Genome wide association study for resistance to *Caligus rogercresseyi* in Atlantic salmon (*Salmo salar* L.) using a 50K SNP genotyping array

Katharina Correa a,b, Jean P. Lhorente b, Liane Bassini c, María E. López c, Alex Di Genova d, Alejandro Maass d, William S. Davidson e, José M. Yáñez a,b,*

**A R T I C L E   I N F O**

Article history:
Received 1 October 2015
Received in revised form 6 April 2016
Accepted 11 April 2016
Available online 14 April 2016

Keywords:
*C. rogercresseyi*
Genome wide association study
Pathogen resistance
Atlantic salmon
Single nucleotide polymorphism
Sea lice

**A B S T R A C T**

The sea louse (*Caligus rogercresseyi*) is an external parasite and considered one of the most important health problems in the salmon farming industry. Resistance to conventional chemical treatments has been demonstrated. Sufficient additive genetic variation has been determined to include selection for resistance to this parasite in Atlantic salmon breeding programs. The aim of this study was to perform a Genome Wide Association Study in order to dissect the genetic factors involved in the resistance to *C. rogercresseyi*, one of the most important species of sea lice in the Chilean salmon farming. 2628 Atlantic salmon smolts, which had been experimentally infested with *C. rogercresseyi*, were genotyped using a 50K SNP array. Genome Wide Association Analysis was conducted using a polygenic model. A heritability of 0.12 for resistance to this louse species was estimated using genomic information. This result was consistent with estimates from previous studies which used pedigree records in the same population. Only one SNP, located on chromosome 21, was significant at a local level, explaining 0.5% of the phenotypic variance and 4% of the genotypic heritability for sea lice resistance. This SNP is located in an intronic region of a predicted gene which codes for Collagen alpha-1. Our results suggest that resistance to *C. rogercresseyi* can be considered a polygenic trait, controlled by many variants of relatively small effect. Thus the incorporation of genomic information through genomic selection could be the most appropriate approach for breeding purposes.

*Statement of relevance:* Caligus resistance has a polygenic genetic architecture.

© 2016 Elsevier B.V. All rights reserved.

---

1. Introduction

Sea lice belong to the subclass Copepoda and include several species worldwide. In Chile, the predominant species is *Caligus rogercresseyi*, which was first described in 1997 by Boxshall and Bravo (2000), and now is considered one of the biggest problems in the salmon farming industry. This parasite affects primarily Atlantic salmon (*Salmo salar*) and rainbow trout (*Oncorhynchus mykiss*), while Coho salmon (*Oncorhynchus kisutch*) has been found to be resistant (Hamilton-West et al., 2012).

Sea lice infestations produce great economic losses due to carcass devaluation, the cost of treatments and increased susceptibility to diseases like Salmon Rickettsial Syndrome (Costello, 2009; Johnson et al., 2004). Salmon infestation by sea lice may cause stress, loss of appetite, depression of the immune system and skin damage (Boxaspen, 2006; Fast et al., 2006; Finstad et al., 2000; Lhorente et al., 2014; Tully and Nolan, 2002). Annual losses caused by different species of sea lice are estimated at more than $178 M USD worldwide (Costello, 2009).

Several studies have been conducted in order to dissect the genetic variation of the resistance to different species of sea lice (Glover et al., 2005; Kolstad et al., 2005; Lhorente et al., 2012). Lhorente et al. (2012) estimated moderate heritabilities for *C. rogercresseyi* resistance in an Atlantic salmon breeding population, showing that there is enough additive genetic variation for selection to be applied to improve resistance to this species of sea lice. Selection for disease resistance may be a strategy to improve fish health, productivity and sustainability in salmon farming (Bishop and Woolliams, 2014; Yáñez et al., 2014b).

Usually, selection for resistance is conducted without knowledge of the causative mutations or genes involved. This process includes information from full- or half-siblings challenged with the pathogen of interest and information from the pedigree (Odegård et al., 2011; Yáñez and Martínez, 2010; Yáñez et al., 2014).

The development of dense Single Nucleotide Polymorphisms (SNP) genotyping panels will contribute to a better understanding of the
genetic factors involved in economically important traits in aquaculture species (Yáñez et al., 2015). Thus, it is possible to dissect the genetic basis of resistance through the use of Genome Wide Association Studies (GWAS), where a group of phenotyped individuals are genotyped using a dense panel of SNPs, and association between a marker and a trait is tested (Goddard and Hayes, 2009). The associated markers can be used to select better breeders through Marker Assisted Selection (MAS). MAS can use a causative mutation that has been identified in a gene or regulatory region that has a major effect; or can use SNPs that are in linkage disequilibrium (LD) with a Quantitative Trait Locus (QTL) (Goddard and Hayes, 2007).

The aim of this study was to conduct a GWAS for resistance to *C. rogercresseyi* in Atlantic salmon, in order to determine both the genetic factors and architecture explaining differences in the susceptibility to this parasite.

2. Materials and methods

2628 Atlantic salmon smolts belonging to 118 maternal full-sib families (40 paternal half-sib families) from the breeding population of Salmones Chaicas, Xth Region, Chile, were challenged with *C. rogercresseyi*. The population used in the present study belongs to the year-class 2010, which has had four generations of selective breeding in Chilean farming conditions. The average number of fish per family was 22, ranging from 9 to 24. Average weight was 274.89 g (SD = 90.6 g). The fish were PIT (Passive Integrated Transponder) tagged, acclimated and distributed in three replicated tanks as described in previous studies (Lhorente et al., 2012; Yáñez et al., 2014c). Fish were negative for Infectious Salmon Anemia Virus, Infectious Pancreatic Necrosis Virus, *Piscirickettsia salmonis*, *Renibacterium salmoninarum* and *Flavobacterium* spp.

The infestation with the parasite was carried out using 13 to 24 copepods per fish, stopping the water flow for 6 hours after infestation. The challenge lasted 6 days, and after this period fish were euthanized and fins from each fish were collected and fixed for processing and lice counting. The resistance trait was defined as the count of sessile lice per fish on all fins after the infestation period, as the number of lice on fins can be used as a good estimator of the total number of lice on the fish (Llloreente et al., 2012). Fish tank and final body weight were recorded for each fish.

Genomic DNA was extracted from fin clips using a commercial kit (DNeasy Blood & Tissue, Qiagen), controlling for DNA quality and quantity. Fish were genotyped using a 50 K SNP Affymetrix® Axiom® myDesign™ Genotyping Array designed by Aquainnovo and the University of Chile (Correa et al., 2015; Yáñez et al., 2016). The markers included in this 50 K array were chosen from a 200 K SNP array previously developed and validated in six commercial populations and two wild populations (Yáñez et al., 2014a, 2016). The 50K SNP selection from the 200 K array is described in detail by (Correa et al., 2015). SNP positions were obtained as described by (Yáñez et al., 2016). Briefly, we positioned the SNPs in the latest genome assembly of Atlantic salmon (GenBank Accession number: GCA_000233375.4) generated by the International Collaboration to Sequence the Atlantic Salmon Genome (ICSASG) (Davidson et al., 2010). SNP probes of 71 base pairs length were aligned to the reference genome and those having a unique location were used to determine SNP positions in the genome assembly. SNP data from the 50 K Array used in this study have been deposited in the SalmonDB database (Di Génova et al., 2011) [http://salmondb.cmm.uchile.cl/download/Array-Aquainnovo-UChile/ARRAY/50K-SNPs_flanking_sequences.txt.gz].

The genotypes were obtained using the Affymetrix Genotyping Console and the SNPPolisherR package following the Axiom® Genotyping Solution Data Analysis Guide (Affymetrix, 2013, 2014). The obtained genotypes were quality-controlled by considering Hardy-Weinberg Disequilibrium (p value <1×10⁻³), Minor Allele Frequency (MAF > 0.01) and genotyping rate for SNP and samples >0.95. For the quality control...
of the genotypes the R statistical software and the GENAbel library (Aulchenko et al., 2007, 2015) were used.

A polygenic mixed model was used to determine the heritability of the resistance trait (lice count), using the polygenic function implemented in GenABEL. Whole genome data were used to estimate kinship and to take into account population structure. Tank was included as factor in the statistical model. A family-based association test was then performed using the Family-based Score Test for Association (FASTA) (Chen and Abecasis, 2007) implemented in GenABEL. We used the maximum likelihood estimates of the intercept, $\alpha$, the proportion of the variance explained by the polygenic component, $\sigma^2_p$, and the residual variance, $\sigma^2_e$, to compute the FASTA test statistics. The significance threshold was determined at a genome wide and chromosome-wide level by means of Bonferroni corrections. At the genome wide level, a SNP was considered significant if its p value was <0.05/M, where M is the total number of markers tested in the GWAS. At the chromosome-wide level, a SNP was considered significant if its p value was <0.05/Mc, where Mc is the number of SNPs tested in a particular chromosome. The proportion of the heritability explained by each marker was determined by comparing heritability values estimated with and without the significant marker genotype as cofactor (Korte and Farlow, 2013). The proportion of phenotypic variance explained by each marker was estimated by multiplying the proportion of the heritability explained times the heritability value. The effect of the favorable allele was obtained from the FASTA test.

### Table 1
Significant marker found for Caligus rogercresseyi resistance in Salmo salar. \( f(A_1) \) and \( f(A_2) \) are the major and minor allele frequencies, respectively. Eff(A1) is the effect of the favorable allele and SE(EffA2) is its respective standard error.

<table>
<thead>
<tr>
<th>Marker</th>
<th>Chr</th>
<th>Position (Chr)</th>
<th>A1</th>
<th>A2</th>
<th>p val</th>
<th>p val*</th>
<th>p valb</th>
<th>p valc</th>
<th>f(A1)</th>
<th>f(A2)</th>
<th>Eff(A1)</th>
<th>SE(EffA2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AQL_Uch-93383729</td>
<td>21</td>
<td>6684025</td>
<td>G</td>
<td>A</td>
<td>7.98E-06</td>
<td>0.315</td>
<td>0.00820</td>
<td>0.54</td>
<td>0.46</td>
<td>−0.5614</td>
<td>0.1257</td>
<td></td>
</tr>
</tbody>
</table>

\( a \) p value obtained with FASTA

\( b \) p value corrected by Bonferroni at genome wide level.

\( c \) p value corrected by Bonferroni at chromosome-wide level.

3. Results

39,508 markers and 2277 samples passed all quality control criteria. The number of SNPs per chromosome ranged from 348 to 2890 (average = 1269, SD = 544). The average lice count was 5.09 (SD = 4.4) and the average final weight was 281 g (SD = 92.8 g). The average lice count per family ranged from 2.3 to 11.3 (Fig. 1). We observed highly resistant families (the ones with lower lice count) and highly susceptible families (the ones with higher lice count). The heritability estimated from the genotype data was 0.12 (p value <0.00001) for the parasite count, matching the one estimated from pedigree data, assessed in the same population using the same fish (Yáñez et al., 2014c), indicating the presence of sufficient genetic variation for this trait. The final weight did not differ significantly from the initial weight.

The results obtained from the GWAS analysis indicate evidence of significance on chromosome 21. The $-\log_{10}(p)$ value for each SNP across the chromosomes was plotted to summarize the GWAS results. P values were corrected for possible inflation caused by family structure. After multiple testing corrections we did not find any significant SNP at a genome wide level (Fig. 2). Only one SNP (AQL_Uch-93383729, in the sequence: TCCCCAGTCGCAAGGCAAATGGTTTCTGCTT[A/G]CTCCATTAAAGGG CTTGAAACTGCTGATGTG) showed significance at a chromosome-wide level (Table 1). The proportion of phenotypic variance explained by this marker is 0.0049 and the proportion of heritability explained is 0.041. The frequency of the favorable allele (A) in this population is intermediate (0.46) and could be increased through selection. This marker is located in an intronic region of a predicted gene (Di Génova et al., 2011), which codes for Collagen alpha-1, suggesting that this gene may be involved in the resistance to sea lice in Atlantic salmon. The quantile–quantile plot of the observed versus expected $-\log_{10}(p)$ values from the association analysis is presented in Fig. 3. We can observe that most of the SNPs (dots) do not show more statistical significance than would be randomly expected given the null hypothesis of no true association (line), suggesting that there are no variants highly associated with the trait.

4. Discussion

Our results confirm the presence of significant genetic variation for resistance against C. rogercresseyi with a relatively low heritability for the trait in this particular population of Atlantic salmon. The heritability estimated with genotypes from a dense SNP array was in accordance with the one estimated from pedigree data in the same population (Yáñez et al., 2014c). The heritability presented here is in the lower limit (0.12) of the range of values reported for resistance against the sessile stage of C. rogercresseyi (0.12–0.34) in a previous study (Lhorente et al., 2012). This may be due to the fact, that the studies were performed in two independent populations, which may implicate differing levels of both environmental and genetic variation. Another cause can be attributed to the different size of the fish used in both studies. The average weight of the fish used in (Lhorente et al., 2012) was lower (130 g), compared to this study (281 g). Nevertheless, the heritability value presented here is within the range of previously reported heritabilities for resistance to another sea louse, Lepeopthirius salmonis: 0.074 (Glover et al., 2005), 0.03–0.19 (Kolstad et al., 2005), 0.13–0.14 (Odegård et al., 2014). It is important to mention that our results confirm that the genetic improvement for C. rogercresseyi resistance is plausible.

The results obtained from the GWAS suggest that C. rogercresseyi resistance has a polygenic genetic architecture. Similar results have been reported for resistance against L. salmonis in Atlantic salmon (Houston et al., 2014), which may indicate that a similar architecture for the genetic resistance against these two ectoparasite species exists in Atlantic salmon. The most significant SNP is located in chromosome 21. More specifically, this variant occurs in an intronic region of a predicted gene (Di Génova et al., 2011), which codes for Collagen alpha-1. Collagen type I can be classified as DAMP (Damage Associated Molecular Pattern).
and initiator of inflammatory cytokine signaling (Castillo-Briecé et al., 2009, 2011). This gene may be involved in the resistance to sea lice, as previous studies have shown that differential expression of collagen genes is present in the response against sea lice in Atlantic salmon (Krasnov et al., 2012; Skugor et al., 2008).

Previous studies suggest that sea louse resistance in Atlantic salmon is a polygenetic trait, with no presence of QTLs of large effect for resistance against L. salmin (Houston et al., 2014; Ødegaard et al., 2014). We did not find any genome wide significant marker associated with C. rogercresseyi resistance, and the amount of variance explained by the locally significant marker was small. Thus, our results also suggest that C. rogercresseyi resistance might be controlled by several alleles of small to moderate effect.

Many genes usually affect quantitative traits and consequently the benefit of conducting MAS depends on the proportion of the variance explained for each QTL (Meuwissen et al., 2001). The results obtained from this study indicate that there may not be a major QTL contributing to the genetic variance for this trait, and thus, the implementation of MAS will probably not be successful. However, molecular information from dense SNP arrays may be incorporated in breeding programs through the application of genomic selection (Goddard and Hayes, 2007), where effects of all genotyped SNPs are included, without the need of surpassing a determined threshold of significance (Meuwissen et al., 2001). Such an approach should be evaluated in order to determine the usefulness of high density SNP arrays for this particular trait.

5. Conclusion

This is the first study aimed at characterizing the genomic architecture of the resistance to C. rogercresseyi in Atlantic salmon. Further studies are necessary to better understand the genomic regulation of this trait. Approaches such as genomic selection should be evaluated to efficiently incorporate molecular information to predict genetic merit of breeding candidates for C. rogercresseyi resistance in order to accelerate genetic progress for this trait.

Acknowledgments

KC (21130669) & MEL (21120382) want to acknowledge the National Commission of Scientific and Technologic Research (CONICYT) for funding through the National PhD funding program. KC wants to acknowledge the Government of Canada for funding through the Canada-Chile Leadership Exchange Scholarship. LB wish to thank CAPES scholarship for PhD studies, Government of Brazil. This project was partially funded by the projects CORFO Innova-Chile (111IE-12843) and U-Visión, Universidad de Chile. This work was partially funded by the grant FONDEF NEWTON-PICARTE (IT14I10100), by CONICYT (Government of Chile) and by the Newton Fund – The British Council (Government of United Kingdom). The authors thank René Figueroa for his help in developing informatics tools. We are grateful to the National Center for High Performance Computing at CMM, Universidad de Chile.

References


