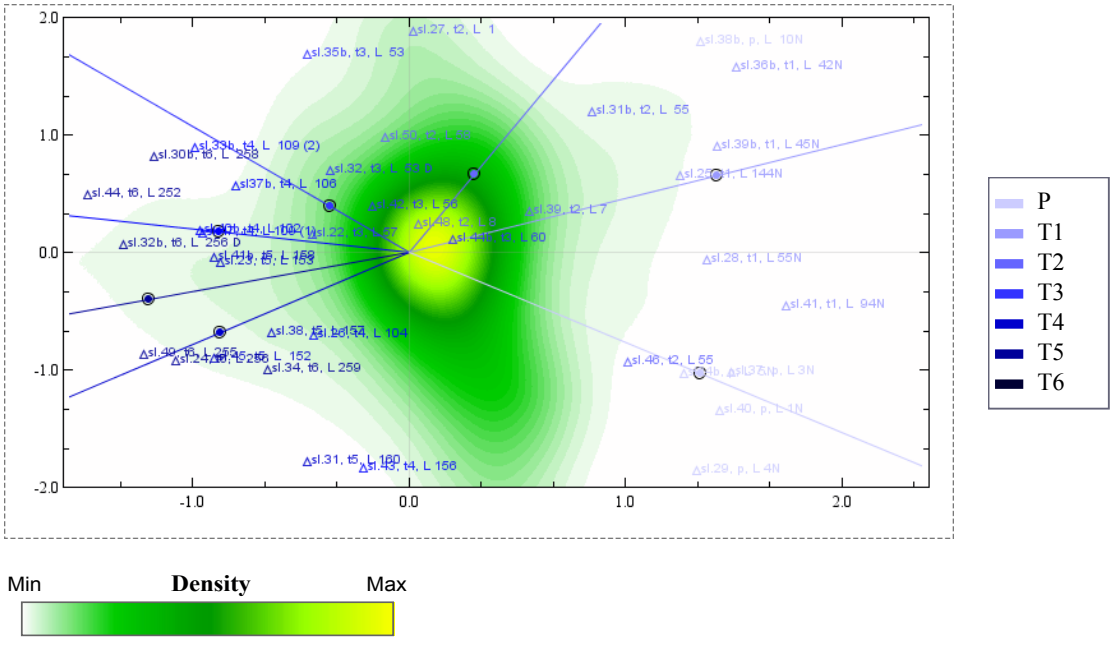

Figure 6: Correspondence analysis including all samples. Samples have color representing each development group (color codes are provided to the right). A vector showing the mean of each development group is also added. The green/yellow color in the background indicates gene density.

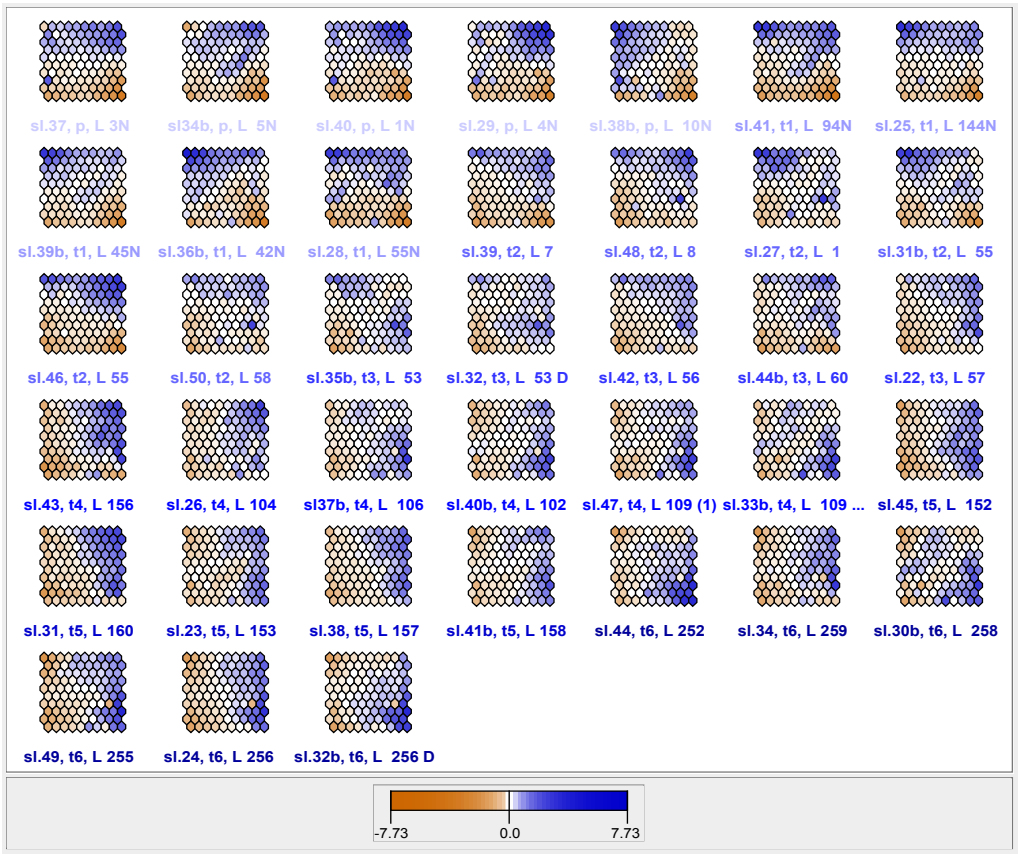

Figure 7: Self-organizing map (10x10 neuron mean heat map). For each sample, the SOM creates 10x10 cells where each cell corresponds to the mean expression value for one neuron in the map. Similarity in color pattern can be confirmed for most samples within each development group. Some outliers exist, but patterns in these often resemble patterns from groups adjacent in the time series (e.g the pattern for slide 46, t2 looks more like a t1 sample (see patterns for slide 41 and 25) which agrees with the full CA analysis, figure 6).

Group 1 defined from the 5x5 SOM (clusters 1, 2 and 5) contains 573 ESTs while group 2 (clusters 24 and 25) contains 267 ESTs. The data underlying the corresponding expression profiles were re-examined for quality (probe size double-bands, signal to noise ratio) and 414 ESTs were kept (277 up, 137 down). The analysis showed that only two of the ESTs that pass the quality criteria encode typical housekeeping genes (one 16S rRNA mt and one putative ribosomal protein). Since the array contains a high number of 16S rRNA mt probes and ESTs encoding ribosomal proteins, this demonstrates a low number of false positives from this analysis. In addition, within contigs, almost all of the individual ESTs that encode the same gene had the same transcription profile (determinded by visual inspection). Due to this low error rate all quality-checked singletons were also included in further analysis.

In addition to the two groups defined from the SOM analysis, the CA plot (Figure 3) revealed the existence of a third group of probes with a distinct expression profile corresponding to relatively high expression in pread2 and lower expression in T1-T6.

We also performed an analysis where all ESTs belonging to the same contig were combined. The two main groups contained a total of 111 contigs and singletons (of which approximately 68% were downregulated) while the third group contained 13 contigs and

singletons. All these contigs and singletons were *in silico* re-annotated after manually inspection of the sequences. ORFs larger than 100 AA were submitted to BlastP searches against the non-redundant (nr) database and the entire cDNA sequences were used in translated (BlastX) search against the nr database. The highest scoring sequences were used to indicate of the function for *L. salmonis* transcripts with particular emphasis on identify groups of biological function.

The number of novel genes was high in the major groups, 49% and 39% respectively (Figure 4). The majority of annotated genes in group 2 and 3 gave hits with cuticula proteins (26% in group 2) or other genes involved in growth related processes while in group 1 two large contigs (c552 and c276) encode two vitellogenins (hit with arthropod vitellogenins) and are obviously involved in reproduction. Furthermore contig (c256) were similar to a nucleolar protein in starfish (Asterina pectinifera) oocyte transcribed in growing oocytes only and are therefore highly likely to be involved in reproduction.

Only 1 of the 76 contigs in group 2 contained more than 10 ESTs while in group 1, 10 of the 35 contigs, consist of more than 10 ESTs indicating a generally higher transcription level in group 1 genes compared to group 2 genes. The two vitellogenin contigs contains approximately 100 ESTs which are in compliance with the high transcription level typically seen for vitellogenins that are to be deposited inn eggs. One novel gene contig in group 1 contained 106 ESTs indicating an expression level comparable to the vitellogenins. The microarray transcription profiles were confirming by Northern blot analysis using 4 genes in Group 1 (c387,c12, c318,c5new) and 5 genes in Group 2 (c508, c34new, c488a, c529, c533).

DISCUSSION

There is a large difference in the number of obtained ESTs from the different libraries with an overrepresentation of ESTs from mature adult female cDNA library. This bias is due to the nature of the present study where a main goal was to identify transcripts regulated during sexual maturation in female lice.

Approximately 6% of the ESTs encoded mitochondrial transcripts (Figure 1). Of these a large proportion (81.6 %) encoded 16S rRNA from the mitochondrium. The 16S rRNA region in *L. salmonis* mitochondria is A/T rich [12], which enables binding of the polydT primer during first strand cDNA synthesis. A similar problem, but at a larger magnitude has been seen in *Drosophila* (ref). However, since less than 5% of the ESTs in our libraries came from this mitochondrial transcript no action was taken to overcome the problem. It should be noted that in the female-male subtracted cDNA library, no mitochondrial sequences were identified.

A striking feature of the full length *L. salmonis* transcripts fully characterised during evaluation of the microarray experiments is the presence of two or more overlapping ORFs larger than the cut-off at 100 AA. Of the 111 contigs/singletons in list 2 and 3 showing significant regulation in the microarray analysis, 30 had two or more ORFs (data not shown). Two of the sequences had 4 ORFs, two on each strand and in many instances several of the ORFs gave significant hit in BlastP searches.

The two vitellogenins up regulated during the maturation of adult females, demonstrates that group 1 from the microarray analysis includes genes involved in reproduction. Furthermore the high number of ESTs in many of the contigs indicates a high transcription levels as typically seen for genes involved in reproduction. Considering that reproduction processes have been intensively studied in many organisms, the number of highly transcribed novel (49%) in group1, therefore likely to be involved in reproduction, was surprisingly high.

Induction of the genes in group 1 appears to be at timepoint T2 when the animal is still in an early stage of the post molting growth. This is in correlation with the fact that eggs are observed in the gonade compex (GC) at T3 before the GC is fully grown (Figure 2). Whether the GC stretches as a consequence of eggs entering the GC is not known but our results, and the fact that the abdomen also doubles in size, indicates that the growth of female adult salmon louse is "programmed" and that necessary genes (group 2 genes) are induced prior to the genes involved in reproduction (group 1 genes). How these processes are linked, if linked at all, remains unknown and will have to be revealed in ongoing knock down studies. In other animals egg production is often induced by external signals (blodfeeding, fertilisering artropods, copepods, etc.) but this appears not to be the case for salmon louse. In our laboratory animals it has been evident that normal development including egg production is induced even in a female only populations. However, adult males can identify females already in the pre-adult II stages and tend to identify and hold on to females in a pre-copula and deposit their sperm after the females molt into the adult stage.

A common feature of many of the genes in group 2 (down regulated from T1 to T4) was that as the annotation suggested they are involved in growth related processes (9%) or encodes proteins similar to cuticula proteins (26%) (Figure 4). All the different cuticula-like proteins were compared to each other and based on the size of the encoded protein they can be separated into two sub-groups that could be divided into several other groups based on sequence similarity. Some of the shortest proteins (HA753, HA755, contig 137, contig 529) resemble BCS-1 from *Balanus amphitrite* (barnacles) [13]. Similar short proteins are found in *Drosophila* and other insects. The *bcs* genes are believed to be involved in larval attachment and metamorphosis in barnacles [13], where as most of the insect sequences in Genbank have not been functionally characterised. Another group of proteins (Contig 488a, c, Contig 136, Contig 34 new (HA878) and HA942/HA925) are similar to a group of cuticular proteins from insects. Aligning these sequences to a selection of sequences from insects shows some degree of conservation (a few stretches of conserved amino acids) indicating similar functions. A group consisting of Contig 166 and 527 have a conserved domain, a chitin binding Peritrophin-A domain. This domain (pfam 01607.11) is particularly found in chitin binding proteins in the perotrophic matrix proteins of insects and animal chitinases.

METHODS

Animal material

A laboratory-reared strain of the salmon louse serves as the main source for animals. Infectious copepodids are produced in small buckets with flow-through water and they are allowed to infect naïve Atlantic salmon. The salmon louse is then reared on Atlantic salmon in flow through water tanks. At relevant time points after infection, the salmon will be anesthetised and salmon louse of interest removed and conserved for later use. Adult fertilized individuals of *L salmonis* were sampled during the period of maturation from preadult II to mature eggproducing adult (app 14 days), photographed and stored on RNA later® according to the manufactures recommendations. Animals were divided into six different morphological groups (T1-T6) based on the cephalotorax/gonadal complex length (mm) ratio (see figure 2) and the degree of gonadal filling/egg-string development. Five (in one case four) lice from each group were used as biological parallels in the experiment. In addition 5 preadult II females were defined as T0.

cDNA library construction and EST sequencing

Total RNA was extracted by Trizol (Sigma) and by RNeasy Mini Kit (Qiagen) followed by an enrichment of polyA RNA by Poly(A) Pure ™ or Poly(A) Purist ™ (Ambion) respectively.

Two cDNA libraries (mRNA from female intestine and mRNA from the entire mature female) were constructed in lambda ZAP as described by Kvamme et al. (2004) [7]. In addition, two cDNA libraries (pre-adult II female and adult male) were constructed directly in the pBluscript SK+ vector as described by the manufactures (Stratagene). A subtracted cDNA library was constructed by using the PCR-select cDNA subtraction kit (Clontech). Mature adult female lice were used as the source for tester cDNA where as adult males were used as driver cDNA. The lambda ZAP libraries were mass excised according to the manufactures instructions (Stratagene). All clones were blue-white screened and white clones were picked randomly from all the different libraries for plasmid purification. Bacteria were grown overnight in 96 well dishes (Millipore) and plasmids were purified according to the recommendations from the manufacturer (Millipore). Clones were sequenced used vector primers T3, m13f or m13f and BigDye chemistry (Applied Biosystems). The Vector NTI software package was used for sequence handling and analysis unless otherwise stated. After retrieval, the EST sequences were trimmed for vector sequence and quality assessment using options in the ContigExpress. Each EST was trimmed until there were less than 3 ambiguous bases per. 25 bases. All approved ESTs were further analysed by NCBI BlastX and BlastN searches in GenBank using the non-redundant database. In the primary annotaion process a significant hit was defined as an Expectation value (*E*-value) less than 5.5×10^{-5}. Based on this analysis the ESTs were annotated as similar to the sequence showing the highest score. All ESTs were then clustered by the Contig Express module in the Vector NTI package using a minimum overlap of 50 bp, an identity of 0.9 and a cut off score of 20. Prior to the clustering, identified mitochondrial genes were removed.

Microarray design and production

CDNA probes were amplified from the individual cDNA clones by PCR using pBluescript-specific primers (TAATACGACTCACTATAGGGATAGGGCGAATTGGGTACCG and TAATACGACTCACTATAGGGAAAGGGAACAAAAGCTGGAGC). PCR reactions (100µl) contained 10 µl 10x reaction buffer (Promega), 160 µM $MgCl_2$, 100 µM dNTPs, 15 µg each primer and 2.5 U Taq U polymerase (Promega). An initial 2 min denaturation was followed by 30 cycles of 94ºC (30sec), 15 sec annealing at 60ºC and 2 min elongation at 72ºC, before a final i10 min at 72ºC. The PCR products were purified using Millipore Montage^{TM} $PCR_{\mu 96}$ Plate according to manufactures instructions. All purified probes were checked for size and purity by gel-electrophoresis (Invitrogen E-gel 96well 2% (GP)). Furthermore, probe concentrations (70 ±24ng ng/µl) were estimated using Pico Green® dsDNA quantitation reagent (Molecular Probes) according to the manufactures protocol and measured on a Fluorstar Optima (BMG Labtechnoologies).
Probes (70-120ng/ul) in 50% DMSO were printed on Aminosilane coated slides (Corning® UltraGaps^{TM}) at 20-22°C and 45-55% RH using a BioRobotics, Micro Grid II arrayer (Genomic Solutions®) with Mikrospot 10K split pins. Slides were dried in a desiccator cabinet for 24-48 hours and DNA crosslinked at 350 mJ/cm2 using a UV-Stratalinker 2400, (Stratagene Inc.). The 7776 probes were printed in 24 subarrays, 7008 *L. salmonis* probes, 72 (triplicate on each subarray) *L. salmonis* reference gene *EF1α* [14], 360 were negative controls mouse ESTs (15 on each subarray) and the rest were “no well spots”.

RNA isolation, cDNA labelling for microarray hybridization

RNA was isolated for individual animals using the RNeasy Mini kit (Qiagen) according to the manufactures recommendations. The RNA was DNAse treated by TURBO DNA-free^{TM} (Ambion) according to the supplied protocol and superase (Ambion) was added. The RNA

samples were frozen at –80°C until analysis. One aliquot was used for RNA integrity and quantity measures using the Agilent 2100 Bioanalyzer and NanoDrop Spectrophotometer, respectively. Another aliquot was used for cDNA synthesis and labelling using Fair Play® Microarray Labelling Kit (Stratagene) according to the manufactures instructions. Typically, 10 μg total RNA was used for cDNA generation. In a few cases less RNA (down to 5 μl) were used due to lower RNA yields, typically from the smaller single pre-adult lice. Samples (Cy5 labelled) where hybridised against a common standard (Cy3 labelled) reverse transcribed and labelled in paralell. The standard RNA was pooled and aliquoted mRNAs from adult females, pre-adult females, adult males and pre-adult males (1:1:1:1), quality and quantity tested identical as the samples. Labelling efficiency and quantity of labelled cDNA was determined using the NanoDrop Spectrophotometer and identically quantity of sample and standard was used in all hybridisations. Slides were pre-hybridised in 20xSSC, 10% SDS and 1%BSA for about 45 min at 65ºC followed by washing twice in water and once in isopropanol. Slides were centrifuged dry in a mini-centrifuge. For probe cleanup sample and standard was unified, diluted by Tris pH 8.0 and centrifuged using microcon YM-30 columns (Millipore). After sample denaturation (100°C, 2min), hybridisation was performed at 60ºC over night with rotation using Agilent 2x hybridisation buffer (250 μl) in Agilent hybridisation chambers. The slides were set into 2xSSC/0.1%SDS at 65ºC to remove gasket slide and then washed for 5 min in 1xSSC at 65ºC, for 5 min in 0.2xSSC at RT, for 45 sec in 0.05xSSC at RT, and spinned dry (mini-centrifuge). Dye swap quality control experiments were performed with two biological samples (louse 53 and 256) and technical replicates (duplicate hybridizations) performed with five lice (louse 1, 3, 52, 55 and 109).

Microarray analysis

Slides were scanned directly after the washing procedure using an Agilent scanner at a resolution of 10μm with default settings. The scanned microarray images were analyzed using the GenePix Pro 6.0 software package and exported as image quantitation files (gpr-files). Each gpr file was then further processed in the J-Express (version 2.7) software package [15] and organized into a gene expression matrix where each row represents a clone, each column represents a sample (salmon louse) and each cell contains a $\log_2$ (sample vs. reference) ratio. The processing of image quantitation data was performed by removing (1) all "empty" and "control" spots and all spots flagged by the GenePix software package as "bad", "absent" or "not found" and (2) all spots with a signal to noise ratio (reported by GenePix) below 3.0 in both channels before a global lowess normalization was applied to all remaining probes. To prepare the gene expression matrix for further downstream analysis, we applied a KNN (k=10) [16] imputation to make sure an expression value existed for all genes in all samples. Prior to imputation, we removed all genes having more than 80% missing values, giving an expression matrix with 6653 clones. To remove ESTs not in the scope of this project, we used a self-organizing map [9, 10] to group ESTs according to expression profiles. All time courses of red/green ratios for all ESTs were sorted via self-organizing maps with Euclidian distance. The self-organizing map shows clusters of ESTs sorted by cluster variance. To identify candidate genes, which are most regulated over time the clusters with highest within cluster variance were selected. A combined dataset was created by using a trimmed group mean of samples from the same development stage. To further explore gene expression similarities corresponding to development stage, we applied a correspondence analysis to the complete dataset and to a group-combined expression matrix (samples within a development group was combined using a trimmed mean). Developmental group membership was also explored by a self-organizing map (10 x 10 neuron map). This way of using a SOM differs from the

previous version in that we can study the mean expression for each neuron in the map, for each sample and look for patterns shared between the samples [17].

All spots from the chosen clusters were quality checked after analysis. First, all probes with double bands or smear in the agarose gel electrophoresis and all short insert clones (<400 bp) were removed. Secondly, the spot quality was examined and cut off for average values of signal to noise ratio (SNR 635, 532), spot intensity (F635, F532) for both red and green channel and the percentage of featured pixels with intensities more than two standard deviations above the background intensity (at wavelength 635 or 532 (% > B635+2SD)) were selected by validating contigs with more than 20 ESTs. These cut offs were: SNR<40, F<100 and >10000, B+2SD<60. All other ESTs belonging to the same contig and that were not on the lists were quality assessed as well and compared to the other ESTs from the same contig.

Edited consensus sequences of contigs, and re-sequenced singletons (full length sequences), showing significant regulation were checked for ORFs and re-annotated. Any significant hit to functional domains and/or putative functions were included as annotation.

Northern blots

Total RNA from preAII, T1, T3, T5 and adult male lice (3.2µg) was mixed with Northern Max Formaldehyde Loading Dye (8552, Ambion), denaturated (10 min, 80°C), ethidiumbromide added and samples run on a 1% denaturating agarose gel with MOPS-buffer. Quality and quantity of RNA was evaluated under UV light before RNA was blotted onto Hybond-N nylon membrane (Amersham) using standard upward blotting technique in 10xSSC blotting buffer and crosslinked at 120 mJ/cm2 using a UV Stratalinker 2400 (Stratagene). PCR probes were produced as previously described for microarray probes and analysed on agarose gel. The individual probes were cut out and PCR product purified from gel using DNA Gelextraction (Millipore). 25ng of each PCR product was labelled with ^{32}P 6000 Ci/mmol using StripEZ®DNA (Ambion) and cleaned with QIAquick Nucleotide Removal Kit (28304 Qiagen) according to manufactures instructions. After denaturation (90°C, 10min) individual probes were hybridised to individual membranes at 68°C over night (Perfect Hyb ^{4M} Plus hybridisation buffer (Sigma), 5ml per filter). The membranes were washed with 2xSSC/0.1% SDS (2x5min RT), 1xSSC/0.1% SDS (1x15min RT), 0.5xSSC/0.1% SDS (2x10min 68°C) followed by exposure on Kodak BioMax MS for 1 to 3 days. Stripping has been done also with StripEZ®DNA (Ambion) after manufactures instructions

Acknowledgements

This project was financed by NFR grant: 153171/130 and supported by NFR's Salmon Genome Project and the functional genomics program FUGE, technology platform for microarrays. The microarray printing service was provided by the National Microarray Platform (www.microarray.no), supported by the FUGE program in the Research Council of Norway. We are grateful to Heidi Kongshaug for technical help in the laboratory.

References

1. Brooks KM: **The effects of water temperature, salinity, and currents on the survival and distribution of the infective copepodid stage of sea lice (Lepeophtheirus salmonis) originating on Atlantic salmon farms in the Broughton Archipelago of British Columbia, Canada**. *Reviews In Fisheries Science* 2005, **13**(3):177-204.
2. Grimnes A, Jakobsen PJ: **The physiological effects of salmon lice infection on post-smolt of Atlantic salmon**. *Journal Of Fish Biology* 1996, **48**(6):1179-1194.
3. Chang ES, Oconnor JD: **Invitro Secretion And Hydroxylation Of Alpha-Ecdysone As A Function Of Crustacean Molt Cycle**. *General And Comparative Endocrinology* 1978, **36**(1):151-160.
4. Ritchie G, Mordue AJ, Pike AW, Rae GH: **Morphology and ultrastructure of the reproductive system of Lepeophtheirus salmonis (Kroyer, 1837) (Copepoda: Caligidae)**. *Journal Of Crustacean Biology* 1996, **16**(2):330-346.
5. Anstensrud M: **Effects Of Mating On Female Behavior And Allometric Growth In The 2 Parasitic Copepods Lernaeocera-Branchialis (L 1767) (Pennellidae) And Lepeophtheirus-Pectoralis (Muller, 1776) (Caligidae)**. *Crustaceana* 1990, **59**:245-258.
6. Heuch PA, Nordhagen JR, Schram TA: **Egg production in the salmon louse [Lepeophtheirus salmonis (Kroyer)] in relation to origin and water temperature**. *Aquaculture Research* 2000, **31**(11):805-814.
7. Kvamme BO, Skern R, Frost P, Nilsen F: **Molecular characterisation of five trypsin-like peptidase transcripts from the salmon louse (Lepeophtheirus salmonis) intestine**. *Int J Parasitol* 2004, **34**(7):823-832.
8. Kvamme BO, Kongshaug H, Nilsen F: **Organisation of trypsin genes in the salmon louse (Lepeophtheirus salmonis, Crustacea, copepoda) genome**. *Gene* 2005, **352**:63-74.
9. Kohonen T: **Self-organizing maps**, 3rd edn. New York: Springer; 2001.
10. Toronen P, Kolehmainen M, Wong G, Castren E: **Analysis of gene expression data using self-organizing maps**. *FEBS Lett* 1999, **451**(2):142-146.
11. Fellenberg K, Hauser NC, Brors B, Neutzner A, Hoheisel JD, Vingron M: **Correspondence analysis applied to microarray data**. *Proc Natl Acad Sci U S A* 2001, **98**(19):10781-10786.
12. Tjensvoll K, Hodneland K, Nilsen F, Nylund A: **Genetic characterization of the mitochondrial DNA from Lepeophtheirus salmonis (Crustacea; Copepoda). A new gene organization revealed**. *Gene* 2005, **353**(2):218-230.
13. Okazaki Y, Shizuri Y: **Effect of inducers and inhibitors on the expression of bcs genes involved in cypris larval attachment and metamorphosis of the barnacles Balanus amphitrite**. *Int J Dev Biol* 2000, **44**(5):451-456.
14. Frost EE, Nielsen JA, Le TQ, Armstrong RC: **PDGF and FGF2 regulate oligodendrocyte progenitor responses to demyelination**. *J Neurobiol* 2003, **54**(3):457-472.
15. Dysvik B, Jonassen I: **J-Express: exploring gene expression data using Java**. *Bioinformatics* 2001, **17**(4):369-370.
16. Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, Botstein D, Altman RB: **Missing value estimation methods for DNA microarrays**. *Bioinformatics* 2001, **17**(6):520-525.

17. Wang J, Delabie J, Aasheim H, Smeland E, Myklebost O: **Clustering of the SOM easily reveals distinct gene expression patterns: results of a reanalysis of lymphoma study**. *BMC Bioinformatics* 2002, **3**:36.