Genome Wide Association Analysis for Resistance to Sea Lice in Atlantic Salmon: Application of a Dense SNP Array

R. D. Houston[1], S.C. Bishop[1], D. R. Guy[2], A. E. Tinch[2], J. B. Taggart[3], J. E. Bron[3], A. Downing[4], M. J. Stear[5], K. Gharbi[4], and A. Hamilton[2]

[1]The Roslin Institute and R(D)SVS, University of Edinburgh, UK [2]Landcatch Natural Selection, Stirling, UK, [3]University of Stirling, UK [4]Edinburgh Genomics, University of Edinburgh, UK, [5]University of Glasgow, UK.

ABSTRACT: The salmon louse (*Lepeophtheirus salmonis*) is an ectoparasitic copepod that is a major threat to the farming of salmonid species. Herein we describe the application of our newly-developed dense SNP array to investigate the genetic basis of resistance to louse infestation. Using a sample of ~600 lice-challenged and genotyped salmon, consistent heritability estimates for lice counts were obtained using pedigree-based and genomic-matrix-based models (h^2 = 0.20-0.24). A GWAS suggested that resistance has a polygenic genetic architecture. The most significant SNP (P~10^{-5}) occurred in a gene which has been shown to differ in expression between louse-infected salmon and uninfected controls.

Keywords:
Salmon Lice
SNP chip
GWAS

Introduction

Salmon lice (*L. salmonis*) present a large and ongoing threat to the production of salmonid fish through aquaculture. The louse copepodids typically attach to the skin of the salmon host, where they feed on its tissue and reproduce. This feeding erodes the epidermis resulting in lesions which, in turn, can lead to osmotic problems and secondary infections (Boxaspen (2006)). Control of lice is reliant on frequent chemotherapeutic interventions. The associated economic and environmental costs are substantial; approximately €300M worldwide in treatment costs alone (Costello (2009)). Encouragingly, genetic variation in host resistance to lice infestation has been demonstrated, with estimates of heritability under experimental challenges of ~0.25 [e.g. Kolstad et al. (2006); Gjerde (2011)]. A QTL affecting lice resistance has been identified on linkage group 6 (Gharbi et al. (2009)), but the genome-wide genetic architecture of resistance is not well characterized.

Extensive genomic tools are available for Atlantic salmon and its genome is in the process of being sequenced and assembled (Davidson et al. (2010)). We have recently developed a dense SNP genotyping array comprising >130K validated SNPs dispersed across the salmon genome (Houston et al. 2014). The SNP density offered by this new Affymetrix array facilitates the application of LD-based, genome-wide association studies and genomic selection for traits of economic importance. In the current study, we used this array to genotype a sample of ~600 Atlantic salmon smolts from a sea lice challenge experiment with the aims of investigating (i) the heritability of resistance under pedigree-based and genomic-matrix-based models and (ii) the genetic architecture of lice resistance via a GWAS.

Materials and Methods

Animals and Lice Challenge. Eggs from the 2007 cohort of LNS broodstock fish were hatched and reared in separate family tanks. At one year post-hatch fish from 62 full sibling families were PIT-tagged and transferred to a single tank. The infection trial was carried out from April 27 to May 14, 2009. The fish (N = 1,479) were infected at the Marine Environmental Research Laboratory (Machrihanish, UK) with a moderate dose of copepodid larvae (96 per fish), and monitored daily until the majority of parasites had moulted into chalimus I. At 7 days post infection (dpi), approximately half of the fish (N = 725) were sampled over a 10-hour period following euthanasia with benzocaine. Each fish was identified via PIT tag, weighed, measured and an adipose fin clip was archived in ethanol. Individual lice counts were obtained from the right side of the fish body using a stereo-microscope**.** From the samples taken at 7 dpi, parents and offspring of families represented by a minimum of 6 fish in the population (61 families) were selected for genotyping.

Genotyping and Quality Control. DNA was extracted using DNeasy-96 tissue DNA extraction kits (Qiagen, Crawley, UK). A total of 712 samples were successfully genotyped using the Affymetrix Axiom SNP array as described in Houston et al. (2014). Starting with the 132K previously-validated SNPs, filtering of SNP data was performed using the Plink software (Purcell et al. (2007)) to remove individuals and SNPs with excessive Mendelian errors and SNPs with minor allele frequency <0.05 in this dataset. A total of 111,259 SNPs genotyped across 624 fish (534 offspring, 29 sires and 61 dams) remained for further analysis. The phenotypic sex of the offspring was unknown and, therefore, the Y-specific probes on the array were used to predict the genetic sex of the fish based on the putative sex determining gene (Yano et al. (2013)), as described in Houston et al. (2014).

Heritability Estimation. The genomic (pairwise IBS) relationship matrix for these fish was calculated in the software Genabel (Aulchenko et al. (2007)). This G-Matrix was then used in the ASReml V3.0 software (Gilmour et al. 2009) to estimate the heritability of lice abundance and compare it to the pedigree-based estimate of heritability, also using ASReml. The animal model fitted included the fixed effect of (genetic) sex, the covariate of body weight and the random effect of the pedigree-based relationship

matrix or the G-matrix. The heritability was then calculated as $h^2a = \sigma^2a/\sigma^2p$, where σ^2a is the additive genetic variance and σ^2p is the total phenotypic variance.

Genome-Wide Association Study. The association of all 111,259 QC-filtered SNPs with lice abundance counts was performed in Genabel, using the G-matrix model described above, followed by the 'mmscore' function on the residuals of this model. A Q-Q plot was generated and P-values were corrected for the genomic inflation factor λ, which accounts for any systematic deviations of observed vs expected P-values.

Results and Discussion

The main aim of this study was to utilize a newly-developed dense SNP array to investigate the genetic architecture of resistance to sea lice in farmed Atlantic salmon. A sample of ~600 lice-challenged salmon and their parents were chosen for analysis. High-quality SNP data were generated for ~111K SNPs segregating in this population (MAF > 0.05). All fish were categorized as males or females based on their mean score of all Y-specific probes on the array. This facilitated the testing of a sex effect in the analysis.

The summary statistics for the lice count data at 7 days post-challenge are given in Table 1. The heritability estimates for lice counts at 7 days post-infection obtained using the pedigree-based relationship matrix (0.24 ± 0.08) and the genomic relationship matrix (0.20 ± 0.07) were significant and consistent with each other (Table 1). This consistency is an encouraging sign that the SNP array can be used effectively *in lieu* of pedigree data to predict the relationship between individual fish and, therefore, genetic parameters. These estimates are also quite consistent with previous studies of experimental lice infections in Atlantic salmon [e.g. Kolstad et al. (2006); Gjerde (2011)]. Body weight had a highly significant effect on lice counts in both models ($P < 0.001$), while there was no significant effect of sex in either model.

The location of the array SNPs on the published Atlantic salmon reference genome assembly (NCBI Assembly GCA_000233375.1) is known, but the assembly is very fragmented (N50 ~ 9.3 KB) and contigs are typically not yet assigned to chromosomal location. Therefore, each SNP was tested for its association with lice abundance independently. This multiple testing resulted in a very stringent Bonferroni-corrected P value threshold ($\sim 10^{-7}$) and no individual SNPs surpassed this threshold. The observed chi—square (1 df) test statistic scores followed the expected test statistic scores closely, with the highest observed test scores typically rising slightly above the expected line on the Q-Q plot (Figure 1). However, this plot also indicates that lice abundance as a proxy for lice resistance is likely to be a polygenic trait controlled by many variants of relatively small effect. Therefore, given its economic importance and difficulty to measure on a large scale, it may be a promising candidate trait for genomic selection. The significant genomic-matrix based heritability suggests that genomic prediction should, in principle, be possible in these data.

Just over one third (36,109) of the SNPs used in the GWAS had previously been mapped to a known chromosome using linkage mapping (Houston et al. 2014). Therefore, for these SNPs it was possible to crudely examine the genomic distribution of the SNPs with the lowest P values (Figure 2). This chart also demonstrates that the majority of these SNPs are as yet unmapped on the salmon genome. By utilizing the lack of recombination observed in male salmon, it will be possible to map a high proportion of these SNPs to chromosome using this dataset. The most significant of the mapped SNPs were found on chromosomes 9 and 22, however the most highly significant SNPs are yet to be mapped (Figure 2).

The SNP array contains a bias towards SNPs from transcribed regions of the salmon genome, due to approximately one third of the validated SNPs being sourced from an RNA-Seq experiment in salmon fry. If interesting SNPs from a GWAS study happen to be derived from RNA-Seq then this facilitates the identification of a putative candidate gene. In the current study, the SNP with the lowest P value (1.3×10^{-5}) was located within a gene that may differ in expression levels between the skin of lice-infected and control fish (Krasnov et al. (2012)). The frequency of the allele associated with higher lice abundance was low (0.12) and, using ASReml, the additive effect was estimated at approximately 5 lice (counted on the right side of the fish only). This is one of several interesting candidate genes for follow-up studies.

Conclusion

We have used a dense SNP array to evaluate the genetic architecture of resistance to sea lice in a challenged, pedigreed population of farmed Atlantic salmon. The estimates of heritability for lice count at 7 days post-challenge were consistent between pedigree (0.24) and genomic relationship (0.20) based analyses. The GWAS demonstrated that the underlying genetic architecture of lice resistance is likely to be polygenic and, as such, may be a good candidate trait for genomic selection using the SNP array. The SNP with the lowest P value (1.3×10^{-5}) was derived originally from RNA-Seq and occurs in a gene which may be important in the skin response of salmon to lice infection. These results provide an example of the utility of the dense SNP array to dissect the genetic basis of complex traits in Atlantic salmon.

Acknowledgements: The authors would like to thank Richard Talbot (Edinburgh Genomics) and Fiona Brew (Affymetrix) for assistance with the SNP array production and genotyping, Bill Roy and Matt Tinsley at the Marine Environmental Research Laboratory for assistance with lice challenges and Oswald Matika for assistance with the analysis. We acknowledge the Technology Strategy Board (project TP 5771-40299), BBSRC (project BB/H022007/1), and DEFRA/Scottish Goverment (LK691) for funding.

Literature Cited

Aulchenko, Y. S., Ripke, S., Isaacs, A. et al. (2007). Bioinformatics, 23:1294-1296.

Boxaspen, C. (2006). ICES Journal of Marine Science, 63: 1304-1316.

Costello, M. (2009). J. Fish Dis., 32:115-118.

Davidson, W. S., Koop, B. F., Jones, S. J. M. et al (2010). Genome Biol. 11:403.

Gharbi, G, Glover, K. A., Stone, L. C. et al. (2009). BMC Genetics, 10:20.

Gilmour, A. R., Gogel, B. J., Cullis B. R. et al. (2009). ASReml user guide release 3.0.

Gjerde, B., Odegard, J. and Thorland, I. (2011). Aquaculture, 314:66-72

Houston, R. D., Taggart, J. B., Cezard, T. et al. (2014). BMC Genomics 15:90.

Kolstad, K., Heuch, P.A., Gjerde, B. et al. (2005). Aquaculture, 247:145-151.

Krasnov, A., Skigor, S., Todorcevic, M. et al. (2012). BMC Genomics, 13:130.

Purcell, S., Neale, B., Todd-Brown, K. et al. (2007). Am. J. Hum. Genet, 81:559-575

Yano, A., Guyomard, R., Nicol, B. et al. (2012). Curr. Biol. 22:1423-1428

Table 1: Summary statistics and heritability estimates for lice counts at 7 days post-infection

Mean	Standard Deviation	h^2P (SE)*	h^2G (SE)*
25.6	12.4	0.24 (0.08)	0.20 (0.07)

*h^2P = heritability estimate using pedigree
h^2G = heritability estimate using genomic relationship

Figure 1: Q-Q plot of observed (circles) versus expected (red line) test statistics in the GWA analysis

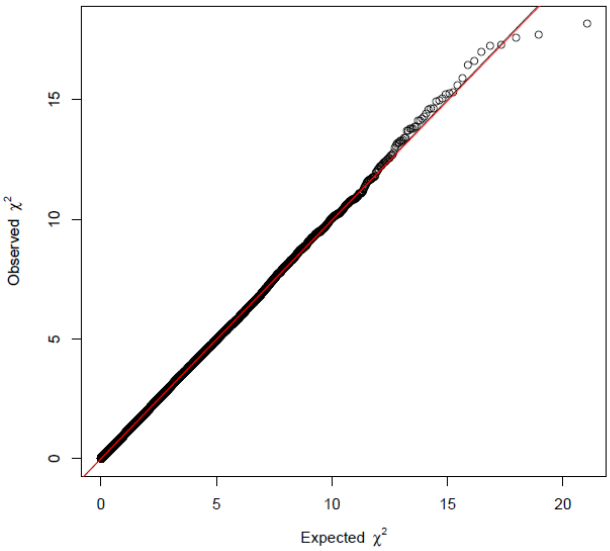

Figure 2: Results of the GWA analysis sorted by chromosome where known (U = Unknown)