# A 200K SNP chip reveals a novel Pacific salmon louse genotype linked to differential efficacy of emamectin benzoate

Amber M. Messmer, Jong S. Leong, Eric B. Rondeau, Anita Mueller, Cody A. Despins, David R. Minkley, Matthew P. Kent, Sigbjørn Lien, Brad Boyce, Diane Morrison, Mark D. Fast, Joseph D. Norman, Roy G. Danzmann, Ben F. Koop

2018

Marine Genomics 40 (2018) 45–57

Contents lists available at ScienceDirect

Marine Genomics

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

Method paper

# A 200K SNP chip reveals a novel Pacific salmon louse genotype linked to differential efficacy of emamectin benzoate

Amber M. Messmer[a,1], Jong S. Leong[a,1], Eric B. Rondeau[a,1], Anita Mueller[a], Cody A. Despins[a], David R. Minkley[a], Matthew P. Kent[b], Sigbjørn Lien[b], Brad Boyce[c], Diane Morrison[c], Mark D. Fast[d], Joseph D. Norman[e,f], Roy G. Danzmann[e], Ben F. Koop[a,g,*,1]

[a] Department of Biology, University of Victoria, Victoria V8W 2Y2, BC, Canada
[b] Centre for Integrative Genetics (CIGENE), Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, Ås, Norway
[c] Marine Harvest Canada, Campbell River, BC, Canada
[d] Department of Pathology and Microbiology, University of Prince Edward Island, Charlottetown, PEI C1A 4P3, Canada
[e] Department of Integrative Biology, University of Guelph, Guelph, ON N1G 2W1, Canada
[f] Present address: The Hospital for Sick Children, 686 Bay St., Toronto, ON M5G 0A4, Canada
[g] Centre for Biomedical Research, University of Victoria, Victoria, BC V8W 3N5, Canada

ARTICLE INFO

*Keywords:*
Sea lice
*Lepeophtheirus salmonis*
EMB
SLICE
Resistance
Salmon farm

ABSTRACT

Antiparasitic drugs such as emamectin benzoate (EMB) are relied upon to reduce the parasite load, particularly of the sea louse *Lepeophtheirus salmonis,* on farmed salmon. The decline in EMB treatment efficacy for this purpose is an important issue for salmon producers around the world, and particularly for those in the Atlantic Ocean where widespread EMB tolerance in sea lice is recognized as a significant problem. Salmon farms in the Northeast Pacific Ocean have not historically experienced the same issues with treatment efficacy, possibly due to the relatively large population of endemic salmonid hosts that serve to both redistribute surviving lice and dilute populations potentially under selection by introducing naïve lice to farms. Frequent migration of lice among farmed and wild hosts should limit the effect of farm-specific selection pressures on changes to the overall allele frequencies of sea lice in the Pacific Ocean. A previous study using microsatellites examined *L. salmonis oncorhynchi* from 10 Pacific locations from wild and farmed hosts and found no population structure. Recently however, a farm population of sea lice was detected where EMB bioassay exposure tolerance was abnormally elevated. In response, we have developed a Pacific louse draft genome that complements the previously-released Atlantic louse sequence. These genomes were combined with whole-genome re-sequencing data to design a highly sensitive 201,279 marker SNP array applicable for both subspecies (90,827 validated Pacific loci; 153,569 validated Atlantic loci). Notably, kmer spectrum analysis of the re-sequenced samples indicated that Pacific lice exhibit a large within-individual heterozygosity rate (average of 1 in every 72 bases) that is markedly higher than that of Atlantic individuals (1 in every 173 bases). The SNP chip was used to produce a high-density map for Atlantic sea louse linkage group 5 that was previously shown to be associated with EMB tolerance in Atlantic lice. Additionally, 478 Pacific louse samples from farmed and wild hosts obtained between 2005 and 2014 were also genotyped on the array. Clustering analysis allowed us to detect the apparent emergence of an otherwise rare genotype at a high frequency among the lice collected from two farms in 2013 that had reported elevated EMB tolerance. This genotype was not observed in louse samples collected from the same farm in 2010, nor in any lice sampled from other locations prior to 2013. However, this genotype was detected at low frequencies in louse samples from farms in two locations reporting elevated EMB tolerance in 2014. These results suggest that a rare genotype present in Pacific lice may be locally expanded in farms after EMB treatment. Supporting this hypothesis, 437 SNPs associated with this genotype were found to be in a region of linkage group 5 that overlaps the region associated with EMB resistance in Atlantic lice. Finally, five of the top diagnostic SNPs within this region were used to screen lice that had been subjected to an EMB survival assay, revealing a significant

* Corresponding author at: Department of Biology, University of Victoria, Victoria V8W 2Y2, BC, Canada.
*E-mail addresses:* amessmer@uvic.ca (A.M. Messmer), jong@uvic.ca (J.S. Leong), erondeau@uvic.ca (E.B. Rondeau), anita.mueller77@yahoo.de (A. Mueller), cdespins@uvic.ca (C.A. Despins), dminkley@uvic.ca (D.R. Minkley), matthew.peter.kent@nmbu.no (M.P. Kent), sigbjorn.lien@nmbu.no (S. Lien), brad.boyce@marineharvest.com (B. Boyce), diane.morrison@marineharvest.com (D. Morrison), mfast@upei.ca (M.D. Fast), jonorman@gmail.com (J.D. Norman), rdanzman@uoguelph.ca (R.G. Danzmann), bkoop@uvic.ca (B.F. Koop).
[1] Contributed equally to this work.

https://doi.org/10.1016/j.margen.2018.03.005
Received 13 September 2017; Received in revised form 28 February 2018; Accepted 27 March 2018
Available online 16 April 2018

association between these SNPs and EMB treatment outcome. To our knowledge this work is the first report to identify a genetic link to altered EMB efficacy in *L. salmonis* in the Pacific Ocean.

## 1. Introduction

The salmon aquaculture industry is a major global economic contributor, with a production of 3.2 million tonnes in 2013 worth nearly US$18 billion (Food and Agriculture Organization of the United Nations and Fisheries and Aquaculture Department, 2014). In Canada, the fourth largest farmed salmon producer in the world, salmon represent the majority of finfish production, with the value to the Canadian economy of $816 CAD million in 2013, of which 60% was produced in British Columbia (Production Statistics, 2005).

One of the most publicized concerns with aquaculture farms in British Columbia in the last decade surround infestations of the parasitic copepod *Lepeophtheirus salmonis*, commonly referred to as salmon or sea lice. Sea louse infestations are a major concern to the global salmon farming industry, with the most recent estimates of the cost to industry between 4 and 10% of the global value of salmon production in 2008 (Mustafa et al., 2001). In British Columbia in particular a major public focus has been on the role salmon farming may play in increasing the parasite load on vulnerable wild salmon, given the common placement of farms in sheltered inlets which lay along the migratory path of juvenile wild salmon (e.g. (Krkosek et al., 2005)).

The parasite load on farms is managed in part to reduce the risk of transfer of lice from farmed to wild salmon. Reductions in the incidence of louse transfer have been largely attributed to the optimization of louse removal treatment schedules on farms (Peacock et al., 2013). Although multiple treatment methods exist worldwide, the avermectin emamectin benzoate (EMB; trade name SLICE®, Merck Animal Health) has historically been the only method permitted for use in BC. EMB's mode of action is to bind to glutamate-gated chloride channels, which significantly increases neuron permeability and leads to paralysis and death of the parasite (Arena et al., 1995; Stone et al., 1999). Similar to antibiotic resistance in human pathogens, reliance on a single antiparasitic drug may increase the likelihood that populations of drug-tolerant lice will emerge with repeated exposure to EMB. After industry requests for access to an EMB alternative, partially driven by observations of reduced EMB efficacy in some farms, permits were granted in BC for limited regional use of hydrogen peroxide baths in 2014 (http://www.marineharvest.ca/about/news-and-media/container/april-10-2014/).

Resistance to EMB was first detected in Europe in the mid 2000's (Lees et al., 2008) and was identified in Atlantic Canada by the end of that decade (Jones et al., 2013, 2012). The establishment of the resistance phenotype in Atlantic sea lice is mainly attributed overreliance on EMB, though the lack of strong host populations outside the farm environment has likely contributed as well. The selective advantage of surviving drug treatment becomes stronger when alternative host environments, where lice are not exposed to EMB, are less common. Frequent emigration of lice into farms from wild hosts should help counteract the establishment of any particular genotype on farms.

In contrast to the declining effectiveness of EMB treatment in Atlantic Ocean salmon farms, treatment on farms in the Pacific is not associated with widespread EMB resistance (Saksida et al., 2013, 2010), although elevated bioassay EC50s (EC50: half maximal effective concentration) were reported on one farm population in 2013 (Saksida et al., 2013). It has been suggested that the continued effectiveness of EMB in Pacific Ocean farms is partly due to the large and diverse Pacific salmonid population where lice face selection pressures largely unrelated to those in a farm environment. In addition, Pacific lice on farm hosts are typically exposed to EMB much less frequently than those in the Atlantic Ocean with the result that characteristics favouring EMB resistance occur less often than in Atlantic Ocean farms. The frequent influx of lice from wild Pacific salmonids to farms should oppose the adaptive tolerance that farm-site sea lice may acquire to EMB due to genetic drift in founder populations. Population sub-structuring does not appear to exist among geographically distinct groups of lice in the Pacific Ocean (Messmer et al., 2011). The lack of detectable genetic structure in Pacific sea lice suggests that migration among large wild salmonid and other host populations is sufficient to maintain extensive gene flow among lice infecting both farmed and wild salmon. This is consistent with the pattern observed in the Atlantic ocean, where redistribution of lice mediated by wild hosts is also sufficient to maintain panmixia (Besnier et al., 2014; Glover et al., 2011). A reduced number of wild hosts in the Atlantic could, however, limit the degree to which incoming lice can dilute out a selectively advantageous mutation, even as a small number of hosts could rapidly spread resistant lice from a local to oceanic scale through host mixing in common feeding grounds (Besnier et al., 2014; Jacobsen et al., 2012). It could be hypothesized that dilution by large numbers of naïve lice from returning hosts could inhibit the establishment and spread of any such advantageous mutation should it emerge in the Pacific.

Atlantic and Pacific sea lice are allopatric subspecies separated by 2.5–11 million years (Skern-Mauritzen et al., 2014; Yazawa et al., 2008). Many of the molecular tools currently available for the species as a whole, including both the current reference genome (https://licebase.org/) and a 6000-marker single nucleotide polymorphism (SNP) chip (Besnier et al., 2014), were generated using Atlantic sea lice. While the two subspecies are quite close genetically (Yazawa et al., 2008), the addition of tools designed specifically for the Pacific subspecies will allow for more in-depth analysis of Pacific-specific genome evolution and of the ability of these lice to adapt to the multiple hosts that they can infect.

Evidence for recent selective sweeps throughout Atlantic sea lice populations appear to be associated with a drug resistance phenotype (Besnier et al., 2014). Found primarily on sea lice linkage group 5 (abbreviated to SL05 (Besnier et al., 2014)), with minor contributions on SL01 and SL14, it appears that these regions have been under heavy and recent selection and that the genotype linked to resistance has spread quickly throughout the Atlantic. Given the isolation of the two sea louse subspecies it is highly unlikely that this genotype could be introduced to the Pacific Ocean through migration. Any detection of similar genetic signals related to EMB treatment survival in Pacific populations is therefore likely due to convergent or parallel evolution, or the retention of an ancestral or atavistic trait.

In this study we sought to examine whether any distinct genetic characteristics could be detected in lice collected in Pacific regions that have recently shown decreased EMB sensitivity. To further this goal, we produced a genome assembly for a single louse from the Pacific subspecies of *L. salmonis*. This genomic resource was combined with whole-genome re-sequencing information from 43 additional individuals from both Pacific and Atlantic subspecies to identify over 200,000 SNPs; these markers were used to construct a SNP array expected to be useful in studies of both subspecies. Using the re-sequenced samples initially obtained for SNP chip construction, we established that Pacific lice possess markedly higher within-individual heterozygosity than do Atlantic lice. SNP chip genotyping of an additional 478 individuals enabled the identification of sex-linked markers, both within Prohibtin-2 as previously described as well as additional gene-candidates. With removal of sex-linked markers, we observed a recently-emerged genotype in lice collected from those Pacific populations that demonstrated differential EMB efficacy. We demonstrated that this unique genotype is

found at elevated frequencies in population samples where EMB-tolerant phenotypes were present in 2013. While our sampling prior to 2013 is limited, we did not find any individuals with this unique genotype in available samples from earlier years. We show that this genomic region overlaps with the major sweep observed on SL05 in Atlantic sea lice, suggesting a similar mechanism for differential drug efficacy. Finally, we show that for five SNPs in this region of interest, the genotype is significantly linked to EMB treatment survival.

## 2. Methods

### 2.1. Pacific reference genome

A Pacific sea louse reference genome was generated using a single *L. salmonis* individual collected from a pink salmon (*Oncorhynchus gorbuscha*) near Sidney, BC. DNA extraction was performed using a phenol-chloroform (Sambrook et al., 2000) prior to sequencing on Illumina GAIIx and Roche 454 platforms. Sequencing libraries were generated for Illumina PE (400 bp fragments, 501 M total reads from 7 lanes × 101 bp and 1 lane × 76 bp reads; ~80× coverage) and 454 data (13.8 M total reads, 350 bp reads, 8× genome coverage). 454 reads were assembled with Newbler (v2.6) and Illumina reads were assembled with ABySS (v1.3.0) (Simpson et al., 2009), both using default settings. The resulting 333,893 contigs from Illumina (> 500 bp) and 321,836 contigs from Newbler (> 500 bp) (Roche) were combined with PCAP (v06/07/05) (Huang et al., 2003) and screened for contaminates using BLASTN (1e-25) by comparing to GenBank's non-redundant nucleotide database. Potential non-invertebrate sequence matches (primarily bacterial and salmonid) were flagged and manually inspected to determine classification.

### 2.2. SNP array design

To identify SNPs, DNA was obtained from a total of 37 additional Pacific and Atlantic *L. salmonis* libraries (43 individuals - 31 single-individual samples and 6 two-individual pooled Atlantic samples) using either the DNeasy Blood and Tissue Kit or the Gentra Puregene genomic extraction kit (Qiagen). Pooling was performed where individual lice yielded quantities of DNA to low for library preparation. A single lane of whole-genome re-sequencing was performed on each sample using the Illumina HiSeq2000 platform (PE 100 bp - BGI). Atlantic samples were obtained from EMB-sensitive (7 single individuals and one pool of 2) or -resistant lab-reared populations (6 single individuals + 3 pools of two) originally retrieved from the Bay of Fundy or from Atlantic salmon taken from the coast of Newfoundland and Labrador (4 individuals and 2 pools of 2). Pacific samples were retrieved off salmonids from various locales: Hokkaido, Japan (n = 2); Port Moller, Alaska (n = 1); Pedder Bay, BC (n = 1); Klemtu, BC (n = 10); see Table S1. The individual used for the Pacific genome was also used for SNP identification (Sydney, BC: n = 1). Sequencing data for all individuals was aligned to either the Pacific reference (above) or the Atlantic sea lice reference genome (https://licebase.org/) using Burrows-Wheeler Aligner (BWA) (Li and Durbin, 2010). SNPs were called using SamTools (Li et al., 2009) and Integrated genomics viewer (IGV) (Robinson et al., 2011) was used to visually inspect and verify SNP data.

SNPs chosen for the array fell into one of five categories. Group 1 consisted of SNPs designed by (Besnier et al., 2014). These can be further broken down into four subcategories described by (Besnier et al., 2014): i) SNPs at 100 Kb intervals, from the largest available scaffolds ("Gdist" in Table S2); ii) SNP probes proximal to genes, based on EST sequences ("EST" in Table S2); iii) SNP probes that should discriminate between Atlantic populations ("Diag" in Table S2); and, iv) SNP probes across a selection of 10 Kb intervals within single scaffolds obtained in an effort to provide information about linkage disequilibrium ("LD" in Table S2). Group 2 probes were designed to genotype EMB-related loci. The majority of the Group 2 probes (n = 4191) were designed for SNPs that distinguished EMB-sensitive from EMB-tolerant Atlantic lice ("EMB" in Table S2). The rest of the Group 2 probes (n = 768; "EMB-P" in Table S2) targeted SNPs that discriminate between populations with varying EMB susceptibilities within Pacific lice. The putatively-tolerant Pacific lice were taken from the Klemtu Goat Cove farm in February 2013, where the previous EMB treatment had an estimated 5.5% efficacy. The suspected sensitive lice were taken either from wild fish prior to 2010, or from the Klemtu Sheep Pass farm in February 2013, where previous EMB treatment had 60.8% efficacy. Group 3 probes were designed for SNPs found in putative Atlantic and Pacific exons inferred from aligned transcripts, with targets selected at approximately 8 Kb intervals ("EXON" in Table S2). Group 4 probes were designed to avoid known exon regions. Targets were selected in approximately 4 Kb intervals ("NonTrans" in Table S2). Group 5 probes were designed for SNPs present in fewer than 30% of Atlantic lice and were therefore more likely to represent recent mutations. Loci were selected at approximately 10 Kb intervals ("New" in Table S2).

### 2.3. Heterozygosity analysis

Within-individual polymorphism rates for the 31 single-individual (non-pooled) samples used for SNP discovery, as well as for the Pacific reference genome individual, were estimated using the kmer histogram method described by (Vij et al., 2016). Following this method a frequency histogram of all kmers (k = 25) was created for the reads from each sample using the Jellyfish tool (Marçais and Kingsford, 2011). In the resulting distributions two distinct peaks were generally apparent, with one centered at half the kmer frequency value of the other. Such a distribution, in which the first peak is populated by kmers from heterozygous loci and the second with those from homozygous loci, is indicative of a highly heterozygous genome. By comparing the relative numbers of kmers contributing to the two different peaks the polymorphism rate within the genome was estimated.

In order to determine the number of kmers contributing to each peak a mixture model consisting of two negative binomial components, as well as scaling parameters, was fitted to each individual's kmer histogram using the nls function in R (See eq. 1 in (Vij et al., 2016)). Counts of kmers occurring at very low frequencies (those less frequent than the first local minima) largely correspond to sequencing errors and were thus ignored for the purposes of model fitting. In all cases the mixture model fit the data well (see Fig. S1 for example). After fitting, the number of kmers contributing to each peak was calculated by separately summing the area under each of the two model components. Eq. 2 (see (Vij et al., 2016)) was then used to calculate the rate of heterozygosity, the inverse of which was used to establish the polymorphism rate.

### 2.4. Linkage group SL05

A preliminary linkage map was constructed with Atlantic sea lice to assign SNP markers to linkage groups. Eight families of lab-raised Atlantic Canada sea lice consisting of two sets of two paternal half-sib families were used, resulting in an average of 37 offspring per father [range: 24–49, total progeny 296] which genotyped on the 200K SNP chip. LepMap (Rastas et al., 2013) was initially used to remove markers showing extreme deviations from Mendelian expectations. Additionally, all markers for which both parents were heterozygous were removed from the analysis as the phase assignments for heterozygous progeny arising from such crosses cannot be accurately assessed. Assignments of markers to the respective sea lice linkage groups were made at a LOD = 6.0 threshold, a relatively stringent value that allowed for identification of the expected number of linkage groups while ensuring inclusion of only the most confidently placed markers. Initial linkage map construction was made using only zero recombination clusters (ZRC) within the linkage group and singleton markers (i.e.

located to a unique position in the map) that were present in two or more of the mapping parents. Due to the extremely low recombination rate in females across the genome, the genetic map was constructed based on male meioses only. Initial orders were established by removing all markers with any missing genotypes as markers with gaps in their sequence phases may cause a spurious joining between ZRC. ZRC orders were identified for each family using the recombination ordering algorithm (record) within the OneMap software package (Margarido et al., 2007), using a majority-rule consensus map following 5–15 iterations of the core map positions. The record option was found to generate maps with the smallest number of detectable adjacent double crossover positions in the final map order, and therefore, was deemed the most reliable algorithm. A seed marker representing each ZRC was used to establish the map order as this approach greatly decreases computing time (Table S3a). To orient the markers onto the final map, ZRC segments were considered analogous to domains or bins and the overlaps in these domains were oriented using a VB script (Table S3b). This script and several others within the LINKMFEX software platform were utilized to convert data formats and restructure the data into process specific steps (RG Danzmann, http://www.uoguelph.ca/~rdanzman/software.htm). SNPs mapping to SL05 were identified by common markers with the 6 K map (Besnier et al., 2014). SNPs were located on Atlantic genomic scaffolds using BLASTN (Altschul et al., 1990) and the scaffolds were ordered using both individual families and the final merged linkage group into a proximate chromosome SL05 (Table S3c). SNPs that were not mapped in the eight mapping families were placed, where possible, into an approximate position based on the closest SNP on the scaffold to expand the total number of array markers assigned to SL05 (Table S3d).

### 2.5. Genotyping Pacific sea lice populations

DNA extracted from 480 individual Pacific sea lice was genotyped using the 200K SNP array. Individuals screened on the array do not include the re-sequenced individuals used to develop the array. Louse collection information is detailed in Table 1 with host species indicated. Samples were collected from Marine Harvest farms in the Klemtu and Quatsino regions of BC, where differential EMB efficacy had been detected. The highest density of samples was collected in the Klemtu region, with four farm samples and two wild samples (n = 17–48). Quatsino samples included a single farm sample and two wild samples (n = 16–34). Single farm samples near Campbell River, BC (n = 48) and the Broughton Archipelago, BC (n = 35) as well as two wild samples from Pedder Bay, BC (n = 19–29) were similarly genotyped. Outgroups from Japan (n = 12), Bering Sea (n = 41), and Alaska (n = 5–6) were also sampled. DNA was extracted from ethanol or RNA-later-preserved tissue using either DNeasy 96-well extraction plates (Qiagen) or an NaCl extraction protocol (Aljanabi, 1997). The cephalothorax was dissected and used for DNA extractions of female *L. salmonis*, while the entire body was used for extractions of male *L. salmonis*. DNA quantity was estimated using the Qubit HS-DNA kit (Thermo Fisher) and quality was verified by agarose gel electrophoresis.

Genotype filtering and analysis was done in R (R Core Team Computing, 2017) and mainly relied on R/ADEGENET 2.0.1 (Jombart, 2008; Jombart et al., 2010). Genotype calls were categorized by results quality, with those loci possessing off-target variants, a low call rate (< 0.97), no polymorphism or otherwise identified as being generally problematic were excluded from downstream analysis.

### 2.6. Sex linked loci

SNPs were next filtered to remove loci below a conservative global (across all samples) minor allele frequency (MAF) threshold of 1% (> 9 instances of the minor allele). SNPs linked to sex were identified in the data set by grouping individuals into male (n = 140), female (n = 301) and unknown sex (n = 37) groups and using principal component analysis (PCA) followed by a discriminant analysis of principal components (DAPC, retained the top 20 principal components and two discriminant functions) to identify SNPs above a variable contribution threshold of 0.0015. Following this DAPC, the sex of all individuals was confirmed, and in some cases reassigned using the membership probability from this DAPC. To confirm this reassignment, the DAPC was rerun (20 principal components, 1 discriminant function) using reassigned sex. Because sex presented the strongest discriminating signal among individuals, sex-linked loci were removed from subsequent analyses to detect more subtle effects.

### 2.7. Population analysis

Hardy-Weinberg Equilibrium (HWE) exact tests were done for the remaining loci using R/HARDY-WEINBERG ($\alpha = 0.001$) (Graffelman and Weir, 2016). Tests for HWE were done separately for samples from each collection site and one-tailed tests for heterozygote excess and heterozygote deficiency were performed for each locus in each sample group. An excess of missing SNP calls per individual (i.e., genotype failure) of > 3% was used as a criteria to identify and remove individuals with low quality DNA (Anderson et al., 2010).

A PCA followed by DAPC was performed using all high-quality individuals (n = 462), following which the top seven principal components (PC; number chosen by inflection point of scree-plot) and two discriminant functions (DF) retained. This was done both including and excluding the “EMB-P" SNPs, which were designed to differentiate

**Table 1**
Collection sites for sea lice run on the SNP array. Brackets in “Number of lice” column indicate total lice on SNP-array, prior to removal of low-quality individuals.

| Region | Collection site (abbreviation) | Number of lice | Collection date | Last EMB treatment efficacy | Farm/Wild | Host type |
|---|---|---|---|---|---|---|
| Japan | Yoichi, Hokkaido (Ja05) | 12 | 2005 | | Wild | Pacific |
| Bering Sea | Bering Sea (Bs05) | 39 | 12-Jul-05 | | Wild | Pink, Chinook |
| Alaska | Port Kodiak (Pk05) | 6 | 2005 | | Wild | Pacific |
| Alaska | Port Moller (Pm05) | 5 | 2005 | | Wild | Pacific |
| BC | Pedder Bay, Metchosin (PBPK09) | 15 (19) | 22-Aug-09 | | Wild | Pink |
| BC | Pedder Bay, Metchosin (PBCK09) | 28 (29) | 22-Aug-09 | | Wild | Chinook |
| BC | Farside Farm, Campbell River (FSAT10) | 43 (48) | 19-Oct-10 | Unknown | Farm | Atlantic |
| BC | Goat Cove Farm, Klemtu (KGAT10) | 30 (32) | 02-Sep-10 | 96.1% | Farm | Atlantic |
| BC | Sheep Pass Farm, Klemtu (KSAT13) | 45 (48) | 27-Feb-13 | 60.8% | Farm | Atlantic |
| BC | Goat Cove Farm, Klemtu (KGAT13) | 47 (48) | 27-Feb-13 | 5.5% | Farm | Atlantic |
| BC | Carter River, Klemtu (KCPK13) | 33 | 20-Aug-13 | NA | Wild | Pink |
| BC | Goat Cove Farm, Klemtu (KGAT14) | 35 | 12-Nov-14 | 98% | Farm | Atlantic |
| BC | Kid Bay, Klemtu (KKCO14) | 17 | 07-Aug-14 | NA | Wild | Coho |
| BC | Cliff Point, Quatsino Sound (QCCO13) | 16 | 30-Jul-13 | NA | Wild | Coho |
| BC | Cliff Point, Quatsino Sound (QCCK13) | 22 | 30-Jul-13 | NA | Wild | Chinook |
| BC | Mahatta East Farm, Quatsino (QMAT14) | 34 | 09-Sep-14 | 79% | Farm | Atlantic |
| BC | Midsummer Farm, Broughton (BMAT14) | 35 | 03-Nov-14 | 100% | Farm | Atlantic |

between KGAT13 (population abbreviations in Table 1) and all other populations (version "A" with EMB-P loci = 33,194, version "B" without EMB-P loci = 32,806). All loci that contributed to the differences found in DF1 above a variable contribution threshold of 0.001 were retained for further analysis. This set of SNPs of interest (SOI) is referred to as SOIa or SOIb based on the presence or absence of EMB-P SNPs.

An alternative approach using PCA followed by genotype-based K-means clustering of individuals ignored the collection site population and was used to identify any major sample groupings and the genotypes which differentiated them. The number of genotype clusters that best described the data was identified based on the minimum Bayesian Information Criterion (BIC), determined using R/ADEGENET and retaining all principal components, and cluster number selected on a minimum Bayesian Information Criterion (BIC) for clusters greater than one. Following genome-wide cluster prediction DAPC was used to identify the SNPs that best discriminate between individuals that were assigned to different clusters. This cluster-based analysis was performed using both the SOIa and SOIb datasets. SNPs from the SOIa dataset which exceeded a variable contribution threshold of 0.001 were retained for further work; this dataset was labelled SOIc.

### 2.8. Sub-clustering

To further characterize the genotype patterns found in the SNPs of interest (SOI) loci, a second round of K-means clustering was performed using the SOIb loci (see Section 2.7 for description of SNP lists). Because the SOIa and SOIc lists were composed of many SNPs that were specifically designed to differentiate between populations with putatively different EMB tolerances, there existed a possibility that these "over-represented" SNPs would presuppose the principal factors driving differences between populations (e.g., EMB tolerance). The SOIb SNPs were thus separately analysed as they were assumed to be unbiased toward population-level distinctions in the Pacific lice. Following this round of clustering, 200 PCs were retained and the number of groups (K-value) corresponding to the lowest BIC was chosen. For the subsequent DAPC step, four PCs (chosen through SCREE-plot selection) and two DFs were retained. To visualize the patterns of each cluster, SNPs from SOIa, SOIb, and SOIc were pooled into a single list (SOIabc) and SNP genotypes were plotted according to their approximate physical positions (by linkage map, followed by physical position in scaffold), sorted by post-DAPC cluster group membership. Mean observed heterozygosity (Ho) for each group was calculated using the mean locus-by-locus Ho with R/HIERFSTAT (Goudet, 2005). Locus-by-locus $F_{ST}$ (Masatoshi, 1987) was compared among population pairs using R/HIERFSTAT (Goudet, 2005).

### 2.9. Sequencing SNPs in EMB exposed lice

To confirm the association of SNP alleles that drove the initial cluster separations with EMB efficacy, primers were designed to target the 96 SNPs of interest, as identified in the cluster analysis as those with high variable contribution in DAPC. In October 2015, sea lice were collected from the Jackson Pass and Lime Point farms in Klemtu, BC; decreased EMB efficacy had been detected in lice from the previous generation from Lime Point. In an EMB challenge assay (Saksida et al., 2013), sampled lice were exposed to EMB concentrations varying from 62.5 ppb to 1500 ppb. One group of lice remained untreated. Two subsets were identified from the EMB-treated lice: those that died after a treatment of less than or equal to 250 ppb EMB, and those that lived following a treatment of at least 250 ppb. Lice were subsequently stored in RNAlater until DNA extraction using the DNeasy 96 Blood and Tissue kit (Qiagen). Samples from untreated, dead and surviving lice were PCR amplified for SNPs of interest in 10 µl reactions using Promega HotStart GoTaq in the following proportions: 1X buffer, 2.0 mM $MgCl_2$, 0.2 mM each dNTP, 0.5 µM each primer, 0.5 U of GoTaq and ~10 ng of DNA. Reactions were run on a Techne TC-412 using an initial denaturation of 95 °C for 4 min, followed by 40 cycles at 95 °C for 30 s, 52 °C for 30 s and 72 °C for 1 min with a final 72 °C extension for 10 min and a 4 °C hold. Samples were purified using Exonuclease I and Fast Alkaline phosphatase (Thermo Fisher). Sequencing reactions were performed using BigDye3.1 cycle sequencing (ABI) kit with 95 °C for 1 min, followed by 30 cycles at 95 °C for 30 s, 50 °C for 15 s and 60 °C for 3 min with a final 60 °C extension for 5 min and a 4 °C hold. Sequencing reactions were purified by ethanol precipitation and sequenced on a 3730 DNA analyzer (Applied Biosystems) and sequence files were aligned and scored using Geneious Pro v8.1.7 (Kearse et al., 2012). Wild-type and novel alleles were defined from the relative frequency of each in SNP-chip data, and the allele distribution was quantified in each of the groups of lice looking for maximum retention of the novel allele in surviving lice to confirm linkage of the novel genotype to louse survival of EMB treatment. Several samples with degraded DNA led to inconsistent amplification, reflected by variable numbers of lice included in allele counts for each primer pair. Significance in the difference of allele counts for each marker between the groups (alive vs. dead; alive vs. control, dead vs. control were determined by contingency table chi-square tests.

## 3. Results and discussion

### 3.1. Genome sequencing, SNP identification and SNP chip validation

In total, eight lanes of Illumina PE and one lane of Roche 454 sequencing were assembled into 149,469 contigs > 500 bp in size, with a total length of 790,051,552 bp (N50 = 9.7 kb). This Pacific sea louse genome was submitted to NCBI under accession GenBank: ADND00000000.2 - this assembly represents Version 2 of the Pacific *Lepeophtheirus salmonis* WGS.

By mapping the additional 37 libraries (NCBI Bioproject: PRJNA447894) to the Atlantic and Pacific reference sequences, the SNPs necessary for SNP chip construction were identified: 5198 SNPs were included from previous studies (Group 1), 4959 SNPs were EMB response-linked (Group 2), 62,404 SNPs were present in putative transcripts from both Atlantic and Pacific lice (Group 3), 96,082 were from non-coding regions in Atlantic sea lice (Group 4) and 32,636 SNPs were rare Atlantic SNPs (Group 5). Out of a total of 201,279 SNPs included on the array, chip validation yielded 153,569 SNPs useful in Atlantic lice and 90,827 SNPs useful in Pacific lice; see Table S2 for additional information on the SNPs.

The sea louse, *Lepeophtheirus salmonis*, is thought to have first migrated to the Pacific roughly 5 million years ago (Yazawa et al., 2008). In the limited time since this event, they have been able to successfully adapt to the Pacific environment in part by parasitizing new salmonid hosts as well as the non-salmonid threespine stickleback (Jones et al., 2006), a strategy that has not been observed in the Atlantic. Evolutionary divergence of Pacific and Atlantic *L. salmonis* is reflected in their genomes, with 3.2% divergence in nuclear genes and 7.1% divergence in the mitochondrial genome between the two subspecies (Yazawa et al., 2008). As a comparison, the divergence between *L. salmonis* subspecies is greater than the species-level divergence found between humans and orangutans (Glazko and Nei, 2003). For the current study, it was important to include SNP probes designed from both Atlantic and Pacific sea louse genomes for the SNP array to be useful for research on both divergent subspecies. We expected that probes designed exclusively for one subspecies would be less useful in the other subspecies due to genetic drift and geographic isolation. Because both Atlantic and Pacific probes were included on the SNP array, it will be possible to directly compare the specific polymorphisms detected in both subspecies to investigate these differences at a very fine scale. The generation of the first Pacific *L. salmonis* genome sequence allows for the analysis of genome rearrangements, coding and promoter changes that may affect downstream expression differences in comparison to the

more ancestral Atlantic *L. salmonis* genome.

The generated SNP array facilitates alignments of specific genetic variants (i.e., the Pacific subspecies) to genomic scaffolds. Only 38% of the probes reused from the previous 6 K chip had a minor allele frequency > 10% in our Pacific louse study samples. The updated 200K chip provides a SNP an average of once every 8700 bases in Pacific lice and every 5100 bases in Atlantic lice, thereby providing much higher-resolution information about the sea louse genome. Furthermore, because we chose to select new SNP probes that are proximal to or within coding sequences, the 200K SNP array allows for the association of significant SNPs to genes, providing primary candidates for future study of functional differences and selection pressures.

A kmer-based within-individual polymorphism analysis indicated a high degree of heterozygosity for both subspecies (Table S4). Polymorphism was particularly pronounced in Pacific individuals, where the median rate was one heterozygous locus every 72 bases (15 individuals; max: 1/41, min: 1/81), while Atlantic lice had a lesser, but still-notable median rate of 1/171 (17 individuals; Max: 1/128, Min: 1/212) – an example for both subspecies can be found in Fig. S1. An analysis of the Illumina reads which contributed to the creation of the Pacific louse reference sequence evinced a polymorphism rate of 1/64 bp.

Future improvements to this genome assembly are likely to be extremely difficult given the observed polymorphism rate as high as one in every 41 bases. The variation implied by such a high degree of heterozygosity would impede the improvement of genomic assemblies using conventional next-generation sequencing technologies unless significant attention is paid to selecting an individual that is highly homozygous in future sequencing efforts, as reducing variation is known to significantly improve genome assembly contiguity (e.g. (Zhang et al., 2015)). New genome assembly programs developed specifically for highly polymorphic genomes (e.g. dipSPAdes (Safonova et al., 2015)) may aid in improving existing assemblies with existing sequencing data. However, it is likely that while third-generation sequencers and library preparation technologies that allow phased assembly of heterozygous genomes (reviewed in (Jiao and Schneeberger, 2017)) that significant improvements will be seen in such polymorphic species. Until then, the assembly presented within this work will serve as a basis for SNP discovery and gene annotation within this subspecies of salmon lice.

The high rate of polymorphism within the Pacific sea lice analysed is striking. While this rate of heterozygosity is high, it is by no means the most extreme observation in an animal, which is most likely represented by the sea squirt *Ciona savignyi* (Leffler et al., 2012; Small et al., 2007). It is also not necessarily unexpected that small, short-lived, high fecundity species would have a high genetic diversity (Romiguier et al., 2014). Intriguingly this high rate of polymorphism may also imply an increased reservoir of adaptive potential, with large numbers of unique genotypes and haplotypes available to act as a selective substrate. What does seem to be odd is the degree to which heterozygosity in the Atlantic subspecies is low relative to the Pacific subspecies. Whether this difference may be relative to population size, host-diversity or other unknown mechanism is unknown, but will certainly warrant further study in the future.

### 3.2. Linkage group SL05

Linkage mapping produced a new preliminary SNP-dense SL05 linkage group which was incorporated into the previously available linkage groups (Besnier et al., 2014). In all, 7546 markers could be assigned to this linkage group from the array (Table S3b), providing a provisional assignment of 390 scaffolds from the Atlantic reference assembly (Table S3c). The SL05 map utilized in this work is male-specific, as females show a very reduced recombination rate in the families. While several scaffolds were shifted to SL05 in this work relative to the assignment within Besnier et al. (2014), a comprehensive analysis of chimeric scaffold frequency cannot be performed until the full map is available. Only two scaffolds, LSalAtl2s1278 and LSal800, appeared to be chimeric within the linkage group as pieces of the scaffolds were mapped to distinct portions of SL05 – these have been split in Table S3c. We are continuing to work on a comprehensive map that includes the remaining data produced from the families scored on the SNP arrays (Danzmann et al., in review).

While the focus in much of this work is on the Pacific subspecies of lice, the utility of this array in Atlantic lice is experimentally demonstrated in this study through the production of the new linkage group. As the basic principle behind linkage mapping requires recombination to disrupt linkage (Sturtevant, 1913), high resolution linkage mapping in sea lice is hampered by the very low levels of recombination detected in females. This is not exclusive to *L. salmonis* – a lack of female recombination is seen in other copepods such as *Tigriopus californicus* (Edmands, 2008) and lack of male recombination was first described in *Drosophila* over 100 years ago (Morgan, 1914). There are sex-specific differences in the location of recombination events along a chromosome in many species, with some evidence that proximity to the centromere (Singer et al., 2002) or chromosome length (Tortereau et al., 2012) may frequently be the primary determinant of sex-specific recombination frequency. Sex-biased recombination diminishes the number of recombinant events detected in the heterogametic sex, requiring a larger number of offspring to be genotyped to detect additional recombination nodes.

Eight of the 10 males used to sire families for linkage mapping generated enough offspring to produce an informative number of recombination events. Markers for each family were first mapped to scaffolds and the linkage map was then used to order and orientate scaffolds along linkage groups. Markers that were monomorphic in any given family could still be placed on the map by using their proximity to other polymorphic markers within the scaffold, allowing the construction of a pseudo-chromosome sequences for *L. salmonis* SL05. The generation of a new dense linkage map with this SNP chip should provide a useful reference for validating future attempts (e.g., comparisons between Pacific and Atlantic population genomic orders) to create scaffolds for the sea louse genome or to generate region-specific maps to evaluate structural differences between the subspecies.

### 3.3. Pacific sea lice populations and the SNP-chip

The genotypes of 478 of the 480 lice passed initial quality filters and were included for further evaluation. Of the 201,279 SNPs included on the array, 52,677 were categorized as high quality and polymorphic in the Pacific lice. Of these loci, 19,456 were present at a MAF < 1% and were consequently removed from the data set, which left 33,221 SNPs for downstream analysis. Sex-linked SNPs were identified by DAPC comparing between male and female Pacific sea lice, which revealed 27 SNPs associated with sex (Table S5), although one male individual consistently associated with the female cluster suggesting it was initially misidentified. Using this DAPC, the incorrectly sexed louse as well as 33 individuals of unknown sex were reassigned as female (n = 335) and three individuals of unknown sex were reassigned as male (n = 143).

After sex-assignment corrections were made, eight of the 27 sex-linked markers (AX-98433607, AX-98455181, AX-98454911, AX-98367792, AX-98436466, AX-98439922, AX-98428721, AX-98373092) were perfectly associated with sex, while an additional six disagreed at only < 3 individuals (AX-98371709, AX-98427605, AX-98435998, AX-98434653, AX-98436879, AX-98443698), although the miscalled individuals were not consistent among markers. The top markers followed a ZZ-ZW profile consistent with previous results (Carmichael et al., 2013a). Approximately half of these markers are well within or just upstream of expressed sequences. The sequence containing AX-98439922 and AX-98427605 was found around Prohibitin-2, markers AX-98455181, AX-98402462 and AX-98422119 are annotated as

Kinase Suppressor of Ras 2, marker AX-98434653 is annotated as G1/S-specific cyclin-E and marker AX-98436879 as Luc7-like protein 3. Only one marker, AX-98373092, was found on a scaffold anchored to a linkage group; however, additional mapping suggests this portion of the scaffold is likely chimeric and this locus is best assigned to the pool of markers that remain unlinked (data not shown).

A sex-specific marker was previously identified in Atlantic sea lice within the Prohibitin-2 gene (Carmichael et al., 2013a). Unfortunately, this marker was not included on the array, although four markers from this gene were. Two of these markers, AX-98427605 (152 bp away) and AX-98439922 (314 bp away) were among the 27 SNPs identified as being linked to sex. If we ignore a single ambiguously sexed individual, AX-98427605 was observed to have one allelic recombinant with sex, while no recombinants were observed with AX-98439922. This strengthens the case for Prohibitin-2 as a sex-determining candidate in both subspecies of louse, while the report that higher Prohibitin expression diminishes androgen receptor expression (Dart et al., 2009) provides a possible mechanism coupled to sex determination. Our analysis identified three additional gene candidates including a transcript annotated as Kinase Repressor of Ras 2 which incorporated the marker AX-98455181 and correctly assigned sex for all individuals, as well as the transcripts G1/S-specific cyclin-E and Luc7-like protein 3, in which genotypes disagreed with sex in two individuals. Six additional markers on small scaffolds not associated with a transcript were also able to correctly distinguish sex, suggesting that as the genome assembly improves it is possible that other gene candidates will emerge. While this work has been performed in the Pacific subspecies of lice, it is highly likely that with overlapping SNPs-of-interest of the Prohibitin-2 gene that these gene-candidates would also extend to the Atlantic subspecies.

While the identification of sex-linked markers was not the focus of this work, it was necessitated as phenotypic sex was the basis for clustering of populations as identified in the k-means analysis (Section 3.4) when using the full dataset. This “sex” effect has been previously identified in genotyping-by-sequencing based analyses of both American lobster and Arctic charr (Benestan et al., 2017). While not strictly an NGS-based method, the number of markers utilized between the two methods is comparable and thus not surprising that such effects would also be observed in SNP-chip based analyses, again reinforcing the point made by Benestan et al. (2017) that accounting for sex is very important in population-based analyses of non-model organisms.

Removal of the sex-linked markers left 33,194 SNPs (16.5% of all SNPs) for population-based analysis. Thirty loci displayed heterozygote excess in at least one population ($\alpha = 0.001$; Table S6). Of these loci, 25 were flagged in a single population and all others were flagged in two or more populations. A disproportionate number of these loci (11/30) were found on SL05 and eight of those 11 loci were in heterozygote excess only in KGAT13. It is possible that the loci exhibiting heterozygote excess in multiple populations originate from probes that unintentionally target duplicated sequences or repetitive elements. 2514 loci displayed heterozygote deficiency in at least one population and 842 loci displayed heterozygote deficiency in two or more populations ($\alpha = 0.001$; Table S7).

Sixteen lice were identified with > 3% missing SNP calls. These lice included one individual from PBCK09, four individuals from PBPK09, five individuals from FSAT10, two individuals from KGAT10, one individual from KGAT13, and 3 individuals from KSAT13 (see Fig. S2). These individuals were removed from the dataset prior to further analysis, leaving 462 lice for further analysis.

### 3.4. Clustering Pacific populations

The PCA/DAPC analysis for all populations, with geographic locations separated by year, indicated that some individuals from KGAT13, KSAT13, KCPK13, KKCO13, KGAT14, and QMAT14 could be distinguished from all other individuals using SNPs that contributed to DF1. This was true whether EMB-P loci were included in the analysis or not (Fig. S3, panel A and B). In version A of the analysis (SNPs of Interest (SOI)a dataset; all SNP categories; Fig. S3, panel A)), 281 SNP loci contributed to DF1 above the 0.001 loading threshold. Of these loci, 217 belonged to the EMB-P design group, 57 belonged to the EXON design group, and seven belonged to the NonTrans design group. All but three of these loci mapped between 84.62 cM and 95.18 cM on SL05 with one outlier at 37.13 cM; the remaining three loci were unmapped.

In version B of this analysis in which EMB-P loci were excluded, 216 SNPs contributed to DF1 above the loading threshold (SOIb dataset; Fig. S3, panel B)). All but nine of these SNPs mapped between 83.3 cM and 97.81 cM on SL05 and the remaining nine loci were unmapped. Most of these SNPs (n = 189) belonged to the EXON design group, with the remainder split between Gdist, New and NonTrans categories. The highest density of SNPs mapped to 84.62 cM on SL05 in both versions of the analysis. Although each of version A and B could distinguish the same patterns among individuals, the EMB-P SNPs included in version A showed a much clearer pattern distinguishing these individuals from all others (Fig. S3 panel A and B, far right).

To further resolve the differences detected in the initial DAPC analyses, a K-means clustering approach was used to group individuals based on their genotypic, rather than geographic origin (R/ADEGENET::find.clusters). Clustering based on genotype can be useful for detection of cryptic population structure within samples. When K-means clustering was performed using the version A dataset, all individuals were found to belong to one cluster (K = 1; BIC = 3500.354), however, this analysis found only slightly lower support for two clusters (K = 2, BIC = 3501.612; for K = 3–17, BIC = 3505.100–3572.465). A DAPC using the group assignments made with K = 2 identified 297 SNPs above a loading threshold of 0.001 (SOIc dataset; Fig. S3, panel C). Most of these SNPs (273/281) were common to DF1 in both the population group-based DAPC (SOIa) and the cluster group-based DAPC (SOIc). Six of the DF1 cluster SNPs were unmapped while 275 were mapped to SL05 between 84.62 cM and 95.81 cM, with one outlier at 37.13 cM. The individual group assignments after DAPC are illustrated in Fig. 1. This DAPC assigned 392 individuals to Cluster 1 and 70 individuals to Cluster 2 across the entire sample set (Fig. 1 and Table 2). Some uncertainty in group assignment existed for certain individuals. Five individuals (three from KGAT13, one from KSAT13 and one from QMAT14) had membership probabilities lower than 90% assigning them to Cluster 1 or 2. One of these KGAT13 individuals initially assigned to Cluster 2 was reassigned to Cluster 1 following DAPC (Fig. 1).

K-means clustering analysis performed with the version B dataset found less support for two or more clusters than the analysis done using the version A dataset (K = 1, BIC = 3495.010; K = 2, BIC = 3498.194). SNPs identified from the DAPC run using K = 2 group assignments with the version B genotypes were located outside of SL05 and did not define any useful groups for the samples collected (data not shown).

When considering version A of the data (above), all Pacific individuals were assigned to one of the two main clusters, 392 individuals belonged to the first cluster (Cluster 1) and 70 individuals belonging to the second cluster (Cluster 2) after DAPC. None of the lice collected prior to 2013 possessed the Cluster 2 genotype. Fig. 2 illustrates the apparent emergence of the Cluster 2 genotype. The initial detection of Cluster 2 was made in 48% of the lice collected from Klemtu farms in 2013, while no other 2013 lice possessed the Cluster 2 genotype. In 2013, the distribution of Cluster 2 was clearly different between lice collected from farmed and wild salmon in Klemtu; 55 of the 92 lice collected on Klemtu farms (KGAT13 and KSAT13) were assigned to Cluster 2, and one of the 33 lice collected from wild Klemtu salmon (KCPK13) were assigned to Cluster 2. A low abundance of Cluster 2 lice was detected in all sample collections from 2014, including Klemtu (KGAT14) and regions along both the inner and outer coasts of Vancouver Island. Similar proportions of lice were assigned to Cluster 2 in Quatsino and Klemtu farms from 2014 (QMAT14: 5/34 lice, KGAT14:

**Table 2**
The distribution of lice in each of the two clusters described in Section 3.4, for each of the 17 groups of lice in Table 1, identified by population code. Cluster 1 identifies individuals with SL05 identified in all years, with Cluster 2 identifying individuals that share genotype that first emerged in 2013 on farms with decreased efficacy of EMB.

| | BMAT 14 | Bs05 | FSAT 10 | Ja05 | KCPK 13 | KGAT 10 | KGAT 13 | KGAT 14 | KKCO 14 | KSAT 13 | PBCK 09 | PBPK 09 | Pk05 | Pm05 | QCCK 13 | QCCO 13 | QMAT 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Clust 1 | 34 | 39 | 43 | 12 | 32 | 30 | 17 | 28 | 16 | 20 | 28 | 15 | 6 | 5 | 22 | 16 | 29 |
| Clust 2 | 1 | 0 | 0 | 0 | 1 | 0 | 30 | 7 | 1 | 25 | 0 | 0 | 0 | 0 | 0 | 0 | 5 |

7/35 lice). Single instances of the Cluster 2 genotype were found in lice collected from the two wild Klemtu samples, Carter river Pink salmon in 2013 (KCPK13) and Kid Bay Coho in 2014 (KKCO14), as well as the Broughton 2014 farm sample (BMAT14).

Locus-by-locus $F_{ST}$ (Fig. 3) revealed elevated genetic differentiation in SL05. The elevated $F_{ST}$ values were not present when comparing populations sampled in or prior to 2010 (e.g., Fig. 3, panel 1) and were most extreme when comparing 2010 (or earlier) samples to KGAT13 or KSAT13 samples (e.g., Fig. 3, panel 2 and 3). The signal of elevated $F_{ST}$ decreased for sample comparisons made between pre-2013 and post-2013 samples (e.g. Fig. 3, panel 4 and 5). 293 loci had an $F_{ST} > 0.1$ in the most extreme comparison, between KGAT10 and KGAT13. Of these 293 loci, 218 belonged to the EMB-P category, 64 belonged to the EXON category, and 1, 3 and 7 SNPs belonged to each of the Gdist, New, and NonTrans categories. 267 of the SNPs of interest from the DAPC-based analysis (SOIc dataset) had FST values above 0.1 in the comparison of KGAT10 to KGAT13.

The pooled list of interesting SNPs, SOIabc, included 437 loci (Table S8). Most of these loci were found at 84.62 cM on SL05 (Fig. 4). Accounting for relative cM changes between the prior linkage map and the updated map for SL05 made presented here, the region of SL05 that includes the SOIabc SNPs overlaps the primary region that was previously associated with EMB resistance in Atlantic sea lice (Besnier et al., 2014). The SOIabc list is made up of 221 EMB-P SNPs, 189 EXON SNPs, three Gdist SNPs, eight New SNPs, and 16 NonTrans SNPs. At the variable contribution threshold of 0.001 from which the list was generated, 17 out of 68 scaffolds overlapped with the Atlantic results: LSalAtl2s scaffolds 65, 237, 241, 313, 390, 438, 521, 556, 589, 596, 676, 702, 798, 856, 873, 895 and 1003. At a higher threshold of 0.003, 7/18 scaffolds overlapped with the Atlantic results: LSalAtl2s scaffolds 65, 237, 241, 702, 798, 856, 873, and 1003. All indications from these results suggest the region identified in this work in the Pacific subspecies overlaps with the region identified as the major sweep on SL05 in the Atlantic.

Sub-clustering of the main cluster groups identified above was done using SOIb loci using groups of lice collected in 2013 and 2014 (groups defined in Table 1). Sub-clustering identified three to six genotype clusters to which individuals could be assigned (BIC ≈ 854.8 for K = 3–6), though the minimum BIC supported five clusters slightly more than other possibilities (BIC = 853.210). Two of these five groups had members from all 17 population samples (Table S9, Fig. S4 Group 2, 4) and three of these groups were made up of individuals from a subset of populations including KGAT13, KSAT13, KCPK13, KGAT14, KKCO14, QMAT14, and BMAT14 (Table S9, Fig. S4; sub-cluster 1,3,5). Sub-clusters 2 and 4 had reduced heterozygosity across the SOIabc loci (mean Ho = 0.1945, 0.2021 respectively) and these genotypes were generally homozygous for the more common allele. Sub-cluster 3 also had low heterozygosity across SOIabc loci (mean Ho = 0.2365) but these genotypes were dominantly homozygous for the minor allele. Group 1 and 5 were highly heterozygous for SOIabc loci (mean Ho = 0.6095, 0.7150 respectively).

The removal of the sex-linked markers, which interfered with our ability to interpret population-level differences, allowed us to directly address the following question: do patterns of non-neutral genomic differentiation exist among groups of Pacific sea lice? As shown in Fig. 4, we found a narrow region of linkage group 5 containing a multilocus SNP genotype that could be used to sort lice into two distinct clusters (Fig. 1). Fig. S3 panel C illustrates the genotypes that most strongly differentiate these two clusters. Cluster 1 is likely the ancestral genotype, as this is both the dominant genotype and the genotype shared among all lice collected prior to 2013. While both alleles of individual SNPs were found prior to 2013, the consistent association of alleles defining the novel Cluster 2 genotype was first observed in our samples in 2013 in the Klemtu region, where almost half of the sampled lice from farms belonged to this cluster (48%). This change in genotype frequency coincides with a period of reduced EMB treatment success in

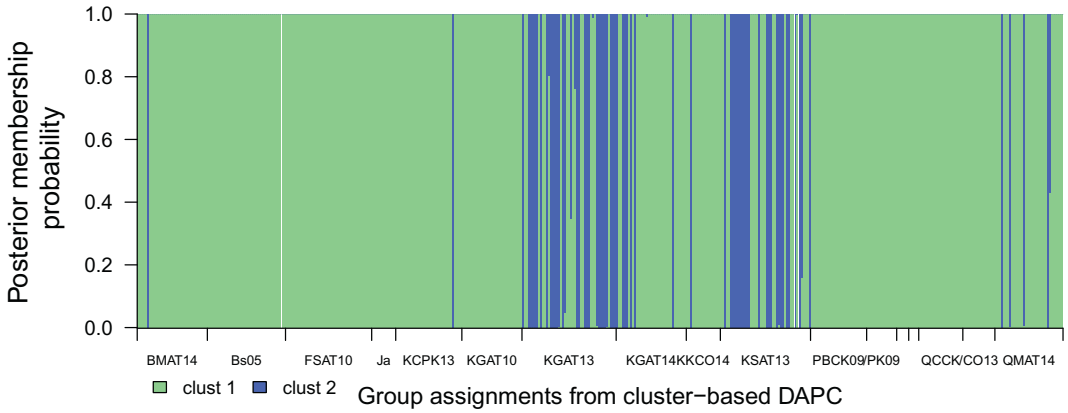


**Fig. 1.** Membership probability of all individuals to 2 clusters, based on genome wide K-means clustering and DAPC using all good quality SNPs (n = 33,194). Individuals are ordered by population. Unless otherwise noted, population codes are as in Table 1. Ja = Ja05, /PK09 = PBPK09, /CO13 = QCCO13. The two unlabeled populations are Pk05 and Pm05. Population information can be found in Table 1.

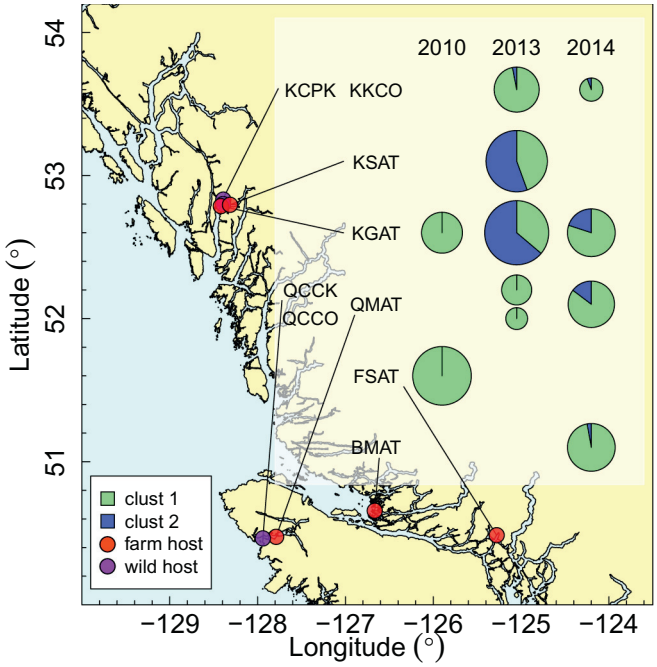


**Fig. 2.** Sample collection sites and lice cluster distributions. Map generated using PBSmapping (Schnute et al., 2008) displays the northern end of Vancouver Island and the Central Coast of British Columbia. Population codes are abbreviated from those in Table 1. Diameter of each pie is proportional to population sample size. All individuals from Ja05, Bs05, Pm05, Pk05, PBCK09, PBPK09 were assigned to Cluster 1 (not shown). KCPK = Klemtu Carter River Wild; KGAT = Klemtu Goat Cove Farm; KSAT = Klemtu Sheep Pass Farm; KKCO = Klemtu Kid Bay Wild; QCCK/QCCO = Quatsino Cliff Point Farm; QMAT = Quatsino Mahatta East Farm; FSAT = Farside Farm Campbell River area; BMAT = Broughton Archipelago Midsummer Farm. Additional population information can be found in Table 1.

Klemtu farms - a historically uncommon occurrence in Pacific Ocean farms. In 2014, Klemtu EMB treatments were more effective. Coinciding with increased susceptibility to EMB, we detected a low frequency of the Cluster 2 genotype in lice collected from this farm and the surrounding area. Our analysis suggests that the prevalence of the Cluster 2 genotype on Klemtu farms in 2013 may be due to a selective sweep following EMB exposure. The frequency of Cluster 2 genotypes may be influenced by the time since last EMB exposure and the subsequent rate of immigration of lice from wild to farm salmon. It is possible that Cluster 2 lice were present in the Pacific Ocean outside Klemtu prior to 2013 and our sample sizes or sampling locations were not adequate to detect it. The strong emergence of this genetic cluster of lice occurring during the same time frame as the appearance of limited EMB treatment efficacy would seem non-random, but certainly cannot be linked directly without more evidence. As Cluster 2 lice dispersed, the frequency of this genotype decreased strongly, likely due to interbreeding with the prevalent Cluster 1 lice and, assuming the Cluster 2 genotype confers an increased tolerance to EMB, the lack of selective advantage of the Cluster 2 genotype when not under the strong selective pressure generated by EMB exposure. The reduction in Cluster 2 genotypes in samples from 2014 relative to those from 2013 may also reflect the introduction of hydrogen peroxide treatments relieving selection pressure for EMB. In 2014, differential EMB efficacy was detected in Quatsino farms, along with an increased presence of the Cluster 2 genotype in this region, although at much lower frequency than observed in Klemtu 2013 samples. This could be due to a dispersal of the Cluster 2 genotype into the larger population directly from Klemtu, or low levels of Cluster 2 genotype found in the general population of lice that have re-emerged in another location following repeated EMB treatments. With these samples we cannot distinguish between the possibilities. Ideally, lice would have been screened from Klemtu, Quatsino and Broughton regions in subsequent years; however, as is illustrated by the number of lice available per sample (Table 1), we had to balance the limitations of sample sizes and sample availability with the costs of screening samples on the SNP array platform.

### 3.5. SNPs and survival assays

To link EMB efficacy with genotype at higher resolution, the top 96 SNPs that defined the Cluster 2 genotype were evaluated using Klemtu farm lice that had undergone an EMB assay. Allele distribution in untreated, survived after treatment and dead after treatment lice were recorded, with allele frequencies summarized in Table 3 for the five most predictive SNPs. Allele frequencies trended toward survival when the louse was a Cluster 2 (novel) homozygote, while treatment was generally lethal when louse was a Cluster 1 (wild-type) homozygote. When lice were heterozygous, survival could not be easily predicted, but overall the proportion of Cluster 2 alleles in surviving lice was significantly elevated based on chi-squared and Fisher's exact probability tests. As can be seen in Table 3, the proportion of Cluster 1 and Cluster 2-linked alleles were significantly different between dead and surviving lice for all five markers, which clearly identifies a strong region of interest in the search for genes involved in the efficacy of EMB in sea lice.

The high abundance of Cluster 2 genotypes in the Klemtu 2013 sample as well as in Klemtu 2014 and Quatsino 2014 samples clearly

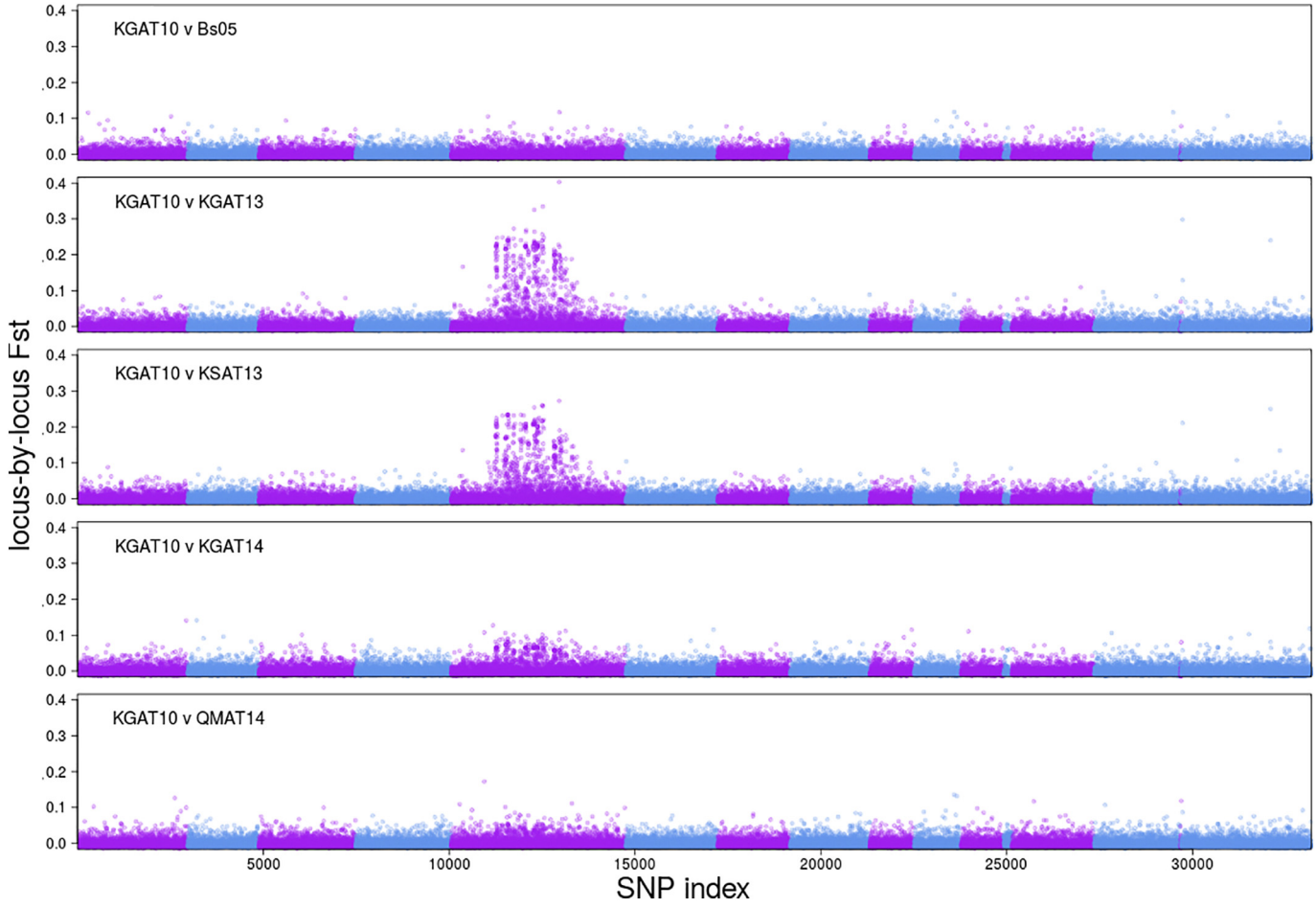


**Fig. 3.** Locus-by-locus pairwise FST for key population comparisons. Blue and purple color blocks represent alternating linkage groups on SL05 with unmapped loci in last block on right. SNPs are ordered by approximate cM position within each block. $F_{ST}$ calculated from (Masatoshi, 1987). Calculations were done for all good quality SNPs (n = 33,194). KGAT10 = Klemtu Goat Cove Farm 2010; KGAT13 = Klemtu Goat Cove Farm 2013; KGAT14 = Klemtu Goat Cove Farm 2014; KSAT13 = Klemtu Sheep Pass Farm 2013; Bs05 = Bering Sea 2005; QMAT14 = Quatsino Mahatta East Farm 2014. Additional population information can be found in Table 1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

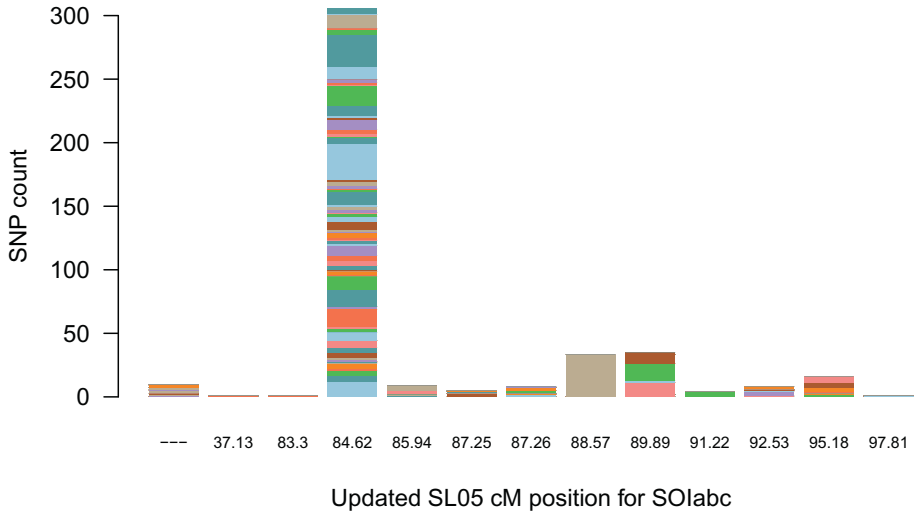


**Fig. 4.** SNP loci above variable contribution threshold (0.001) from three DAPC analyses by their approximate centimorgan positions on the updated SL05. Color blocks in each bar represent genomic scaffold of origin for each SNP. Individual SNP IDs can be found in Table S8, with scaffold associations in Table S2, and centimorgan positions in Table S3.

distinguishes these populations from the background genotype pattern found in other populations at these loci. We designed primers from the top 96 SNPs contributing to the Cluster 2 group to screen directly against a population of Pacific lice where differential reactions to EMB exposure was assessed by bioassay. A summary of the best five SNPs in Table 3 shows that the allele ratios between surviving and susceptible lice were significantly different. While no one marker could be considered diagnostic, the allele frequencies of AX-98352256, AX-98394355, AX-98457619, AX-98417689 and AX-98409177 were all highly significant. ($p < 0.001$).

**Table 3**
EMB assay results using best five SNPs associated with survival in Pacific lice described in Section 2.9 and 3.5.

| SNP ID: | AX-98352256 | AX-98394355 | AX-98457619 | AX-98417689 | AX-98409177 |
|---|---|---|---|---|---|
| Scaffold Atl/ Pac | 3732/Contig_033184 | 65/Contig_011616 | 353/Contig_036060 | 1352/Contig_106101 | 1042/Contig_046982 |
| Atl scaff. position | 6361 bp | 16,055 bp | 93,814 bp | 69,867 bp | 28,075 bp |
| Linkage Group | SL05 | SL05 | SL05 | SL05 | SL05 |
| Position | 84.62 cM | 84.62 cM | 84.62 cM | 84.62 cM | 84.62 cM |
| Forward primer | AAACTAAAGACTCTGCAGCA | GTGCTGTAGACTGGTTTGTA | GGAACTTCATGACGTCCAAGT | TGAGGTTGTGTAGGTGGATA | CGTCATCATCATCCTCAAGT |
| Reverse primer | CTTCTTTTATGGCGTTGGAC | GCTCAGAATGAAATCCAAGC | ACAACCTGCATCATTGTAGT | TTCATTGAGCGAGGAAAGAA | GCAGGGACACACTATAAGAG |
| n AA/AB/BB Pac | 392/47/22 | 349/93/21 | 17/76/371 | 18/75/374 | 19/67/380 |
| n AA/AB/BB Atl | 1577/0/0 | | 0/0/1575 | 0/0/1577 | |
| Allele distribution | | | | | |
| Untreated A/B | 63/27 | 61/27 | 7/35 | 30/58 | 33/63 |
| Dead A/B | 86/16 | 98/20 | 9/79 | 22/98 | 21/101 |
| Alive A/B | 73/71 | 75/79 | 17/21 | 80/68 | 91/71 |
| Chi-squared Test p-values (* ≤ 0.05) | | | | | |
| Untreated to dead | 0.0176* | 0.0202* | 0.296 | 0.3755 | 0.0036* |
| Untreated to alive | 0.0036* | 0.0019* | 0.0062* | 0.0216* | 0.0007* |
| Dead to alive | < 0.0001* | < 0.0001* | < 0.0001* | 0.0007* | < 0.0001* |

Atl = Atlantic *L. salmonis* genome assembly and SNP genotypes, Pac = Pacific *L. salmonis* genome assembly and SNP genotypes. A and B refer to alternate alleles at each SNP locus. In the top section of the table, the position of the SNP within the Atlantic sea lice reference genome (https://licebase.org). Primers designed to amplify the region containing the SNP are presented. The proportion of each genotype from individual lice run on the SNP array presented in this work is also give (n AA/AB/BB) for both Pacific and Atlantic lice. In section Allele Distribution, ratio of A and B alleles observed in the Untreated; Treated and Dead (Dead); and Treated and Alive (Alive) are given for lice collected in EMB survival assays. In section Chi-Squared Test, *p*-values are given between each of the three grouping of lice, with asterisk (*) indicating significant at $p \leq 0.05$.

As these markers all fall into a very narrow region of linkage group 5, we sought to identify genes in this genomic locality. Unfortunately, many of the transcripts in this region remain uncharacterized or poorly annotated. However, given the mode of action of the drug EMB (Carmichael et al., 2013b), a few possible candidates were identified. These include a GABA receptor on scaffold LSalAtl2s575 and a potassium voltage-gated transporter on scaffold LSalAtl2s873 of the Atlantic sea louse reference genome. Preliminary RT-qPCR in unchallenged sea lice with Cluster 1 or Cluster 2 genotypes found a 4-fold increase in the expression of the Potassium voltage-gated transporter in lice with cluster 2 genotypes ($n = 6$ each, $p = 0.003$; data not shown), although no difference could be detected in the GABA receptor. Future improvements to the genome as well as additional populations collected over multiple year classes will hopefully allow for more precise positioning of the peak of association on SL05.

Future work on refining the linkage map, genome and transcriptome for both the Pacific and Atlantic subspecies of *L. salmonis* is needed to generate a more accurate marker order for this important genomic region, as well as a better functional annotation of all sea louse genes as many transcripts in this region remain without annotation. In the future we hope to perform RNA-seq experiments with lice from Cluster 1 and 2 to quantify baseline gene expression differences in all loci from this region of linkage group 5 and improve the list of candidate genes associated with EMB efficacy in Pacific sea lice.

## 4. Conclusions

In this work we presented the first genome assembly for the *L. salmonis oncorhynchi* subspecies. By mapping 37 louse re-sequencing libraries, from both Pacific and Atlantic subspecies, to the Atlantic and Pacific *L. salmonis* reference genomes, we identified genome-wide variation within both sub-species and develop a 200K SNP-chip with markers validated for both. The SNP-chip is used with several Atlantic sea louse families to expand the number of markers on SL05. The chip is also used on Pacific sea lice collected across a wide geographic and temporal range to reveal a novel genotype, first detected in 2013, timed simultaneously with the reduced efficacy of the delousing drug emamectin benzoate (EMB). The novel genotype is in the same genomic region that has been linked to the primary EMB sweep identified in Atlantic sea lice. Finally, we linked this novel genotype directly to the likelihood of treatment survival for a population of Pacific sea lice that show reduced sensitivity to EMB.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.margen.2018.03.005.

## Acknowledgments

AMM performed all extractions prior to sequencing and SNP-chip analysis. JSL, MPK and SL performed genomic alignments, SNP identifications, SNP chip design and validation, and SNP calling. DRM performed heterozygosity analyses. MDF bred families for linkage mapping while RGD performed linkage mapping and EBR performed pseudo-chromosome construction. BB and DM collected samples for SNP chip analysis and EMB assays. AMM drafted original R-script, which was further modified by EBR and AM. CAD and EBR designed primers and analysed genotyped lice from EMB assays. EBR, AMM and DRM drafted the manuscript, and BFK conceived of and coordinated the study. We would like to thank Elanco Animal Health Canada, Marine Harvest Canada, the Atlantic Canada Opportunities Agency - Atlantic Innovation Fund (199308), Compute Canada, the University of Victoria, the University of Prince Edward Island Atlantic Veterinary College and the University of Guelph for financial or material contributions to this work. We would like to thank staff at the BC Centre for Aquatic Health sciences for performing EMB assays, and we would like to thank Ben Sutherland, Scott Pavey and Ryosuke Yazawa for collecting additional lice that contributed to this work.

## References

Aljanabi, S., 1997. Universal and rapid salt-extraction of high quality genomic DNA for

PCR- based techniques. Nucleic Acids Res. 25, 4692–4693. http://dx.doi.org/10.1093/nar/25.22.4692.
Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410. http://dx.doi.org/10.1016/S0022-2836(05)80360-2.
Anderson, C.A., Pettersson, F.H., Clarke, G.M., Cardon, L.R., Morris, A.P., Zondervan, K.T., 2010. Data quality control in genetic case-control association studies. Nat. Protoc. 5, 1564–1573. http://dx.doi.org/10.1038/nprot.2010.116.
Arena, J.P., Liu, K.K., Paress, P.S., Frazier, E.G., Cully, D.F., Mrozik, H., Schaeffer, J.M., 1995. The mechanism of action of avermectins in *Caenorhabditis elegans*: correlation between activation of glutamate-sensitive chloride current, membrane binding, and biological activity. J. Parasitol. 81, 286–294. http://dx.doi.org/10.2307/3283936.
Benestan, L., Moore, J., Sutherland, B.J.G., Le Luyer, J., Maaroufi, H., Rougeux, C., Normandeau, E., Rycroft, N., Atema, J., Harris, L.N., Tallman, R.F., Greenwood, S.J., Clark, F.K., Bernatchez, L., 2017. Sex matters in massive parallel sequencing: evidence for biases in genetic parameter estimation and investigation of sex determination systems. Mol. Ecol. 26, 6767–6783. http://dx.doi.org/10.1111/mec.14217.
Besnier, F., Kent, M., Skern-Mauritzen, R., Lien, S., Malde, K., Edvardsen, R.B., Taylor, S., Ljungfeldt, L.E.R., Nilsen, F., Glover, K.A., 2014. 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, 937. http://dx.doi.org/10.1186/1471-2164-15-937.
Carmichael, S.N., Bekaert, M., Taggart, J.B., Christie, H.R.L., Bassett, D.I., Bron, J.E., Skuce, P.J., Gharbi, K., Skern-Mauritzen, R., Sturm, A., 2013a. Identification of a sex-linked SNP marker in the Salmon louse (Lepeophtheirus salmonis) using RAD sequencing. PLoS One 8, e77832. http://dx.doi.org/10.1371/journal.pone.0077832.
Carmichael, S.N., Bron, J.E., Taggart, J.B., Ireland, J.H., Bekaert, M., Burgess, S.T., Skuce, P.J., Nisbet, A.J., Gharbi, K., Sturm, A., 2013b. Salmon lice (*Lepeophtheirus salmonis*) showing varying emamectin benzoate susceptibilities differ in neuronal acetylcholine receptor and GABA-gated chloride channel mRNA expression. BMC Genomics 14, 408. http://dx.doi.org/10.1186/1471-2164-14-408.
Core Team Computing, R., 2017. R: A Language and Environment for Statistical Computing. (ISBN 3–900051–07-0).
Dart, D.A., Spencer-Dene, B., Gamble, S.C., Waxman, J., Bevan, C.L., 2009. Manipulating prohibitin levels provides evidence for an in vivo role in androgen regulation of prostate tumours. Endocr. Relat. Cancer 16, 1157–1169. http://dx.doi.org/10.1677/ERC-09-0028.
Edmands, S., 2008. Recombination in interpopulation hybrids of the copepod *Tigriopus californicus*: release of beneficial variation despite hybrid breakdown. J. Hered. 99, 316–318. http://dx.doi.org/10.1093/jhered/esn008.
Food and Agriculture Organization of the United Nations, Fisheries and Aquaculture Department, 2014. The State of World Fisheries and Aquaculture: Opportunities and Challenges. Food and Agriculture Organization of the United Nations, Rome.
Glazko, G.V., Nei, M., 2003. Estimation of divergence times for major lineages of primate species. Mol. Biol. Evol. 20, 424–434. http://dx.doi.org/10.1093/molbev/msg050.
Glover, K.A., Stolen, A.B., Messmer, A., Koop, B.F., Torrissen, O., Nilsen, F., 2011. Population genetic structure of the parasitic copepod *Lepeophtheirus salmonis* throughout the Atlantic. Mar. Ecol. Ser. 427, 161–172. http://dx.doi.org/10.3354/meps09045.
Goudet, J., 2005. Hierfstat, a package for r to compute and test hierarchical F-statistics. Mol. Ecol. Resour. 5, 184–186. http://dx.doi.org/10.1111/j.1471-8286.2004.00828.x.
Graffelman, J., Weir, B.S., 2016. Testing for Hardy–Weinberg equilibrium at biallelic genetic markers on the X chromosome. Heredity (Edinb). 116, 558–568. http://dx.doi.org/10.1038/hdy.2016.20.
Huang, X., Wang, J., Aluru, S., Yang, S.P., Hillier, L., 2003. PCAP: a whole-genome assembly program. Genome Res. 13, 2164–2170. http://dx.doi.org/10.1101/gr.1390403.
Jacobsen, J.A., Hansen, L.P., Bakkestuen, V., Halvorsen, R., Reddin, D.G., White, J., Ó Maoiléidigh, N., Russell, I.C., Potter, E.C.E. (Ted), Fowler, M., Smith, G.W., Mork, K.A., Isaksson, A., Oskarsson, S., Karlsson, L., Pedersen, S., 2012. Distribution by origin and sea age of Atlantic salmon (*Salmo salar*) in the sea around the Faroe Islands based on analysis of historical tag recoveries. ICES J. Mar. Sci. 69, 1598–1608.
Jiao, W.-B., Schneeberger, K., 2017. The impact of third generation genomic technologies on plant genome assembly. Curr. Opin. Plant Biol. 36, 64–70. http://dx.doi.org/10.1016/j.pbi.2017.02.002.
Jombart, T., 2008. Adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24, 1403–1405. http://dx.doi.org/10.1093/bioinformatics/btn129.
Jombart, T., Devillard, S., Balloux, F., 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. 11, 94. http://dx.doi.org/10.1186/1471-2156-11-94.
Jones, S.R.M., Prosperi-Porta, G., Kim, E., Callow, P., Hargreaves, N.B., 2006. The occurrence of *Lepeophtheirus salmonis* and *Caligus clemensi* (Copepoda: Caligidae) on Three-Spine Stickleback *Gasterosteus aculeatus* in Coastal British Columbia. J. Parasitol. 92, 473–480. http://dx.doi.org/10.1645/GE-685R1.1.
Jones, P.G., Hammell, K.L., Dohoo, I.R., Revie, C.W., 2012. Effectiveness of emamectin benzoate for treatment of *Lepeophtheirus salmonis* on farmed Atlantic salmon *Salmo salar* in the Bay of Fundy, Canada. Dis. Aquat. Org. 102, 53–64. http://dx.doi.org/10.3354/dao02517.
Jones, P.G., Hammell, K.L., Gettinby, G., Revie, C.W., 2013. Detection of emamectin benzoate tolerance emergence in different life stages of sea lice, *Lepeophtheirus salmonis*, on farmed Atlantic salmon, *Salmo salar* L. J. Fish Dis. 36, 209–220. http://dx.doi.org/10.1111/jfd.12022.
Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649. http://dx.doi.org/10.1093/bioinformatics/bts199.
Krkosek, M., Lewis, M.A., Volpe, J.P., 2005. Transmission dynamics of parasitic sea lice from farm to wild salmon. Proc. R. Soc. B Biol. Sci. 272, 689–696. http://dx.doi.org/10.1098/rspb.2004.3027.
Lees, F., Baillie, M., Gettinby, G., Revie, C.W., 2008. The efficacy of emamectin benzoate against infestations of Lepeophtheirus salmonis on Farmed Atlantic Salmon (*Salmo salar* L) in Scotland, 2002-2006. PLoS One 3, e1549. http://dx.doi.org/10.1371/journal.pone.0001549.
Leffler, E.M., Bullaughey, K., Matute, D.R., Meyer, W.K., Ségurel, L., Venkat, A., Andolfatto, P., Przeworski, M., 2012. Revisiting an old riddle: what determines genetic diversity levels within species? PLoS Biol. 10, e1001388.
Li, H., Durbin, R., 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26, 589–595. http://dx.doi.org/10.1093/bioinformatics/btp698.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079. http://dx.doi.org/10.1093/bioinformatics/btp352.
Marçais, G., Kingsford, C., 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27, 764–770. http://dx.doi.org/10.1093/bioinformatics/btr011.
Margarido, G.R.A., Souza, A.P., Garcia, A.A.F., 2007. OneMap: software for genetic mapping in outcrossing species. Hereditas 144, 78–79. http://dx.doi.org/10.1111/j.2007.0018-0661.02000.x.
Masatoshi, N., 1987. Molecular Evolutionary Genetics. Columbia University Press.
Messmer, A.M., Rondeau, E.B., Jantzen, S.G., Lubieniecki, K.P., Davidson, W.S., Koop, B.F., 2011. Assessment of population structure in Pacific *Lepeophtheirus salmonis* (Kroyer) using single nucleotide polymorphism and microsatellite genetic markers. Aquaculture 320, 183–192. http://dx.doi.org/10.1016/j.aquaculture.2010.09.033.
Morgan, T.H., 1914. No crossing over in the male of *Drosophila* of genes in the second and third Pairs of chromosomes. Biol. Bull. 26, 195–204. http://dx.doi.org/10.2307/1536193.
Mustafa, A., Rankaduwa, W., Campbell, P., 2001. Estimating the cost of sea lice to salmon aquaculture in eastern Canada. Can. Vet. J. 42, 54–56. http://dx.doi.org/10.1111/j.1365-2761.2008.01011.x.
Peacock, S.J., Krkosek, M., Proboszcz, S., Orr, C., Lewis, M.A., 2013. Cessation of a salmon decline with control of parasites. Ecol. Appl. 23, 606–620. http://dx.doi.org/10.1890/12-0519.1.
Production Statistics, 2005. Farmed Salmon [WWW Document]. DFO, URL. http://www.dfo-mpo.gc.ca/aquaculture/sector-secteur/species-especes/salmon-saumon-eng.htm.
Rastas, P., Paulin, L., Hanski, I., Lehtonen, R., Auvinen, P., Brudno, M., 2013. Lep-MAP: fast and accurate linkage map construction for large SNP datasets. Bioinformatics 29, 3128–3134. http://dx.doi.org/10.1093/bioinformatics/btt563.
Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P., 2011. Integrative genomics viewer. Nat. Biotechnol. 29, 24–26. http://dx.doi.org/10.1038/nbt.1754.
Romiguier, J., Gayral, P., Ballenghien, M., Bernard, A., Cahais, V., Chenuil, A., Chiari, Y., Dernat, R., Duret, L., Faivre, N., Loire, E., Lourenco, J.M., Nabholz, B., Roux, C., Tsagkogeorga, G., Weber, A.A.-T., Weinert, L.A., Belkhir, K., Bierne, N., Glémin, S., Galtier, N., 2014. Comparative population genomics in animals uncovers the determinants of genetic diversity. Nature 515, 261.
Safonova, Y., Bankevich, A., Pevzner, P.A., 2015. dipSPAdes: assembler for highly polymorphic diploid genomes. J. Comput. Biol. 22, 528–545. http://dx.doi.org/10.1089/cmb.2014.0153.
Saksida, S.M., Morrison, D., Revie, C.W., 2010. The efficacy of emamectin benzoate against infestations of sea lice, *Lepeophtheirus salmonis*, on farmed Atlantic salmon, *Salmo salar* L., in British Columbia. J. Fish Dis. 33, 913–917. http://dx.doi.org/10.1111/j.1365-2761.2010.01192.x.
Saksida, S.M., Morrison, D., Mckenzie, P., Milligan, B., Downey, E., Boyce, B., Eaves, A., 2013. Use of Atlantic salmon, *Salmo salar* L., farm treatment data and bioassays to assess for resistance of sea lice, *Lepeophtheirus salmonis*, to emamectin benzoate (SLICE®) in British Columbia, Canada. J. Fish Dis. 36, 515–520. http://dx.doi.org/10.1111/jfd.12018.
Sambrook, J., Fritsch, E.F., Maniatis, T., 2000. Molecular Cloning: A Laboratory Manual, 3rd ed. Cold Spring Harbor Laboratory Press.
Schnute, J.T., Boers, N.M., Haigh, R., Couture-Beil, A., 2008. PBS Mapping 2.57: User's Guide, Canadian Technical Report of Fisheries and Aquatic Sciences. Fisheries and Oceans Canada, Science Branch, Pacific Region, Pacific Biological Station.
Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J.M., Birol, I., 2009. ABySS: a parallel assembler for short read sequence data. Genome Res. 19, 1117–1123. http://dx.doi.org/10.1101/gr.089532.108.
Singer, A., Perlman, H., Yan, Y., Walker, C., Corley-Smith, G., Brandhorst, B., Postlethwait, J., 2002. Sex-specific recombination rates in zebrafish (*Danio rerio*). Genetics 160, 649–657.
Skern-Mauritzen, R., Torrissen, O., Glover, K., 2014. Pacific and Atlantic Lepeophtheirus salmonis (Krøyer, 1838) are allopatric subspecies: *Lepeophtheirus salmonis salmonis* and *L. salmonis oncorhynchi* subspecies novo. BMC Genet. 15, 32. http://dx.doi.org/10.1186/1471-2156-15-32.
Small, K.S., Brudno, M., Hill, M.M., Sidow, A., 2007. Extreme genomic variation in a natural population. Proc. Natl. Acad. Sci. 104, 5698 LP–5703.
Stone, J., Sutherland, I.H., Sommerville, C.S., Richards, R.H., Varma, K.J., 1999. The efficacy of emamectin benzoate as an oral treatment of sea lice, *Lepeophtheirus salmonis* (Kroyer), infestations in Atlantic salmon, *Salmo salar* L. J. Fish Dis. 22, 261–270. http://dx.doi.org/10.1046/j.1365-2761.1999.00176.x.
Sturtevant, A.H., 1913. The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. J. Exp. Zool. 14, 43–59. http://dx.doi.org/10.

1002/jez.1400140104.
Tortereau, F., Servin, B., Frantz, L., Megens, H.-J., Milan, D., Rohrer, G., Wiedmann, R., Beever, J., Archibald, A.L., Schook, L.B., Groenen, M.A., 2012. A high density recombination map of the pig reveals a correlation between sex-specific recombination and GC content. BMC Genomics 13, 586. http://dx.doi.org/10.1186/1471-2164-13-586.
Vij, S., Kuhl, H., Kuznetsova, I.S., Komissarov, A., Yurchenko, A.A., Van Heusden, P., Singh, S., Thevasagayam, N.M., Prakki, S.R.S., Purushothaman, K., Saju, J.M., Jiang, J., Mbandi, S.K., Jonas, M., Hin Yan Tong, A., Mwangi, S., Lau, D., Ngoh, S.Y., Liew, W.C., Shen, X., Hon, L.S., Drake, J.P., Boitano, M., Hall, R., Chin, C.S., Lachumanan, R., Korlach, J., Trifonov, V., Kabilov, M., Tupikin, A., Green, D., Moxon, S., Garvin, T., Sedlazeck, F.J., Vurture, G.W., Gopalapillai, G., Kumar Katneni, V., Noble, T.H., Scaria, V., Sivasubbu, S., Jerry, D.R., O'Brien, S.J., Schatz, M.C., Dalmay, T., Turner, S.W., Lok, S., Christoffels, A., Orbán, L., 2016. Chromosomal-level assembly of the Asian Seabass genome using long sequence reads and multi-layered scaffolding. PLoS Genet. 12, e1005954. http://dx.doi.org/10.1371/journal.pgen.1005954.
Yazawa, R., Yasuike, M., Leong, J., Von Schalburg, K.R., Cooper, G.A., Beetz-Sargent, M., Robb, A., Davidson, W.S., Jones, S.R.M., Koop, B.F., 2008. EST and mitochondrial DNA sequences support a distinct Pacific form of salmon louse, *Lepeophtheirus salmonis*. Mar. Biotechnol. 10, 741–749. http://dx.doi.org/10.1007/s10126-008-9112-y.
Zhang, H., Tan, E., Suzuki, Y., Hirose, Y., Kinoshita, S., Okano, H., Kudoh, J., Shimizu, A., Saito, K., Watabe, S., Asakawa, S., 2015. Dramatic improvement in genome assembly achieved using doubled-haploid genomes. Sci. Rep. 4, 6780. http://dx.doi.org/10.1038/srep06780.