The salmon louse genome: copepod features and parasitic adaptations.

Supplementary files are available here: **DOI: 10.5281/zenodo.4600850**

Rasmus Skern-Mauritzen[§a,1], Ketil Malde*[1,2], Christiane Eichner*[2], Michael Dondrup*[3], Tomasz Furmanek[1], Francois Besnier[1], Anna Zofia Komisarczuk[2], Michael Nuhn[4], Sussie Dalvin[1], Rolf B. Edvardsen[1], Sindre Grotmol[2], Egil Karlsbakk[2], Paul Kersey[4,5], Jong S. Leong[6], Kevin A. Glover[1], Sigbjørn Lien[7], Inge Jonassen[3], Ben F. Koop[6], and Frank Nilsen[§b,1,2].

[§]Corresponding authors: rasmus@hi.no[§a], frank.nilsen@uib.no[§b]

*Equally contributing authors

[1]Institute of Marine Research, Postboks 1870 Nordnes, 5817 Bergen, Norway

[2]University of Bergen, Thormøhlens Gate 53, 5006 Bergen, Norway

[3]Computational Biology Unit, Department of Informatics, University of Bergen

[4]EMBL-The European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, CB10 1SD, UK

[5] Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AE, UK

[6] Department of Biology, University of Victoria, Victoria, British Columbia, V8W 3N5, Canada

[7] Centre for Integrative Genetics (CIGENE), Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, Oluf Thesens vei 6, 1433, Ås, Norway

Abstract

Arthropods comprise the most populous groups of animals and show an astonishing diversity. Aquatic arthropods belonging to the sub-class Copepoda encompass a range of ecological roles from endo- and ectoparasites to grazers of phytoplankton that link primary producers to higher trophic levels. Despite the important role of copepods in central ecosystem services and their impact as parasites, representative genomic data and genome assemblies are scarce. This limits our opportunities to understand both the specific biology of individual species and unifying copepod genomic features that may govern their capacities to, for example, adapt to a changing environment. Among the copepod parasites we find *Lepeophtheirus salmonis,* an important ectoparasite that represents a threat to wild salmonid stocks and causes large annual losses to the salmon farming industry. Here we present the salmon louse genome - the first genome of a parasitic copepod that is fully sequenced and annotated. The 695.4 Mbp assembly was validated by a genetic linkage map and comprises 13 autosomes that recombine almost exclusively in males, one autosome that is shielded from recombination in both sexes and a ZZ-ZW style sex chromosome system. The genome assembly contains approximate 60% repetitive regions and comprise 13081 annotated predicted protein-coding genes. The predicted gene set appears to be quite complete as 92.4% of the expected Arthropod genes were found by a BUSCO. The gene annotations were validated by transcriptome sequencing that corresponds to the expected function of selected tissues. Transcriptome sequencing further revealed a marked shift in the gene expression pattern at the transition from the planktonic dispersal phase to the parasitic lifestyle after host attachment. Among other features, genes related to circadian rhythm are downregulated upon attaching to a host - probably reflecting abandoning a planktonic life with diurnal migration. The genome shows several evolutionary signatures including a large expansion of FNII domains, commonly considered vertebrate specific, and an incomplete heme homeostasis pathway suggesting that heme proteins are obtained from the host. The salmon louse has repeatedly demonstrated a large capacity to develop resistance against chemical treatments. Nonetheless, it exhibits low reduced numbers of several genes commonly involved in detoxification; cytochrome P450, ATP-Binding Cassette type transporters and Glutathione S-transferases. Interestingly, only one gene family with a putative detoxification role was expanded: the major vault proteins. Finally, the salmon louse has lost the ability to sustain peroxisomes, a loss apparently shared in the Caligid family but not among copepods in general.

Keywords:

Genome assembly, FNII expansion, Arthropod, Peroxisome, repetitive DNA, comparative genomics.

1 Introduction

Genomes may be seen as sensitive cellular organs that monitor and respond to environmental challenges [1], and consequently increasing our understanding of genomes will improve our understanding of the harboring organisms, and *vice versa*. Yet, extracting meaning from genetic code has proven challenging, and the large increase in the availability of genomes has raised as many questions as answers [2, 3]. One of the reasons for this is that while genomes constrain the range of possible phenotypes, they do not necessarily determine the phenotypes within that range [4, 5]. This flexible relationship between genotypes and phenotypes implies that interpretation of phenotype-genotype relationships requires biological data and sequence information from numerous species representing a wide range of environmental backgrounds and evolutionary lineages. The most valuable data will therefore come from taxa or ecological adaptations with previous scarce coverage. Copepods are aquatic arthropods with central ecological roles as grazers linking primary producers to higher trophic levels [6], vectors for disease [7] and parasitic pathogens [8]. Despite their widespread importance only few copepod genome assemblies have been published, and annotated genomes are, with the exception of *Caligus rogercresseyi* [9], restricted to free-living species [9-12].

Copepods date back to the Cambrian Period, and have since then diversified into numerous lineages, including several types of fish parasites such as the ectoparasitic sea lice (family Caligidae). Sea lice spend most of their life on their hosts and parasites that does so evolve physiological and behavioral adaptations to their host habitat, including the hosts immune defenses [13, 14]. Although host-parasite adaptations go both ways, parasites, including sea lice, can be deleterious to the host, as exemplified by the effects of *Lepeophtheirus salmonis*, *Caligus elongatus*, and *C. rogercresseyi* on salmonids [15-19]. While the *Caligus* species typically are generalists infecting many unrelated host species, *L. salmonis* depends exclusively on salmonid hosts for successful reproduction. Such host-parasite relationships are commonly manifested through functional adaptations reflected in the genomes [20, 21]. Hence a genome

assembly from a host specialist such as *L. salmonis* would be valuable for comparison to the recently published genome of the host generalist *C. rogercresseyi* [9].

A reliable genome assembly for *L. salmonis* will also be of instrumental value in its own right. Since its inception in the 1970's, intensive salmon farming has dramatically increased host density, which facilitate salmon lice transmission and population growth [22]. This is problematic for both wild and farmed salmonids since salmon lice cause lesions when feeding, which can entail reduced growth, osmotic imbalance, secondary infections and increased mortality [8, 23-25]. *L. salmonis* has a direct transmission cycle without intermediate hosts allowing it to respond directly to the abundance of hosts [8]. The life cycle consists of eight stages separated by molts [26, 27]. Adult males fertilize adult females which carry the eggs until they hatch into planktonic nauplius larvae. The planktonic larvae pass through two molts before reaching the infective copepodid stage. The remaining 5 stages (chalimus I and II, preadult I and II and adults) are parasitic on the host. Adult females may live for more than a 452 days [28]and continuously produce egg strings containing hundreds of eggs [29]. This life history strategy assures high fecundity and wide copepodid dispersal and may be regarded as an adaptation to historical low densities of the anadromous hosts. Because of the salmon farming driven population increase, salmon lice have until recently been controlled primarily using chemotherapeutics, but resistances have appeared and spread repeatedly [30-32] leading to a shift towards mechanical delousing strategies, such as low pressure washing and warm baths, which significantly compromise animal welfare [33]. Salmon louse infestations therefore remain a main obstacle for sustainable salmon farming, representing a threat to wild salmonid stocks and causing annual losses estimated to be around one billion EUR [29]. There has therefore been an instrumental need for a high-quality salmon louse genome assembly for studies of basal biology, studies uncovering resistance mechanisms, and studies aiming at vaccine development.

Here, we present the annotated LSalAtl2s genome assembly of the Atlantic subspecies of salmon louse, *Lepeophtheirus salmonis salmonis* [34] which has proven a valuable tool for exploring genome evolution, gene regulation and gene function in this highly adapted parasite. The genome may prove particularly valuable in conjunction with the genome of one of its hosts: the Atlantic salmon [35]. Our analysis expands the current knowledge of genome diversity in arthropods in general and in copepods in particular. It also reveals a set of features

reflecting the parasitic lifestyle, including loss of protein families, loss or reduction of metabolic pathways, and even loss of the entire peroxisome organelle.

2 Materials and Methods

2.1 Sequencing, assembly and annotation

The Ls1a strain [28] of *L. salmonis salmonis* inbred for 27 generations was sequenced to 181-fold assembly coverage in a hybrid approach using Illumina, 454 life sciences and Sanger sequencing to facilitate construction of a *de novo* assembly (see Supplementary Material section S1 for details). Several experimental assemblies of the genomic sequence data were constructed, and after an evaluation process (see Supplementary Material section S2), the final assembly process was decided on. To produce the final scaffolded LsAtl2s assembly, the 454 life sciences reads were mapped to the salmon louse mitochondrial genome [36] using BWA [37], and matching reads were removed. The remaining reads were assembled using Newbler [38], version 2.6, using the `-large` option. In order to adjust for homopolymer errors that are common artifacts of the 454 pyrosequencing process [38], the Illumina reads were mapped to the contigs and a new consensus sequence was produced using samtools mpileup [39] to collect mapping information, bcftools view `-cg` - to generate per position variant information, and vcfutils.pl vcf2fq [40] to call the consensus assembly. The assembly was scaffolded by SSPACE [41] in a series of iterations using libraries with increasing read pair distances (paired reads with 260 bps distance, then paired reads with 500 bps distance and finally mate pair reads with 3-6 kbps distance). All scaffolding was performed with parameters `-k 3 -a 0.7`, except for illumina mate pair data, where the parameters were `-k 5 -a 0.3`. The scaffolds were aligned to the Swissprot and UniProt90 databases [42], and scaffolds that were found to contain genes with bacterial annotation, and with no RNASeq mapping *and* fragmented/incomplete DNA mapping were regarded as contamination and removed. Finally, scaffolds had terminal N's removed and were filtered for length, removing all scaffolds with fewer than 200 nucleotides with mapped Illumina gDNA reads and all scaffolds shorter than 500 nucleotides without mapped Illumina gDNA reads.

Protein-coding gene models were constructed using Maker (v2.27). Firstly, a *de novo* repeat library was generated using the program RepeatModeler (v1.0.5), which was subsequently used by Maker to mask repetitive regions of the genome. To enhance gene predictions, transcriptome

data derived from the inbred salmon louse strain was used. Samples for Illumina sequencing were derived from all stages; unfertilized eggstrings, early developing eggstrings, late developing eggstrings, nauplius I, nauplius II, copepodids, chalimus I and II, preadult I and II females, preadult I and II males, adult females and adult males. Resulting RNA-seq data was mapped to the genome with Tophat (v1.3.2) and assembled using Cufflinks (v1.1.0) to provide supporting evidence for the gene build. Expressed Sequence Tag (EST) data from the parasitic copepods *L. salmonis salmonis*, *Caligus clemensi*, *C. rogercresseyi* and *Lernaeocera branchialis* [43] was used by Maker in its est2genome prediction mode to predict an initial set of genes, which was then used, in conjunction with all known *Daphnia pulex* proteins, to train the gene finder SNAP. SNAP was run on the salmon louse genome and retrained with the resulting genes. The salmon louse EST set was used to train the gene finder Augustus (v2.5.5). Maker was then run on the repeat masked sequence using the trained SNAP and Augustus programs, the EST alignment data, and all protein sequences from the phylum Arthropoda available from the UniProt Knowledgebase (downloaded 17. May 2017). A second gene set was derived by running Maker on the genome without prior repeat masking. InterProScan5 (RC7) was run to identify protein domains and to map GeneOntology terms to salmon louse genes. The final gene set comprises all genes from the first run, together with genes from the second run containing InterPro domains and not overlapping with genes from the first run. The BUSCO dataset (downloaded 22. February 2021) for Arthropoda was used to check for presence of genes that are expected to be conserved across all arthropod species. OrthoMCL (v2.0.8) was used to compare the gene set with other species. Homology relationships between salmon louse genes and genes from other species were detected using the Ensembl Compara Gene Trees pipeline [44]. KEGG Orthology assignments and KEGG pathway maps were generated by submitting the Ensembl predicted protein sequences to the KEGG Automatic Annotation Server, KAAS [45].

2.2 Repeat analysis

Comparative repeat analysis of crustaceans was done as follows: repeat families were modelled *de novo* from each genome assembly using RepeatModeler 2.0.1 [46] with the pipeline extensions for classification of LTR elements enabled. Then, RepeatMasker version 4.1.1 [47] was run in sensitive mode using the generated repeat families, both programs were used with rmblastn version 2.10.0+. For salmon lice, repeat families were generated based on the *L.*

salmonis salmonis LSalAtl2s assembly only, these were used for all four salmon louse assemblies. The following assemblies were used: *Acartia tonsa* - GCA_900241095.1 [11], *Caligus rogercresseyi* - GCA_013387185.1 [9], *Daphnia pulex* - GCA_900092285.2 [48], *Eurytemora affinis* - GCF_000591075.1 (I5K initiative, unpublished), *L. salmonis salmonis* – LsalAtl2s (this manuscript), *L. salmonis* female - GCA_001005205.1 (Leong et al., unpublished), *L. salmonis* male - GCA_001005235.1 (Leong et al., unpublished), *Leopeophtheirus salmonis onchorhynchi* (Pacific subspecies of *L. salmonis* [34])- GCA_000181255.2 (GiLS, unpublished), *Oithona nana* - GCA_900157175.1 [10], *Tigriopus californicus* - GCA_007210705.1 (Scripps Institution of Oceanography, unpublished), *Tigriopus kingsejongensis* - GCA_012959195.1 [12] and *T. kingsejongensis* [49].

A recent unpublished copepod assembly found in GenBank (*L. salmonis onchorhynchi* - GCA_016086655.1) was excluded from whole genome analyses in compliance with the responsibilities for data users set forth in the Fort Lauderdale Agreement, Section C.2 [50].

2.3 Genomic structure and recombination

Using sequence information from related project sequencing salmon louse from multiple locations in the Northern Atlantic Ocean [30] Single Nucleotide Polymorphism (SNP) markers were identified and subsequently used to construct a linkage map. The data consisted of 5098 SNP markers genotyped on 12 full sib families [30, 51], each consisting of two parents and 46 offspring. Both genotype and pedigree data were handled in LepMAP3 [52]. The data was first checked and corrected for erroneous or missing genotypes with the ParentCall2 function. SNPs were then assigned to chromosomes with the SeparateChromosomes2 function. The default parameter of SeparateChromosomes2 lead to 13 autosomes, but previous work based on a larger number of markers [53], showed that the genome of *L. salmonis* comprises 14 autosomes. The SeparateChromosomes2 function was therefor run again with a higher threshold parameter (`LODlimit=15`), identifying 14 autosomes and one sex chromosome. Finally, SNP order and sex specific recombination distances were estimated on each chromosome separately by using the OrderMarkers2 function with default parameters.

2.4 Expression analysis of an Spo11 endonuclease ortholog

Triplicate pairs of testes and ovaries from adult inbred Ls1a males and females were dissected and placed into RNAlater, extracted using Qiagen RNeasy kits and cDNA was made using

Affinity script (Agilent) and random hexamer primers. QPCR on dilution series was run using target (mRNA) *LsSpo11* and control (mRNA) *LseEF1α* assays in triplicate 10 µl reactions (for details see Supplementary Materials section S6). The PCR efficiencies of the *LsSpo11*and *LseEF1α* [54] quantitative PCR assays were comparable in the entire dilution range. The assays were further evaluated for sample or assay specific trends in efficiencies (eg. induced by PCR inhibitors) observing no such trends. Comparative analysis of six gonad samples (three testes and three ovaries) was conducted in triplicate reactions on a three step dilution series. Relative gene expression was calculated using the established *LseEF1α* standard gene [54] and ovary as calibrator tissue according to the $\Delta\Delta C_T$ method [55]. C_T values diverging by more than 0.5 cycles from the average of triplicate reactions were omitted and the 95% confidence intervals were calculated using average ΔC_T values for each of the triplicate reactions.

2.5 Phylogenomic analysis

Based on the Ensembl Compara analysis 173 1:1 orthologs were identified of the species included in the analysis. These protein sequences were used to create a concatenated sequence from each species that were aligned using default settings in ClustalX [56]. This alignment was used for phylogenetic inference using MrBayes [57] implementing a GTR + I + G model, run with 4 simultaneous chains for 1000000 generations. Full names of the included species: *Ciona intestinals*, *Danio rerio*, *Homo sapiens*, *Mus musculus*, *Schistosoma mansoni*, *Trichinella spiralis*, *Brugia malayi*, *Caenorhabditis elegans*, *Ixodes scapularis*, *D. pulex*, *L. salmonis*, *Pediculus humanus*, *Tribolium castaneum*, *Drosophila melanogaster*, *Aedes aegypti* and *Anopheles gambiae*.

2.6 Gene presence and absence

Annotations for all included species (Fig. 3) were retrieved from Ensembl (release 101) and Ensembl Metazoa (release 48) BioMart either manually or via the R package biomaRt [58]. Protein families and domain expansions and - deletions were assessed based on the *L.salmonis* InterProScan annotations compared to previous annotations stored in Ensembl. Pathway analysis was based on KEGG pathways [59]. Gene presence and absence calls were based on reciprocal best BlastP hits between the predicted *L. salmonis* proteome and UniprotKB or GenBank, requiring a best hit within the same group of orthologs or EC-number, but not necessarily the same species. To call a sequence likely absent, we also required absence of representative InterPro hits and performed reciprocal TBlastN/BlastX searches between the

Lsaltl2s assembly and UniProtKB/GenBank. Blast searches were performed using the NCBI Blast suite versions 2.6.0+ and 2.9.0+ [60, 61] with E-value threshold < 1E-6. Additional analytical steps and accessionsline number of templates used in Blast searches are described in the Supplementary Material.

2.7 Transcriptomic analysis of antennae and gut for annotation validation.

For transcriptome analysis of different tissues lice from Ls1a strain, also used for gDNA sequencing, were utilized. Ovaries, testis and intestine were dissected from adult lice. Antennae were sampled from adult females, adult males and copepodids. Copepodid antennae were taken from planktonic copepodites. Feet were dissected from both male and female adult lice. For a more detailed description see Supplementary Material section S3. In addition to the stage samples described above, attached copepodids were also sampled each day from one to six days after infection. For RNA sequencing, purified RNA from copepodids was combined from two following sampling days each. RNA was extracted and purified as described in Supplementary Material section S4. Due to low RNA quantity in the adult female antenna sample, this RNA was amplified, using the SeqPlex RNA Amplification Kit (Sigma Aldrich). The following libraries were generated: HBR-1: ovaries, pooled sample; HBR-2: testis, pooled sample; HBR-3: feet adult male and female pooled; HBR-4: antenna adult male, pooled sample; HBR-5: antenna copepodid, pooled sample; HBR-6: Antenna adult female, pooled sample; HBR-7: copepodids sampled 1 and 2 days post infection at 10 °C; HBR-8: copepodites sampled 3 and 4 days post infection at 10 °C; HBR-9: copepodids sampled 5 and 6 days post infection at 10 °C; HBR-10: intestine, adult female, pooled sample.

Library preparation and sequencing was conducted by Fasteris SA (Geneva, Switzerland). A more detailed description is available in Supplementary Materials section S3. Libraries were prepared for 50 bp single-end reads and were sequenced in multiplexed mode on Illumina HiSeq 2000 with up 10 million reads per sample. Illumina's stranded RNA-seq protocol (TruSeq, with polyA selection) with forward direction primer were used for all libraries, except for library HBR-4N which was prepared using total RNA-protocol with normalization by Duplex-specific Nuclease (DSN) due to low RNA concentration. Data from library HBR-6 is not strand specific due to the amplification process applied.

RNA-sequences were aligned and counted with respect to the Ensembl genome annotation. Samples were quality clipped and eventual adapter sequences removed using

TRIMMOMATIC [62]. Trimmed sequences were aligned against the LSalAtl2s assembly with the splice-aware aligner STAR [63] using the Ensembl transcript models for constructing the genome build. Read counts were obtained for all transcripts in the Ensembl annotation using the Bioconductor package easyRNAseq [64]. Library-size normalized read counts were calculated using the Bioconductor package edgeR [65] and are given in counts per million (CPM). Further details on methods used for creating Fig. 1 are given in Supplementary Material section S3.

2.8 Transcriptomic analysis of transition from planktonic to attached lifestyle

For planktonic *L. salmonis* copepodids, published data were used [66]. Briefly, these copepodids of the LsGulen strain were hatched from three different egg-string pairs at second day after molting to the copepodid stage. To obtain attached copepodids, farmed Atlantic salmon reared at 10 °C, were infected with copepodids from LsGulen strain. Experimental procedures were performed in accordance with Norwegian animal welfare legislation (permit ID7704, no 2010/245410). Lice were sampled from six different fish (sextuple samples) at one, three and five days after infection. For details see supplementary section S12. Library preparation, RNA sequencing and data processing was implemented in the same way as for planktonic copepodids as previously described [66]. Sequences were aligned to the LSalAtl2s assembly using STAR, and transcript counts were obtained using featureCounts [67]. Differential gene expression analysis was implemented by DESEq2 using Galaxy [68]. An adjusted p-value of 0.05 was taken as cut off.

3 Results and discussion

3.1 Assembly and annotation

The final LsalAtl2s assembly has a size of 695.4 Mbp distributed among 36095 scaffolds with a scaffold N50 of 478 Kbp (contig N50: 6044). Annotation results in prediction of 13081 protein coding and 482 non-coding RNA genes. The BUSCO dataset for Arthropoda, Metazoa and Eukaryota were used to assess completeness of the genome by checking for presence of genes that are expected to be conserved across all arthropod species. Among the genes in the BUSCO V 5.0.0 Arthoroda dataset 92.4% were complete (90.4% as single copy genes, 2.0% as duplicated genes) while 4.4 % of the BUSCO genes were missing and 3.2% were fragmented

(results for Metazoa and Eukaryota are shown in Supplementary_Table_GenomeStats). Previous transcriptomics studies that used LSalAtl2s as reference genome also reported high average mapping rates for RNA-sequencing reads (>93%) and high rates of unique mapping to a single genomic location (>88%) when using a highly sensitive splicing-aware aligner [66, 69].

We used transcriptomic data to validate the gene annotations by functionally characterizing copepodid antenna and adult intestine *L. salmonis* tissues by Gene Ontology (GO) terms: (Fig. 1). For copepodid antenna, detection and response to stimulus as well as signaling are main GO-terms, while for the adult intestine a more variable distribution of GO-terms is found. Proteolysis and hemolysis in other organism involved in symbiotic interaction, as well as in transport, are found. Through this approach, we confirmed the sensory function of antennae and the digestive function of the intestine, thus lending support to the soundness of the annotation.

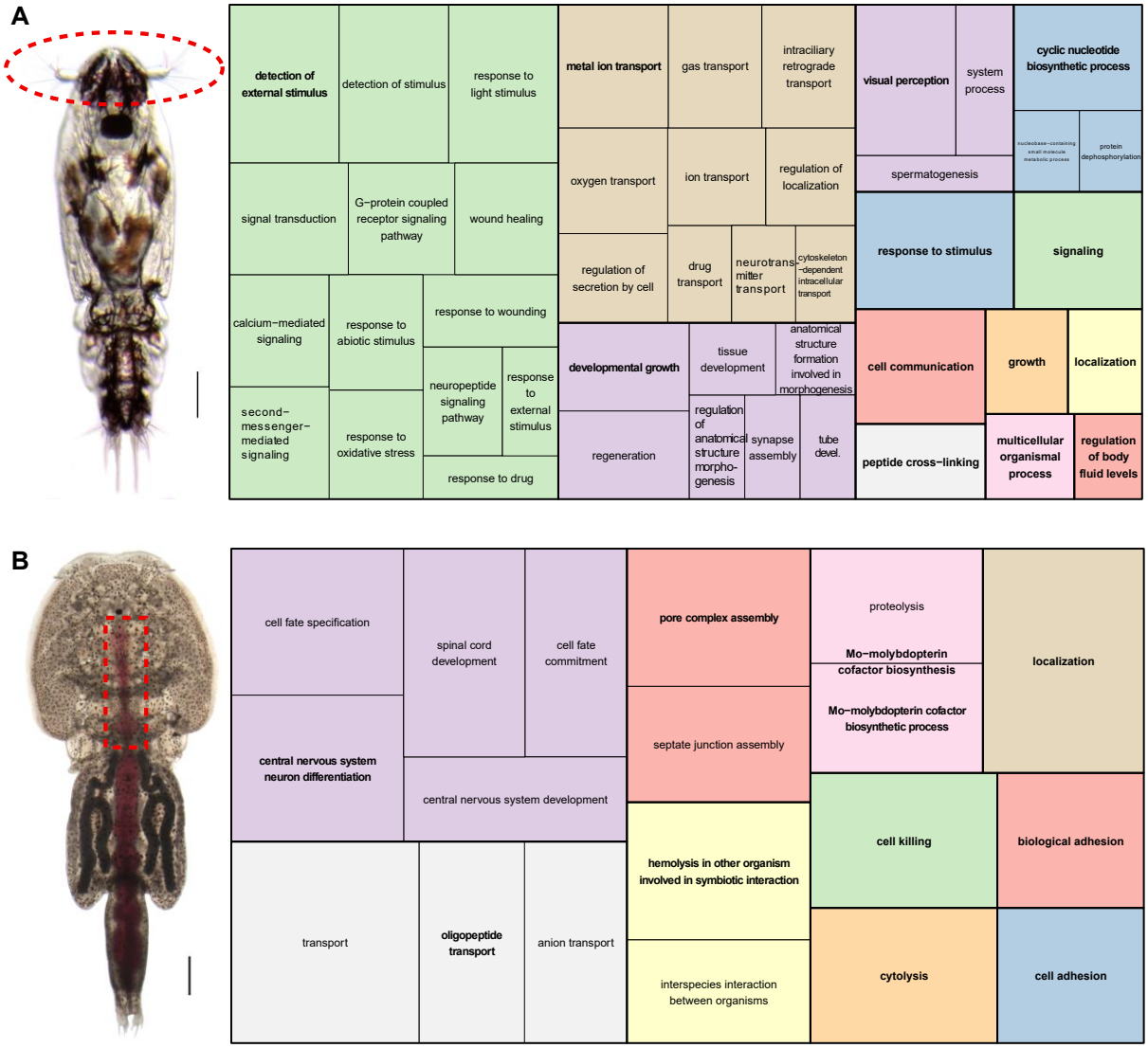


Figure 1. GO Enrichment analysis of transcripts expressed in A) copepodid antenna, B) adult intestine (anterior). The approximate dissection areas are indicated by red dashed lines on the photographs of the respective stages to the left: A) free-living copepodid, scale bar 0.1mm; B) adult female, with visible blood-filled intestinal lumen, scale bar 1mm. Significantly enriched GO-terms from the BP ontology are depicted in the form of a tree-map to the right, with clusters of semantically similar terms indicated by map colours. Centroids are typeset in bold. Note, GO-terms related to the central nervous system in B) may be caused by inclusion of ganglia that localize too close to the anterior intestinal lumen to separate.

Genome statistics for the salmon louse and selected organisms illustrate that the salmon louse has a relatively compact genome with an unexceptional number of predicted genes (Table 1). Copepod genome sizes may reach extreme values of 32 pg (≈ 32 Gb) haploid DNA per cell as found in *Paraeuchaeta norvegica* [70]. Of the examined species, the genome of the free-living

copepod *T. californicus* is more compact than the salmon louse with comparable gene content while both *T. konsejongensis* and *C. rogercresseyi* are reported to have smaller genomes and around twice as many genes (Table 1).

Table 1. Comparison of genome metrics of selected ecdysozoan species and representative assemblies. The LsalAtl2s assembly has N50 comparable to assemblies based on Sanger sequencing and generally higher N50 than pure Illumina-based assemblies. Assembly names, depth and method information was taken from GenBank or the associated publication if not available in GenBank. Scaffold N50 was computed if not found in GenBank, GC%, and Repeat% were computed from the genome sequence. Repeat content is given in percent of total bases masked by RepeatMasker. Species are (from left): Lepheoptheirus salmonius, Caligus rogercresseyi, Tigriopus kingsejongensis, Tigriopus californicus, Acartia tonsa, Pediculus humanus, Caenorhabditis elegans. An extended table with additional species, information and accession numbers is available (Supplementary_Table_GenomeStats).

Species	*L. sal.*	*C. rog.#*	*T. kin.*	*T. cal.*	*A. ton.*	*P. hum.*	*C. ele.*
Assembly Mb	695.4	478.2	338.6	191.1	989	110.8	100.3
Assembly level	Scaffold	Chromo-some	Scaffold	Chromo-some	Scaffold	Scaffold	Chromo-some
Sequencing technology	Illumina, 454, Sanger	PacBio, Hiseq4000	PacBio	Illumina, PacBio	Illumina	Sanger	Sanger
Depth	175	155	70	180	50	8.5	NA
#Scaffolds	36 095	21	938	459	351 850	1 882	7
Contig N50	6 044	38 017	NA	44 438	3 244	34 097	NA
Scaffold N50 Kbp	478	27 803	1 473	15 806	4	497	17 494
GC%	31	34	47	42	32	27	35
Repeat %	60*	57*	25*	28*	45*	20	18*
# genes	13 085	25 399	25 470	15 577	NA	10 785	20 191
BUSCO Arthropoda	C:92.4%, F:3.2%	C:61.0%, F:10.9%	C:95.5%, F:2.3%	C:93.1%, F:3.6%	C:57.8%, F:24.6%	C:97.1%, F:2.3%	NA

*De novo RepeatModeler run was used to first generate repeat families and then RepeatMasker was run with the output. In *P. humanus*, RepeatMasker was used with the public DFAM repeat library and taxon information.
#Numbers are taken from GenBank; slight deviations were found between these and the values in [9].

3.2 Extensive repetitive elements in L. salmonis.

De-novo detection by RepeatModeler yielded 4076 unique repeat families in the LSalATL2s assembly. For comparison, we included genome assemblies from seven other aquatic crustaceans, as well as all four public assemblies of *L. salmonis* (Supplementary Material Fig.

S4-1). Despite differences in sequencing depth, assembly software, and fragmentation, both GC-content in -, and the proportion of repeat families is very similar between all four *L. salmonis sp.* assemblies providing confidence that the observed repeat frequency is not affected by assembly parameters. For all assemblies analyzed, the majority of repeats could not be assigned to any known family (Supplementary Material section S4 and Fig. S4-1).

The total proportion of masked sequences was ≈60% in the *L. salmonis salmonis* LSalAtl2s assembly (≈61% - 62,5% in assemblies of *L. salmonis onchorhynchi*, see Supplementary Material section S4), making it the highest repeat content of all sequenced crustacean genomes to date. The only other available Caligid genome presently available, *C. rogercresseyi*, is second with 57%. Gallardo-Escárate [9] reported a lower repeat content (51.9%) possibly due to differences in parameter settings (e.g. running RepeatMasker in sensitive mode). Caligid repeat content is among the highest in all sequenced arthropod genomes. Notwithstanding recent advances in methods for repeat detection, Caligid repeat content is also comparable to much larger genomes, for example the Atlantic salmon (*Salmo salar*, 58–60%, 2.97 Gb [35]), locust (*Locusta migratoria*, ~60%, 6.5 Gb, [71]), and Axolotl, the largest sequenced animal genome (*Ambystoma mexicanu*, 65.6%, 18-30Gb [72]).

Evidently, *C. rogercresseyi* and *L. salmonis spp.* are rich in autonomous mobile genetic elements, DNA transposons and retro-elements, and these represent much larger fractions of the genome, than in the sequenced free-living crustaceans. The classifiable Transposable Elements (TEs) with highest copy number in *L. salmonis* belong to the Tc1 and Mariner families of transposons (Supplementary Material section S4 and Supplementary_data_TableS4-1.gz). These families are extant in various copy-numbers in eukaryote lineages, and their expansion in *L. salmonis* might have a beneficial effect on genome plasticity and evolution [73, 74]. There is growing evidence for horizontal transposon transfer (HTT) between different species, including fish and their parasites [75-77] and future studies are needed to establish whether some caligid TEs are active and potentially acquired via HTT from their host.

3.3 Genomic structure and recombination

A linkage map was constructed based on 5098 SNP markers previously identified [30]. Of these markers 5062 were assigned to linkage groups. Using Blat, 4786 markers were mapped unambiguously to 1250 scaffolds with a total size of 534 Mbp (77% of the assembly). 4127

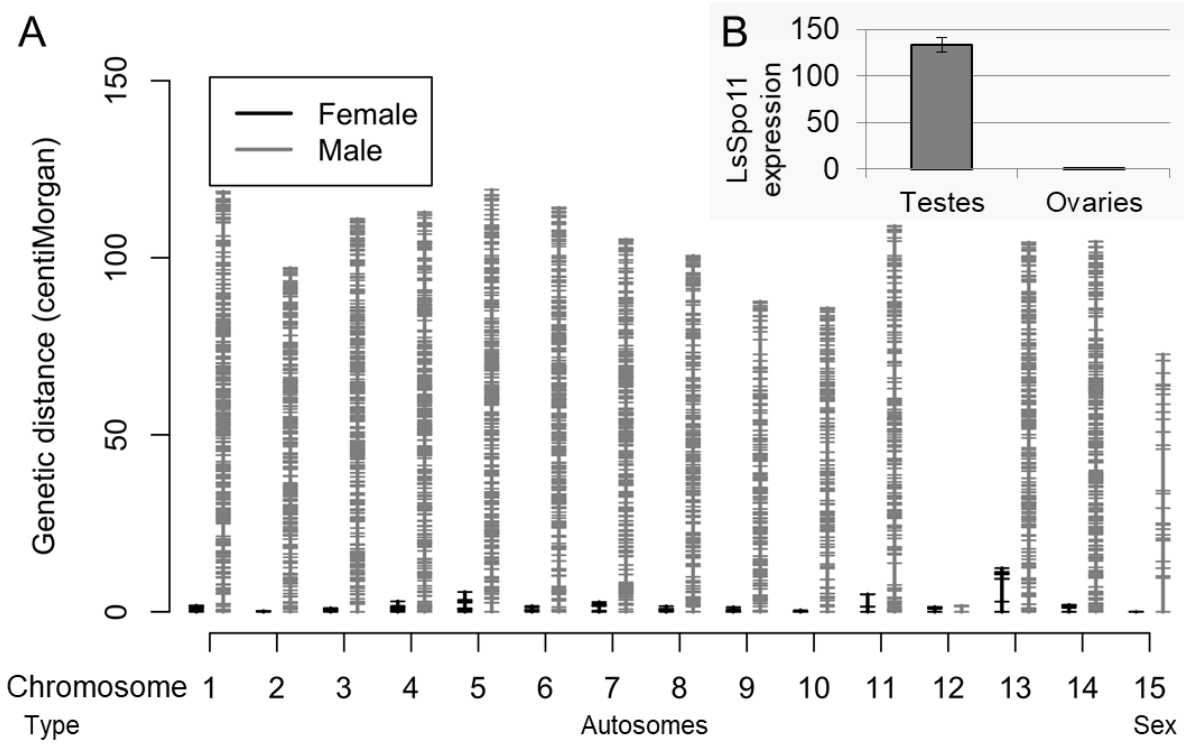
markers mapped to scaffolds with markers from only one linkage group. Of these, 398 scaffolds contained only one marker, and 777 scaffolds contained >2 markers from the same linkage group (in total 3729 markers). 75 scaffolds with a total size of 72 Mbp (10.4% of the genome assembly) contained markers from multiple linkage groups (659 markers). These results suggest that synteny of app. 10% of the assembly may be affected by errors in assembly, scaffolding or linkage group assignment.

Linkage map analysis showed that the *L. salmonis salmonis* genome comprises 15 linkage groups (LGs) ranging in size from 3 to 157 cM (Fig. 2). A total of 462 Mbp (66% of the 695.5 Mbp assembly) was assigned to linkage groups through scaffolds without ambiguous LG assignment. LG12 exhibited extremely low recombination rates in both sexes while the remaining LGs exhibit frequent recombination in males and extremely low recombination rates in females wherefore marker distances were calculated from a male map (Fig. 2). A lack of heterozygous markers was observed in LG15 in females, whereas heterozygosity was common in males for the same LG (Fig. 2B) confirming that LG15 is a sex chromosome as previously suggested [53]. These observations functionally explain the 1:1 male:female ratio in salmon lice and are compatible with both a ZZ-ZW and a ZZ-Z0 sex chromosome configuration and alternative chromosome configurations with several sex chromosomes as previously reported in both deuterostome and protostome animals [78, 79].

Independent sequencing of an Atlantic *L. salmonis* male (NCBI sequence read archive SRX976782) allowed independent mapping of male and female reads to the assembly. This showed that the majority of scaffolds received average mapping of both male and female reads indicating them to be autosomal (Supplementary Material Fig. S5-1). A smaller proportion of scaffolds received average mapping of male reads and 0.5x average mapping of female reads indicating them as a Z-type sex chromosome (present in two copies in males and a single copy in females). All of the scaffolds assigned to LG15, save one, fell in this category indicating LG15 to be a Z chromosome. Finally, a number of scaffolds received no mapping of male reads and 0.5x average mapping of female reads indicating them as belonging to a W-type sex chromosome (present as single copy in females only). This group contained no scaffolds assigned to LGs as markers not present in both sexes were omitted from the LG-analysis. In summary *L. salmonis salmonis* is heterogametic with a ZW type chromosome system possibly comprising more Z and/or W chromosomes. A single scaffold assigned to LG15 received

average mapping from both males and females suggesting it to represent a homologous region shared between the Z and W chromosomes. These findings agree with the conclusions drawn by [53].

The salmon louse is heterochiasmic as genetic recombination is predominantly found in the homogametic males. As the topoisomerase-like Spo11 is involved in genetic recombination by inducing double stranded breaks [80, 81] and is present as a single copy gene in the salmon louse genome (indicated by Blast searches against LSalAtl2s), we investigated the possible involvement of the LsSpo11 protein in the recombination profile observed. The *LsSpo11* gene is highly expressed in testes compared to ovaries with a difference exceeding 100-fold (Fig. 2B). These results are in keeping with the anticipated function of LsSpo11 in recombination and may explain the almost absent recombination in females. Although recombination in only one sex is unusual, a similar pattern is found in other species including *D. melanogaster* that has XX-XY type sex chromosomes, two regular chromosomes that recombine in the females only and a small "dot" chromosome that does not recombine in either sex [82-85]. Interestingly, in similarity to the *D. melanogaster* 'dot' chromosome the salmon louse also has a chromosome sheltered from recombination in both sexes; the chromosome corresponding to LG12 (Fig. 2 and [53]).



SNPs	498	409	505	488	453	431	362	360	261	201	193	44	399	419	39
♂cM	119	97	111	113	119	114	105	101	88	86	109	2	104	105	73
♀cM	2	1	1	3	6	2	3	2	1	0	5	1	12	2	0
Recomb.	70	81	111	39	21	71	39	67	67	215	22	1	9	55	n/a
Genes	1385	658	1088	866	1157	1066	742	890	577	394	486	215	791	672	18
Mbps	50.1	41.2	50	48.2	55	48.7	40.3	35.9	28.4	20.2	19.7	8.9	43.2	40	1.4

Figure 2 Linkage map, chromosome statistics and LsSpo11 expression. A: Linkage map showing the relative size of male and female linkage groups and marker distribution. The table shows chromosome type for each LG, male and female chromosome sizes in centimorgan (cM), the male/female recombination ratio, the number of genes located to the individual LGs, and the number of base pairs comprised in unambiguously assigned contigs (total base pairs comprised in assigned scaffolds indicated in brackets). Note that the total number of LG assigned genes and basepairs are lower than the total figures for the assembly as not all scaffolds are assigned to a LG. B: Relative expression of LsSpo11, involved in double stranded break formation during recombination.

3.4 Phylogenomic analysis.

The official gene set predicted by Ensembl comprised 13081 coding genes, and among these the Ensembl Compara pipeline identified 173 1:1 orthologs across 16 selected species that were used in a whole genome phylogenetic reconstruction (Fig. 3). *L. salmonis* groups together with

D. pulex and these two “crustacean” species occur as a sister group to the included insects where the tick (*I. scapularis*) represents the most basal arthropod branch of the included species (Fig. 3). The observed grouping pattern is in line with previous studies [86, 87]. The included parasites have fewer genes (14285 genes in average) compared to the non-parasites (19708 genes in average) which could suggest a general gene loss in the parasite species. Similar observations have been found in bacteria [88] and it could be hypothesized that parasites may require fewer genes than their free-living relatives since parasites can exploit their host for pre-processed nutrients, metabolic intermediates etc. However, parasites are also likely to have additional requirements related to parasite-host interactions and it should be considered that phylogenetic bias may influence the interpretation. Hence comprehensive studies with additional taxa should be conducted before firm conclusions regarding trends in gene numbers and parasitism are made.

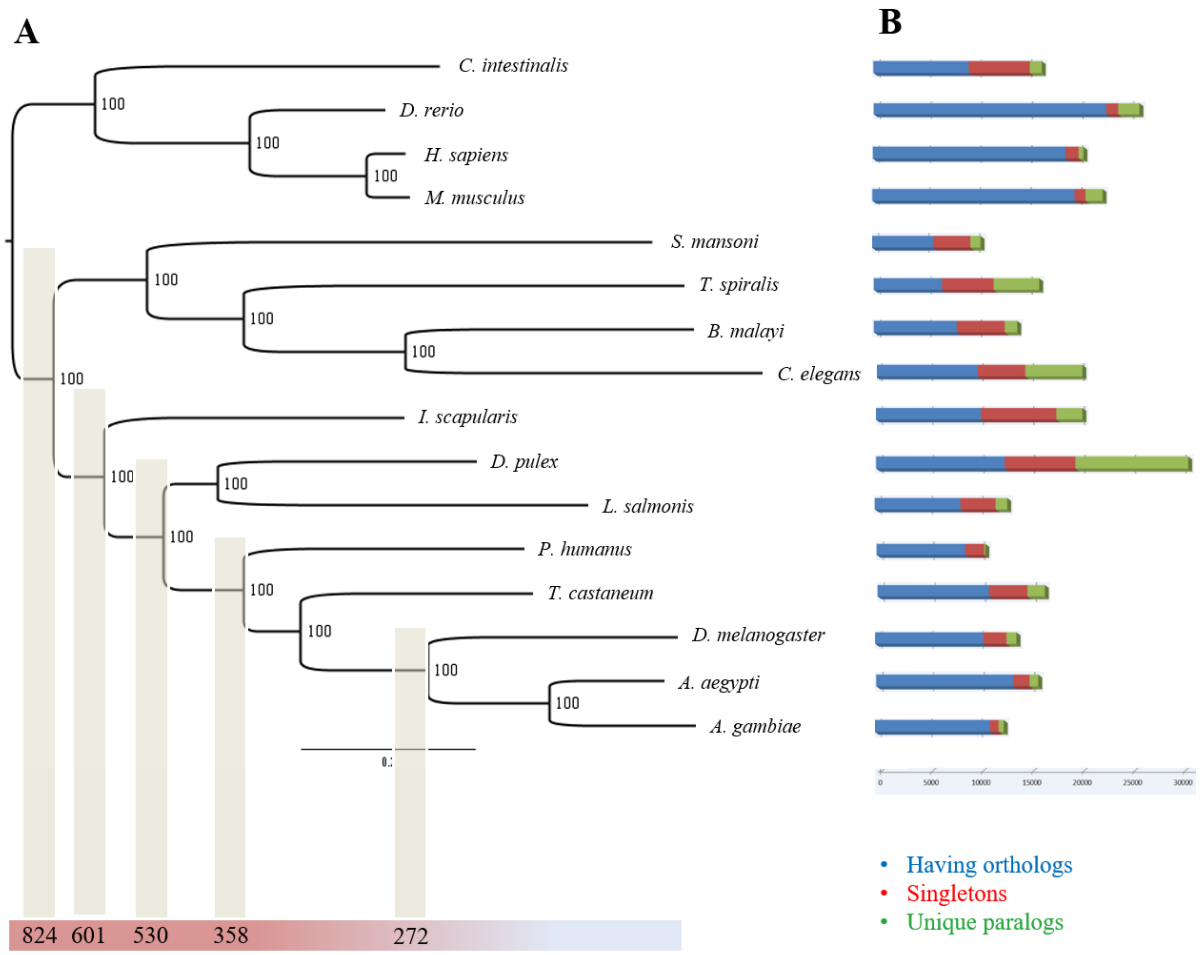


Figure 3. Whole genome phylogeny based on 173 1:1 orthologous (A) and the number of genes in each species having orthologs, singletons and unique paralogous (B). The number on the branches represents posterior probabilities and all showed 100% support for the given branching pattern. The divergence time is given below some of the main lineages and is based on estimates from TimeTree [89]. Full names of the included species: Ciona intestinals, Danio rerio, Homo sapiens, Mus musculus, Schistosoma mansoni, Trichinella spiralis, Brugia malayi, Caenorhabditis elegans, Ixodes scapularis, Daphnia pulex, Lepeophtheirus salmonis, Pediculus humanus, Tribolium castaneum, Drosophila melanogaster, Aedes aegypti and Anopheles gambiae.

3.5 Gene losses and expansions

The gene presence and absence analysis revealed some protein groups and pathway elements to be missing, and others to be found in surprisingly high or low numbers. A comprehensive overview can be found in Supplementary_Table_Compara_Genes. Among the most conspicuous findings are: A complete lack of annexins, a large expansion of SHK domains, an incomplete heme homeostasis pathway, loss of the genes needed to sustain peroxisomes, a reduction in most genes expected to be involved in detoxification, and a large number of FNII

domains - a domain previously considered to be vertebrate specific. These findings are treated in further detail below.

3.5.1 L. salmonis lacks annexins. Annexins are present in most organisms and are involved in a range of basic biological processes such as calcium metabolism, cell adhesion, growth and differentiation. Cantasessi [90] assessed 35 species of invertebrates for annexins finding these in all but the nematode *T. spiralis* and the mollusc *Oncomelania hupensis*. Surprisingly annexin domains (PF00191) were not found in the *L. salmonis* genome and could not be detected in the *C. rogercresseyi* (GCA_013387185) or *E. affinis* (GCF_000591075) genomes either. Annexin was, however, found in the *T. kingsejongensis* genome showing that its absence is not a shared copepod trait.

3.5.2 L. salmonis has a large expansion of the SHK domains (PF01549). The SHK domain was first identified in a potassium channel inhibitor from a sea anemone [91, 92]. A total of 125 SHK domains in 24 genes were identified in *L. salmonis* and represent an extreme number for any arthropod species. *T. kingsejongensis* has only two SHK domains and the cladoceran *D. pulex* has 10 SHK-domains whereas some nematodes have a large number of SHK-domains (Supplementary_Table_Compara_Domains). It has been shown that SHK binds to potassium channels in human B and T lymphocytes and prevent activation of these [93]. We speculate that the SHK-domain containing proteins in the salmon louse could be important for immunomodulation of the host and it warrants further investigation.

3.5.3 Incomplete heme homeostasis pathway. Host blood is an important part of the salmon louse diet and is rich in nutrients, among them proteins, lipids, and trace metals. Heme is an iron-containing macrocycle and an essential prosthetic group in aerobic life. Its biosynthesis is achieved via a highly conserved 8-step enzymatic pathway. It has recently been shown that 7 of the 8 genes coding for this pathway are absent from the genome of the salmon louse. Instead, it contains an intestinally expressed gene coding for a putative heme receptor linked to heme absorption [69]. Furthermore, the genome encodes genes for several requisite heme-binding proteins such as cytochrome c and a recently characterized heme-peroxidase [94] and the mitochondrial and nuclear genomes contain all components of the mitochondrial electron transfer chain. Considering that excess heme is accessible from salmon blood, we conclude that the salmon louse is a natural heme auxotroph (see Supplementary Material section S7 for details). This feature combined with the initial free-living stages of the lifecycle, calls for

further investigation into the mechanisms of heme trafficking and storage, in particular within oocytes. Homologous loci related to heme-metabolism, -binding and -trafficking detected in the genome are listed in the Supplementary Material file Supplementary_Table_Heme. Taking into account heme and porphyrin toxicity (reviewed in [95]), the absence of any known detoxification mechanism is surprising. The classical conserved two step pathway via heme oxigenase, biliverdin-reductase is lacking from the genome unlike in other arthropods. Other mechanisms of heme detoxification such as formation of hemozoin or retaining heme in the peritrophic matrix as described for hematophagous insects [96] currently lacks evidence in the salmon louse. Interestingly, out of five conserved steps of heme biosynthesis the genome contains only an orthologue of an enzyme responsible for one: Coproporphyrinogen oxidase (Cpox, EMLSAG08964). In humans, mutations in genes encoding heme biosynthesis, including CPOX, can cause autosomal dominant porphyria via accumulation of toxic intermediates of porphyrins (reviewed by [97]). The preservation of Cpox in the salmon louse may indicate a role in clearing such compounds. Finally, since *L. salmonis* lack Hemeoxygenase (Hmox) known to be responsible for Heme degradation we hypothesize that the *L. salmonis* has a hitherto unknown heme detoxification pathway or - more generally - resistance mechanism against heme toxicity.

3.5.4 L. salmonis lack peroxisomes. Peroxisomes are organelles with a common evolutionary origin and conserved traits in biogenesis [98]. Despite their shared phylogenetic origin, they are involved in metabolic processes that may diverge significantly between species and tissues [99, 100]. Four Pex genes are conserved over a wide taxonomic range and may be considered marker genes required for presence of peroxisomes: Pex3, Pex10, Pex12 and Pex19 [101]. A fifth Pex gene with bacterial orthologs, Pex5, is also ubiquitously present in peroxisome containing organisms [101]. InterProScan analysis indicated that none of the 5 core Pex genes were present in *L. salmonis*. To reassert that the core Pex genes were indeed missing in the salmon louse, the LSalAtls2s assembly was also scrutinized by Blast analysis using a large number of Pex genes as query (Supplementary Material section S6). Absence of the Pex core genes show that the salmon louse is unable to sustain peroxisomes [101]. To investigate if lack of peroxisomes is a conserved feature among copepods, we searched available genome assemblies for *C. rogercresseyi, Calanus finmarchicus, T. californicus and E. affinis* (Supplementary Material section S6). We found that both *T. californicus* (order Harpactgocoida), *E. affinis* and *C. finmarchicus* (order Calanoida) possess all core elements of

peroxisomes while the fish louse *C. rogercresseyi* (order Siphonostomatoida, same family as *L. salmonis*) lack core Pex genes (Supplementary Material section S6). This indicates that the lack of peroxisomes may be an adaptation to parasitism among members of the family Caligida, such as *L. salmonis* and *C. rogercresseyi*, or maybe even shared among members of order Siphonostomatoida in general as suggested by the lack of Pex core gene encoding ESTs from other siphonostomatoids presently represented in GenBank (*L. branchialis*; 14927 ESTs, and *C. clemensi;* 46858 ESTs). While peroxisomes are generally considered ubiquitous organelles, their absence have previously been reported in various taxa, including parasitic platyhelminths and nematodes, and even the free-living appendicularian *Oikopleura dioica* [102]. A key function of peroxisomes is to facilitate catalase mediated reduction of reactive H2O2 to oxygen and water. This capacity does not appear to be lost as a single catalase gene is retained and expressed in the salmon louse (see licebase.org; EMLSAG00000007315). This gene is, unsurprisingly, devoid of PTS1/2 peroxisome targeting signal motifs (cf. Islinger et al. [100]).

3.5.5 Peptidases in L. salmonis. We compared different types of peptidase domains across the included species (Supplementary Material section S9) with particular reference to species that are blood feeding. For most species including *L. salmonis*, serine peptidases (e.g. Trypsin PF00089) form the most abundant peptidase domain. However, there are some striking differences between the different species where the tick deviates from the other included blood feeders by having a much larger proportion of M13 peptidases (139 N-terminal domains and 106 C-terminal domains) than any other of the included species. Ticks have intracellular digestion and the high proportion of M13 peptidases could be a signature for this property [103]. On the other hand, the salmon louse has the highest proportion of M12A (astacin) peptidase domains (67 domains in total) of the five species included in the InterProScan analysis (see Supplementary_Table_Compara_Domains and Supplementary Material Figure S9-1). This is more than twice as many as for *D. pulex*.

3.5.6 Detoxification and stress-response. It has been shown that the salmon louse possesses a remarkable capacity for adaptation to most chemical delousing agents [30-32], including hydrogen peroxide [104, 105]. We therefore hypothesized a potential expansion of gene-families related to xenobiotics. Interestingly, the opposite is the case. When comparing the frequency of gene-families with putative xenobiotics related roles to other arthropods and

vertebrate species in Ensembl, three important gene-families are strongly reduced while only one is slightly expanded with a total of four members (Supplementary_Table_Detox).

Using the LSalAtl2s assembly and additional transcriptome data, Humble *et al.*, [106] reported that the family of cytochrome P450 (Cyp) genes is the most compact compared of all arthropods. Besides their role in detoxification of various substrates, enzymes of the Cyp family contribute, among others, to metabolism of steroids, fatty acids, and vitamins. Notably, all *L. salmonis* Cyp genes fall into sub-family class E, group I (IPR002401), and none is classified as class E, group II (containing major insect Cyp genes with detoxification ability) by the InterProScan analysis.

The second group of significantly underrepresented families are transporters of the ATP-Binding Cassette type (ABC). The ABC transporter family in *L. salmonis* has been extensively surveyed [107].

The third compacted gene family in the salmon louse are the Glutathione S-transferases (GSTs); one of the most important families of detoxifying enzymes in nature (reviewed by Oakley [108]). There is growing evidence that GSTs are induced by and may contribute to the clearing of synthetic pyrethroids such as cypermethrin in terrestrial arthropods [109] and aquatic crustaceans [110]. Amelioration of pyrethroid toxicity by GSTs may be due to their antioxidant capacity [111]. Despite their importance, the salmon louse genome carries only 7 and 13 hits to the conserved C- and N-terminal GST domains respectively, compared to 55 hits to the N-terminal domain (IPR004045) in *C. elegans*. Again, this is the lowest number of GSTs in the arthropod genomes in our comparison.

We found only one expanded gene-family that might be related to detoxification. We identified four paralogs of the major vault protein (Mvp), which is so far un-reported in crustaceans and hexapods. We further detected single copy orthologs in other copepods (*C. finmarchicus, T. kingsejongensis* and *T. californicus*, see Supplementary_Table_Vault_Genes and Supplementary Material section S10). Mvp is highly conserved throughout many eukaryote linages and the major constituent of vaults, the largest cytoplasmic ribonucleoprotein complex in eukaryotes known. We detected orthologs of the minor vault proteins Tep1, and several Parp-like orthologs. (Supplementary_Table_Vault_Genes). The less conserved ribonucleotide components, vault-associated RNA (vtRNA), could not be detected.

The function of vault particles is still elusive, but Mvp and other vault components have been implicated in a variety of functions, such as signaling pathways, regulation of apoptosis, autophagy, inflammation, nuclear-cytoplasmic transport, and multi-drug resistance (MDR) in cancer [112, 113]. In a recent study on *L. salmonis*, *Mvp* was upregulated together with *Cyp* genes and other stress responsive genes under cypermethrin exposure [114]. Mammalian MVP is induced by or responds to biotic and abiotic stressors and toxins in vivo and vitro [115-117]. Unlike most eukaryote genomes which harbor either none or a single copy of the *Mvp* gene, *L. salmonis* contains four paralogous Mvp-like sequences, likely stemming from gene duplication events. Implication in resistance and stress-response render Mvp and vaults interesting proteins for future studies in sea lice. However, it should be noted that contribution to MDR is debated (reviewed by Park [118]) and that inferences from cancer cells and mammalian models to an aquatic parasite should be made with caution. Therefore, more studies are needed to elucidate a potential connection of vault duplication to lifestyle or detoxification.

3.5.7 Large expansion of FN2-domains. Fibronectin II domains (FNII-domain, PF00040), commonly considered to be vertebrate specific, is – surprisingly - the largest of the *L. salmonis* expanded protein domain families with 192 domain copies. FNII domains are present in 74 annotated genes, mostly alone but regularly in combination with other domains, mainly trypsin (see [119] and Supplementary Material section S11). While FNII domains have not previously been reported in other arthropods, they are described in tubularians [120] and a few other invertebrate groups (see Supplementary Material Table S11-1). Invertebrates commonly have Kringle domains (PF00051) that have been suggested to be ancestral to FNII domains [121]. In addition to FNII domains we also identified five different proteins containing a single Kringle domain in combination with other domains in *L. salmonis*. *T. californicus*, belonging to a different order than *L. salmonis*, also has proteins with both Kringle and FNII domains indicating that the divergence of Kringle and FNII domains occurred earlier than previously suggested and probably in a common metazoan ancestor.

While the exact function(s) of genes with FNII-domains in *L. salmonis* is unknown, the expression profile for all transcripts containing FNII domains in the RNA sequencing samples (Supplementary Material Fig. S11-2) reveals that these genes are expressed at different stages and different tissues, demonstrating that the FNII-domain is widely used in the salmon louse proteome.

3.5.8 Reduced diversity in chemosensory molecules in the salmon louse. It has previously been shown that salmon lice respond to chemical cues from the Atlantic salmon host [122] and chemical sensing has been proposed to be important for host identification [123]. In Arthropods chemical sensing is mediated by gustatory receptors (GRs), odorant receptors (ORs) and ionotropic receptors (IRs) and we assessed the genome for orthologs for these proteins. No GRs or ORs were detected in the salmon louse genome whereas 26 IRs were identified (see Supplementary Material section S12) and it has been shown that some of these IRs play a crucial role in host recognition by copepodids [124]. In the Crustacea *D. pulex* genome 58 GRs were detected [125]. Furthermore, we identified nine putative GRs in *T. californicus* and two in *T. kingsejongensis* whereas we could not confidently identify any GRs in *C. rogercresseyi* (see Supplementary Material section S12). This indicates that the loss of GRs may be a signature of Caligidae or Siphonostomatoidea. As expected, many genes/sequences were detected in both *T. kingsejongensis* and *C. rogercresseyi* that contained the PF00060 and PF10613 domains (i.e. iGluR, see Supplementary Material section S12). The more conserved co-receptors (i.e. IR25a, IR8 and IR93a) were detected in both species and for *C. rogercresseyi* five specific IRs (i.e. IR328, IR329, IR333, IR337, IR339) were also found that have orthologs in *L. salmonis*. We could not detect two IR8 in *Tigriopus* or *Caligus*, indicating that this could be a *Lepeophtheirus* specific feature. However, due to the divergent nature of the IRs all sequences should be experimentally validated before firm conclusions are made (see Table S12-1 in Supplementary_Table_IR).

3.6 Gene expression during transition to a parasitic lifestyle. The infective copepodid stage can be investigated in either their non-parasitic planktonic phase or in their host bound parasitic phase. By assessing the transcriptome of planktonic versus attached copepodids we aim to compare these different lifestyles. Gene expression of planktonic copepodids were compared to the expression of parasitic copepodids collected one, three or five days after attachment. The copepodids from the different groups differ in their overall gene expression (Supplementary Material section S13). We found 6273 differently expressed genes, 3205 transcripts that are up-regulated between planktonic and attached lice consistently in all three comparisons and 3068 that are consistently down-regulated (Supplementary Material section S13). The most relevant GO-terms during this transition are shown in Fig. 4. When attaching to the host metabolism increases and sensory abilities decrease. All enriched GO annotations can be found in the Supplementary Material section S13. Planktonic copepodids do not feed but rely on

internal stores for survival until they find a host [8, 29]. The attachment, and by that the access to food, induces a switch to another nourishment and triggers the start to growth and development. The results show that sensing is strongly down regulated after attachment, while cell cycle, protein synthesis and different types of metabolism are enriched (Fig. 4). Lipid metabolism is also down regulated after attachment presumably due to the change of nutrition from stored lipids [126] to ingested host tissues. Furthermore, several copies of a gene with blast hit protein Timeless [127], known for regulating circadian rhythm in *Drosophila*, are among the down regulated transcripts. Among the downregulated genes we also find three genes known for regulating circadian rhythm in *Drosophila*; Shab, Shaw and Cycle [128, 129],. This may reflect the transformation from being a planktonic organism with diurnal migration [130] to a parasitic lifestyle. There are also several downregulated muscle related proteins (Myosin, Troponin, Tropomyosin, Actin, Titin and the regulator of muscle contraction Twitchin [131]), suggesting a shift to a more sessile live style. The upregulated genes in attached copepodids include ribosomal proteins, proteosome proteins, transcription and translation factors. The ionotropic receptors probably involved in host recognition [124] are down regulated, while genes with FNII domains (see above) are among the strongest up regulated genes in parasitic copepodids.

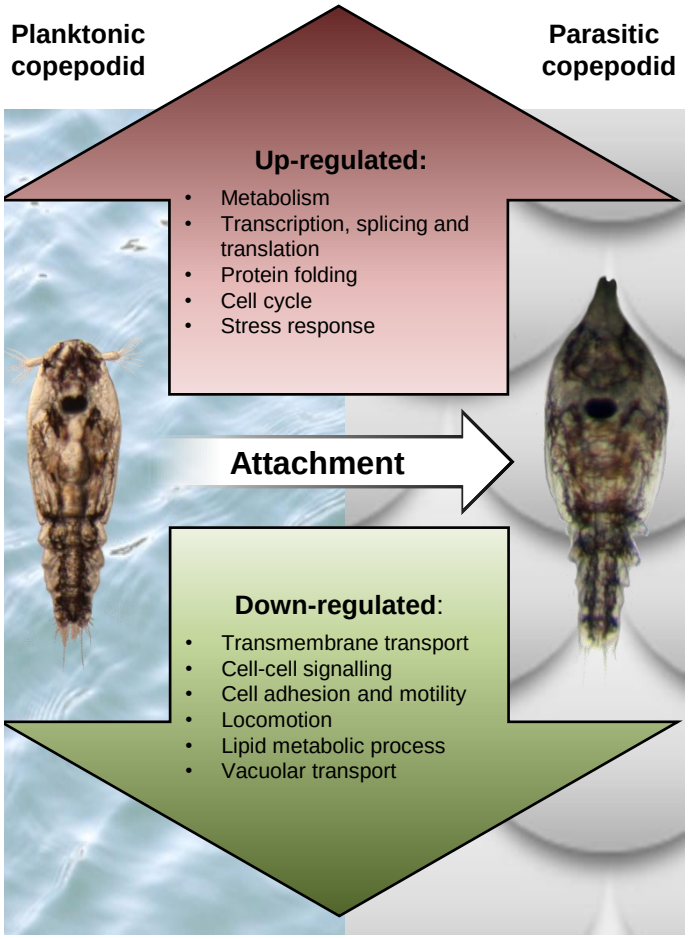


Figure 4. Main biological processes regulated in copepodids after attachment. Gene ontology enrichment in the up- and down-regulated genes in copepodids after infecting a host.

4 Conclusions

The LSalAtl2s draft genome assembly in conjunction with the gene models in Ensembl Metazoa is the first fully annotated genome of a parasitic copepod and represent an invaluable tool for studies of copepods in general and the salmon louse in particular. The genome assembly sequence and annotation have been validated by linkage analysis and RNA-sequencing and are reasonably complete with respect to the coding gene inventory. Many genes have already been curated or experimentally validated. The genome is useful as a reference for future genomics and transcriptomics studies, and together with the genome of one of its mains hosts, the Atlantic salmon, it may facilitate investigations into host-parasite interactions. Taken together, the Atlantic salmon louse genome provides a useful tool for further investigations into the genome biology and evolution of copepods, an important crustacean group that is still only sparsely covered by genome assemblies.

5 Data Availability

Sequencing raw-data and transcriptomics data have been deposited in SRA and are available in GenBank under BioProjects with accession nos. PRJNA705827, PRJNA413461 and PRJNA577842. The LSalAtl2s scaffold-level assembly and genome annotation including gene models and InterProScan and Compara results are available at Ensembl Metazoa (https://metazoa.ensembl.org/Lepeophtheirus_salmonis). An updated annotation and associated RNA-interference data are kept in LiceBase (https://licebase.org). Original output files from the comparative repeats analyses of various organisms and the BUSCO analysis are available upon request.

6 Acknowledgements

We are thankful to Dr. E. Birney (European Bioinformatics Institute) and O. Torrissen (Institute of Marine Research) for their assistance and enthusiasm. The project was funded by The Ministry of Trade and Fisheries, The Norwegian Seafood Research Fund (project 900400) and the Research Council Norway, SFI-Sea Lice Research Centre, grant number 203513/ O30. We are further thankful to Professor Richard Reinhart from Max-Planck Genome Centre for his help and extensive involvement in the early phases of the project.

References

[1] B. McClintock, The Significance of Responses of the Genome to Challenge, Science, 226 (1984) 792-801.
[2] E.A. Boyle, Y.I. Li, J.K. Pritchard, An Expanded View of Complex Traits: From Polygenic to Omnigenic, Cell, 169 (2017) 1177-1186.
[3] E. Callaway, New concerns raised over value of genome-wide disease studies, Nature, 546 (2017) 463-463.
[4] R. Noble, D. Noble, Was the Watchmaker Blind? Or Was She One-Eyed?, Biology-Basel, 6 (2017).
[5] K.N. Laland, T. Uller, M.W. Fellman, K. Sterelny, G.B. Muller, A. Moczek, E. Jablonka, J. Odling-Smee, The extended evolutionary synthesis: its structure, assumptions and predictions, P Roy Soc B-Biol Sci, 282 (2015).
[6] J.M. Aarflot, H.R. Skjoldal, P. Dalpadado, M. Skern-Mauritzen, Contribution of Calanus species to the mesozooplankton biomass in the Barents Sea, Ices J Mar Sci, 75 (2018) 2342-2354.
[7] G.C. de Magny, R. Murtugudde, M.R.P. Sapiano, A. Nizam, C.W. Brown, A.J. Busalacchi, M. Yunus, G.B. Nair, A.I. Gil, C.F. Lanata, J. Calkins, B. Manna, K. Rajendran, M.K.

Bhattacharya, A. Huq, R.B. Sack, R.R. Colwell, Environmental signatures associated with cholera epidemics, P Natl Acad Sci USA, 105 (2008) 17676-17681.
[8] A.W. Pike, S.L. Wadsworth, Sealice on salmonids: Their biology and control, Advances in parasitology, 44 (1999) 233-337.
[9] C. Gallardo-Escarate, V. Valenzuela-Munoz, G. Nunez-Acuna, D. Valenzuela-Miranda, A.T. Goncalves, H. Escobar-Sepulveda, I. Liachko, B. Nelson, S. Roberts, W. Warren, Chromosome-scale genome assembly of the sea louse Caligus rogercresseyi by SMRT sequencing and Hi-C analysis, Sci Data, 8 (2021) 60.
[10] M.A. Madoui, J. Poulain, K. Sugier, M. Wessner, B. Noel, L. Berline, K. Labadie, A. Cornils, L. Blanco-Bercial, L. Stemmann, J.L. Jamet, P. Wincker, New insights into global biogeography, population structure and natural selection from the genome of the epipelagic copepod Oithona, Mol Ecol, 26 (2017) 4467-4482.
[11] T.S. Jørgensen, B. Petersen, H.C.B. Petersen, P.D. Browne, S. Prost, J.H. Stillman, L.H. Hansen, B.W. Hansen, The Genome and mRNA Transcriptome of the Cosmopolitan Calanoid Copepod Acartia tonsa Dana Improve the Understanding of Copepod Genome Size Evolution, Genome Biol Evol, 11 (2019) 1440-1450.
[12] M.C. Lee, B.S. Choi, M.S. Kim, D.S. Yoon, J.C. Park, S. Kim, J.S. Lee, An improved genome assembly and annotation of the Antarctic copepod Tigriopus kingsejongensis and comparison of fatty acid metabolism between T. kingsejongensis and the temperate copepod T. japonicus, Comp Biochem Physiol Part D Genomics Proteomics, 35 (2020) 100703.
[13] R.M. Maizels, H.J. McSorley, Regulation of the host immune system by helminth parasites, J Allergy Clin Immun, 138 (2016) 666-675.
[14] S.K. Wikel, Modulation of the host immune system by ectoparasitic arthropods - Blood-feeding and tissue-dwelling arthropods manipulate host defenses to their advantage, Bioscience, 49 (1999) 311-320.
[15] C.W. Revie, G. Gettinby, J.W. Treasurer, G.H. Rae, The epidemiology of the sea lice, Caligus elongatus Nordmann, in marine aquaculture of Atlantic salmon, Salmo salar L., in Scotland, J Fish Dis, 25 (2002) 391-399.
[16] A. Grimnes, P.J. Jakobsen, The physiological effects of salmon lice infection on post-smolt of Atlantic salmon, J Fish Biol, 48 (1996) 1179-1194.
[17] B. Finstad, P.A. Bjorn, A. Grimnes, N.A. Hvidsten, Laboratory and field investigations of salmon lice [Lepeophtheirus salmonis (Kroyer)] infestation on Atlantic salmon (Salmo salar L.) post-smolts, Aquac Res, 31 (2000) 795-803.
[18] K. Birkeland, P.J. Jakobsen, Salmon lice, Lepeophtheirus salmonis, infestation as a causal agent of premature return to rivers and estuaries by sea trout, Salmo trutta, juveniles, Environ Biol Fish, 49 (1997) 129-137.
[19] P.A. Bjørn, B. Finstad, R. Kristoffersen, Salmon lice infection of wild sea trout and Arctic char in marine and freshwaters: the effects of salmon farms, Aquac Res, 32 (2001) 947-962.
[20] A.J. Reid, Large, rapidly evolving gene families are at the forefront of host-parasite interactions in Apicomplexa, Parasitology, 142 (2015) S57-S70.
[21] A. Coghlan, R. Tyagi, J.A. Cotton, N. Holroyd, B.A. Rosa, I.J. Tsai, D.R. Laetsch, R.N. Beech, T.A. Day, K. Hallsworth-Pepin, H.M. Ke, T.H. Kuo, T.J. Lee, J. Martin, R.M. Maizels, P. Mutowo, P. Ozersky, J. Parkinson, A.J. Reid, N.D. Rawlings, D.M. Ribeiro, L.S. Swapna, E. Stanley, D.W. Taylor, N.J. Wheeler, M. Zamanian, X. Zhang, F. Allan, J.E. Allen, K. Asano, S.A. Babayan, G. Bah, H. Beasley, H.M. Bennett, S.A. Bisset, E. Castillo, J. Cook, P.J. Cooper, T. Cruz-Bustos, C. Cuellar, E. Devaney, S.R. Doyle, M.L. Eberhard, A. Emery, K.S. Eom, J.S. Gilleard, D. Gordon, Y. Harcus, B. Harsha, J.M. Hawdon, D.E. Hill, J. Hodgkinson, P. Horak, K.L. Howe, T. Huckvale, M. Kalbe, G. Kaur, T. Kikuchi, G. Koutsovoulos, S. Kumar, A.R.

Leach, J. Lomax, B. Makepeace, J.B. Matthews, A. Muro, N.M. O'Boyle, P.D. Olson, A. Osuna, F. Partono, K. Pfarr, G. Rinaldi, P. Foronda, D. Rollinson, M.G. Samblas, H. Sato, M. Schnyder, T. Scholz, M. Shafie, V.N. Tanya, R. Toledo, A. Tracey, J.F. Urban, L.C. Wang, D. Zarlenga, M.L. Blaxter, M. Mitreva, M. Berriman, I.H.G. Consortium, Comparative genomics of the major parasitic worms, Nat Genet, 51 (2019) 163-+.
[22] T. Forseth, B.T. Barlaup, B. Finstad, P. Fiske, H. Gjoaester, M. Falkegard, A. Hindar, T.A. Mo, A.H. Rikardsen, E.B. Thorstad, L.A. Vollestad, V. Wennevik, The major threats to Atlantic salmon in Norway, Ices J Mar Sci, 74 (2017) 1496-1513.
[23] P.G. Fjelldal, T.J. Hansen, O. Karlsen, D.W. Wright, Effects of laboratory salmon louse infection on Arctic char osmoregulation, growth and survival, Conserv Physiol, 7 (2019) coz072.
[24] P.G. Fjelldal, T.J. Hansen, O. Karlsen, Effects of laboratory salmon louse infection on osmoregulation, growth and survival in Atlantic salmon, Conservation Physiology, 8 (2020).
[25] S.E. Barker, I.R. Bricknell, J. Covello, S. Purcell, M.D. Fast, W. Wolters, D.A. Bouchard, Sea lice, Lepeophtheirus salmonis (Kroyer 1837), infected Atlantic salmon (Salmo salar L.) are more susceptible to infectious salmon anemia virus, Plos One, 14 (2019) e0209178.
[26] L.A. Hamre, C. Eichner, C.M.A. Caipang, S.T. Dalvin, J.E. Bron, F. Nilsen, G. Boxshall, R. Skern-Mauritzen, The Salmon Louse Lepeophtheirus salmonis (Copepoda: Caligidae) Life Cycle Has Only Two Chalimus Stages, Plos One, 8 (2013).
[27] S.C. Johnson, L.J. Albright, The Developmental Stages of Lepeophtheirus-Salmonis (Kroyer, 1837) (Copepoda, Caligidae), Can J Zool, 69 (1991) 929-950.
[28] L.A. Hamre, K.A. Glover, F. Nilsen, Establishment and characterisation of salmon louse (Lepeophtheirus salmonis (Kroyer 1837)) laboratory strains, Parasitol Int, 58 (2009) 451-460.
[29] A.J. Brooker, R. Skern-Mauritzen, J.E. Bron, Production, mortality, and infectivity of planktonic larval sea lice, Lepeophtheirus salmonis (Kroyer, 1837): current knowledge and implications for epidemiological modelling, Ices J Mar Sci, 75 (2018) 1214-1234.
[30] F. Besnier, M. Kent, R. Skern-Mauritzen, S. Lien, K. Malde, R.B. Edvardsen, S. Taylor, L.E. Ljungfeldt, F. Nilsen, K.A. Glover, Human-induced evolution caught in action: SNP-array reveals rapid amphi-atlantic spread of pesticide resistance in the salmon ecotoparasite Lepeophtheirus salmonis, BMC Genomics, 15 (2014) 937.
[31] S.M. Aaen, K.O. Helgesen, M.J. Bakke, K. Kaur, T.E. Horsberg, Drug resistance in sea lice: a threat to salmonid aquaculture, Trends Parasitol, 31 (2015) 72-81.
[32] K. Kaur, F. Besnier, K.A. Glover, F. Nilsen, V.T. Aspehaug, H.B. Fjortoft, T.E. Horsberg, The mechanism (Phe362Tyr mutation) behind resistance in Lepeophtheirus salmonis pre-dates organophosphate use in salmon farming, Sci Rep-Uk, 7 (2017).
[33] K. Overton, T. Dempster, F. Oppedal, T.S. Kristiansen, K. Gismervik, L.H. Stien, Salmon lice treatments and salmon mortality in Norwegian aquaculture: a review, Rev Aquacult, 11 (2019) 1398-1417.
[34] R. Skern-Mauritzen, O. Torrissen, K.A. Glover, Pacific and Atlantic Lepeophtheirus salmonis (Kroyer, 1838) are allopatric subspecies: Lepeophtheirus salmonis salmonis and L. salmonis oncorhynchi subspecies novo, Bmc Genet, 15 (2014).
[35] S. Lien, B.F. Koop, S.R. Sandve, J.R. Miller, M.P. Kent, T. Nome, T.R. Hvidsten, J.S. Leong, D.R. Minkley, A. Zimin, F. Grammes, H. Grove, A. Gjuvsland, B. Walenz, R.A. Hermansen, K. von Schalburg, E.B. Rondeau, A. Di Genova, J.K. Samy, J. Olav Vik, M.D. Vigeland, L. Caler, U. Grimholt, S. Jentoft, D.I. Vage, P. de Jong, T. Moen, M. Baranski, Y. Palti, D.R. Smith, J.A. Yorke, A.J. Nederbragt, A. Tooming-Klunderud, K.S. Jakobsen, X. Jiang, D. Fan, Y. Hu, D.A. Liberles, R. Vidal, P. Iturra, S.J. Jones, I. Jonassen, A. Maass, S.W.

Omholt, W.S. Davidson, The Atlantic salmon genome provides insights into rediploidization, Nature, 533 (2016) 200-205.
[36] K. Tjensvoll, K. Hodneland, F. Nilsen, A. Nylund, Genetic characterization of the mitochondrial DNA from Lepeophtheirus salmonis (Crustacea; Copepoda). A new gene organization revealed, Gene, 353 (2005) 218-230.
[37] H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform, Bioinformatics, 25 (2009) 1754-1760.
[38] M. Margulies, M. Egholm, W.E. Altman, S. Attiya, J.S. Bader, L.A. Bemben, J. Berka, M.S. Braverman, Y.J. Chen, Z. Chen, S.B. Dewell, L. Du, J.M. Fierro, X.V. Gomes, B.C. Godwin, W. He, S. Helgesen, C.H. Ho, G.P. Irzyk, S.C. Jando, M.L. Alenquer, T.P. Jarvie, K.B. Jirage, J.B. Kim, J.R. Knight, J.R. Lanza, J.H. Leamon, S.M. Lefkowitz, M. Lei, J. Li, K.L. Lohman, H. Lu, V.B. Makhijani, K.E. McDade, M.P. McKenna, E.W. Myers, E. Nickerson, J.R. Nobile, R. Plant, B.P. Puc, M.T. Ronan, G.T. Roth, G.J. Sarkis, J.F. Simons, J.W. Simpson, M. Srinivasan, K.R. Tartaro, A. Tomasz, K.A. Vogt, G.A. Volkmer, S.H. Wang, Y. Wang, M.P. Weiner, P. Yu, R.F. Begley, J.M. Rothberg, Genome sequencing in microfabricated high-density picolitre reactors, Nature, 437 (2005) 376-380.
[39] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, S. Genome Project Data Processing, The Sequence Alignment/Map format and SAMtools, Bioinformatics, 25 (2009) 2078-2079.
[40] P. Danecek, A. Auton, G. Abecasis, C.A. Albers, E. Banks, M.A. DePristo, R.E. Handsaker, G. Lunter, G.T. Marth, S.T. Sherry, G. McVean, R. Durbin, G. Genomes Project Analysis, The variant call format and VCFtools, Bioinformatics, 27 (2011) 2156-2158.
[41] M. Boetzer, C.V. Henkel, H.J. Jansen, D. Butler, W. Pirovano, Scaffolding pre-assembled contigs using SSPACE, Bioinformatics, 27 (2011) 578-579.
[42] C. UniProt, UniProt: a worldwide hub of protein knowledge, Nucleic Acids Res, 47 (2019) D506-D515.
[43] M. Yasuike, J. Leong, S.G. Jantzen, K.R. von Schalburg, F. Nilsen, S.R.M. Jones, B.F. Koop, Genomic Resources for Sea Lice: Analysis of ESTs and Mitochondrial Genomes, Mar Biotechnol, 14 (2012) 155-166.
[44] J. Herrero, M. Muffato, K. Beal, S. Fitzgerald, L. Gordon, M. Pignatelli, A.J. Vilella, S.M.J. Searle, R. Amode, S. Brent, W. Spooner, E. Kulesha, A. Yates, P. Flicek, Ensembl comparative genomics resources, Database-Oxford, (2016).
[45] Y. Moriya, M. Itoh, S. Okuda, A.C. Yoshizawa, M. Kanehisa, KAAS: an automatic genome annotation and pathway reconstruction server, Nucleic Acids Res, 35 (2007) W182-185.
[46] J.M. Flynn, R. Hubley, C. Goubert, J. Rosen, A.G. Clark, C. Feschotte, A.F. Smit, RepeatModeler2 for automated genomic discovery of transposable element families, Proc Natl Acad Sci U S A, 117 (2020) 9451-9457.
[47] A.F.A. Smit, R. Hubley, P. Green, RepeatMasker 4.1.1. http://www.repeatmasker.org, accessed Oct 7. 2020, in, 2020.
[48] J.K. Colbourne, M.E. Pfrender, D. Gilbert, W.K. Thomas, A. Tucker, T.H. Oakley, S. Tokishita, A. Aerts, G.J. Arnold, M.K. Basu, D.J. Bauer, C.E. Caceres, L. Carmel, C. Casola, J.H. Choi, J.C. Detter, Q. Dong, S. Dusheyko, B.D. Eads, T. Frohlich, K.A. Geiler-Samerotte, D. Gerlach, P. Hatcher, S. Jogdeo, J. Krijgsveld, E.V. Kriventseva, D. Kultz, C. Laforsch, E. Lindquist, J. Lopez, J.R. Manak, J. Muller, J. Pangilinan, R.P. Patwardhan, S. Pitluck, E.J. Pritham, A. Rechtsteiner, M. Rho, I.B. Rogozin, O. Sakarya, A. Salamov, S. Schaack, H. Shapiro, Y. Shiga, C. Skalitzky, Z. Smith, A. Souvorov, W. Sung, Z. Tang, D. Tsuchiya, H. Tu, H. Vos, M. Wang, Y.I. Wolf, H. Yamagata, T. Yamada, Y. Ye, J.R. Shaw, J. Andrews, T.J.

Crease, H. Tang, S.M. Lucas, H.M. Robertson, P. Bork, E.V. Koonin, E.M. Zdobnov, I.V. Grigoriev, M. Lynch, J.L. Boore, The ecoresponsive genome of Daphnia pulex, Science, 331 (2011) 555-561.
[49] S. Kang, D.H. Ahn, J.H. Lee, S.G. Lee, S.C. Shin, J. Lee, G.S. Min, H. Lee, H.W. Kim, S. Kim, H. Park, The genome of the Antarctic-endemic copepod, Tigriopus kingsejongensis, Gigascience, 6 (2017) 1-9.
[50] Anon., Sharing Data from Large-scale Biological Research Projects: A System of Tripartite Responsibility, in, 2003.
[51] L.E.R. Ljungfeldt, P.G. Espedal, F. Nilsen, M. Skern-Mauritzen, K.A. Glover, A common-garden experiment to quantify evolutionary processes in copepods: the case of emamectin benzoate resistance in the parasitic sea louse Lepeophtheirus salmonis, Bmc Evol Biol, 14 (2014).
[52] P. Rastas, Lep-MAP3: robust linkage mapping even for low-coverage whole genome sequencing data, Bioinformatics, 33 (2017) 3726-3732.
[53] R.G. Danzmann, J.D. Norman, E.B. Rondeau, A.M. Messmer, M.P. Kent, S. Lien, O. Igboeli, M.D. Fast, B.F. Koop, A genetic linkage map for the salmon louse (Lepeophtheirus salmonis): evidence for high male:female and inter-familial recombination rate differences, Mol Genet Genomics, 294 (2019) 343-363.
[54] P. Frost, F. Nilsen, Validation of reference genes for transcription profiling in the salmon louse, Lepeophtheirus salmonis, by quantitative real-time PCR, Vet Parasitol, 118 (2003) 169-174.
[55] K.J. Livak, T.D. Schmittgen, Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method, Methods, 25 (2001) 402-408.
[56] M.A. Larkin, G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, D.G. Higgins, Clustal W and Clustal X version 2.0, Bioinformatics, 23 (2007) 2947-2948.
[57] F. Ronquist, M. Teslenko, P. van der Mark, D.L. Ayres, A. Darling, S. Hohna, B. Larget, L. Liu, M.A. Suchard, J.P. Huelsenbeck, MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space, Syst Biol, 61 (2012) 539-542.
[58] S. Durinck, P.T. Spellman, E. Birney, W. Huber, Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt, Nat Protoc, 4 (2009) 1184-1191.
[59] M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes, Nucleic Acids Res, 28 (2000) 27-30.
[60] S.F. Altschul, T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res, 25 (1997) 3389-3402.
[61] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T.L. Madden, BLAST+: architecture and applications, BMC Bioinformatics, 10 (2009) 421.
[62] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics, 30 (2014) 2114-2120.
[63] A. Dobin, C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T.R. Gingeras, STAR: ultrafast universal RNA-seq aligner, Bioinformatics, 29 (2013) 15-21.
[64] N. Delhomme, I. Padioleau, E.E. Furlong, L.M. Steinmetz, easyRNASeq: a bioconductor package for processing RNA-Seq data, Bioinformatics, 28 (2012) 2532-2533.

[65] M.D. Robinson, D.J. McCarthy, G.K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, Bioinformatics, 26 (2010) 139-140.
[66] C. Eichner, M. Dondrup, F. Nilsen, RNA sequencing reveals distinct gene expression patterns during the development of parasitic larval stages of the salmon louse (Lepeophtheirus salmonis), J Fish Dis, 41 (2018) 1005-1029.
[67] Y. Liao, G.K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features, Bioinformatics, 30 (2014) 923-930.
[68] E. Afgan, D. Baker, M. van den Beek, D. Blankenberg, D. Bouvier, M. Cech, J. Chilton, D. Clements, N. Coraor, C. Eberhard, B. Gruning, A. Guerler, J. Hillman-Jackson, G. Von Kuster, E. Rasche, N. Soranzo, N. Turaga, J. Taylor, A. Nekrutenko, J. Goecks, The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update, Nucleic Acids Res, 44 (2016) W3-W10.
[69] E.I. Heggland, C. Eichner, S.I. Stove, A. Martinez, F. Nilsen, M. Dondrup, A scavenger receptor B (CD36)-like protein is a potential mediator of intestinal heme absorption in the hematophagous ectoparasite Lepeophtheirus salmonis, Sci Rep, 9 (2019) 4218.
[70] H.P. Leinaas, M. Jalal, T.M. Gabrielsen, D.O. Hessen, Inter- and intraspecific variation in body- and genome size in calanoid copepods from temperate and arctic waters, Ecol Evol, 6 (2016) 5585-5595.
[71] X.H. Wang, X.D. Fang, P.C. Yang, X.T. Jiang, F. Jiang, D.J. Zhao, B.L. Li, F. Cui, J.N. Wei, C.A. Ma, Y.D. Wang, J. He, Y. Luo, Z.F. Wang, X.J. Guo, W. Guo, X.S. Wang, Y. Zhang, M.L. Yang, S.G. Hao, B. Chen, Z.Y. Ma, D. Yu, Z.Q. Xiong, Y.B. Zhu, D.D. Fan, L.J. Han, B. Wang, Y.X. Chen, J.W. Wang, L. Yang, W. Zhao, Y. Feng, G.X. Chen, J.M. Lian, Q.Y. Li, Z.Y. Huang, X.M. Yao, N. Lv, G.J. Zhang, Y.R. Li, J. Wang, J. Wang, B.L. Zhu, L. Kang, The locust genome provides insight into swarm formation and long-distance flight, Nat Commun, 5 (2014) 1-9.
[72] S. Nowoshilow, S. Schloissnig, J.F. Fei, A. Dahl, A.W.C. Pang, M. Pippel, S. Winkler, A.R. Hastie, G. Young, J.G. Roscito, F. Falcon, D. Knapp, S. Powell, A. Cruz, H. Cao, B. Habermann, M. Hiller, E.M. Tanaka, E.W. Myers, The axolotl genome and the evolution of key tissue formation regulators, Nature, 554 (2018) 50-55.
[73] G. Bourque, K.H. Burns, M. Gehring, V. Gorbunova, A. Seluanov, M. Hammell, M. Imbeault, Z. Izsvak, H.L. Levin, T.S. Macfarlan, D.L. Mager, C. Feschotte, Ten things you should know about transposable elements, Genome Biology, 19 (2018).
[74] C. Feschotte, E.J. Pritham, DNA transposons and the evolution of eukaryotic genomes, Annu Rev Genet, 41 (2007) 331-368.
[75] A. Pocwierz-Kotus, A. Burzynski, R. Wenne, Family of Tc1-like elements from fish genomes and horizontal transfer, Gene, 390 (2007) 243-251.
[76] S. Kuraku, H. Qiu, A. Meyer, Horizontal Transfers of Tc1 Elements between Teleost Fishes and Their Vertebrate Parasites, Lampreys, Genome Biology and Evolution, 4 (2012) 929-936.
[77] A. Palazzo, P. Lorusso, C. Miskey, O. Walisko, A. Gerbino, C.M.T. Marobbio, Z. Ivics, R.M. Marsano, Transcriptionally promiscuous "blurry" promoters in Tc1/mariner transposons allow transcription in distantly related genomes, Mobile DNA-Uk, 10 (2019).
[78] M.J.D. White, Animal Cytology and Evolution, 3. ed ed., Cambrdge University Press, Cambridge, 1973.
[79] M.J.D. White, The origin and evolution of multiple sex-chromosome mechanisms, Journal of Genetics, 40 (1940) 303-336.

[80] G. Mirzaghaderi, E. Horandl, The evolution of meiotic sex and its alternatives, P Roy Soc B-Biol Sci, 283 (2016).
[81] C.J. Underwood, K. Choi, Heterogeneous transposable elements as silencers, enhancers and targets of meiotic recombination, Chromosoma, 128 (2019) 279-296.
[82] L. Sandler, P. Szauter, The effect of recombination-defective meiotic mutants on fourth-chromosome crossing over in Drosophila melanogaster, Genetics, 90 (1978) 699-712.
[83] M.D. Adams, S.E. Celniker, R.A. Holt, C.A. Evans, J.D. Gocayne, P.G. Amanatides, S.E. Scherer, P.W. Li, R.A. Hoskins, R.F. Galle, R.A. George, S.E. Lewis, S. Richards, M. Ashburner, S.N. Henderson, G.G. Sutton, J.R. Wortman, M.D. Yandell, Q. Zhang, L.X. Chen, R.C. Brandon, Y.H. Rogers, R.G. Blazej, M. Champe, B.D. Pfeiffer, K.H. Wan, C. Doyle, E.G. Baxter, G. Helt, C.R. Nelson, G.L. Gabor, J.F. Abril, A. Agbayani, H.J. An, C. Andrews-Pfannkoch, D. Baldwin, R.M. Ballew, A. Basu, J. Baxendale, L. Bayraktaroglu, E.M. Beasley, K.Y. Beeson, P.V. Benos, B.P. Berman, D. Bhandari, S. Bolshakov, D. Borkova, M.R. Botchan, J. Bouck, P. Brokstein, P. Brottier, K.C. Burtis, D.A. Busam, H. Butler, E. Cadieu, A. Center, I. Chandra, J.M. Cherry, S. Cawley, C. Dahlke, L.B. Davenport, P. Davies, B. de Pablos, A. Delcher, Z. Deng, A.D. Mays, I. Dew, S.M. Dietz, K. Dodson, L.E. Doup, M. Downes, S. Dugan-Rocha, B.C. Dunkov, P. Dunn, K.J. Durbin, C.C. Evangelista, C. Ferraz, S. Ferriera, W. Fleischmann, C. Fosler, A.E. Gabrielian, N.S. Garg, W.M. Gelbart, K. Glasser, A. Glodek, F. Gong, J.H. Gorrell, Z. Gu, P. Guan, M. Harris, N.L. Harris, D. Harvey, T.J. Heiman, J.R. Hernandez, J. Houck, D. Hostin, K.A. Houston, T.J. Howland, M.H. Wei, C. Ibegwam, M. Jalali, F. Kalush, G.H. Karpen, Z. Ke, J.A. Kennison, K.A. Ketchum, B.E. Kimmel, C.D. Kodira, C. Kraft, S. Kravitz, D. Kulp, Z. Lai, P. Lasko, Y. Lei, A.A. Levitsky, J. Li, Z. Li, Y. Liang, X. Lin, X. Liu, B. Mattei, T.C. McIntosh, M.P. McLeod, D. McPherson, G. Merkulov, N.V. Milshina, C. Mobarry, J. Morris, A. Moshrefi, S.M. Mount, M. Moy, B. Murphy, L. Murphy, D.M. Muzny, D.L. Nelson, D.R. Nelson, K.A. Nelson, K. Nixon, D.R. Nusskern, J.M. Pacleb, M. Palazzolo, G.S. Pittman, S. Pan, J. Pollard, V. Puri, M.G. Reese, K. Reinert, K. Remington, R.D. Saunders, F. Scheeler, H. Shen, B.C. Shue, I. Siden-Kiamos, M. Simpson, M.P. Skupski, T. Smith, E. Spier, A.C. Spradling, M. Stapleton, R. Strong, E. Sun, R. Svirskas, C. Tector, R. Turner, E. Venter, A.H. Wang, X. Wang, Z.Y. Wang, D.A. Wassarman, G.M. Weinstock, J. Weissenbach, S.M. Williams, WoodageT, K.C. Worley, D. Wu, S. Yang, Q.A. Yao, J. Ye, R.F. Yeh, J.S. Zaveri, M. Zhan, G. Zhang, Q. Zhao, L. Zheng, X.H. Zheng, F.N. Zhong, W. Zhong, X. Zhou, S. Zhu, X. Zhu, H.O. Smith, R.A. Gibbs, E.W. Myers, G.M. Rubin, J.C. Venter, The genome sequence of Drosophila melanogaster, Science, 287 (2000) 2185-2195.
[84] N.C. Riddle, C.D. Shaffer, S.C.R. Elgin, A lot about a little dot - lessons learned from Drosophila melanogaster chromosome 4, Biochem Cell Biol, 87 (2009) 229-241.
[85] B. Hochman, The fourth chromosome of Drosophila melanogaster, in: M. Ashburner, E. Novitski (Eds.) The Genetics and Biology of Drosophila, Academic Press, New York, 1976, pp. 903-928.
[86] O. Rota-Stabelli, L. Campbell, H. Brinkmann, G.D. Edgecombe, S.J. Longhorn, K.J. Peterson, D. Pisani, H. Philippe, M.J. Telford, A congruent solution to arthropod phylogeny: phylogenomics, microRNAs and morphology support monophyletic Mandibulata, Proc Biol Sci, 278 (2011) 298-306.
[87] G. Giribet, G.D. Edgecombe, The Phylogeny and Evolutionary History of Arthropods, Curr Biol, 29 (2019) R592-R602.
[88] S. Sridhar, A. Sharma, H. Kongshaug, F. Nilsen, I. Jonassen, Whole genome sequencing of the fish pathogen Francisella noatunensis subsp. orientalis Toba04 gives novel insights into Francisella evolution and pathogenecity, BMC Genomics, 13 (2012) 598.

[89] S. Kumar, G. Stecher, M. Suleski, S.B. Hedges, TimeTree: A Resource for Timelines, Timetrees, and Divergence Times, Mol Biol Evol, 34 (2017) 1812-1819.
[90] C. Cantacessi, J.M. Seddon, T.L. Miller, C.Y. Leow, L. Thomas, L. Mason, C. Willis, G. Walker, A. Loukas, R.B. Gasser, M.K. Jones, A. Hofmann, A genome-wide analysis of annexins from parasitic organisms and their vectors, Sci Rep-Uk, 3 (2013).
[91] O. Castaneda, V. Sotolongo, A.M. Amor, R. Stocklin, A.J. Anderson, A.L. Harvey, A. Engstrom, C. Wernstedt, E. Karlsson, Characterization of a potassium channel toxin from the Caribbean Sea anemone Stichodactyla helianthus, Toxicon, 33 (1995) 603-613.
[92] S.J. Chang, S. Bajaj, K.G. Chandy, ShK toxin: history, structure and therapeutic applications for autoimmune diseases, WikiJournal of Science, 1 (2018).
[93] V. Chi, M.W. Pennington, R.S. Norton, E.J. Tarcha, L.M. Londono, B. Sims-Fahey, S.K. Upadhyay, J.T. Lakey, S. Iadonato, H. Wulff, C. Beeton, K.G. Chandy, Development of a sea anemone toxin as an immunomodulator for therapy of autoimmune diseases, Toxicon, 59 (2012) 529-546.
[94] A.C. Øvergard, C. Eichner, F. Nilsen, S. Dalvin, Molecular characterization and functional analysis of a salmon louse (Lepeophtheirus salmonis, Kroyer 1838) heme peroxidase with a potential role in extracellular matrixes, Comp Biochem Phys A, 206 (2017) 1-10.
[95] A.V. Graca-Souza, C. Maya-Monteiro, G.O. Paiva-Silva, G.R. Braz, M.C. Paes, M.H. Sorgine, M.F. Oliveira, P.L. Oliveira, Adaptations against heme toxicity in blood-feeding arthropods, Insect Biochem Mol Biol, 36 (2006) 322-335.
[96] V. Pascoa, P.L. Oliveira, M. Dansa-Petretski, J.R. Silva, P.H. Alvarenga, M. Jacobs-Lorena, F.J.A. Lemos, Aedes aegypti peritrophic matrix and its interaction with heme during blood digestion, Insect Biochem Molec, 32 (2002) 517-523.
[97] J.D. Phillips, Heme biosynthesis and the porphyrias, Mol Genet Metab, 128 (2019) 164-177.
[98] T. Gabaldon, Peroxisome diversity and evolution, Philos T R Soc B, 365 (2010) 765-773.
[99] L. Pieuchot, G. Jedd, Peroxisome assembly and functional diversity in eukaryotic microorganisms, Annu Rev Microbiol, 66 (2012) 237-263.
[100] M. Islinger, M.J.R. Cardoso, M. Schrader, Be different-The diversity of peroxisomes in the animal kingdom, Bba-Mol Cell Res, 1803 (2010) 881-897.
[101] A. Schlüter, S. Fourcade, R. Ripp, J.L. Mandel, O. Poch, A. Pujol, The evolutionary origin of peroxisomes: An ER-peroxisome connection, Molecular Biology and Evolution, 23 (2006) 838-845.
[102] V. Zarsky, J. Tachezy, Evolutionary loss of peroxisomes - not limited to parasites, Biol Direct, 10 (2015).
[103] D. Sojka, Z. Franta, M. Horn, C.R. Caffrey, M. Mares, P. Kopacek, New insights into the machinery of blood digestion by ticks, Trends Parasitol, 29 (2013) 276-285.
[104] J.W. Treasurer, S. Wadsworth, A. Grant, Resistance of sea lice, Lepeophtheirus salmonis (Kroyer), to hydrogen peroxide on farmed Atlantic salmon, Salmo salar L., Aquac Res, 31 (2000) 855-860.
[105] K.O. Helgesen, H. Romstad, S.M. Aaen, T.E. Horsberg, First report of reduced sensitivity towards hydrogen peroxide found in the salmon louse Lepeophtheirus salmonis in Norway, Aquacult Rep, 1 (2015) 37-42.
[106] J.L. Humble, G. Carmona-Antonanzas, C.M. McNair, D.R. Nelson, D.I. Bassett, I. Egholm, J.E. Bron, M. Bekaert, A. Sturm, Genome-wide survey of cytochrome P450 genes in the salmon louse Lepeophtheirus salmonis (Kroyer, 1837), Parasite Vector, 12 (2019).

[107] G. Carmona-Antonanzas, S.N. Carmichael, J. Heumann, J.B. Taggart, K. Gharbi, J.E. Bron, M. Bekaert, A. Sturm, A Survey of the ATP-Binding Cassette (ABC) Gene Superfamily in the Salmon Louse (Lepeophtheirus salmonis), Plos One, 10 (2015).
[108] A. Oakley, Glutathione transferases: a structural perspective, Drug Metab Rev, 43 (2011) 138-151.
[109] S.M. Valles, Toxicological and biochemical studies with field populations of the German cockroach, Blattella germanica, Pestic Biochem Phys, 62 (1998) 190-200.
[110] B.T.G. Gowland, C.F. Moffat, R.M. Stagg, D.F. Houlihan, I.M. Davies, Cypermethrin induces glutathione S-transferase activity in the shore crab, Carcinus maenas, Mar Environ Res, 54 (2002) 169-177.
[111] J.G. Vontas, G.J. Small, J. Hemingway, Glutathione S-transferases as antioxidant defence agents confer pyrethroid resistance in Nilaparvata lugens, Biochem J, 357 (2001) 65-72.
[112] W. Berger, E. Steiner, M. Grusch, L. Elbling, M. Micksche, Vaults and the major vault protein: Novel roles in signal pathway regulation and immunity, Cell Mol Life Sci, 66 (2009) 43-61.
[113] M. Amort, B. Nachbauer, S. Tuzlak, A. Kieser, A. Schepers, A. Villunger, N. Polacek, Expression of the vault RNA protects cells from undergoing apoptosis, Nat Commun, 6 (2015).
[114] J.D. Poley, L.M. Braden, A.M. Messmer, S.K. Whyte, B.F. Koop, M.D. Fast, Cypermethrin exposure induces metabolic and stress-related gene expression in copepodid salmon lice (Lepeophtheirus salmonis), Comp Biochem Phys D, 20 (2016) 74-84.
[115] Y. Tajitsu, R. Ikeda, Y. Nishizawa, H. Mataki, X.F. Che, T. Sumizawa, M. Nitta, T. Yamaguchi, M. Yamamoto, S. Tabata, S.I. Akiyama, K. Yamada, T. Furukawa, Y. Takeda, Molecular basis for the expression of major vault protein induced by hyperosmotic stress in SW620 human colon cancer cells, Int J Mol Med, 32 (2013) 703-708.
[116] S.J. Ryu, H.J. An, Y.S. Oh, H.R. Choi, M.K. Ha, S.C. Park, On the role of major vault protein in the resistance of senescent human diploid fibroblasts to apoptosis, Cell Death Differ, 15 (2008) 1673-1680.
[117] P.C. Lara, M. Pruschy, M. Zimmermann, L.A. Henriquez-Hernandez, MVP and vaults: a role in the radiation response, Radiat Oncol, 6 (2011).
[118] K. Park, The role of major vault protein (MVP) in drug resistance, J Control Release, 163 (2012) 266-266.
[119] E. Harasimczuk, A.C. Overgard, S. Grotmol, F. Nilsen, S. Dalvin, Characterization of three salmon louse (Lepeophtheirus salmonis) genes with fibronectin II domains expressed by tegumental type 1 glands, Mol Biochem Parasit, 219 (2018) 1-9.
[120] I.W. Chalmers, K.F. Hoffmann, Platyhelminth Venom Allergen-Like (VAL) proteins: revealing structural diversity, class-specific features and biological associations across the phylum, Parasitology, 139 (2012) 1231-1245.
[121] O.A. Ozhogina, M. Trexler, L. Banyai, M. Llinas, L. Patthy, Origin of fibronectin type II (FN2) modules: Structural analyses of distantly-related members of the kringle family identify the kringle domain of neurotrypsin as a potential link between FN2 domains and kringles, Protein Sci, 10 (2001) 2114-2122.
[122] G.J. Devine, A. Ingvarsdottir, W. Mordue, A.W. Pike, J. Pickett, I. Duce, A.J. Mordue, Salmon lice, Lepeophtheirus salmonis, exhibit specific chemotactic responses to semiochemicals originating from the salmonid, Salmo salar, J Chem Ecol, 26 (2000) 1833-1847.
[123] A.J. Mordue, M.A. Birkett, A review of host finding behaviour in the parasitic sea louse, Lepeophtheirus salmonis (Caligidae: Copepoda), J Fish Dis, 32 (2009) 3-13.

[124] A.Z. Komisarczuk, S. Grotmol, F. Nilsen, Ionotropic receptors signal host recognition in the salmon louse (Lepeophtheirus salmonis, Copepoda), Plos One, 12 (2017).
[125] D.C. Penalva-Arana, M. Lynch, H.M. Robertson, The Chemoreceptor Gene Family of the Waterflea Daphnia Pulex: Many GRS but No ORS, Chem Senses, 33 (2008) S41-S42.
[126] M.T. Khan, S. Dalvin, Q. Waheed, F. Nilsen, R. Male, Molecular characterization of the lipophorin receptor in the crustacean ectoparasite Lepeophtheirus salmonis, Plos One, 13 (2018).
[127] A. Sehgal, Genetic Dissection of the Circadian Clock - a Timeless Story, Semin Neurosci, 7 (1995) 27-35.
[128] P. Smith, E. Buhl, K. Tsaneva-Atanasova, J.J.L. Hodge, Shaw and Shal voltage-gated potassium channels mediate circadian changes in Drosophila clock neuron excitability, J Physiol-London, 597 (2019) 5707-5722.
[129] J.E. Rutila, V. Suri, M. Le, W.V. So, M. Rosbash, J.C. Hall, CYCLE is a second bHLH-PAS clock protein essential for circadian rhythmicity and transcription of Drosophila period and timeless, Cell, 93 (1998) 805-814.
[130] P.A. Heuch, A. Parsons, K. Boxaspen, Diel Vertical Migration - a Possible Host-Finding Mechanism in Salmon Louse (Lepeophtheirus-Salmonis) Copepodids, Can J Fish Aquat Sci, 52 (1995) 681-689.
[131] Y. Matsunaga, H. Hwang, B. Franke, R. Williams, M. Penley, H. Qadota, H. Yi, L.T. Morran, H. Lu, O. Mayans, G.M. Benian, Twitchin kinase inhibits muscle activity, Mol Biol Cell, 28 (2017) 1591-1600.