Genomics 113 (2021) 3666–3680

Contents lists available at ScienceDirect

Genomics

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

The salmon louse genome: Copepod features and parasitic adaptations

Rasmus Skern-Mauritzen [a,*], Ketil Malde [a,b,1], Christiane Eichner [b,1], Michael Dondrup [c,1], Tomasz Furmanek [a], Francois Besnier [a], Anna Zofia Komisarczuk [b], Michael Nuhn [d], Sussie Dalvin [a], Rolf B. Edvardsen [a], Sven Klages [e], Bruno Huettel [f], Kurt Stueber [f], Sindre Grotmol [b], Egil Karlsbakk [b], Paul Kersey [d,g], Jong S. Leong [h], Kevin A. Glover [a,b], Richard Reinhardt [f], Sigbjørn Lien [i], Inge Jonassen [c], Ben F. Koop [h], Frank Nilsen [a,b,*]

[a] Institute of Marine Research, Postboks 1870 Nordnes, 5817 Bergen, Norway
[b] Sea Lice Research Centre. Department of Biological Sciences, University of Bergen, Thormøhlens Gate 53, 5006 Bergen, Norway
[c] Computational Biology Unit, Department of Informatics, University of Bergen, Thormøhlens Gate 55, 5008 Bergen, Norway
[d] EMBL-The European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, CB10 1SD, UK
[e] Sequencing Core Facility, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany
[f] Max Planck Genome Centre Cologne, Carl von Linné Weg 10, D-50829 Köln, Germany
[g] Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AE, UK
[h] Department of Biology, University of Victoria, Victoria, British Columbia V8W 3N5, Canada
[i] Centre for Integrative Genetics (CIGENE), Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, Oluf Thesens vei 6, 1433 Ås, Norway

ARTICLE INFO

Keywords:
Genome assembly
FNII expansion
Arthropod
Peroxisome
Repetitive DNA
Major vault
Comparative genomics

ABSTRACT

Copepods encompass numerous ecological roles including parasites, detrivores and phytoplankton grazers. Nonetheless, copepod genome assemblies remain scarce. *Lepeophtheirus salmonis* is an economically and ecologically important ectoparasitic copepod found on salmonid fish. We present the 695.4 Mbp *L. salmonis* genome assembly containing ≈60% repetitive regions and 13,081 annotated protein-coding genes. The genome comprises 14 autosomes and a ZZ-ZW sex chromosome system. Assembly assessment identified 92.4% of the expected arthropod genes. Transcriptomics supported annotation and indicated a marked shift in gene expression after host attachment, including apparent downregulation of genes related to circadian rhythm coinciding with abandoning diurnal migration. The genome shows evolutionary signatures including loss of genes needed for peroxisome biogenesis, presence of numerous FNII domains, and an incomplete heme homeostasis pathway suggesting heme proteins to be obtained from the host. Despite repeated development of resistance against chemical treatments *L. salmonis* exhibits low numbers of many genes involved in detoxification.

1. Introduction

Genomes were portrayed as sensitive cellular organs that monitor and respond to environmental challenges by Barbara McClintock [1], and accordingly increasing our understanding of genomes will improve our understanding of the harboring organisms, and vice versa. While genomes constrain the range of possible phenotypes, they do not necessarily determine the phenotypes within that range [2,3,155]. Biological interpretation of a sequenced genome is therefore often challenging [4,5] and requires comparable genomic sequences from species representing a diversity of lifestyles and evolutionary lineages. Furthermore, comparing restricted evolutionary lineages can improve interpretation of genomic signatures that would otherwise be difficult to interpret (e.g. [6,7]). Copepods are aquatic arthropods with central ecological roles as predators [8], detrivores [9], grazers linking primary producers to higher trophic levels [10], vectors for disease [11] and parasitic pathogens [12]. Their monophyletic origin and ecological diversity make copepods ideal for comparative genomic analyses. Alas, despite their widespread importance only a limited number of copepod genome assemblies have been published and annotated. Furthermore, the available genomes are, with the exception of *Caligus rogercresseyi* [13], restricted to free-living and largely marine species (e.g. *A. tonsa*

* Corresponding authors at: Institute of Marine Research, Postboks 1870 Nordnes, 5817 Bergen, Norway.
E-mail addresses: rasmus@hi.no (R. Skern-Mauritzen), frank.nilsen@uib.no (F. Nilsen).
[1] Equally contributing authors.

https://doi.org/10.1016/j.ygeno.2021.08.002
Received 26 March 2021; Received in revised form 6 July 2021; Accepted 3 August 2021
Available online 14 August 2021

[14], *Eurytemora affinis* [15], *Oithona nana* [16], *Tigriopus kingsejongensis* [17], *Tigriopus californicus* [18] and *Tigriopus japonicus* [17,19]). Consequently, the addition of complete genome sequences for freshwater and parasitic copepods will be of particular value.

The oldest copepod fossils are from the Carboniferous period, but copepods likely date back to the Cambrian [20], and have since then diversified into numerous lineages, including several types of fish parasites such as the ectoparasitic sea lice (family Caligidae). Sea lice are parasites that spend most of their life on their hosts and are expected to have evolved physiological and behavioral adaptations to their host habitat, including the host's immune defenses [21,22]. Sea lice can be deleterious to the host, as exemplified by the effects of *Lepeophtheirus salmonis, Caligus elongatus,* and *C. rogercresseyi* on salmonids [23–27] where infections can entail reduced growth, osmotic imbalance, secondary infections and increased mortality [12,28–30]. 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 [31,32]. Hence genome assemblies from host specialists such as L. *salmonis* would be valuable for comparison to the genomes of host generalists (such as the genome of *C. rogercresseyi* [13]) since such comparisons may contribute to our understanding of general parasitic adaptations (e.g. gene numbers) as well as more host-specific metabolic adaptations.

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 facilitates salmon lice transmission and population growth [33]. Consequentially, *L. salmonis* abundance has increased and is now regarded as one of the most important factors limiting growth of salmon aquaculture, and the salmon louse infection risk is used directly in regulating farmed biomass [34,35]. *L. salmonis* has a direct transmission cycle without intermediate hosts allowing it to respond directly to the abundance of hosts [12]. The life cycle consists of eight stages separated by molts [36,37]. 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 452 days [38] and continuously produce clutches of hundreds of eggs [39]. 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. Salmon farming has driven L. *salmonis* population increases that in turn have resulted in need for salmon louse population control. This has until recently been achieved using chemotherapeutics, but resistances have appeared and spread repeatedly [40–42], leading to a shift towards mechanical delousing strategies, such as low pressure washing and warm baths, which significantly compromise animal welfare [43]. 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 [39]. There has therefore been an instrumental need for a high-quality salmon louse genome assembly for studies of general salmon louse biology, studies uncovering resistance mechanisms, and vaccine development.

Here, we present the annotated LSalAtl2s genome assembly of the Atlantic subspecies of salmon louse, *Lepeophtheirus salmonis salmonis* [44] which has proven to be 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 its hosts, e.g. Atlantic salmon [45]. 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, and loss or reduction of metabolic pathways, and even the apparent loss of the entire peroxisome organelle.

2. Materials and methods

2.1. Sequencing, assembly and annotation

The Ls1a strain [38] of *L. salmonis salmonis* inbred for 27 generations was sequenced to 181-fold assembly coverage in a hybrid approach using Illumina, 454 pyrosequencing 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 LSalAtl2s assembly, the 454 pyrosequencing reads were mapped to the salmon louse mitochondrial genome [46] using BWA [47], and matching reads were removed. The remaining reads were assembled using Newbler [48], version 2.6, using the `-large` option. In order to adjust for homopolymer errors that are common artifacts of the 454 pyrosequencing process [48], the Illumina reads were mapped to the contigs and a new consensus sequence was produced using samtools mpileup [49] to collect mapping information, bcftools view `-cg` - to generate per position variant information, and vcfutils.pl vcf2fq [50] to call the consensus assembly. The assembly was scaffolded by SSPACE [51] 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 [52], and scaffolds that were found to contain genes with bacterial annotation, and with no RNA-seq mapping *and* fragmented or incomplete mapping of Illumina gDNA reads 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 genomic DNA (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 eggs, early developing eggs (pooled eggs obtained 0–24 h after fertilization), late developing eggs (pooled eggs obtained 2–7 days after fertilization), 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 were 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* [53] were 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 2013). A second gene set was derived by running Maker on the genome without prior repeat masking. InterProScan 5 (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 Benchmarking Universal Single-Copy Orthologs (BUSCO) V 5.0.0 datasets (downloaded 22. February 2021) for Arthropoda, Metazoa, and Eukarya (arthropoda_odb10,

metazoa_odb10, eukarya_odb10) were used to check for presence of genes that are expected to be conserved across the included lineages. 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 [54]. KEGG Orthology assignments and KEGG pathway maps were generated by submitting the Ensembl predicted protein sequences to the KEGG Automatic Annotation Server (KAAS) [55].

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 [56] with the pipeline extensions for classification of LTR elements enabled. Then, RepeatMasker version 4.1.1 [57] 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: *A. tonsa* - GCA_900241095.1 [14], *Caligus rogercresseyi* - GCA_013387185.1 [13], *Daphnia pulex* - GCA_900092285.2 [58], *Eurytemora affinis* - GCF_000591075.1 [15], *L. salmonis salmonis* – LSalAtl2s (this manuscript), *L. salmonis salmonis* female - GCA_001005205.1 (Leong et al., unpublished), *L. salmonis salmonis* male - GCA_001005235.1 (Leong et al., unpublished), *Leopeophtheirus salmonis onchorhynchi* (Pacific subspecies of L. *salmonis* [44]) - GCA_000181255.2 (GiLS, unpublished), *Oithona nana* - GCA_900157175.1 [16], *Tigriopus californicus* - GCA_007210705.1 [18], *Tigriopus kingsejongensis* - GCA_012959195.1 [17] and *T. kingsejongensis* [59].

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 [60].

2.3. Genomic structure and recombination

Using data from a related project sequencing salmon lice from multiple locations in the Northern Atlantic Ocean [40], 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 [40,61], each consisting of two parents and 46 offspring. Both genotype and pedigree data were handled in LepMAP3 [62]. The data were 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 identified 13 autosomes, but previous work based on a larger number of markers [63], 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. Quantitative real time PCR (qPCR) on dilution series was run using target (mRNA) *LsSpo11* and control (mRNA) *LseEF1α* assays in triplicate 10 μl reactions (for details see Supplementary Material section S6). The PCR efficiencies of the *LsSpo11*and *LseEF1α* [64] qPCR 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 [64] and ovary as calibrator tissue according to the $\Delta\Delta C_T$ method [65]. 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. Ortholog identification and phylogenomic analysis

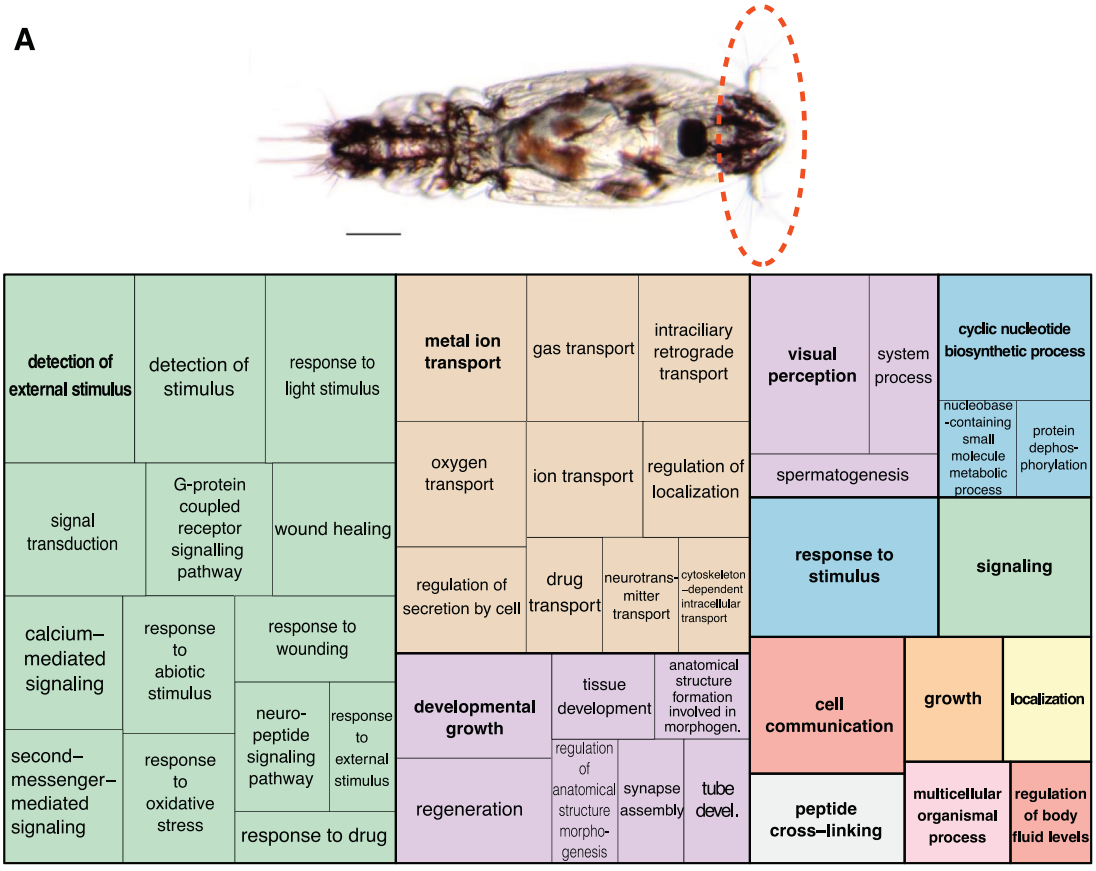
Predicted protein sequences from metazoan genomes were downloaded from GenBank and Ensembl Metazoa. In the case of multiple isoforms per gene, FASTA files were filtered using custom Perl scripts to retain only the longest isoform per gene. Orthologous groups were identified with OrthoFinder [66] using DIAMOND [67] as the search engine with default parameters. 180 single copy orthologous groups were identified from the species included in the analysis. These protein sequences in each orthologous group were aligned using Muscle with maxiter = 32 and otherwise default parameters [68,69]. The resulting multiple sequences alignments were concatenated using a custom Perl script and the resulting alignment was further trimmed for gaps and informative blocks were selected with Gblocks with default parameters [70]. Phylogenetic inference was done using the MCMC method in MrBayes [71]. MrBayes was run in mixed mode allowing for invariant sites and 4 rate categories for the gamma distribution resulting in a Rtrev + I + Γ model, run with 8 simultaneous chains for 2 million generations. Full names of the species and accession numbers of the GenBank genomes included in this analysis can be found in Supplementary_Table_GeneProtein-stats.

2.6. Statistical analysis of predicted gene number and proteome length

The same species and annotation data were used as in the ortholog identification and phylogenomic analysis. An extended set ($n = 25$) was also generated for Supplementary Section S7. We manually annotated the species for lifestyle and classified them as either belonging to the host-associated or free-living group of organisms. Protein sequences were filtered to retain only the longest isoform per gene in the same way as in Section 2.5. Sequence statistics for each FASTA file were calculated with SeqKit's stats command [72]. Statistical inference and plots were computed in R [73]. Statistical inference was conducted using the nonparametric Wilcoxon-Mann-Whitney two-sample rank-sum test with the one-sided alternative hypotheses that gene count and proteome length are larger in the free-living organisms (true location shift is greater than 0).

2.7. Gene presence and absence

Annotations for selected species (*Aedes aegypti*, *Anopheles gambiae*, *Brugia malayi*, *Caenorhabditis elegans*, *Danio rerio*, *Daphnia pulex*, *Drosophila melanogaster*, *Homo sapiens*, *Ixodes scapularis*, *Mus musculus*, *Pediculus humanus*, *Schistosoma mansoni*, *Tribolium castaneum* and *Trichinella spiralis*) were retrieved from Ensembl (release 101) and Ensembl Metazoa (release 48) using BioMart either manually or via the R package biomaRt [74]. Protein families and domain expansions and deletions were assessed based on the L. *salmonis* InterProScan annotations compared to sixteen existing genome annotations stored in Ensembl (see Supplementary_Table_Compara_Domains). Pathway analysis was based on KEGG pathways [75]. 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 InterProScan hits and performed reciprocal TBlastN/



(caption on next page)

Fig. 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 above the terms; A) planktonic copepodid, scale bar 0.1 mm; B) adult female with blood-filled intestine, scale bar 1 mm. Significantly enriched GO-terms from the Biological Process ontology are depicted in the form of a REVIGO tree-map. Clusters of semantically similar terms (by SimRel) as indicated by map colors. Centroids are typeset in bold. Box sizes for each term are proportional to $|\log_{10}(p)|$. Note, classification of genes as either expressed or not expressed in a tissue is based on non-replicated RNA-sequencing libraries. The full list of enriched GO terms and p-values is available as Supplementary_Table_GO. GO-terms related to the central nervous system in B) may be caused by inclusion of ganglia that is localized close to the anterior intestinal lumen. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

BlastX searches between the LSalAtl2s assembly and UniProtKB/GenBank. Blast searches were performed using the NCBI Blast suite versions 2.6.0+ and 2.9.0+ [76,77] with *E*-value threshold <1E-6. Additional analytical steps and accessions of queries used in Blast searches are described in the Supplementary Material.

2.8. Transcriptomic analysis of antennae and gut for annotation validation

For transcriptome analysis of different tissues, salmon lice from Ls1a strain, also used for gDNA sequencing, were utilized. Ovaries, testes, and intestine were dissected from adult salmon lice. Antennae were sampled from adult females, adult males and copepodids. Copepodid antennae were taken from planktonic copepodites. Legs were dissected from both male and female adult salmon 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. The purified RNA from these copepodids was pooled to create three samples each corresponding to two consecutive sampling days. RNA was extracted and purified as described in Supplementary Material section S3. 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: ovaries, pooled sample; testis, pooled sample; legs adult male and female pooled; antenna copepodid, pooled sample; Antenna adult female, pooled sample; copepodids sampled 1 and 2 days post infection at 10 °C; copepodites sampled 3 and 4 days post infection at 10 °C; copepodids sampled 5 and 6 days post infection at 10 °C; intestine, adult female, pooled sample.

Library preparation and sequencing was conducted by Fasteris SA (Geneva, Switzerland). A more detailed description is available in Supplementary Material 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 which was prepared using total RNA-protocol with normalization by Duplex-specific Nuclease (DSN) due to low RNA concentration. Data from the antenna adult female library 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 adapter sequences were removed using Trimmomatic, removing low quality (below quality 3) leading and trailing bases and cutting bases below a minimum average quality of 15 over a sliding window of size 4 bp (LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15) [78]. Trimmed sequences were aligned against the LSalAtl2s assembly with the splice-aware aligner STAR [79] 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. [80]. Library-size normalized read counts were calculated using the Bioconductor package edgeR [81] and are given in counts per million (CPM). Further details on methods used for creating Fig. 1 are given in Supplementary Material section S3.

2.9. Transcriptomic analysis of transition from planktonic to attached lifestyle

For planktonic *L. salmonis* copepodids, published data obtained from biological triplicates were used [82]. Briefly, these copepodids of the LsGulen strain were hatched from three different egg-string pairs and sampled on second day after molting to the copepodid stage. To obtain attached copepodids, farmed Atlantic salmon reared at 10 °C were infected with copepodids from the same strain [38]. Experimental procedures were performed in accordance with Norwegian animal welfare legislation (permit ID7704, no 2010/245410). Salmon lice were sampled from six different fish (sextuple samples) at one, three, and five days after infection. For details see Supplementary Material section S14.

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 (sequencing coverage of the assembly) 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): *Lepeophtheirus salmonis*, *Caligus rogercresseyi*, Tigriopus kingsejongensis, *Tigriopus californicus*, *A. 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 HiSeq 2000, 454, Sanger	PacBio, Illumina Hiseq4000	PacBio	Illumina, PacBio	Illumina	Sanger	Sanger
Depth	175	155	70	180	50	8.5	NA
[#]Scaffolds	36,095	21	938	459	351,850	1882	7
Contig N50	6044	38,017	NA	44,438	3244	34,097	NA
Scaffold N50 Kbp	478	27,803	1473	15,806	4	497	17,494
GC%	31	34	47	42	32	27	35
Repeat %	60[*]	57[*]	25[*]	28[*]	45[*]	20	18[*]
[#] genes	13,081	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 [13].

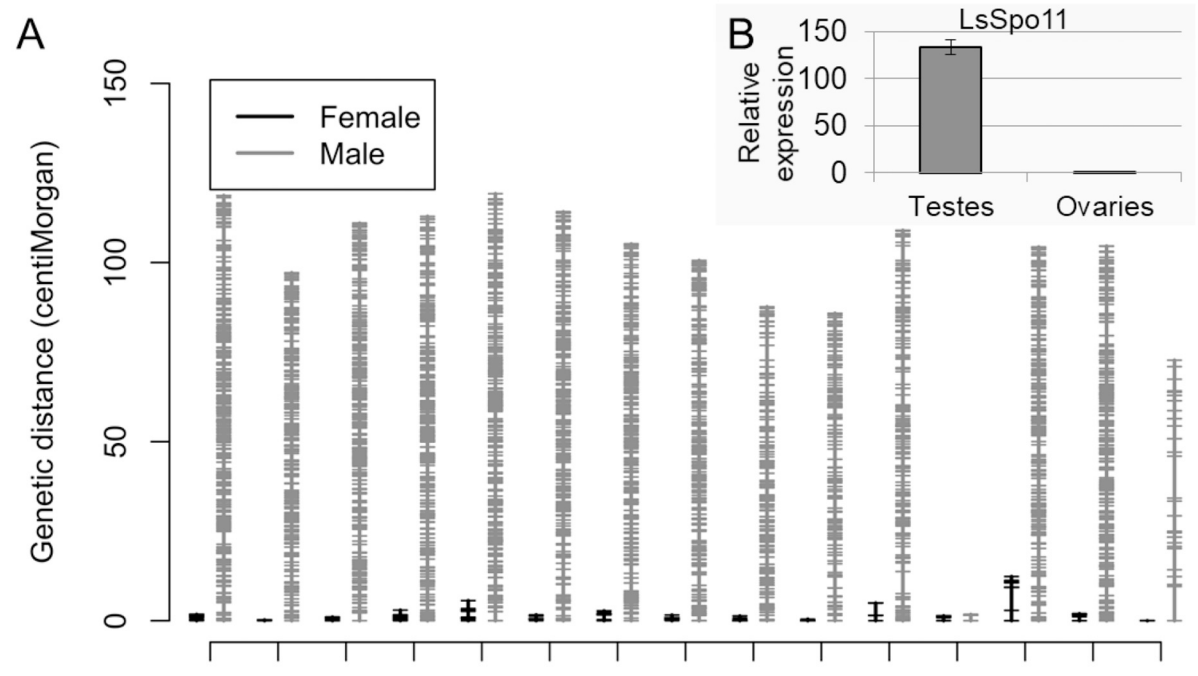
Library preparation, RNA sequencing, and data processing were executed as previously described for planktonic copepodids [82]. Sequences were aligned to the LSalAtl2s assembly using STAR and transcript counts were obtained using featureCounts [83]. Differential gene expression analysis was implemented by DESEq2 using Galaxy [84,85]. An adjusted *p*-value of 0.05 was taken as cut off. Gene ontology analysis was done by GOEnrichment (Version 2.0.1) [86] and conversion of the GO annotation to GOSlim terms was done by GOSlimmer [87] (Galaxy Version 1.0.1).

3. Results and discussion

3.1. Assembly and annotation

The final LSalAtl2s assembly has a size of 695.4 Mbp distributed among 36,095 scaffolds with a scaffold N50 of 478 Kbp (contig N50: 6044). Annotation results in prediction of 13,081 protein coding and 482 non-coding RNA genes. The BUSCO datasets 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 these taxa. Among the genes in the BUSCO V 5.0.0 Arthropoda 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 [82,88].

We used transcriptomic data to support the general validity of 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 among the significantly enriched GO-terms. For the adult



Chromosome	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Type	Autosomes														Sex
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	1.7	104	105	73
♀cM	1.7	1.2	1.0	2.9	5.6	1.6	2.7	1.5	1.3	0.4	4.9	1.3	12.3	1.9	0
Recomb.	70	81	111	39	21	71	39	67	67	215	22	1.3	8.5	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

Fig. 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 genetic distance in centimorgan (cM) does not translate directly to physical distance but is derived from recombination events. The shorter appearance of female compared to male chromosomes is caused by very low recombination frequency in females. The table shows chromosome type for each LG, the number of SNP markers included in the map, male and female chromosome sizes (cM), the male/female recombination ratio (Recomb.), 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 genes and base pairs assigned to LGs 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. Expressions were measured by qPCR with 3 biological replicates and calculated by the $\Delta\Delta C_T$ method using ovaries as calibrator (i.e. defining female expression as 1). Error bars show 95% confidence intervals.

intestine a more variable distribution of GO-terms was found, including "Proteolysis and hemolysis in other organism involved in symbiotic interaction" and"transport". 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.

Comparison among genome assemblies of the salmon louse and other ecdysozoans indicates that the salmon louse genome is unexceptional with regard to size and the 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* [89]. Of the examined species, the free-living harpactocoid copepod *T. californicus* [18] has a more compact genome with a predicted gene number comparable to *L. salmonis*. In contrast two other free-living harpactocoid copepods (*T. kingsejongensis* [17] *and T. japonicus* [19]) and the siphonomastoid copepod *C. rogercresseyi* [13] also have smaller genomes, but are reported to have approximately twice as many genes. It should be noted that an older *T. kingsejongensis* genome assembly [59] was predicted to have far less (12,772 vs. 25,470) protein coding genes than the more recent assembly [17] shown in Table 1. However, the more recent assembly is 10% larger and has a far higher BUSCO score suggesting it to be more complete. The calanoid copepod *A. tonsa* is indicated to have a larger genome than *L. salmonis* but a gene count is not reported [14]. In comparison, human lice (*Pediculus humanus*) and the nematode *Caenorhabditis elegans* have smaller genomes but respectively fewer (*P. humanus*) and significantly more (*C. elegans*) genes.

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 *Daphnia pulex* and six other copepod species, as well as all four public assemblies of *L. salmonis* in our analysis (Supplementary Material Fig. S4–1). Despite differences in sequencing depth, assembly software, and fragmentation, both GC-content and the proportion of repeat families are 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). This figure is much higher than the 38% estimate for *L. salmonis* presented by Jørgensen and colleagues [14] and the highest repeat content of all sequenced crustacean genomes presently available. The second highest repeat content is found in the only other caligid genome presently available; *C. rogercresseyi* with 57% repeats. It should be noted that Gallardo-Escárate [13] reported a lower *C. rogercresseyi* repeat content (51.9%), possibly due to differences in parameter settings (e.g. running RepeatMasker in sensitive mode). Our analysis further indicates that caligid repeat content is among the highest found in published arthropod genomes (Table 1 and Supplementary Material section S4). 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 [45]), locust (*Locusta migratoria*, ≈60%, 6.5 Gb, [90]), and Axolotl, the largest sequenced animal genome (*Ambystoma mexicanum,* 65.6%, 18-30Gb [91]). However, in a recent study on gomphocerine grasshoppers using low-coverage reads, the authors arrive at a range of estimated repeat content of 73% to 96% for these species, and 71% for the locust [92]. This may indicate that repeat content in many animal genomes, including the salmon louse, could be even higher than estimated based on assemblies.

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 (Supplementary Material Fig. S4–1 and [18]). 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 [93,94]. There is growing evidence for horizontal transposon transfer (HTT) between different species, including fish and their parasites [95–97] and future studies are needed to establish whether some caligid TEs are active or have been interchanged with their host via HTT.

3.3. Genomic structure and recombination

A linkage map was constructed based on 5098 previously identified SNP markers [40]. 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 approximately 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 centimorgan (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 exhibited frequent recombination in males and extremely low recombination rates in females wherefore marker distances were calculated from a male map (Fig. 2). Differences in recombination ratio between sexes (heterochiasmy) is common and earlier studies indicate that the differences commonly range from minor (almost equal) to substantial (eg. ≈3 fold difference) [98,99]. In some instances recombination is entirely absent in one or both sexes [100]. While this phenomenon (achiasmy) is particularly well described in *Drosophila* spp., it is phylogenetically widespread and reported also among copepod species [100–102]. The female salmon louse is not completely achiasmatic since low frequency recombination does occur, but the recombination sex bias is very large (≈35 fold difference). There are several explanatory models for differences in recombination frequencies between sexes but none of them receives universal support, save the principle that achiasmy appears to be restricted to the heterogametic sex [98,99]. Interestingly, in similarity to the *D. melanogaster* 'dot' chromosome [103–106], the salmon louse also has a chromosome sheltered from recombination in both sexes; the chromosome corresponding to LG12 (Fig. 2 and [63]).

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 [63]. Independent sequencing of an Atlantic *L. salmonis* male (NCBI sequence read archive SRX976782) allowed mapping of both 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 section S5). A smaller proportion of scaffolds received average mapping of male reads and 0.5× 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. 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. Finally, a number of scaffolds received no

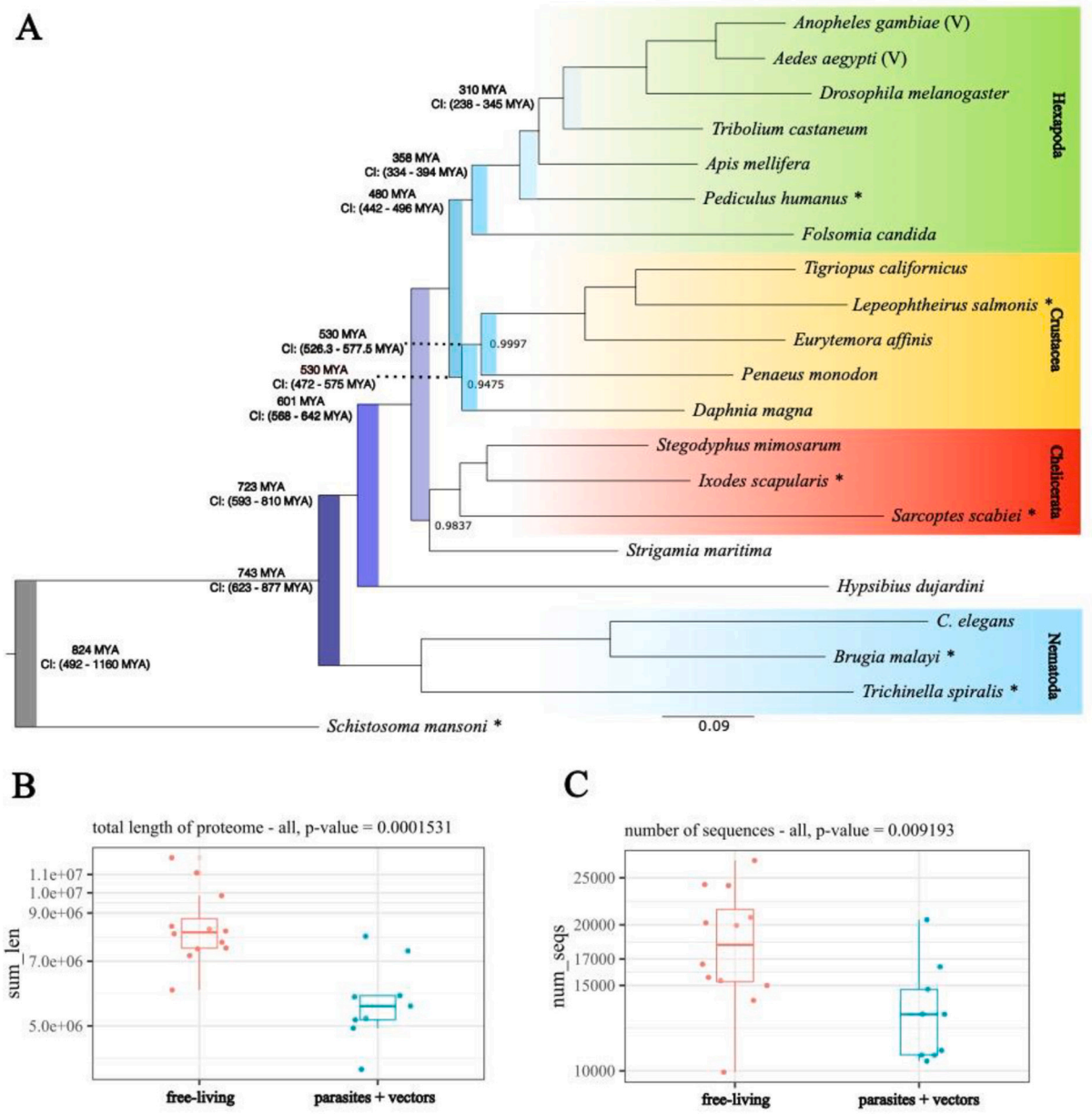


Fig. 3. Whole genome phylogeny from MrBayes based on 180 concatenated single-copy orthologous protein sequence alignments (A), boxplots of the total length of the predicted proteomes using the longest-isoform per gene and the number of protein-coding genes in each species (C). Node support in (A) by posterior probabilities is equal to 1 for all nodes, except where stated. Scale-bar: 0.09 substitutions per site. The divergence time in million years ago (MYA) and the respective confidence interval (CI) are given for some of the major lineages and are based on estimates from the TimeTree database [114]. Major ecdysozoan phyla are highlighted. The taxon *S. mansoni* was used to root the tree. Taxa were classified as either free-living, parasites (*) or vectors (V). The same species as in (A) were used in (B) and (C), and statistics are compared in free-living species versus parasites and vectors. *p*-values are given based on the Wilcoxon–Mann–Whitney two-sample rank-sum test.

mapping of male reads and 0.5× 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. These observations functionally explain the 1:1 male:female ratio in salmon lice and indicate a ZZ-ZW sex chromosome configuration, or alternative chromosome configurations with several W sex chromosomes as previously reported in both deuterostome and protostome animals [107,108].

Since genetic recombination is predominantly found in the homogametic males it was hypothesized that this should be reflected in genes expected to be involved in recombination. The topoisomerase-like Spo11 protein is involved in genetic recombination by inducing double stranded breaks [109,110] and is present as a single copy gene in the salmon louse genome (*LsSpo11*). We therefore 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.

3.4. Phylogenomic analysis

The official gene set predicted by Ensembl comprised 13,081 coding genes, and among these the OrthoFinder pipeline identified 180 1:1 orthologs across 21 selected species that were used in a whole genome phylogenetic reconstruction (Fig. 3). *L. salmonis* groups together with *T. californicus and E. affinis*. In the phylogram, all crustacean species

occur as a sister group to the included Hexapoda where the divergence of the tardigrade (*Hypsibius dujardini*) and the Euarthropoda represents the most basal node within the Panarthropoda and the divergence of nematodes and Panarthropoda is the most basal node within the Ecdysozoa (Fig. 3). The observed grouping is overall in line with previous studies [111,112] with the exception of the Myriapoda (*Strigamia maritima*) appearing to group most closely with Chelicerata in our analysis. It should be noted that there is some controversy about the deepest level of arthropod phylogeny [96].

3.4.1. Parasites have fewer protein-coding genes and smaller proteomes

We analyzed the predicted protein-coding gene-sets of the 21 metazoan genomes in the phylogenomic analysis, consisting of 7 species which we labelled as parasites, 2 species labelled as vectors, and 12 free-living species (Fig. 3A). Proteome sizes range from 3,987,500 residues (*Sarcoptes scabiei*; parasitic itch mite) to 12,004,398 residues (*Folsomia candida*; free-living hexapod), and gene-set size ranges from 9935 (*Apis melifera*; honey bee, social insect) to 27,135 genes (*Stegodyphus mimosarum*; social velvet spider). Of note, the top 7 species with smallest proteomes are all obvious parasites, and the top 7 species with largest proteomes are all free-living (Supplementary_Table_GeneProtein-stats). The included parasites and vectors have significantly fewer genes (median: 13,081 vs. 18,264.5, $p = 0.009193$, 95% CI (difference in location): [1911, ∞], Fig. 3C) and significantly smaller combined predicted proteome sizes (median: 5,544,571 AA vs. 8,137,634 AA, $p = 0.0001531$, 95% CI (difference in location): [1,681,519, ∞] Fig. 3B) compared to the non-parasites. These results are remarkably stable when considering re-classification of single species or addition of further well-annotated species, e.g. the debatable inclusion of vectors as host associated species (Supplementary Material section S7).

Our findings suggest a general pattern of gene loss in parasitic metazoan species. Similar observations have been made in bacteria [113] 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 or other aspects of an organism's lifestyle (e.g. social behavior) or habitat may influence genome evolution. Technically, confounding factors may include methods of sequencing and automatic annotation, availability, and quality of training data for gene predictors, and the amount of attention dedicated to manual curation of genomes. It should be noted though that we did not find any discernable pattern of correlation between gene-set size and sequencing technology, annotation software, or year of publication (data not shown). The most prominent impediment to studying the interplay of host-dependence and genome evolution is still the lack of well-annotated parasite genomes in public databases. Hence, more systematic studies with relevant and diverse taxa should be conducted before firm conclusions regarding trends in gene numbers and parasitism are made.

3.5. Gene losses and expansions

The gene presence and absence analysis revealed some protein groups and pathway elements to be absent, 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 [115] are present in most organisms and are involved in a range of basic biological processes such as calcium metabolism, cell adhesion, growth and differentiation [116,117]. Cantacessi et al. [118] 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 [119,120]. A total of 125 SHK domains in 24 genes were identified in *L. salmonis* which is 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 prevents activation of these [121]. We speculate that the SHK-domain containing proteins in the salmon louse could be important for immunomodulation of the host and warranting 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, including proteins, lipids, and trace metals. Heme is an iron-containing macrocycle and an essential prosthetic group in aerobic life, and its biosynthesis involves 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 [122]. Furthermore, the genome encodes genes for several requisite heme-binding proteins such as cytochrome *c* and a recently characterized heme-peroxidase [123], 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 infer that the salmon louse is a natural heme auxotroph (see Supplementary Material section S8 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 [124]), the absence of any known heme detoxification mechanism is surprising. The classical conserved two step pathway via heme oxigenase and 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 [125], currently lack evidence in the salmon louse. Interestingly, out of the eight conserved steps of heme biosynthesis the salmon louse genome contains only the 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 [126]). The preservation of CPOX in the salmon louse may indicate a role in clearing such compounds. Finally, since *L. salmonis* lacks Heme oxygenase (*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 a conserved set of genes involved in their biogenesis and function [127]. Despite their shared phylogenetic origin, they are involved in metabolic

processes that may diverge significantly between species and tissues [128,129]. Four Pex genes are conserved in eukaryotes (except groups that have lost peroxisomes) and may be considered marker genes required for presence of peroxisomes: *Pex3*, *Pex10*, *Pex12* and *Pex19* [130]. A fifth *Pex* gene with bacterial orthologs, *Pex5*, is also ubiquitously present in peroxisome containing organisms [130]. 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 S9). Absence of the Pex core genes indicates that the salmon louse is unable to sustain peroxisomes [130]. To investigate if lack of peroxisomes is a conserved feature among copepods, we searched available genome assemblies for *C. rogercresseyi, T. californicus and E. affinis* and a *Calanus finmarchicus* transcriptome assembly (Supplementary Material section S9). We found that both *T. californicus* (order Harpactoida) and *C. finmarchicus* (order Calanoida) possess all core elements of peroxisomes while *E. affinis* (order Calanoida) appeared to be missing *Pex19*. In contrast the fish louse *C. rogercresseyi* (order Siphonostomatoida, same family as *L. salmonis*) appears to lack all core *Pex* genes (Supplementary Material section S9). This indicates that the lack of peroxisomes may be an adaptation to parasitism among members of the family Caligidae, such as *L. salmonis* and *C. rogercresseyi*, or maybe even shared among members of order Siphonostomatoida in general as suggested by the lack of core PEX protein encoding ESTs from other siphonostomatoids presently represented in GenBank (*L. branchialis*; 14,927 ESTs, and *C. clemensi;* 46,858 ESTs). While peroxisomes are generally considered ubiquitous organelles, their absence has previously been reported in various taxa, including parasitic platyhelminths and nematodes, and even the free-living appendicularian *Oikopleura dioica* [131]. A key function of peroxisomes is to facilitate catalase mediated reduction of reactive H_2O_2 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. [129]).

3.5.5. Peptidases in L. salmonis

We compared different types of peptidase domains across the included species (Supplementary Material section S10) 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 106C-terminal domains). Ticks have intracellular digestion, and the high proportion of M13 peptidases could be a signature for this property [132]. 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 Fig. S10–1). This is more than twice as many as have been found in *D. pulex*.

3.5.6. Detoxification and stress-response

The salmon louse possesses a remarkable capacity for adaptation to most chemical delousing agents [40–42], including hydrogen peroxide [133,134]. We therefore hypothesized an expansion of gene-families related to xenobiotic metabolism. Interestingly, the opposite is the case. When comparing the frequency of gene-families with putative roles related to xenobiotic metabolism 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., [135] reported that of all arthropods surveyed, *L. salmonis* had the most compact family of cytochrome P450 (*Cyp*) genes. Besides their role in detoxification of various substrates, enzymes of the *Cyp* family contribute to metabolism of steroids, fatty acids, and other biomolecules. 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 by Carmona-Antoñanzas et al. [136] who identified close to 30 ABC transporters, which is significantly less than the numbers of ABC-transporters in many other arthropods (including e.g. *D. pulex* (67 genes), *D. melanogaster* (56 genes) and *A. mellifera* (41genes)). The identified ABC transporters comprised 18 genes belonging to subfamilies known to include drug transporters (i.e. subfamilies B, C and G) [136].

The third compacted gene family in the salmon louse is the Glutathione S-transferases (GSTs); one of the most important families of detoxifying enzymes in nature (reviewed by Oakley [137]). There is growing evidence that GSTs are induced by and may contribute to the clearing of synthetic pyrethroids such as cypermethrin in terrestrial arthropods [138] and aquatic crustaceans [139]. Amelioration of pyrethroid toxicity by GSTs may be due to their antioxidant capacity [140]. 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 S11). MVP is highly conserved throughout many eukaryote lineages 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 the vault complex 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 [141,142]. In a recent study on *L. salmonis*, *Mvp* transcription was up-regulated together with *Cyp* genes and other stress responsive genes under cypermethrin exposure [143]. Mammalian MVP is induced by or responds to biotic and abiotic stressors and toxins *in vivo* and *vitro* [144–146]. Unlike most eukaryote genomes in which the *Mvp* gene is absent or present as a single copy, *L. salmonis* contains four paralogous *Mvp*-like sequences, likely stemming from gene duplication events. Implication in drug resistance and stress-response render MVP and other vault components interesting for future studies in sea lice. However, it should be noted that contribution of the vault complex to MDR is debated (reviewed by Park [147]) and that inferences from cancer cells and mammalian models to a parasitic copepod should be made with caution. Therefore, more studies are needed to elucidate potential vault involvement in adaptations to a parasitic lifestyle or detoxification pathways.

3.5.7. Large expansion of FN2-domains

Fibronectin II domains (FNII-domain, PF00040), commonly considered to be vertebrate specific, are – surprisingly - the largest of the expanded *L. salmonis* 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 [148] and Supplementary

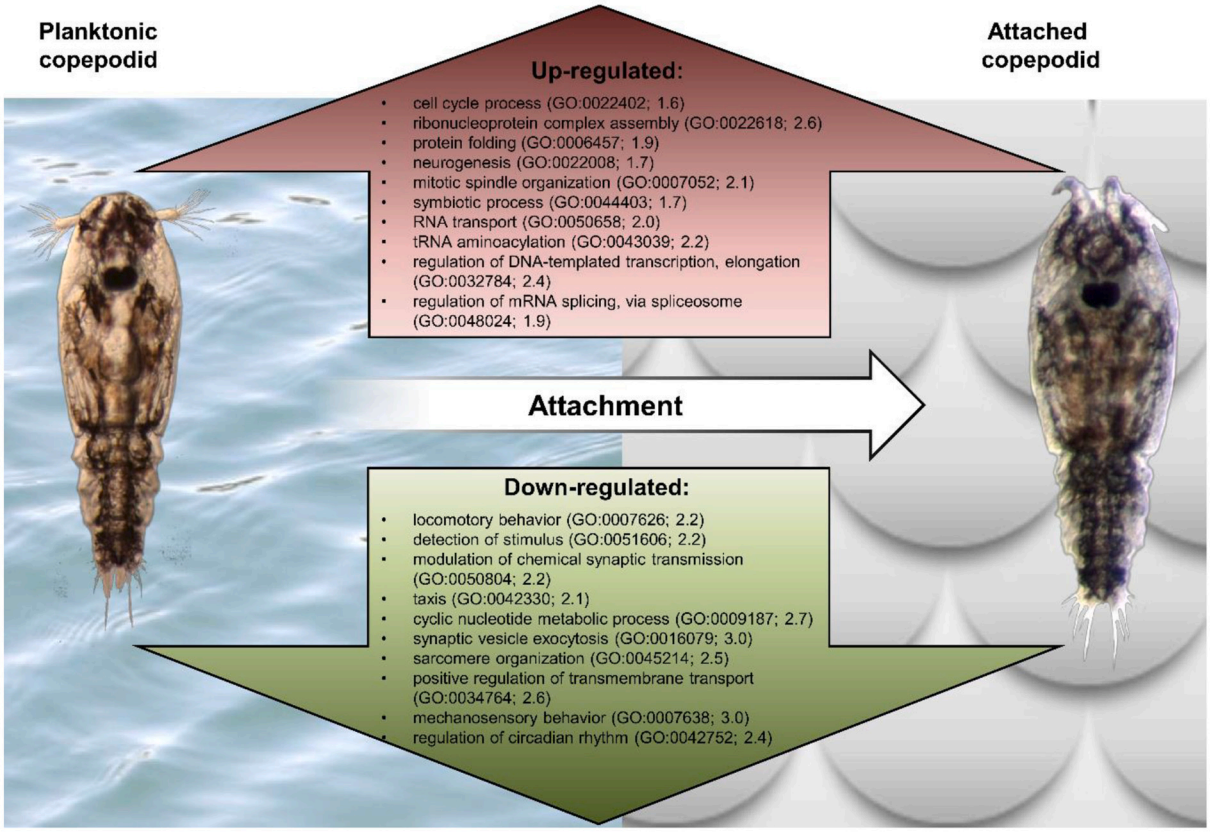


Fig. 4. Enriched GO terms of the genes differentially regulated in copepodids after attachment. Ten of the most significant enriched GO terms (all q-value <0.01, sorted after increasing q-value) of the category Biological Process with more than 1.5 times enrichment were selected from the GO terms enriched under the genes up- and down-regulated after infecting the host. The GO Number and enrichment are quoted in parentheses. A full list of all enriched GO terms can be found in Supplementary_Table_GO_Cops.

Material section S12). While FNII domains have not previously been reported in other arthropods, they are described in tubularians [149] and a few other invertebrate groups (see Supplementary Material Table S12-1). Invertebrates commonly have Kringle domains (PF00051) that have been suggested to be ancestral to FNII domains [150]. In addition to FNII domains, we also identified five different proteins containing a single Kringle domain in combination with other domains in L. *salmonis.* FNII and Kringle domains were found in order Siphonostomatoida (*L. salmonis* and *C. rogercresseyi*), order Harpactoida (*T. californicus*) and order Calanoida (Kringle in *E. affinis*, FNII in *Acartia pacifica*) indicating that the divergence of Kringle and FNII domains occurred earlier than previously suggested and probably in a common metazoan ancestor.

While the exact functions of genes with FNII-domains in *L. salmonis* are unknown, the expression profile for all transcripts containing FNII domains in the RNA sequencing samples (Supplementary Material Fig. S12–2) 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 their host [151] and chemical sensing has been proposed to be important for host identification [152]. In arthropods, chemical sensing is mediated by gustatory receptors (GRs), odorant receptors (ORs), and ionotropic receptors (IRs) and we assessed the genome for orthologs of these proteins. No GRs or ORs were detected in the salmon louse genome whereas 26 IRs were identified (see Supplementary Material section S13) and some play a crucial role in host recognition by *L. salmonis* copepodids [153]. ORs appear to be common in hexapods but absent in pancrustacea [15]. GRs in contrast are common in pancrustacea, and may be numerous as for *D. pulex* where 58 GRs has been identified [154]. We identified GRs in harpactocoid copepods (nine in *T. californicus* and two in *T. kingsejongensis*) which has also been reported by Eyun et al. [15] who furthermore identified GRs in the calanoid copepod *E. affinis*. In contrast no GRs could confidently be identified in *C. rogercresseyi* (see Supplementary Material section S13) or *L. salmonis,* indicating that loss of GRs may be a signature of Caligidae or Siphonostomatoidea. As expected, many iGluR genes (which includes IRs) that contained the PF00060 and PF10613 domains are present both in *T. kingsejongensis* and *C. rogercresseyi* (see Supplementary Material section S13). 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*. In *L. salmonis* IR8 is duplicated, but we could only detect one IR8 in *Tigriopus* or *Caligus*, indicating that this is *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 S13–1 in Supplementary_Table_IR).

3.6. Gene expression during transition to a parasitic lifestyle

The infective copepodid stage can be investigated in either its non-parasitic planktonic or host-bound parasitic phase. By assessing the transcriptome of planktonic versus attached copepodids, we aimed to compare these different lifestyles. Gene expression of planktonic copepodids was compared to the expression of parasitic copepodids collected one, three or five days after attachment. Copepodids from the different groups differ in their overall gene expression (Supplementary Material section S14). We found 6273 differentially expressed genes, 3205

transcripts that were up-regulated between planktonic and attached salmon lice consistently in all three comparisons and 3068 that were consistently down-regulated (Supplementary Material section S14). Highly significantly enriched GO-terms affected by this transition are shown in Fig. 4. All enriched GOSlim-terms can be found in the Supplementary Material section S14, Supplementary Material Fig. S14–3 and enriched GO terms are listed in Supplementary_Table_GO_Cops. These can be summarized as follows: after attaching to the host, metabolism increases and sensory abilities decrease. Planktonic copepodids do not feed but rely on internal stores for survival until they find a host [12,39]. The attachment, and by that the access to food, induces a switch in nourishment and triggers the start to growth and development. Sensing-related genes are strongly down-regulated after attachment (e.g. locomotory behavior (GO:0007626), detection of stimulus (GO:0051606)), while cell cycle (cell cycle (GO:0007049)), protein synthesis (translation (GO:0006412) and other core metabolic functions (e.g. primary metabolic process (GO:0044238), macromolecule biosynthetic process (GO:0009059)) are enriched (Fig. 4). Lipid metabolism is also down-regulated after attachment presumably due to the change of nutrition from stored lipids [152] to ingested host tissues. Furthermore, several copies of a gene with Blast hit to the protein Timeless [153], known for regulating circadian rhythm in Drosophila, are among the down-regulated transcripts. Among the down-regulated genes we also find three other genes known for regulating circadian rhythms in Drosophila; Shab, Shaw and Cycle [154], [155]. This may reflect the transformation from being a planktonic organism with diurnal migration [156] to a parasitic lifestyle. However, as sampling time point can slightly vary between samples, further investigation is necessary to exclude sampling bias. There are also several down-regulated muscle related genes (Myosin, Troponin, Tropomyosin, Actin, Titin and the regulator of muscle contraction Twitchin [157]), suggesting a shift to a more sessile live style. The up-regulated genes in attached copepodids include ribosomal proteins, proteosome proteins, as well as, transcription and translation factors. The ionotropic receptors probably involved in host recognition [150] are down-regulated, while genes with FNII domains (see above) are among the most up regulated genes in parasitic copepodids.

4. Conclusions

The LSalAtl2s draft genome assembly in conjunction with the gene models in Ensembl Metazoa is among the first fully annotated genomes of parasitic copepods and represents 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 main hosts, the Atlantic salmon, it may facilitate investigations into host-parasite interactions. Taken together, the Atlantic salmon louse (*L. salmonis salmonis*) 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.

Data availability

Genomic and transcriptomic sequence 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.

Acknowledgements

We are thankful to Dr. E. Birney (European Bioinformatics Institute) and O. Torrissen (Institute of Marine Research) for their assistance and enthusiasm. We appreciate the constructive criticism from three anonymous reviewers. The project was funded by The Norwegian Ministry of Trade and Fisheries, The Norwegian Seafood Research Fund (project 900400) and the Research Council of Norway, SFI-Sea Lice Research Centre, grant number 203513/O30.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2021.08.002.

References

[1] B. McClintock, The significance of responses of the genome to challenge, Science 226 (1984) 792–801.
[2] R. Noble, D. Noble, Was the watchmaker blind? Or was she one-eyed? Biology-Basel 6 (2017).
[3] 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, Proceed. R. Soc. B Biol. Sci. 282 (2015).
[4] E.A. Boyle, Y.I. Li, J.K. Pritchard, An expanded view of complex traits: from polygenic to omnigenic, Cell 169 (2017) 1177–1186.
[5] E. Callaway, New concerns raised over value of genome-wide disease studies, Nature 546 (2017) 463.
[6] J. Wang, S. Gao, Y. Mostovoy, Y. Kang, M. Zagoskin, Y. Sun, B. Zhang, L.K. White, A. Easton, T.B. Nutman, P.Y. Kwok, S. Hu, M.K. Nielsen, R.E. Davis, Comparative genome analysis of programmed DNA elimination in nematodes, Genome Res. 27 (2017) 2001–2014.
[7] D.B. Stern, C.E. Lee, Evolutionary origins of genomic adaptations in an invasive copepod, Nat. Ecol. Evol. 4 (2020) 1084.
[8] L. Udayanga, T. Ranathunge, M.C.M. Iqbal, W. Abeyewickreme, M. Hapugoda, Predatory efficacy of five locally available copepods on Aedes larvae under laboratory settings: an approach towards bio-control of dengue in Sri Lanka, PLoS One 14 (2019).
[9] H. van der Jagt, I. Wiedmann, N. Hildebrandt, B. Niehoff, M.H. Iversen, Aggregate feeding by the copepods Calanus and pseudocalanus controls carbon flux attenuation in the Arctic Shelf sea during the productive period, Front. Mar. Sci. 7 (2020).
[10] 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.
[11] 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, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 17676–17681.
[12] A.W. Pike, S.L. Wadsworth, Sealice on salmonids: their biology and control, Adv. Parasitol. 44 (1999) 233–337.
[13] 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.
[14] 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.
[15] S.I. Eyun, H.Y. Soh, M. Posavi, J.B. Munro, D.S.T. Hughes, S.C. Murali, J. Qu, S. Dugan, S.L. Lee, H. Chao, H. Dinh, Y. Han, H. Doddapaneni, K.C. Worley, D. M. Muzny, E.O. Park, J.C. Silva, R.A. Gibbs, S. Richards, C.E. Lee, Evolutionary history of chemosensory-related gene families across the Arthropoda, Mol. Biol. Evol. 34 (2017) 1838–1862.
[16] 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.
[17] 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 Genom. Proteomics 35 (2020) 100703.
[18] F.S. Barreto, E.T. Watson, T.G. Lima, C.S. Willett, S. Edmands, W.Z. Li, R. S. Burton, Genomic signatures of mitonuclear coevolution across populations of *Tigriopus californicus*, Nat. Ecol. Evol. 2 (2018) 1250–1257.

[19] C.B. Jeong, B.Y. Lee, B.S. Choi, M.S. Kim, J.C. Park, D.H. Kim, M. Wang, H. G. Park, J.S. Lee, The genome of the harpacticoid copepod *Tigriopus japonicus*: potential for its use in marine molecular ecotoxicology, Aquat. Toxicol. 222 (2020) 105462.
[20] P.A. Selden, R. Huys, M.H. Stephenson, A.P. Heward, P.N. Taylor, Crustaceans from bitumen clast in carboniferous glacial diamictite extend fossil record of copepods, Nat. Commun. 1 (2010) 50.
[21] R.M. Maizels, H.J. McSorley, Regulation of the host immune system by helminth parasites, J. Allergy Clin. Immunol. 138 (2016) 666–675.
[22] 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.
[23] 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.
[24] 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.
[25] 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.
[26] 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.
[27] 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.
[28] 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.
[29] P.G. Fjelldal, T.J. Hansen, O. Karlsen, Effects of laboratory salmon louse infection on osmoregulation, growth and survival in Atlantic salmon, Conserv. Physiol. 8 (2020).
[30] 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.
[31] A.J. Reid, Large, rapidly evolving gene families are at the forefront of host-parasite interactions in Apicomplexa, Parasitology 142 (2015) S57–S70.
[32] 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.
[33] 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.
[34] G.L. Taranger, O. Karlsen, R.J. Bannister, K.A. Glover, V. Husa, E. Karlsbakk, B. O. Kvamme, K.K. Boxaspen, P.A. Bjorn, B. Finstad, A.S. Madhun, H.C. Morton, T. Svasand, Risk assessment of the environmental impact of Norwegian Atlantic salmon farming, ICES J. Mar. Sci. 72 (2015) 997–1021.
[35] M.S. Myksvoll, A.D. Sandvik, J. Albretsen, L. Asplin, I.A. Johnsen, O. Karlsen, N. M. Kristensen, A. Melsom, J. Skardhamar, B. Adlandsvik, Evaluation of a national operational salmon lice monitoring system-from physics to fish, PLoS One 13 (2018).
[36] 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).
[37] S.C. Johnson, L.J. Albright, The developmental stages of Lepeophtheirus-Salmonis (Kroyer, 1837) (Copepoda, Caligidae), Can. J. Zool. 69 (1991) 929–950.
[38] 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.
[39] 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.
[40] 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 ectoparasite *Lepeophtheirus salmonis*, BMC Genomics 15 (2014) 937.
[41] 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.
[42] 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. 7 (2017).
[43] 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. Aquac. 11 (2019) 1398–1417.
[44] 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).
[45] 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.
[46] 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.
[47] H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform, Bioinformatics 25 (2009) 1754–1760.
[48] 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.
[49] 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.
[50] 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.
[51] M. Boetzer, C.V. Henkel, H.J. Jansen, D. Butler, W. Pirovano, Scaffolding pre-assembled contigs using SSPACE, Bioinformatics 27 (2011) 578–579.
[52] C. UniProt, UniProt: a worldwide hub of protein knowledge, Nucleic Acids Res. 47 (2019) D506–D515.
[53] 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.
[54] Javier Herrero, Matthieu Muffato, Kathryn Beal, Stephen Fitzgerald, Leo Gordon, Miguel Pignatelli, Albert J. Vilella, Stephen M.J. Searle, Ridwan Amode, Simon Brent, William Spooner, Eugene Kulesha, Andrew Yates, Paul Flicek, Ensembl comparative genomics resources, Database 2016 (2016), https://doi.org/10.1093/database/bav096.
[55] 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–W185.
[56] 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.
[57] A.F.A. Smit, R. Hubley, P. Green, RepeatMasker 4.1.1. http://www.repeatmasker.org, 2020 accessed Oct 7. 2020.
[58] 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.
[59] 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.
[60] Anon, Sharing Data from Large-scale Biological Research Projects: A System of Tripartite Responsibility, 2003.
[61] 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).
[62] P. Rastas, Lep-MAP3: robust linkage mapping even for low-coverage whole genome sequencing data, Bioinformatics 33 (2017) 3726–3732.
[63] 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. Gen. Genomics. 294 (2019) 343–363.

[64] 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.
[65] 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.
[66] D.M. Emms, S. Kelly, OrthoFinder: phylogenetic orthology inference for comparative genomics, Genome Biol. 20 (2019) 238.
[67] B. Buchfink, C. Xie, D.H. Huson, Fast and sensitive protein alignment using DIAMOND, Nat. Methods 12 (2015) 59–60.
[68] 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.
[69] R.C. Edgar, MUSCLE: a multiple sequence alignment method with reduced time and space complexity, Bmc Bioinf. 5 (2004) 1–19.
[70] J. Castresana, Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis, Mol. Biol. Evol. 17 (2000) 540–552.
[71] 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.
[72] W. Shen, S. Le, Y. Li, F.Q. Hu, SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation, PLoS One 11 (2016).
[73] R.-C. Team, A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2021.
[74] 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.
[75] M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes, Nucleic Acids Res. 28 (2000) 27–30.
[76] 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.
[77] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden, BLAST+: architecture and applications, Bmc Bioinf. 10 (2009) 421.
[78] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics 30 (2014) 2114–2120.
[79] 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.
[80] N. Delhomme, I. Padioleau, E.E. Furlong, L.M. Steinmetz, easyRNASeq: a bioconductor package for processing RNA-Seq data, Bioinformatics 28 (2012) 2532–2533.
[81] 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.
[82] 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.
[83] 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.
[84] 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.
[85] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol. 15 (2014).
[86] D. Faria, GOEnrichment, in: GitHub, GitHub, Github Repository, 2017.
[87] D. Faria, GOSlimmer, in: GitHub, GitHub, Github Repository, 2017.
[88] E.I. Heggland, M. Dondrup, F. Nilsen, C. Eichner, Host gill attachment causes blood-feeding by the salmon louse (*Lepeophtheirus salmonis*) chalimus larvae and alters parasite development and transcriptome, Parasit. Vectors 13 (2020).
[89] 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.
[90] 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.
[91] 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.
[92] A. Shah, J.I. Hoffman, H. Schielzeth, Comparative analysis of genomic repeat content in gomphocerine grasshoppers reveals expansion of satellite DNA and Helitrons in species with unusually large genomes, Genome Biol. Evol. 12 (2020) 1180–1193.
[93] 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 Biol. 19 (2018).
[94] C. Feschotte, E.J. Pritham, DNA transposons and the evolution of eukaryotic genomes, Annu. Rev. Genet. 41 (2007) 331–368.
[95] A. Pocwierz-Kotus, A. Burzynski, R. Wenne, Family of Tc1-like elements from fish genomes and horizontal transfer, Gene 390 (2007) 243–251.
[96] S. Kuraku, H. Qiu, A. Meyer, Horizontal transfers of Tc1 elements between teleost fishes and their vertebrate parasites, lampreys, Genome Biol. Evol. 4 (2012) 929–936.
[97] 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, Mob. DNA 10 (2019).
[98] T. Lenormand, J. Dutheil, Recombination difference between sexes: a role for haploid selection, PLoS Biol. 3 (2005) 396–403.
[99] J.E. Mank, The evolution of heterochiasmy: the role of sexual selection and sperm competition in determining sex-specific recombination rates in eutherian mammals, Genet. Res. 91 (2009) 355–363.
[100] K. Satomura, N. Osada, T. Endo, Achiasmy and sex chromosome evolution, Ecol. Genet. Genom. 13 (2019) 1000045.
[101] R.S. Burton, M.W. Feldman, S.G. Swisher, Linkage relationships among 5 enzyme-coding gene loci in the copepod *Tigriopus-californicus* - a genetic confirmation of Achiasmatic meiosis, Biochem. Genet. 19 (1981) 1237–1245.
[102] S. Edmands, Recombination in interpopulation hybrids of the copepod *Tigriopus californicus*: release of beneficial variation despite hybrid breakdown, J. Hered. 99 (2008) 316–318.
[103] L. Sandler, P. Szauter, The effect of recombination-defective meiotic mutants on fourth-chromosome crossing over in Drosophila melanogaster, Genetics 90 (1978) 699–712.
[104] 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, K.C. Worley WoodageT, 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.
[105] 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.
[106] 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.
[107] M.J.D. White, Animal Cytology and Evolution, 3, ed ed., Cambrdge University Press, Cambridge, 1973.
[108] M.J.D. White, The origin and evolution of multiple sex-chromosome mechanisms, J. Genet. 40 (1940) 303–336.
[109] G. Mirzaghaderi, E. Horandl, The evolution of meiotic sex and its alternatives, Proceed. R. Soc. B Biol. Sci. 283 (2016).
[110] C.J. Underwood, K. Choi, Heterogeneous transposable elements as silencers, enhancers and targets of meiotic recombination, Chromosoma 128 (2019) 279–296.
[111] 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.
[112] G. Giribet, G.D. Edgecombe, The phylogeny and evolutionary history of arthropods, Curr. Biol. 29 (2019) R592–R602.
[113] 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.
[114] 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.
[115] G.J. Barton, R.H. Newman, P.S. Freemont, M.J. Crumpton, Amino-acid-sequence analysis of the annexin supergene family of proteins, Eur. J. Biochem. 198 (1991) 749–760.
[116] V. Gerke, S.E. Moss, Annexins: from structure to function, Physiol. Rev. 82 (2002) 331–371.
[117] S.E. Moss, R.O. Morgan, The annexins, Genome Biol. 5 (2004).
[118] 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. 3 (2013).
[119] 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.
[120] S.J. Chang, S. Bajaj, K.G. Chandy, ShK toxin: history, structure and therapeutic applications for autoimmune diseases, Wiki J. Sci. 1 (2018).
[121] 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.
[122] 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.
[123] 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. Physiol. A 206 (2017) 1–10.
[124] 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.
[125] 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. Mol. Biol. 32 (2002) 517–523.
[126] J.D. Phillips, Heme biosynthesis and the porphyrias, Mol. Genet. Metab. 128 (2019) 164–177.
[127] T. Gabaldon, Peroxisome diversity and evolution, Philos. Trans. R. Soc. B Biol. Sci. 365 (2010) 765–773.
[128] L. Pieuchot, G. Jedd, Peroxisome assembly and functional diversity in eukaryotic microorganisms, Annu. Rev. Microbiol. 66 (2012) 237–263.
[129] M. Islinger, M.J.R. Cardoso, M. Schrader, Be different-the diversity of peroxisomes in the animal kingdom, Biochim. Biophys. Acta 1803 (2010) 881–897.
[130] A. Schlüter, S. Fourcade, R. Ripp, J.L. Mandel, O. Poch, A. Pujol, The evolutionary origin of peroxisomes: an ER-peroxisome connection, Mol. Biol. Evol. 23 (2006) 838–845.
[131] V. Zarsky, J. Tachezy, Evolutionary loss of peroxisomes - not limited to parasites, Biol. Direct 10 (2015).
[132] 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.
[133] 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.
[134] 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.
[135] 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), Parasit. Vectors 12 (2019).
[136] 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).
[137] A. Oakley, Glutathione transferases: a structural perspective, Drug Metab. Rev. 43 (2011) 138–151.
[138] S.M. Valles, Toxicological and biochemical studies with field populations of the German cockroach, *Blattella germanica*, Pestic. Biochem. Physiol. 62 (1998) 190–200.
[139] 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.
[140] 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.
[141] 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.
[142] 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).
[143] 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. Physiol. D 20 (2016) 74–84.
[144] 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.
[145] 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.
[146] P.C. Lara, M. Pruschy, M. Zimmermann, L.A. Henriquez-Hernandez, MVP and vaults: a role in the radiation response, Radiat. Oncol. 6 (2011).
[147] K. Park, The role of major vault protein (MVP) in drug resistance, J. Control. Release 163 (2012) 266.
[148] 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. Parasitol. 219 (2018) 1–9.
[149] 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.
[150] 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.
[151] 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.
[152] 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.
[153] A.Z. Komisarczuk, S. Grotmol, F. Nilsen, Ionotropic receptors signal host recognition in the salmon louse (*Lepeophtheirus salmonis*, Copepoda), PLoS One 12 (2017).
[154] 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.
[155] Rasmus Skern-Mauritzen, Thomas Nygaard Mikkelsen, The information continuum model of evolution, Biosystems 209 (2021), https://doi.org/10.1016/j.biosystems.2021.104510. ISSN 0303-2647.