Characterising the mechanisms underlying genetic resistance to amoebic gill disease in Atlantic salmon using RNA sequencing

BMC Genomics

RESEARCH ARTICLE

Characterising the mechanisms underlying genetic resistance to amoebic gill disease in Atlantic salmon using RNA sequencing

Diego Robledo[1*] , Alastair Hamilton[2,3], Alejandro P. Gutiérrez[1], James E. Bron[4] and Ross D. Houston[1*]

Abstract

Background: Gill health is one of the main concerns for Atlantic salmon aquaculture, and Amoebic Gill Disease (AGD), attributable to infection by the amoeba *Neoparamoeba perurans,* is a frequent cause of morbidity. In the absence of preventive measures, increasing genetic resistance of salmon to AGD via selective breeding can reduce the incidence of the disease and mitigate gill damage. Understanding the mechanisms leading to AGD resistance and the underlying causative genomic features can aid in this effort, while also providing critical information for the development of other control strategies. AGD resistance is considered to be moderately heritable, and several putative QTL have been identified. The aim of the current study was to improve understanding of the mechanisms underlying AGD resistance, and to identify putative causative genomic factors underlying the QTL. To achieve this, RNA was extracted from the gill and head kidney of AGD resistant and susceptible animals following a challenge with *N. perurans*, and sequenced.

Results: Comparison between resistant and susceptible animals primarily highlighted differences mainly in the local immune response in the gill, involving red blood cell genes and genes related to immune function and cell adhesion. Differentially expressed immune genes pointed to a contrast in Th2 and Th17 responses, which is consistent with the increased heritability observed after successive challenges with the amoeba. Five QTL-region candidate genes showed differential expression, including a gene connected to interferon responses (*GVINP1*), a gene involved in systemic inflammation (*MAP4K4*), and a positive regulator of apoptosis (*TRIM39*). Analyses of allele-specific expression highlighted a gene in the QTL region on chromosome 17, cellular repressor of E1A-stimulated genes 1 (*CREG1*), showing allelic differential expression suggestive of a cis-acting regulatory variant.

Conclusions: In summary, this study provides new insights into the mechanisms of resistance to AGD in Atlantic salmon, and highlights candidate genes for further functional studies that can further elucidate the genomic mechanisms leading to resistance and contribute to enhancing salmon health via improved genomic selection.

Keywords: AGD, Genomics, Amoeba, Gene expression, RNA-seq, Transcriptome, *Salmo salar*, Disease resistance, Allelic specific expression

---

* Correspondence: diego.robledo@roslin.ed.ac.uk;
ross.houston@roslin.ed.ac.uk
[1]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Midlothian EH25 9RG, UK
Full list of author information is available at the end of the article

Background

Gill health is currently one of the major concerns for Atlantic salmon farming worldwide. Fish gills are multifunctional organs fundamental for gas exchange, ionoregulation, osmoregulation, acid-base balance and ammonia excretion, but also play an important role in hormone production and immune defence [1]. Gills are constantly exposed to the marine environment, and are often the first line of defence against pathogens. Gill damage is often observed in Atlantic salmon under farming conditions, and can pose a significant welfare, management and economic burden. While the aetiology of gill disorders is complex, Amoebic Gill Disease (AGD) is currently regarded as a key threat to gill health [2, 3]. This disease adversely affects the gill, and can result in respiratory distress, and ultimately mortality if left untreated. Initially limited to Tasmania, AGD is currently causing major economic and fish welfare burden to Norwegian, Scottish and Australian salmon aquaculture [4]. The causative agent of this disease is the amoeba *Neoparamoeba perurans*, an opportunistic pathogen that typically requires with expensive and laborious fresh water or hydrogen peroxide treatments [5], and there are currently very limited opportunities for prevention.

A promising avenue to decrease the incidence of AGD in farmed Atlantic salmon is to increase genetic resistance of aquaculture stocks to *N. perurans*. There is significant genetic variation in resistance to AGD in commercial Atlantic salmon populations [6–9], therefore selective breeding has potential to improve gill health via a reduction in amoebic load and associated gill damage. The use of genetic markers through genomic selection can expedite genetic gain in aquaculture breeding programmes (e.g. [8, 10–12]), however, the need to genotype a large number of animals and to perform disease challenges in every generation involves a relatively high cost. The discovery of the mechanisms leading to resistance and the underlying causative genetic variants has the potential to reduce this cost via incorporation of functional SNPs into the genomic prediction models.

Discovering the genes and pathways that lead to successful immune responses to pathogens is a major goal in genetics and immunology research. Understanding disease resistance can aid selective breeding via incorporation of putative causative variants with greater weighting in genomic prediction models, which can improve selection accuracy and reduce the need for routine trait recording [13, 14]. Such information can also inform the development of improved disease challenge models, and more successful prevention or treatment strategies through an increased knowledge of host-pathogen interactions. Finally, with the potential role for targeted genome editing (e.g. using CRISPR/Cas9) in

future aquaculture breeding programmes, understanding the functional mechanisms underlying disease resistance traits is key to identifying target genes and variants for editing [15]. Previous studies into AGD-infected Atlantic salmon have suggested that the amoebae might elicit an immunosuppressive effect on the innate response of the host, with a concurrent up-regulation of the adaptive Th2-mediated response [16–18]. Th2 cytokines were also found consistently up-regulated when comparing AGD infected and non-infected samples, and lesion and non-lesion areas [18]. The heritability of resistance to AGD has been shown to increase after successive cycles of disease challenge / treatment [7], which could suggest that the ability to elicit a successful adaptive immune response is partly under genetic control. Finally, a higher expression of genes related to adaptive immunity has been previously reported in more AGD-resistant salmon compared to their more susceptible counterparts using a microarray approach to measure gene expression [19].

In a previous study by our team, several QTL regions with a significant contribution to genetic AGD resistance were identified in Atlantic salmon derived from a commercial breeding programme [8]. In the current study, the gill and head kidney transcriptomes of AGD resistant and susceptible Atlantic salmon from the same population were sequenced and compared. The main goals of the study were a) to assess the differences in local and systemic immune responses between AGD resistant and susceptible Atlantic salmon, and b) to use gene expression data to identify positional and functional candidate genes underlying the previously detected resistance QTL.

Results

Sampling and sequencing

Fish were classified into resistant or susceptible based on their mean gill score and their gill amoebic load. A previous study by our group has shown a high positive genetic correlation between these two traits (higher gill score associated with higher amoebic load), and both are considered indicator traits for resistance to AGD. RNA sequencing (RNA-Seq) was performed on the gill and head kidney of 12 resistant and 12 susceptible fish. Resistant animals had a mean gill score of 2.92 ± 0.13, mean amoebic load (qPCR ct value, high ct value corresponds to low amoebic loads) of 37.12 ± 3.63 and mean weight of 543 ± 116 g at the point of sampling; susceptible animals had a mean gill score of 4.12 ± 0.20, mean amoebic load of 25.99 ± 1.80 and mean weight of 409 ± 96 g. Sequencing of one of the gill samples rendered an extremely low number of reads and therefore was discarded. The remaining samples had an average of 24 M filtered paired-end reads, which were pseudoaligned to transcripts (determine, for each read, which transcript it

is compatible with) in the Atlantic salmon genome (ICSASG_v2; Genbank accession GCF_000233375.1 [20]). Exploratory analyses based on distance measures revealed two head kidney samples as outliers and they were removed (Additional file 1). Therefore, the final dataset comprised of 23 gill and 22 head kidney samples from 24 individuals. The two organs showed clearly distinct patterns of gene expression, as would be expected. The difference in global gene expression pattern between resistant and susceptible samples in both tissues was much less pronounced, but still evident in the gill in particular (Fig. 1). Similar results were described in a Norwegian commercial population [9].

Differential expression

A total of 115 and 42 differentially expressed transcripts (following multiple-testing correction, false discovery rate (FDR) *p*-value < 0.05) were detected between resistant and susceptible samples in gill and head kidney tissues respectively (Fig. 2, Additional file 2). The clearest evidence for differential immune responses was found in

gill, where several differentially expressed immune-related transcripts were detected. Most differentially expressed transcripts in head kidney were not obviously related to AGD or disease resistance. To gain an overall view of the results, a Gene Ontology (GO) enrichment test was performed in both gill and head kidney for sets of differentially expressed transcripts according to three different significance criteria (*p*-value < 0.01, 0.05 and 0.1) (Fig. 3). In the gill, various relevant GO terms were observed, such as "Response to stress", "Cytoskeleton" and "Circulatory system process". A larger number of enriched GO terms were observed in head kidney. While most of them cannot be directly connected to AGD related responses (i.e. "cell proliferation" or "kinase activity"), terms such as "response to stress" or "protein modification process" were observed. For instance, of 22 genes showing *p*-values < 0.01, 15 of them were assigned to "Response to stress". Similar analyses for KEGG pathways did not reveal any significant enrichment.

Detailed inspection of the differentially expressed transcripts in the gill revealed that they can be grouped into
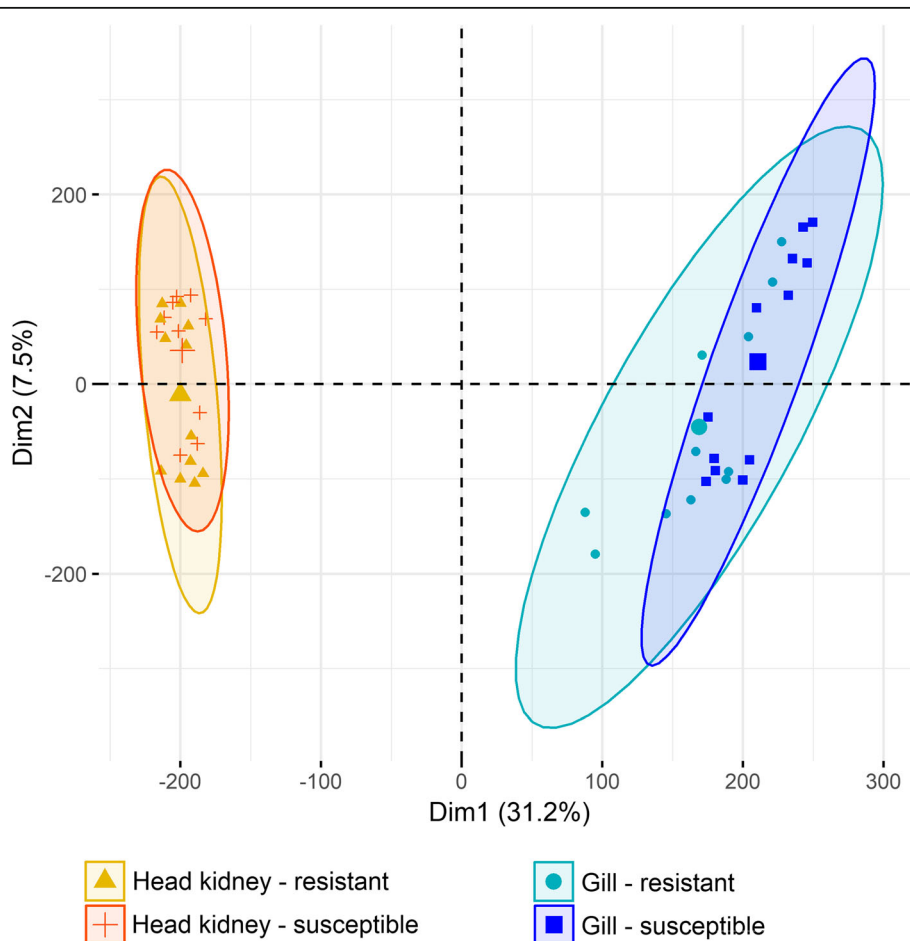
Fig. 1 Principal component analysis. RNA-Seq samples clustered according to their gene expression. The larger symbols represent group means, and ellipses represent 95% confidence intervals for the groups

Fig. 2 Heatmap of differentially expressed genes between resistant and susceptible samples. Heatmaps of all differentially expressed genes in gill (**a**) and head kidney (**b**). Samples and genes were clustered according to gene expression (mean centered and scaled normalized counts)
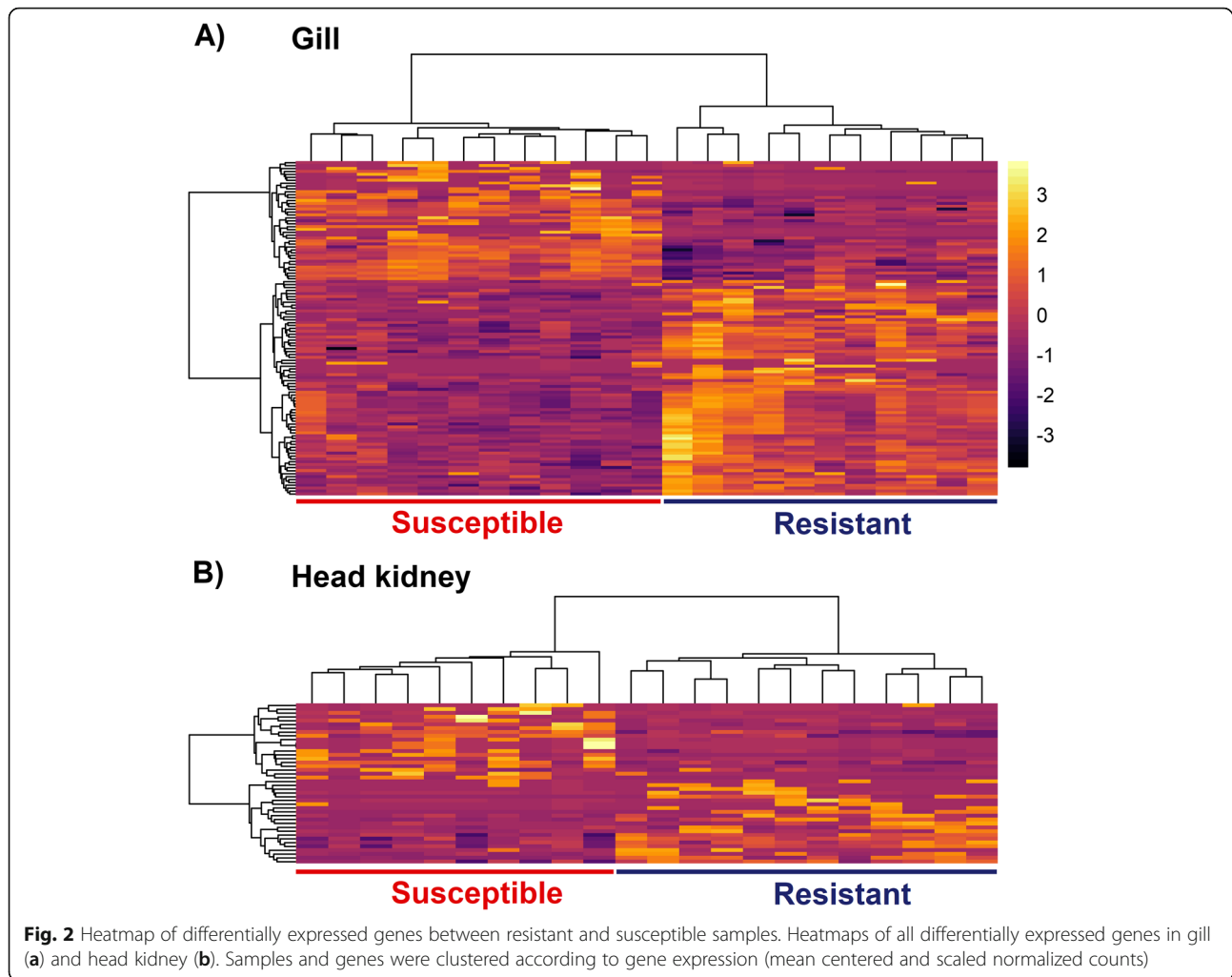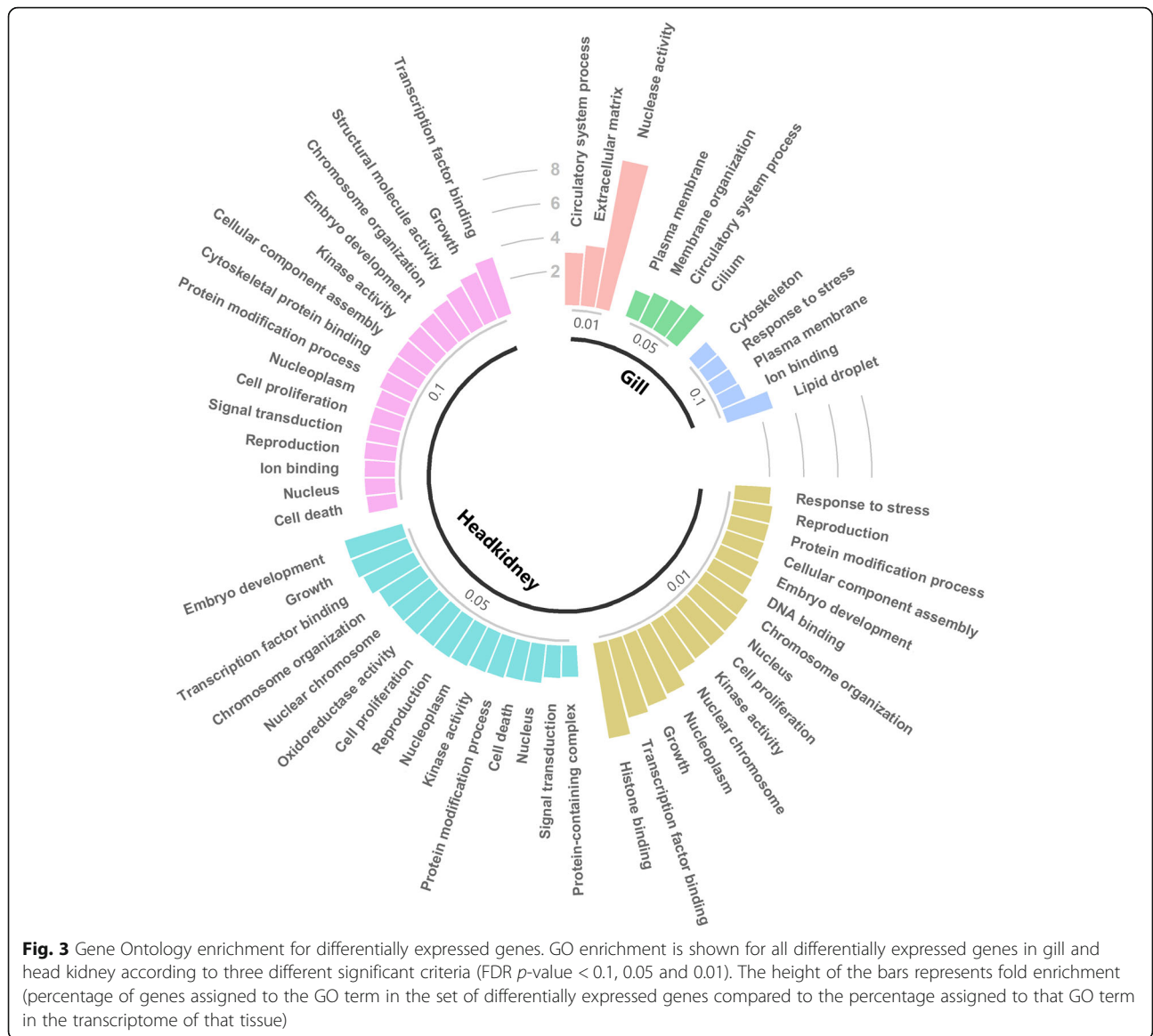
three broad categories concordant with GO enrichment results: 1) immune response ("Response to stress"), 2) red blood cells and coagulation ("Circulatory system process") and 3) cell adhesion or cell shape ("Cytoskeleton").

Amongst the immune-related transcripts showing differential expression in the gill was interleukin-17 receptor E (*IL17RE*), which was highly expressed in resistant animals (Log$_2$ fold change value - logFC = 1.1). In mice *IL17RE* is the receptor for *IL-17C*, which has an essential role in host mucosal defense against infection and is critical for a successful immune response against bacterial infection [21]. The *IL-17C – IL17RE* pair also stimulates T-helper cell 17 responses, which has a proinflammatory effect [22]. *IL-17C* expression was also shown to have a negative correlation with amoebic load in a previous study of Atlantic salmon, and the Th17 pathway in general was found to be significantly down-regulated in response to AGD [16]. This could be a mechanism of immune evasion elicited by the parasite, which might be more effective in susceptible fish than

resistant. Another highly expressed transcript is involved in T-cell function, T-cell specific surface glycoprotein CD28 (*CD28*; logFC = 1.60). *CD28* promotes T-cell survival and proliferation, and enhances the production of multiple cytokines including IL4 [23] *IL4* has been found to be up-regulated in response to AGD [15], and this gene induces differentiation of naïve helper T cells to Th2 cells. The Th2 pathway was found to be up-regulated in late stages of AGD [16]. This pathway is linked to humoral immune responses against extracellular parasites and to tissue repair [24], and therefore is an expected response to AGD. A higher prevalence of this type of response in resistant animals would also be consistent with the observed increase of the heritability of resistance after successive cycles of disease challenge / treatment [7], reflecting genetic variability in the effectiveness of the adaptive response, and / or variation in immune memory.

Several genes connected to red blood cells were found to be differentially expressed, including five different haemoglobin subunit transcripts, which were highly

Fig. 3 Gene Ontology enrichment for differentially expressed genes. GO enrichment is shown for all differentially expressed genes in gill and head kidney according to three different significant criteria (FDR *p*-value < 0.1, 0.05 and 0.01). The height of the bars represents fold enrichment (percentage of genes assigned to the GO term in the set of differentially expressed genes compared to the percentage assigned to that GO term in the transcriptome of that tissue)

expressed and clearly up-regulated in resistant samples in the gill (logFC ~ 2). Haemoglobin α and β subunits have been previously found down-regulated in AGD lesions at the transcript [25] and protein level [26], and reduced hematocrit has been described in AGD infected Atlantic salmon, linked mainly to gill damage [27]. However, it has also been suggested that this haemoglobin dysregulation could be related to antimicrobial peptides derived from haemoglobin β [26], which have been described to have parasiticidal properties in channel catfish [28, 29]. The plasma protease C1 inhibitor gene (*SERPING1*) was also up-regulated (logFC = 1.2). This gene inhibits the complement system and also has anti-inflammatory functions [30]. Complement proteins have been found in gill mucus of AGD infected Atlantic salmon [31]. The lower expression of *SERPING1* in

susceptible samples might simply be a reflection of the higher extent of gill damage in these animals, requiring activation of the complement system and increase of local inflammatory responses.

There are also a few differentially expressed transcripts connected to cell adhesion and cell shape, including a cadherin gene (cadherin-related family member 5; logFC = 4.5) and an actin related gene (actin filament associated protein 1-like 1; logFC = 1.3). Another cadherin gene (Cadherin 1) was previously found dysregulated in response to AGD, along with two additional cell adhesion related genes [25]. The Cdc42 effector protein 2 (*CDC4EP2*; logFC = 0.6) was also up-regulated in resistant fish, and has been associated with roles in actin filament assembly and control of cell shape [32]. A previous study identified an enrichment of cell-adhesion genes in

Robledo *et al. BMC Genomics*   (2020) 21:271

Page 6 of 11

severely affected animals compared to others with healthier gills infected by AGD [9]. These changes are consistent with the epithelial hyperplasia and other structural changes caused by the parasite in the gill of infected animals [26].

In head kidney, most of the differentially expressed (DE) transcripts are seemingly unconnected to biological processes that have previously been related to AGD. Tumor necrosis factor alpha-induced protein 8-like protein 1 (*TNFAIP8L1*; logFC = – 0.9) was more highly expressed in susceptible samples. This gene inhibits apoptosis by suppressing the activity of caspase-8 [33]. The down-regulation of pro-apoptotic genes has been connected to AGD severity [18], however previous studies have not found up-regulation of tumour necrosis factor-alpha (TNA-α), which could potentially suggest some immunomodulation mechanism from the parasite [34]. Nonetheless, the lack of a clear picture in head kidney might reflect the relative importance of the local and systemic immune responses in response to AGD. Previous studies have found that the transcriptomic differences between affected and unaffected gills of AGD infected salmon are similar to those between affected gills of infected salmon and the gills of healthy salmon, suggesting indeed a localized response to AGD [25].

The regulation of transcripts upon infection is a strong indication of the involvement of the gene product in the immune and physiological response of the host to the pathogen, but a comparison between resistant and susceptible animals can offer insight into the mechanisms determining the success of the immune response against the pathogen. The main caveat of this approach is that it is difficult to distinguish cause and consequence, i.e. is the gene differentially expressed because it confers resistance or due to differential disease progression? Additional evidence, such as the co-localization of differentially expressed genes with QTL or the identification of cis regulatory variants in the QTL regions can further contribute to understanding of the mechanisms

of disease resistance, and discover underlying candidate genes.

Integration with previous QTL

The overlap between previously identified putative QTL regions in this population [8] and the differentially expressed genes was explored (Fig. 4). A differentially expressed gene, interferon-induced very large GTPase 1 (*GVINP1*), was found in one of the QTL regions of chromosome 18, which explained ~ 20% of the genetic variance in resistance to AGD (second largest QTL). Very little is known about the function of this gene, but it has been shown to respond to both type I and type II interferon response in mammals [35]. The genes showing FDR corrected *p*-values < 0.1 (a total of 268 genes) were also investigated, and four additional genes were found in these QTL regions. *MAP4K4*, located in a putative QTL region of chromosome 17, surpassed this threshold, and is involved in systemic inflammation in mammals [36], and *TRIM39* in the second QTL region in chromosome 18, a positive regulator of apoptosis [37].

Allele specific expression

To explore potential cis-acting variation underlying the resistance QTL, an allele specific expression (ASE) test was performed for the SNPs in transcripts within the QTL regions (Fig. 5), finding a significant ASE event in a gene in chromosome 17; cellular repressor of E1A-stimulated genes 1 (*CREG1*). In humans this protein is connected to the regulation of cellular proliferation and differentiation [38], and antagonizes the proliferative effects of adenovirus E1A protein [39]. This gene showed a log fold change of 0.75 between resistant and susceptible samples in gill (FDR *p*-value = 0.25). A second significant ASE event was found in an uncharacterised gene in chromosome 18 (LOC106576659; Fig. 5), however this gene showed no differences in fold change between resistant and susceptible samples.
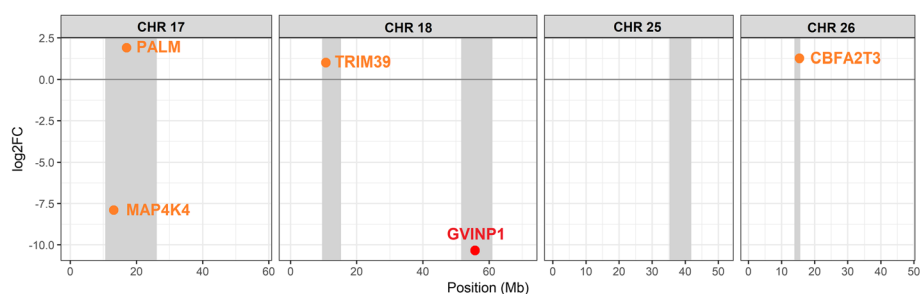
Fig. 4 Differentially expressed genes located in resistance QTL. The location of the QTL regions in the chromosomes are shown in grey. Genes with significance values < 0.05 are in red, those with significance values < 0.1 are in orange. Positive fold changes correspond to higher expression in resistant samples
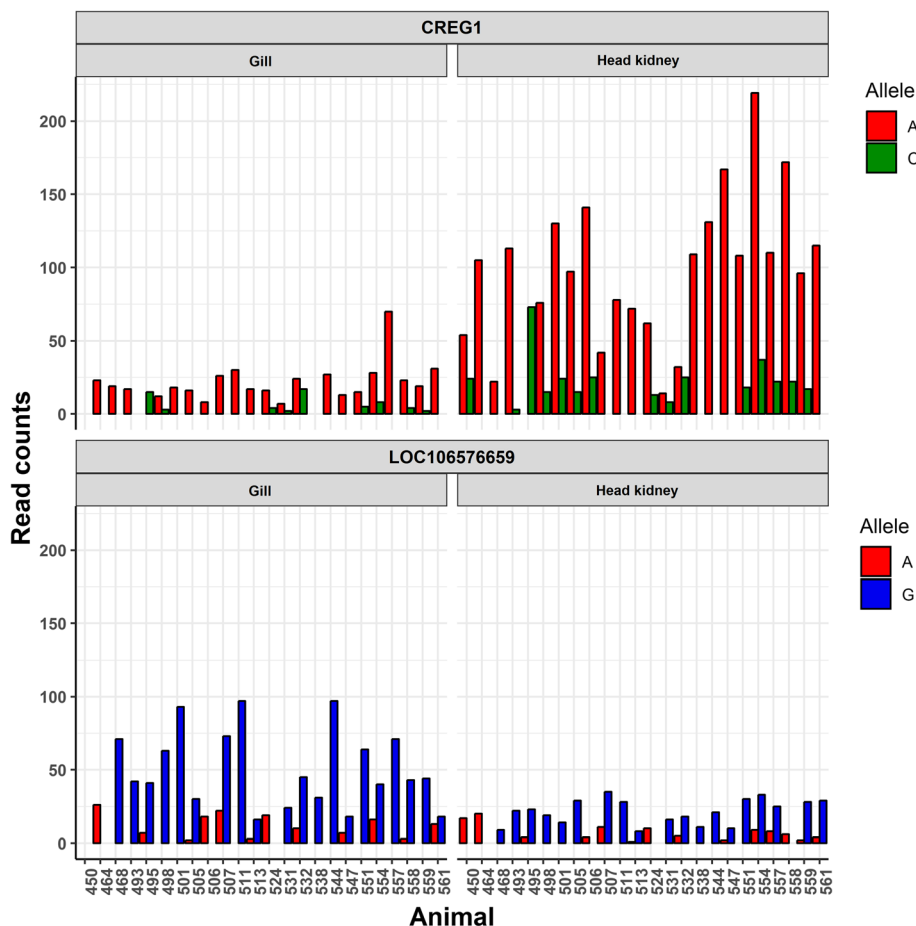
Fig. 5 Allele specific expression CREG1. Barplot showing the read counts for each allele for those SNPs in the QTL regions showing allele specific expression. The two SNPs are located in *CREG1* (Chromosome 17–24,545,527 bp) and the uncharacterized gene LOC106576659 (Chromosome 18–57,163,493 bp)

The polygenic nature of resistance to AGD means that different resistance mechanisms might be operating in each different family. The connection between genotypes and expression, through expression QTL (eQTL) or ASE, can provide strong evidence for functional candidate genes underlying QTL. While eQTL studies require a relatively large number of animals, the advantage of ASE is that the statistical test in performed separately in each heterozygous individual. It is well known that most causative variants are part of regulatory elements and affect gene expression [40, 41], therefore the detection of ASE in a QTL can provide strong evidence linking the function of a gene to the QTL and the phenotype of interest.

Discussion

The potential benefits of the identification of causative variation impacting on complex traits are substantial, ranging from fundamental knowledge of the biology underlying the traits of interest to their application for

enhancing these traits in farmed populations. However, even with the addition of various layers of information such as RNA sequencing, determining the causative gene underlying a QTL is not straightforward, especially because the QTL regions tend to be large and contain a large number of genes, as previously described for sea lice resistance QTL [42].

Eventually, functional assays are necessary to provide actual evidence of its causality. The advent of CRISPR-CAS9 has made this much more feasible in non-model species. Likewise, this technology now provides the opportunity of using this information to introduce or fix favorable alleles in farmed populations [15]. The genetic architecture of quantitative traits usually varies across populations, and indeed AGD resistance QTL seem to vary across different Atlantic salmon commercial populations [6, 7]. While the use of genome editing in farmed animals requires societal and regulatory changes, the transference of causative variants across populations can lead to a rapid increase in disease resistance [15], with

long-lasting effects on animal welfare and food security. Nevertheless, the discovery of causative variants and genes can be used to increase the weight of causative variants in genomic selection, increasing its accuracy and therefore speeding up genetic gain in each generation [13]. More widely, basic knowledge about the pathways leading to resistance to disease can inform drug development or preventive measures such as functional feeds. To summarise, finding the underlying cause(s) of resistance to disease can provide large benefits for aquaculture and society in the form of healthier animals, increased food security and sustainable economic gain, directly through their implementation in breeding schemes in the present and through genome editing in the future. Overlaying genome-wide association studies with gene expression differences between genetically distinct individuals, as performed for resistance to AGD in the current study, is an important step towards identifying these causative mechanisms.

Conclusions

The transcriptomic differences between AGD resistant and AGD susceptible Atlantic salmon are limited, which might not be surprising considering the polygenic nature of the trait. These differences were more evident in the gill than in head kidney, potentially highlighting the importance of the local immune response. Genes involved in immune response (Th2 and Th17 pathways), red blood cells and cell adhesion could be part of the mechanisms leading to AGD resistance, albeit it is difficult to discriminate cause and consequence. The integration of previously discovered QTL and expression data pointed to potential candidate genes of interest, such as *GVINP1*, *MAP4K4* or *TRIM39*. An additional candidate gene, *CREG1*, showed allele specific expression in one of the QTL regions. Follow-up studies to investigate the functional role of these genes in the response to AGD could help improve understanding of the molecular mechanism of resistance to this parasite, and contribute to improving gill health in farmed populations through incorporation of functional data to improve genomic prediction, or potentially via genome editing in the longer term.

Methods

Experimental design

The AGD challenge experiment was performed using 797 Atlantic post-smolt salmon from 132 nuclear families (~ 18 months, mean weight after challenge ~ 464 g) originating from a commercial breeding programme (Landcatch, UK). The challenge experiment was performed as described in [8]. In brief, seeder fish with a similar level of AGD infection (similar gill damage) were produced specifically for this study by cohabitation with

fish infected from an ongoing in vivo culture. The experimental challenge consisted of three separate cycles of infection, established by cohabitation at a ratio of 15% seeder to naïve fish, with a recovery period after the first two infections [7], using a 4 m^3 seawater tank in the experimental facilities of University of Stirling's Marine Environmental Research Laboratory, Machrihanish (Scotland, UK). The fish were kept under a 16-h light and 8-h dark photoperiod, starting at 05:00; the fish were fed Biomar organic salmon feed, automatic every 20 min to approximately 1% biomass; water supply was ambient flow-through filtered to approximately 90 μm, for the duration of the trial, water temperature was between 13 & 14 °C and salinity was 33–35 ppt. Fish were treated with fresh water 21 days after the start of the two first challenges, and fresh water treatment was followed by a week of recovery and the addition of infected seeder fish after that week. The fish were checked visually four times daily during this period. In the third cycle of infection, the disease was allowed to progress until the sampling point, when fish were terminated by an overdose of anaesthetic (Phenoxyethanol, 0.5 mg/L) followed by destruction of the brain according to UK Home Office regulations. All fish were sampled and phenotyped during three consecutive days. Gill damage was recorded for both gills, and scored from 0 to 5 according to the severity of the lesions [43]. A single operator recorded all the gill lesion scores, and the classification was guided by pictures. Additional operators scored some fish, and the scores never differed by > 0.5. Further, one of the gills was stored in ethanol for amoebic load estimation using qPCR with *N. perurans* specific primers. Amoebic load has been used as a suitable indicator of resistance to AGD in Atlantic salmon [7]. All fish in this study had phenotypes for mean gill score (mean of the left gill and right gill scores) and amoebic load (qPCR values using *N. perurans* specific primers, amplified from one of the gills). Twenty-four fish, each from a different full-sib family, were selected for RNA sequencing (Additional file 3) based on high or low levels of resistance according to the measured traits (mean gill score and amoebic load as measured by qPCR). The number of samples was decided based on the availability of animals with extreme phenotypes from different families. The estimated power of the RNA-Seq experiment for this sample size is > 0.8 for genes showing log fold change > |1| and a dispersion of 0.5 [44]. Gill and head kidney samples were obtained from each animal and stored in RNAlater at 4 °C for 24 h, and then at – 20 °C until RNA extraction.

RNA extraction and sequencing

RNA from all the 48 samples was obtained using a standard TRI Reagent RNA extraction protocol. Briefly,

approximately 50 mg of tissue was homogenized in 1 ml of TRI Reagent (Sigma, St. Louis, MO) by shaking using 1.4 mm silica beads, followed by the addition of 100 μl of 1-bromo-3-chloropropane (BCP) for phase separation. Afterwards, 500 μl of isopropanol were added for precipitation, which was followed by subsequent washes with 65–75% ethanol. RNA was resuspended in RNAse-free water and treated with Turbo DNAse (Ambion). Qiagen RNeasy Mini kit columns were used to clean up the samples, and RNA integrity was checked on Agilent 2200 Bioanalyzer (Agilent Technologies, USA). RNA-Seq libraries were prepared using the Illumina Truseq mRNA stranded RNA-Seq Library Prep Kit following standard protocols. Library quantity and quality were quantified using the Bioanalyzer 2100 (Agilent), and then sequenced on three lanes of an Illumina Hiseq 4000 as 75 base paired-end reads at the facilities of Edinburgh Genomics (UK). Raw reads were deposited in NCBI's Sequence Read Archive (SRA) under BioProject accession number PRJNA552604.

Read mapping
Bioinformatic analyses were performed as previously described [45]. Briefly, the quality of the sequencing output was assessed using FastQC v.0.11.5 (http://www.bio-informatics.babraham.ac.uk/projects/fastqc/) and residual adaptors and low quality (< 20) bases were trimmed using Trimmomatic v.0.38 [46]. Only reads where both pairs were longer than 36 bp post-filtering were retained. STAR v.2.6.1a [47] was used to map the filtered reads to the most recent Atlantic salmon genome assembly (ICSASG_v2; Genbank accession GCF_000233375.1 [20]). Kallisto v0.44.0 [48] and the latest Atlantic salmon genome annotation (NCBI *Salmo salar* Annotation Release 100) were used to quantify transcript expression.

Differential expression
Differential expression analyses were performed using R v.3.5.2 (https://www.r-project.org/). Kallisto HDF5 binary files were used obtain differential gene expression estimates with the Bioconductor package DESeq2 v.3.4 [49]. Briefly, the 'median of ratios' method was used to calculate size factors for each sample, and normalization of count data was carried out to account for differences in library depth. After dispersion estimates were fitted to the mean intensity reduced towards the expected dispersion values, the expression of gene was fitted to a negative binomial model and significance assessed using the Wald test. Genes showing Benjamini-Hochberg false discovery rate (FDR) p-values < 0.05 and $\log_2$ fold change values (logFC) > 0.5 were considered differentially expressed. Prior to differential expression the whole dataset was evaluated using hierarchical clustering and

principal component analyses, and outlier samples were removed for downstream analyses. PCA plots were created using the R package factoextra (http://www.sthda.com/english/rpkgs/factoextra/).

Gene Ontology (GO) enrichment analyses were performed in R v.3.5.2 (https://www.r-project.org/) using Bioconductor packages GOstats v.2.48.0 [50] and GSEA-Base v.1.44.0 [51]. GO term annotation for the Atlantic salmon transcriptome was obtained using the R package Ssa.RefSeq.db v1.3 (https://gitlab.com/cigene/R/Ssa.RefSeq.db). The over-representation of GO terms in differentially expressed gene lists compared to the corresponding transcriptomes (gill or head kidney) was assessed with a hypergeometric test. A GO term was considered enriched if it showed ≥5 DE genes assigned and a p-value < 0.05. Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were carried out using KOBAS v3.0.3 [52]. Briefly, salmon genes were annotated against KEGG protein database [53] to determine KEGG Orthology (KO). KEGG enrichment for differentially expressed gene lists was tested by comparison to the whole set of expressed genes in the corresponding tissue using Fisher's Exact Test. KEGG pathways with ≥5 DE genes assigned and showing a Benjamini-Hochberg FDR corrected p-value < 0.05 were considered enriched for differential expression. The reference tissue transcriptome for both GO and KEGG enrichment comprised only those genes with mean normalized counts value > 5.

Allele specific expression
Gene expression estimates and genotypes obtained from the RNA sequencing were used to investigate allele specific expression. The samtools v1.6 software [54] was used to identify SNPs, and call genotypes for those SNPs in individual samples. PCR duplicates, reads with mapping quality < 20 and bases with phred quality scores < 20 were excluded. SNPs within 5 bp of an indel, with quality < 20, MAF < 0.05 or less than 4 reads supporting the alternative allele were discarded. The putative effect of the SNPs was assessed using the official salmon genome annotation (NCBI *Salmo salar* Annotation Release 100) and the SnpEff v.4.2 software [55]. Allelic specific expression was assessed using the R package AllelicImbalance [56]. For every SNP in the regions of interest, read counts were obtained for each allele in heterozygous animals, those with less than 10 reads were filtered, and a binomial test was performed to assess the significance of the allelic differences. Only those genomic positions called as heterozygotes in a minimum of 4 and a maximum of 36 (75%) samples (75%) were considered. An allele specific expression event was considered significant if the mean p-value of all heterozygotes was < 0.05. All significant events were manually inspected.

Robledo *et al. BMC Genomics*   (2020) 21:271

Page 10 of 11

Supplementary information

Additional file 1. Principal component analysis of all RNA sequenced samples. RNA-Seq samples clustered according to their gene expression. Outliers were discarded for further analyses.

Additional file 2. Differentially expressed genes between resistant and susceptible samples. Lists of differentially expressed genes between resistant and susceptible samples in gill and head kidney. Gene ID, position in the Atlantic salmon genome (Chromosome, start and end in base pairs), average expression of the gene, log 2 fold change between resistant and susceptible animals (positive fold changes correspond to higher expression in resistant samples), standard deviation of the fold change, FDR adjusted *p*-value, gene annotation and gene symbol are shown.

Additional file 3. Phenotypes of all samples used for RNA sequencing. All collected phenotypes for the samples used in this study. ID of the sample, tissue, whether it is resistant or susceptible, finclip ID linking to the genotypes (available in Robledo et al. 2018), gill scores for both gills (and mean), weight and length at the end of the challenge, and amoebic load measured by qPCR are shown.

Abbreviations

AGD: Amoebic gill disease; ASE: Allele specific expression; CD28: T-cell specific surface glycoprotein CD28; CDC4EP2: Cdc42 effector protein 2; CREG1: Cellular repressor of E1A-stimulated genes 1; DE: Differential expressed; eQTL: Expression QTL; FDR: False discovery rate; GO: Gene ontology; GVINP1: Interferon-induced very large GTPase 1; IL17RE: Interleukin-17 receptor E; KEGG: Kyoto Encyclopedia of Genes and Genomes; KO: KEGG orthology; logFC: Log$_2$ fold change value; RNA-Seq: RNA sequencing; SERPING1: Plasma protease C1 inhibitor gene; SRA: Sequence Read Archive; TNA- α: Tumour necrosis factor-alpha; TNFAIP8L1: Tumor necrosis factor alpha-induced protein 8-like protein 1

Acknowledgements

Not applicable.

Authors' contributions

RH and DR were responsible for the concept and design of this work and drafted the manuscript. AH and JB were responsible for the disease challenge and collection of samples. AG performed the molecular biology experiments. DR performed bioinformatic and statistical analyses. All authors read and approved the final manuscript.

Funding

Availability of data and materials

The RNA sequencing dataset generated during the current study is available in NCBI's Sequence Read Archive (SRA) under BioProject accession number PRJNA552604. The phenotypes of all sequenced samples are included in this published article (Additional file 3).

Ethics approval and consent to participate

The challenge experiment was performed by the Marine Environmental Research Laboratory (Machrihanish, UK) under approval (LAN13-AH01) of the Institute of Aquaculture Divisional Ethics Committee and Animal Welfare Ethical Review Body of the University of Stirling (Stirling, UK) and according to Home Office license requirements. All animals were reared in accordance with relevant national and EU legislation concerning health and welfare. Landcatch are accredited participants in the RSPCA Freedom Foods standard, the Scottish Salmon Producers Organization Code of Good Practice, and the EU Code-EFABAR Code of Good Practice for Farm Animal Breeding and Reproduction Organizations.

Consent for publication

Not applicable.

Competing interests

A commerical organisation (Landcatch Natural Selection Ltd) was involved in the development of this study. AH works for Landcatch Natural Selection Ltd. The remaining authors declare that they have no competing interests.

Author details

[1]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Midlothian EH25 9RG, UK. [2]Landcatch Natural Selection Ltd., Roslin Innovation Centre, University of Edinburgh, Midlothian EH25 9RG, UK. [3]Hendrix Genetics Aquaculture BV/ Netherlands, Villa 'de Körver', Spoorstraat 69, 5831 CK Boxmeer, Netherlands. [4]Institute of Aquaculture, Faculty of Natural Sciences, University of Stirling, Stirling FK9 4LA, UK.

References

1. Rombough P. The functional ontogeny of the teleost gill: which comes first, gas or ion exchange? Comp Biochem Physiol A Mol Integr Physiol. 2007; 148:732–42.
2. Mitchell SO, Rodger HD. A review of infectious gill disease in marine salmonid fish. J Fish Dis. 2011;34:411–32.
3. Oldham T, Rodger H, Nowak BF. Incidence and distribution of amoebic gill disease (AGD) – an epidemiological review. Aquaculture. 2016;457:35–42.
4. Shinn AP, Pratoomyot J, Bron JE, Paladini G, Brooker EE, Brooker AJ. Economic costs of protistan and metazoan parasites to global mariculture. Parasitology. 2015;142:196–270.
5. Nowak BF, Archibald JM. Opportunistic but lethal: the mystery of paramoebae. Trends Parasitol. 2018;34:404–19.
6. Taylor RS, Wynne JW, Kube PD, Elliott NG. Genetic variation of resistance to amoebic gill disease in Atlantic salmon (*Salmo salar*) assessed in a challenge system. Aquaculture. 2007;272S1:S94–9.
7. Taylor RS, Kube PD, Muller WJ, Elliott NG. Genetic variation of gross gill pathology and survival of Atlantic salmon (*Salmo salar* L.) during natural amoebic gill disease challenge. Aquaculture. 2009;294(3–4):172–9.
8. Robledo D, Matika O, Hamilton A, Houston RD. Genome-wide association and genomic selection for resistance to amoebic gill disease in Atlantic salmon. G3. 2018;8:1195–203.
9. Boison SA, Gjerde B, Hillestad B, Makvandi-Nejad S, Moghadam HK. Genomic and transcriptomic analysis of amoebic gill disease resistance in Atlantic salmon (*Salmo salar* L.). Front Genet. 2019;10:68.
10. Yoshida GM, Bangera R, Carvalheiro R, Correa K, Figueroa R, Lhorente JP, et al. Genomic prediction accuracy for resistance against *Piscirickettsia salmonis* in farmed rainbow trout. G3. 2018;8:719–26.
11. Palaiokostas C, Kocour M, Prchal M, Houston RD. Accuracy of genomic evaluations of juvenile growth rate in common carp (*Cyprinus carpio*) using genotyping by sequencing. Front Genet. 2018;9:82.
12. Palaiokostas C, Cariou S, Bestin A, Bruant JS, Haffray P, Morin T, et al. Genome-wide association and genomic prediction of resistance to viral nervous necrosis in European sea bass (*Dicentrarchus labrax*) using RAD sequencing. Genet Sel Evol. 2018;50:30.
13. MacLeod IM, Bowman PJ, Vander Jagt CJ, Haile-Mariam M, Kemper KE, Chamberlain AJ, et al. Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. BMC Genomics. 2016;17:144.
14. Houston RD. Future directions in breeding for disease resistance in aquaculture species. R Bras Zootec. 2017;46:545–51.
15. Gratacap RL, Wargelius A, Edvardsen RB, Houston RD. Potential of genome editing to improve aquaculture breeding and production. Trends Genet. 2019;35:672–84.

16. Benedicenti O, Collins C, Wang T, McCarthy U, Secombes CJ. Which Th pathway is involved during late stage amoebic gill disease? Fish Shellfish Immunol. 2015;46:417–25.

17. Marcos-López M, Espinosa Ruiz C, Rodger HD, O'Connor I, MacCarthy E, Esteban MA. Local and systemic humoral immune response in farmed Atlantic salmon (*Salmo salar* L.) under a natural amoebic gill disease outbreak. Fish Shellfish Immunol. 2017;66:207–16.

18. Marcos-López M, Calduch-Giner JA, Mirimin L, MacCarthy E, Rodger HD, O'Connor I, et al. Gene expression analysis of Atlantic salmon gills reveals mucin 5 and interleukin 4/13 as key molecules during amoebic gill disease. Sci Rep. 2018;8:13689.

19. Wynne JW, O'Sullivan MG, Stone G, Cook MT, Nowak BF, Lovell DR, et al. Resistance to amoebic gill disease (AGD) is characterized by the transcriptional dysregulation of immune and cell cycle pathways. Dev Comp Immunol. 2008;32:1539–60.

20. Lien S, Koop BF, Sandve SR, Miller JR, Kent MP, Nome T, et al. The Atlantic salmon genome provides insights into rediploidization. Nature. 2016;533: 200–5.

21. Song X, Zhu S, Shi P, Liu Y, Shi Y, Levin SD, et al. IL-17RE is the functional receptor for IL-17C and mediates mucosal immunity to infection with intestinal pathogens. Nat Immunol. 2011;12:1151–8.

22. Chang SH, Reynolds JM, Pappu BP, Chen G, Martinez GJ, Dong C. Interleukin-17C promotes Th17 cell responses and autoimmune disease via interleukin-17 receptor E. Immunity. 2011;35:611–21.

23. Blotta MH, Marshall JD, DeKruyff RH, Umetsu DT. Cross-linking of the CD40 ligand on human CD4+ T lymphocytes generates a costimulatory signal that up-regulates IL-4 synthesis. J Immunol. 1996;156:3133–40.

24. Allen JE, Sutherland TE. Host protective roles of type 2 immunity: parasite killing and tissue repair, flip sides of the same coin. Semin Immunol. 2014; 26:329–40.

25. Young ND, Cooper GA, Nowak BF, Koop BF, Morrison RN. Coordinated down-regulation of the antigen processing machinery in the gills of amoebic gill disease-affected Atlantic salmon (*Salmo salar* L.). Mol Immunol. 2008;45:2581–97.

26. Nowak BF. Neoparamoeba perurans. In: Woo PTK, Buchmann K, editors. Fish parasites, pathobiology and protection, vol. 2012. London: CAB international; 2012. p. 1–18.

27. Hvas M, Karlsbakk E, Mæhle S, Wright DW, Oppedal F. The gill parasite Paramoeba perurans compromisos aerobic scope, swimming capacity and ion balance in Atlantic salmon. Conserv Physiol. 2017;5:cox066.

28. Ullal AJ, Litaker RW, Noga EJ. Antimicrobial peptides derived from hemoglobin are expressed in epithelium of channel catfish (*Ictalurus punctatus*, Rafinesque). Dev Comp Immunol. 2008;32:1301–12.

29. Ullal AJ, Noga EJ. Antiparasitic activity of the antimicrobial peptide HbbetaP-1, a member of the beta-haemoglobin peptide family. J Fish Dis. 2010;33: 657–64.

30. Davis AE, Mejia P, Lu F. Biological activities of C1 inhibitor. Mol Immunol. 2008;45:2057–63.

31. Valdenegro-Vega VA, Crosbie P, Bridle A, Leef M, Wilson R, Nowak BF. Differentially expressed proteins in gill and skin mucus of Atlantic salmon (*Salmo salar*) affected by amoebic gill disease. Fish Shellfish Immunol. 2014; 40:69–77.

32. Higgs HN, Pollard TD. Regulation of actin filament network formation through ARP2/3 complex: activation by a diverse array of proteins. Annu Rev Biochem. 2001;70:649–76.

33. You Z, Ouyang H, Lopatin D, Polver PJ, Wang CY. Nuclear factor-kappa B-inducible death effector domain-containing protein suppresses tumor necrosis factor-mediated apoptosis by inhibiting caspase-8 activity. J Biol Chem. 2001;276:26398–404.

34. Morrison RN, Zou J, Secombes CJ, Scapigliati G, Adams MB, Nowak BF. Molecular cloning and expression analysis of tumour necrosis factor-alpha in amoebic gill disease (AGD)-affected Atlantic salmon (*Salmo salar* L.). Fish Shellfish Immunol. 2007;23:1015–31.

35. Klamp T, Boehm U, Schenk D, Pfeffer K, Howard J. A giant GTPase, very large inducible GTPase-1, is inducible by IFNs. J Immunol. 2003;171:1255–65.

36. Aouadi M, Tesz GJ, Nicoloro SM, Wang M, Chouinard M, Soto E, Ostroff GR, Czech MP. Orally delivered siRNA targeting macrophage Map4k4 suppresses systemic inflammation. Nature. 2009;458:1180–4.

37. Rosenthal CK. Trim39 ligase keeps apoptosis going. Nat Cell Biol. 2012;14: 566.

38. Di Bacco A, Gill G. The secreted glycoprotein CREG inhibits cell growth dependent on the mannose-6-phosphate/insulin-like growth factor II receptor. Oncogene. 2003;22:5436–45.

39. Veal E, Eisenstein M, Tseng ZH, Gill G. A cellular repressor of E1A-stimulated genes that inhibits activation by E2F. Mol Cell Biol. 1998;18:5023–41.

40. Keane TM, Goodstadt L, Danecek P, White MA, Wong K, et al. Mouse genomic variation and its effect on phenotypes and gene regulation. Nature. 2011;477:289–94.

41. Albert FW, Kruglyak L. The role of regulatory variation in complex traits. Nat Rev Genet. 2015;16:197–212.

42. Robledo D, Gutiérrez AP, Barría A, Lhorente JP, Houston RD, Yáñez JM. Discovery and functional annotation of quantitative trait loci affecting resistance to sea lice in Atlantic salmon. Front Genet. 2019;10:56.

43. Taylor RS, Huynh C, Cameron D, Evans B, Cook MT, et al. Gill score guide – amoebic gill disease (AGD) management training document, edited by Evans B. Hobart: Tassal Pty. Ltd.; 2016.

44. Zhao S, Li CI, Sheng Q, Shyr Y. RnaSeqSampleSize: real data based sample size estimation for RNA sequencing. BMC Bioinformatics. 2018;19:191.

45. Robledo D, Gutiérrez AP, Barría A, Yáñez JM, Houston RD. Gene expression response to sea lice in Atlantic salmon skin: RNA sequencing comparison between resistant and susceptible animals. Front Genet. 2018;9:287.

46. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2004;30:2114–20.

47. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21.

48. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016;34:525–7.

49. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.

50. Falcon S, Gentleman R. Using GOstats to test gene lists for GO term association. Bioinformatics. 2007;23:257–8.

51. Morgan M, Falcon S, Gentleman R. *GSEABase:* Gene set enrichment data structures and methods. R package version 1.44.0; 2018.

52. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. Nucleic Acids Res. 2011;39:W316–22.

53. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28:27–30.

54. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map (SAM) format and SAMtools. Bioinformatics. 2009; 25:2078–9.

55. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly (Austin). 2012;6:80–92.

56. Gadin JR, van't Hooft FM, Eriksson P, Folkersen L. AllelicImbalance: an R/ bioconductor package for detecting, managing, and visualizing allele expression imbalance data from RNA sequencing. BMC Bioinformatics. 2015; 16:194.

Publisher's Note