

## ARTICLE OPEN



# A major quantitative trait locus affecting resistance to Tilapia lake virus in farmed Nile tilapia (*Oreochromis niloticus*)

Agustin Barría<sup>1</sup>, Trọng Quốc Trịnh<sup>2</sup>, Mahirah Mahmuddin<sup>2</sup>, Carolina Peñaloza<sup>1</sup>, Athina Papadopoulou<sup>1</sup>, Ophelie Gervais<sup>1</sup>, V. Mohan Chadag<sup>2</sup>, John A. H. Benzie<sup>2</sup> and Ross D. Houston<sup>1</sup>✉

© The Author(s) 2021, corrected publication 2021

Enhancing host resistance to infectious disease has received increasing attention in recent years as a major goal of farm animal breeding programs. Combining field data with genomic tools can provide opportunities to understand the genetic architecture of disease resistance, leading to new opportunities for disease control. In the current study, a genome-wide association study was performed to assess resistance to the Tilapia lake virus (TiLV), one of the biggest threats affecting Nile tilapia (*Oreochromis niloticus*); a key aquaculture species globally. A pond outbreak of TiLV in a pedigreed population of the GIFT strain was observed, with 950 fish classified as either survivor or mortality, and genotyped using a 65 K SNP array. A significant QTL of large effect was identified on chromosome *Oni22*. The average mortality rate of tilapia homozygous for the resistance allele at the most significant SNP ( $P$  value =  $4.51E-10$ ) was 11%, compared to 43% for tilapia homozygous for the susceptibility allele. Several candidate genes related to host response to viral infection were identified within this QTL, including *Igals17*, *vps52*, and *trim29*. These results provide a rare example of a major QTL affecting a trait of major importance to a farmed animal. Genetic markers from the QTL region have potential in marker-assisted selection to improve host resistance, providing a genetic solution to an infectious disease where few other control or mitigation options currently exist.

*Heredity* (2021) 127:334–343; <https://doi.org/10.1038/s41437-021-00447-4>

## INTRODUCTION

Aquaculture is the fastest-growing food production sector worldwide, with an average growth in production of 5.3% per annum during the period 2001–2018 (FAO 2020). Nile tilapia (*Oreochromis niloticus*) has the third-highest production of all finfish species, with more than 4.5 million tons farmed in 2018 (FAO 2020). Tilapia aquaculture represents a critical source of protein and nutrients for human consumption in many low and middle-income countries across Africa, Asia, and America. However, the recent emergence of Tilapia lake virus (TiLV) has presented major threats to the sustainability of tilapia production, affecting both Nile tilapia and interspecific hybrids (Eyngor et al. 2014; Ferguson et al. 2014; Dong et al. 2017), with mass mortalities up to 90% possible at several stages of the production cycle (Dong et al. 2017; Fathi et al. 2017). This disease is produced by an RNA virus which has recently been classified as a member of the *Tilapinevirus* genus, *Amnoonviridae* family (ICTV 2018). There is evidence that TiLV can be transmitted both vertically and horizontally (Eyngor et al. 2014; Jaemwimol et al. 2018; Dong et al. 2020), highlighting the importance of TiLV-free broodstock. Furthermore, there are currently only a few treatments and options for prevention, and with outbreaks possible in very young juvenile fish, the potential of vaccination is limited (Thammatorn et al. 2019).

Improvement of host resistance to infectious disease in farmed fish species via selective breeding has major potential to prevent outbreaks and improve the sustainability of production, with some

notable success stories to date (Yañez et al. 2014; Gjedrem 2015; Houston 2017). Encouragingly, a recent study by our team detected significant and high heritability for host resistance to TiLV in tilapia from a commercial breeding program, as measured by survival during a pond outbreak (Barría et al. 2020). However, while pedigree-based breeding programs are effective at improving target traits, the use of genomic tools can expedite sustainable genetic gain in aquaculture (Houston et al. 2020). The availability of genomic tools for a variety of aquaculture species (e.g. Houston et al. 2014; Palti Gao et al. 2015; Lien et al. 2016; Yañez et al. 2016; Gutierrez et al. 2017; Zeng et al. 2017; Nugent et al. 2019; Yu et al. 2020; Zhou et al. 2020) has also facilitated the identification of quantitative trait loci (QTL) associated with traits such as sexual determination, growth rate, flesh color and disease resistance which are common targets for commercial breeding programs (e.g. Gonen et al. 2015; Palti Vallejo et al. 2015; Tsai et al., 2015; Gutierrez et al. 2018; Mohamed et al. 2019; Fraslin et al. 2020).

The most remarkable example is the identification of a major QTL affecting host resistance to infectious pancreatic necrosis virus (IPNV) in two different Atlantic salmon (*Salmo salar*) populations, explaining between 80 and 100% of the genetic variation for the trait (Houston et al. 2008 2010; Moen et al. 2009). Application of favorable alleles through marker-assisted selection (MAS) has resulted in the number of IPN outbreaks in Norway reducing to near zero (Hjeltnes 2014; Norris 2017). However, QTL of such large effect are rare, and for traits underpinned by a

<sup>1</sup>The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh Easter Bush, Midlothian, United Kingdom. <sup>2</sup>WorldFish, Bayan Lepas, Malaysia. Associate editor: Rowan Barrett ✉email: ross.houston@roslin.ed.ac.uk

Received: 9 March 2021 Revised: 31 May 2021 Accepted: 31 May 2021

Published online: 14 July 2021

polygenic genetic architecture, genomic selection (GS) is a more effective approach (Zenger et al. 2019; Houston et al. 2020).

The development and application of genomic tools have been relatively recent for Nile tilapia. However, there is a high-quality reference genome (Conte et al. 2019) and at least three different medium to high-density SNP arrays (Joshi et al. 2018; Peñaloza et al. 2020; Yáñez et al. 2020). These resources have already been used to highlight the potential improvement in accuracy of prediction of estimated breeding values (EBVs) for growth and disease-resistance related traits compared with a pedigree-based approach (Yoshida Lhorente et al. 2019; Joshi et al. 2020; Joshi Skaaurd 2020), and have also allowed the assessment of linkage disequilibrium (LD) and genetic structure of different GIFT-derived Nile tilapia populations (Yoshida et al. 2019).

Due to the global importance of TiLV, and the evidence for significant and high levels of heritability for host resistance (Barría et al. 2020), there is a major opportunity to utilize genomic tools to investigate the genetic architecture of this resistance. As such, the aims of this study were to (i) dissect the genetic architecture of resistance to TiLV using a high-density SNP array, (ii) map QTLs associated with host resistance to regions of the tilapia genome, and (iii) identify positional candidate genes within those QTL regions that may be associated with this trait. The results of this study will inform the use of genetic markers in tilapia breeding programs aiming to develop tilapia strains with enhanced resistance to TiLV. Furthermore, the identification of candidate genes will lead to research targeting improved knowledge of the functional mechanisms underlying resistance to TiLV, hence offering the opportunity for the development of new control and treatment strategies.

## MATERIAL AND METHODS

### Nile tilapia population

The population sample used in this study is from a genetically improved Nile Tilapia (GIFT) breeding program and had been selected for improved growth rate for 16 generations. The breeding nucleus was based in Jitra, Malaysia, and managed by WorldFish (Penang, Malaysia). The population sample consisted of 124 nuclear families (GIFT generation 16th) produced by crossing 124 dams and 115 sires. Pedigree data for these fish included records of approximately 86,000 fish from generation 16 back to the first generation of GIFT in Malaysia. Each fish from the current generation was tagged using a passive integrated transponder (PIT tag) at an average weight of 4.97 g, which corresponded to an average age of 110.5 days post-hatching. At typical harvest weight, fish were transferred to a single pond where a natural TiLV outbreak was observed shortly thereafter. A summary of the phenotypes measured on these fish, such as weight and age at tagging and harvest for this population can be found in Barría et al. (2020).

### Natural field outbreak

After the fish were transferred to the single pond, a natural field outbreak of TiLV was observed (in February and March 2018). Mortalities were collected and sampled daily for 14 days (22 February to 7 March 2018), and once the mortality levels had returned to baseline, all remaining fish in the pond were euthanized (using 400 mg/l clove oil) and sampled over five days (8–12 March 2018). A total of 1,821 fish were classified as survivors or mortalities, and the phenotypic sex of all fish was determined. On average, each full-sibling family comprised 14 fish (ranging from 2 to 21). Clinical signs of TiLV were observed throughout the outbreak. The presence of TiLV was assessed on the spleen of a subset of 35 fish, through one-step RT-qPCR. Briefly, the EvaGreen assay RT-qPCR was performed using the following primer set; Forward primer: 5'-CTGAGCTAAAGAGGCAAT ATGGATT-3' and Reverse primer: 5'-CGTGCCTACTCGTTTCAGTATAAG TTCT-3'. Finally, the PCR product was detected by measuring the fluorescence generated by the EvaGreen dye bound to dsDNA. More

details about this RT-qPCR assay, including reproducibility and sensitivity, can be found on Tattiyapong et al. (2018). A sample of mortalities was also randomly selected to perform necropsy analyses to further confirm TiLV as the cause of the mortalities. In addition, a caudal fin sample was taken from all survivors and mortalities from the outbreak, placed in 95% ethanol, and stored at  $-20^{\circ}\text{C}$  until further DNA extraction and genetic analysis.

### Genotyping

Total DNA from fin clips of 2,016 fish, including 195 parents and all the 1,821 offspring collected from the outbreak, was extracted using a salt-extraction protocol (Aljanabi and Martinez 1997), with the modifications described by Taslima et al. (2016). The DNA quality and/or quantity from 691 offspring was not sufficient to surpass the initial quality control (QC) criteria required by the genotyping company (Iidentigen, Dublin, Ireland). Therefore, only 1,130 offspring and the 195 parents were genotyped. The extracted DNA samples were genotyped using an Axiom<sup>®</sup> SNP array developed by our team which contains ~65 K SNP markers dispersed throughout the Nile tilapia genome (Peñaloza et al. 2020). The raw array data from the genotyping (CEL intensity files) were imported to the Axiom analysis Suite v4.0.3.3 software for genotype calling and QC. A total of 47 samples with a dish quality control (DQC) and call rate (CR)  $<0.82$  and  $<0.93$ , respectively, were excluded for subsequent analyses. Thus, 187 parents (96%) and 1,091 offspring (97%) passed the thresholds of the Axiom software. Approximately 54 K (78%) of the SNPs on the array were identified as high quality and polymorphic (i.e., "PolyHighResolution") and were retained for further analyses (Peñaloza et al. 2020). Subsequently, a second QC step was applied using Plink v1.09 (Purcell et al. 2007). With an average CR of 99%, all fish surpassed the genotype call rate ( $>0.95$ ). The SNPs with a minor allele frequency (MAF)  $<0.05$ , CR  $<0.95$ , and significant deviation from Hardy-Weinberg equilibrium (HWE) ( $p < 1 \times 10^{-6}$ ) were excluded from further analyses. Thus, 94% of the SNPs (50,710 out 53,811) passed all the QC filters, with most of them being removed due to low MAF (~2 K SNPs). Furthermore, using trio information, the resulting data set was tested for putative Mendelian errors in any fish and SNPs. A total of 166 fish (25 parents and 141 offspring) and 3 K SNPs were excluded for subsequent analyses due to a Mendelian error rate  $>5\%$ , which could reflect either errors in pedigree recording or samples with poor quality genotype data. Following all QC steps, the final filtered dataset comprised 1,112 fish and 47,915 SNPs. This comprised genotype and phenotype data for 950 offspring, and genotype data for 162 parents. From this filtered dataset, the genotyped offspring represent fish from 117 families (108 full-sib families), with an average of 8 fish per family, ranging from 1 to 16.

### Estimation of genetic parameters

Host resistance to TiLV was defined as binary survival (BS) (i.e., dead/alive at the end of the natural field outbreak) and as a time to death (TD). In the case of BS, the survivors and mortalities were designated as 1 or 0, respectively. The TD trait was considered as a continuous trait, with values ranging from 1 up to 14, representing the day of the first observed mortality and the last day before mortality returned to baseline levels, respectively. The heritability for BS and TD in the genotyped population was estimated using a genomic-relationship matrix (GRM) calculated with the genome-wide complex trait analysis (GCTA) software v.1.92.2 (Yang et al. 2011). All SNPs surpassing the QC were used to create the GRM. This matrix was then used to estimate the narrow-sense heritability by using the following linear model:

$$y = \mu + Xb + Zu + e \quad (1)$$

Where  $y$  is the vector of phenotypes (BS or TD records),  $\mu$  is the population mean,  $b$  is the vector of fixed effects (sex as fixed

effect, and weight and age at harvest as covariates),  $u$  is the vector of the additive genetic effects, and  $X$  and  $Z$  are incidences matrices. The following distributions were assumed;  $u \sim N(0, G\sigma_u^2)$  and  $e \sim N(0, I\sigma_e^2)$ . Where  $\sigma_u^2$  and  $\sigma_e^2$  are the additive genetic and residual variance, respectively,  $G$  is the GRM and  $I$  is the identity matrix. Heritability was estimated as the ratio of the additive genetic variance to the phenotypic variance. A genetic correlation was estimated as the ratio of the genetic covariance between BS and TD to the square root of the product of the genetic variance of BS and TD.

### Genome-wide association study (GWAS)

To identify SNPs associated with TiLV resistance (defined as both BS and TD), a mixed linear model using the *leaving-one-chromosome-out* (LOCO) approach was applied using the GCTA v.1.92.2 software. This approach estimates the GRM between individuals by removing the SNPs located in the tested chromosome and including SNPs from all the other chromosomes. The SNP effect from the chromosome being tested is not included twice in the model, and therefore a higher statistical power is achieved (Van den Berg et al. 2019). Subsequently, fitting the GRM allows correction for population structure, which can cause spurious associations in GWAS. The model used for the GWAS was identical to the model described previously for the estimation of the genetic parameters. However, the specific SNP genotypes being assessed were coded as 0, 1, or 2 (reflecting the number of minor alleles at the locus), and included in the model as random effects (Yang et al. 2011). For an SNP to be considered significant at the genome-wide level, it had to surpass the genome-wide Bonferroni-corrected significance threshold for multiple testing, which was defined as  $P < 0.05/47,915$ . This multiple test correction is considered very stringent (Johnson et al. 2010), which reduces the likelihood of any false positive association. To quantify the level of inflation of the obtained  $P$ -values compared with those expected, lambda ( $\lambda$ ) was computed as the median of the quantile  $\chi^2$  distribution of the obtained  $P$ -values/0.455. For practical reasons, SNPs not placed in chromosomes in the reference genome assembly (O\_niloticus\_UMD\_NMBU, Genbank accession number GCA\_001858045.3, Conte et al. 2019), were assigned as being placed on a “pseudo-chromosome” *Oni24*. GWAS results were plotted using the package “CMplot” (Yin 2020) in R (R Core Team 2017).

### Candidate genes in QTL regions

Based on the genome-wide association results, putative candidate genes associated with host resistance to TiLV were identified within a 1 Mb window size (500 Kb upstream and downstream) flanking the significantly associated SNPs, again using the Nile tilapia reference genome assembly (Genbank accession number GCA\_001858045.3).

### Estimation of SNP variances

Following the GWAS, the top three SNPs significantly associated with BS and/or TD on each of the significant chromosomes were tested for the estimation of the additive and dominance effect, by using ASReml v.4.1.0 (Gilmour et al. 2015). Additive ( $a$ ) and dominance ( $d$ ) effect were estimated as follows:  $a = (AA - BB)/2$  and  $d = AB - [AA + BB]/2$  where AA, AB, and BB are the predicted trait value for each genotype. The proportion of genetic variance explained for each of the selected SNPs was estimated as  $[2pq(a + d(q - p))]^2/VA$ , where  $p$  and  $q$  are the frequencies of the SNP alleles, and VA is the total additive genetic variance explained by the model when no individual SNP is fitted.

## RESULTS

### Field outbreak

Throughout the TiLV outbreak, clinical signs related to this viral infection were observed, including damage at the base of fins, hemorrhage, and skin erosion. These were confirmed by a

qualified veterinarian. Subsequently, through an RT-qPCR assay, TiLV was identified in 74% of the randomly sampled fish (26 out of 35). From the 35 sampled fish, 15 were survivors and 20 were mortalities, with 40% of the survivors testing positive and 100% of the mortalities testing positive. The survivors which tested positive had a substantially lower viral load, with 127 viral copies/ $\mu$ l  $\pm$  150, compared to 188,491  $\pm$  233,167 copies/ $\mu$ l for mortalities.

After 14 days of the first observed mortality, the daily mortality level had returned close to the baseline, and all remaining fish in the pond were euthanized and sampled. The total cumulative mortality in the outbreak in the entire sampled population was 39.6%. Further details about the pattern of mortality during the TiLV outbreak can be found in Barriá et al. (2020).

### Genome-wide association study

The estimates of heritability in this dataset based on the GRM, were consistent with the estimates using the pedigree on a larger dataset described in Barriá et al. (2020). Moderate to high heritability values of  $0.38 \pm 0.05$  and  $0.69 \pm 0.09$  were estimated for BS on the observed and underlying scale, respectively, whereas a lower value was estimated for TD ( $0.22 \pm 0.05$ ). The robustness of the estimation of the variance parameters for BS on the underlying scale was confirmed by implementing a threshold model in the Thrgibbs1f90b module of the Blupf90 (Misztal et al. 2015) software (data not shown).

A high genetic correlation was found between both TiLV resistance definitions ( $0.97 \pm 0.02$ , Table 1). A total of 29 SNPs that exceeded the genome-wide significance Bonferroni threshold for BS were identified, and all but 1 were located on chromosome *Oni22* (Table 2), with a clear peak suggestive of a major QTL in the proximal end of this chromosome (Fig. 1). In the case of TD, two SNPs located in the same QTL region of *Oni22* surpassed this genome-wide significance threshold (Fig. S1). The estimated inflation factor ( $\lambda$ ) for BS and TD was 1.19 and 1.11, respectively, suggesting a relatively good concordance between the observed  $P$ -values and the theoretical statistical distribution. The QQ plot associated with the Manhattan plot for both traits is shown in Fig. S2. Interestingly, for both definitions of resistance, the most significant association was found for the same SNP (AX-317616757, located at the position 255,104 bp on *Oni22*) with a  $P$ -value of  $4.5 \times 10^{-10}$  and  $4.8 \times 10^{-07}$ , for BS and TD, respectively (Table 2). All the SNPs located in *Oni22* which were significantly associated with BS are within a genomic region of approximately 9.4 Mb in size. However, this QTL size is reduced to  $\sim 1.7$  Mb when only the top three SNPs are taken into account (AX-317645761 and AX-317617572 located at 239,073 and 1,939,192 bp respectively). The minor allele was associated with resistance to TiLV for 25 out of the 29 significant SNPs. The frequency of these resistance-associated alleles ranges from 0.11 to 0.48 (Table 2), highlighting the potential for increasing its frequency by MAS and hence TiLV resistance at a

**Table 1.** Genetic parameters for host resistance to TiLV in a Nile tilapia (*Oreochromis niloticus*) breeding population.

Parameters <sup>b</sup>	BS <sup>a</sup>	TD <sup>a</sup>
$\sigma_a^2$	0.09(0.02)	5.96(1.47)
$\sigma_e^2$	0.14(0.01)	21.6(1.33)
$\sigma_p^2$	0.22(0.01)	27.57(1.39)
$h^2$	0.38(0.05)	0.22(0.05)
$r_g$	0.97(0.02)	

Standard errors are shown inside brackets.

<sup>a</sup>Host resistance definition: BS binary survival on the observed scale, TD time to death.

<sup>b</sup>Genetic parameters and standard error:  $\sigma_a^2$  = additive genetic variance;  $\sigma_e^2$  = error variance;  $\sigma_p^2$  = phenotypic variance;  $h^2$  = narrow-sense estimated heritability;  $r_g$  = genetic correlation.

**Table 2.** Significant SNPs associated with TiLV resistance as binary survival (BS) and time to death (TD) in a Nile tilapia (*Oreochromis niloticus*) breeding population.

SNP	Oni <sup>a</sup>	BP <sup>b</sup>	P-value <sup>c</sup>	Minor/major allele <sup>d</sup>	Resistance allele freq <sup>f</sup>
<b>TD</b>					
AX-317616757	22	255104	4.80E-07	<b>G/T</b>	0.39
AX-317647630	22	5379664	8.20E-07	<b>G/A</b>	0.11
<b>BS</b>					
AX-317616757	22	255104	4.51E-10	<b>G/T</b>	0.39
AX-317617572	22	1939192	4.32E-09	<b>T/C</b>	0.39
AX-317645761	22	239073	6.17E-09	<b>A/C</b>	0.36
AX-317619074	22	5785936	6.86E-09	<b>G/A</b>	0.37
AX-317617531	22	1888302	1.13E-08	<b>G/A</b>	0.47
AX-317616778	22	276593	1.24E-08	<b>A/G</b>	0.56
AX-317648645	22	6115097	3.30E-08	<b>A/G</b>	0.46
AX-317617852	22	2432164	4.23E-08	<b>T/G</b>	0.40
AX-317616808	22	760642	7.96E-08	<b>T/C</b>	0.48
AX-317647368	22	3478052	8.47E-08	<b>A/G</b>	0.48
AX-317647975	22	5629013	1.60E-08	<b>G/T</b>	0.16
AX-317621349	22	9718148	1.81E-08	<b>G/A</b>	0.72
AX-317617390	22	1758060	2.02E-08	<b>C/T</b>	0.42
AX-317647016	22	2521284	2.13E-08	<b>A/G</b>	0.14
AX-317618485	22	5289967	2.22E-08	<b>T/C</b>	0.39
AX-317649969	22	9299308	2.38E-08	<b>A/G</b>	0.58
AX-317647630	22	5379664	2.70E-08	<b>G/A</b>	0.11
AX-317620863	22	9336674	2.81E-08	<b>C/T</b>	0.45
AX-317647753	22	5479097	3.56E-08	<b>C/A</b>	0.11
AX-317617254	22	1616487	3.92E-08	<b>C/T</b>	0.11
AX-317718855	03	71847333	4.37E-07	<b>C/T<sup>e</sup></b>	0.24
AX-317646938	22	2445275	4.45E-07	<b>A/G</b>	0.10
AX-317646411	22	1714252	4.95E-07	<b>A/G</b>	0.10
AX-317617977	22	2555673	4.97E-07	<b>C/T</b>	0.14
AX-317620317	22	8916395	6.99E-07	<b>C/T</b>	0.13
AX-317063470	22	1441242	7.27E-07	<b>A/G</b>	0.10
AX-317646673	22	1988041	7.86E-07	<b>A/G</b>	0.29
AX-317617288	22	1660924	9.54E-07	<b>T/G</b>	0.11
AX-317648270	22	5898003	9.65E-07	<b>G/A</b>	0.12

<sup>a</sup>Number of chromosome on the *Oreochromis niloticus* genome.

<sup>b</sup>Position of the SNP in the chromosome, in base pairs.

<sup>c</sup>P value of the SNP for the genome-wide association study for host resistance to TiLV.

<sup>d</sup>In bold is the allele conferring the resistant phenotype.

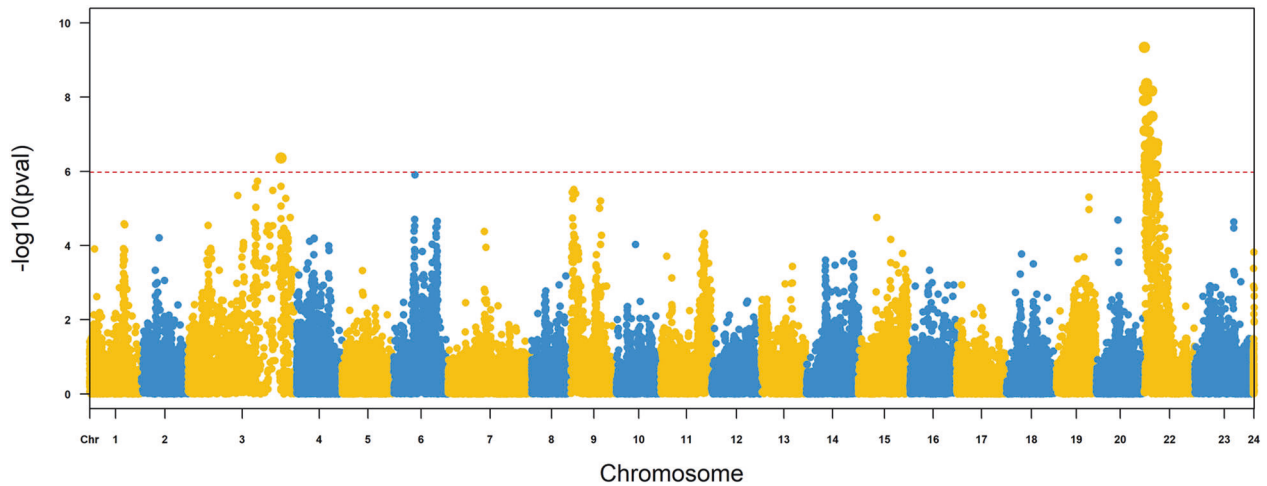
<sup>e</sup>The resistant genotype is heterozygous.

<sup>f</sup>Frequency of the allele conferring the resistance phenotype to TiLV.

population level. In the case of the three SNPs with the lowest P values, the proportion of genetic variance explained ranged from 0.06 to 0.14 (Table 3). The three most significant SNPs on *Oni22* have a substitution effect on TiLV mortality proportion ranging from 0.16 to 0.14 (Table 3 and Fig. 2). For example, at the most significant SNP, the average mortality rate of tilapia carrying two copies of the resistance allele was 11%, compared to 43% for tilapia carrying two copies of the susceptibility allele. Therefore, the predicted difference in mortality between alternate homozygous fish at this single significant QTL is 32%, which can be placed in context by considering that the overall mortality rate in the outbreak was ~40%. In the case of the significant SNP located on *Oni03* (AX-317718855), the equivalent allele substitution effect is 0.07.

### Candidate genes in the QTL regions

The SNP with the lowest P value for BS and TD is located within the second intron of the *Igals17* gene, a member of the galectin family. This family of genes has functions that include antiviral defense. A number of other interesting candidate genes were also located within this QTL region, which has previously been found to be related to host response to a viral infection. For the main QTL on *Oni22* the genes *rnf2*, *vps52*, *cdc42*, and *anp32* were identified. For the secondary QTL on *Oni3*, the *zbed1* (also known as *dref*), *trappc1*, and *psmb6* were identified. The other SNP found to be associated with TD and BS (AX-317647630) is flanked by two genes belonging to the tripartite motif family, *trim21* and *trim29*.



**Fig. 1** Manhattan plot for resistance to Tilapia lake virus (TiLV) in a Nile tilapia (*Oreochromis niloticus*) breeding population. Manhattan plot of GWAS for host resistance, as binary survival, to TiLV. On the y axis is the  $-\log_{10}(P\text{-value})$ . The horizontal dashed red line shows the genome-wide significance threshold. *Oni24* represent SNPs with unknown chromosome location.

**Table 3.** Summary statistics for the most significant genome-wide associated SNPs within each chromosome for host resistance to TiLV.

<i>Oni</i> <sup>a</sup>	SNP	BP <sup>b</sup>	$\alpha$ <sup>c</sup>	<i>P</i> value (a) <sup>d</sup>	Vg
<b>Binary survival (BS)</b>					
22	AX-317616757	0.25	0.16	2.45E−09	0.12
22	AX-317617572	1.93	0.15	1.19E−08	0.10
22	AX-317645761	0.23	0.14	1.90E−07	0.09
3	AX-317718855	71.8	0.07	8.42E−02	0.06
<b>Time to death (TD)</b>					
22	AX-317616757	0.25	−1.37	3.12E−06	0.14
22	AX-317647630	5.37	−1.97	5.45E−02	0.13

<sup>a</sup>Number of chromosome on the *Oreochromis niloticus* genome.

<sup>b</sup>Position of the SNP in the chromosome, in million base pairs.

<sup>c</sup>Additive genetic effect.

<sup>d</sup>*P* value of the Student's *t*-test distribution for the genetic effect.

<sup>e</sup>Proportion of the genetic variance explained by the SNP.

The complete list of genes flanking the SNPs with the strongest association, within each chromosome, for host resistance to TiLV and their position within the QTL regions, are shown in Table 4.

## DISCUSSION

### Data collection and genomic information

Experimental challenges are commonly used as a means to collect measurements of host disease resistance for the purposes of genetic improvement in aquaculture (Ødegård et al. 2011; Yañez et al. 2014; Houston et al. 2020). This is currently difficult for TiLV because experimental challenge models are at a formative stage and not yet routine in breeding programs. However, data and samples from field outbreaks can also be applied for the same purpose if opportunistic sampling can be performed during the course of the outbreak. The data and samples derived from such outbreaks have previously been used to estimate genetic parameters for disease resistance (Lillehammer et al. 2013; Bangerla et al. 2014). In addition, when combined with genomic data, the resulting data can highlight QTL linked to host resistance in a field outbreak (Houston et al. 2008; Boison et al. 2019; Aslam et al. 2020), which is potentially very informative since the trait is a closer representation of the ultimate target trait in the breeding program, whereas equivalent results from experimental challenges would ideally need to be validated in a field setting.

### Heritability and genetic correlation

Together with our recent study (Barriá et al. 2020), the current results provide substantial evidence supporting the feasibility to improve host resistance to TiLV by means of selective breeding, and the potential of genomic tools to assist and expedite this process. Previous heritability estimates for TiLV resistance, using pedigree relationship matrices, ranged from 0.40 to 0.63 for BS depending on the statistical model used. By using genomic relationship matrices, moderate to high heritabilities were estimated for BS within the same population (0.38 and 0.69 for the observed and underlying scale, respectively). These results highlight that rapid genetic gain should be possible by selection for host resistance, using either pedigree or genomic data to model relationships between animals.

A large number of studies have estimated significant heritabilities for host disease resistance in aquaculture species (reviewed in Yañez et al. (2014) and Ødegård et al. (2011)). In the case of Nile tilapia, and by using different statistical models, these estimates range from 0.11 to 0.58 and from 0.14 to 0.30 for the bacterial pathogens *Streptococcus* sp (LaFrentz et al. 2016; Shoemaker et al. 2017; Suebsong et al. 2019; Sukhavachana et al. 2019; Joshi Skaard et al. 2020), and *Flavobacterium columnare* (Wonmongkol et al. 2017), respectively. Thus, our results are within the ranges of previous findings relating to disease resistance in farmed Nile tilapia populations. Furthermore, the magnitude of the genetic correlation between resistance defined as BS and resistance defined as days to death (0.97) is similar to the 0.95 estimated previously by using pedigree data (Barriá et al. 2020). Similar high positive genetic correlations have been observed for host resistance traits defined as continuous or binary traits (Barriá et al. 2018; Bassini et al. 2019), which is likely to be due to the fact that survivors at the end of the experimental challenge or natural field outbreak are typically assigned the same high value as for the days to death trait.

By using pedigree data, a genetic correlation not different from zero was estimated between harvest weight and TiLV resistance (Barriá et al. 2020). The use of genomic data corroborated these results ( $0.06 \pm 0.12$  and  $0.13 \pm 0.14$  for BS and TD regarding HW, respectively). Likewise, there was no evidence for the association of the TiLV-resistance-associated SNPs with harvest weight. Therefore, increasing the frequency of the resistant alleles within the population via family selection or MAS, should not affect genetic gain for body weight at harvest.

### Genome-wide association study

The GWAS comprised 950 fish and 48 K SNPs, which is comparable with previous studies in aquaculture species aimed to identify

genomic regions associated with host resistance to infectious diseases (Gutierrez et al. 2018; Palaiokostas Cariou et al. 2018; Palaiokostas Robledo et al. 2018; Rodríguez et al. 2019; Vallejo et al. 2019; Fraslin et al. 2020). The genetic architecture of resistance to disease in aquaculture species varies substantially, from highly polygenic with no evidence for significant QTLs, to a single major QTL explaining almost all genetic variation in the trait (Fraslin et al. 2020). In the current study, the mapping of several highly significant SNPs to the proximal end of chromosome *Oni22* highlights an interesting QTL associated with resistance to TiLV and another QTL of potential interest on *Oni3*. For the main QTL, the size of the effect is highlighted by the >30% difference in mortality rate between tilapia of alternate homozygous genotypes, which points to the utility of QTL region SNPs in MAS. Therefore, the results suggest that resistance to TiLV in this population of GIFT tilapia is an oligogenic trait, with at least one likely major QTL together with minor effect QTL elsewhere in the genome. It is plausible that additional QTL would be detected with larger sample size and increased statistical power. Furthermore, relatively low levels of LD were detected in the parental population of the fish used for the GWAS (Peñalosa et al. 2020), making it difficult to exclude the possibility of additional QTL that

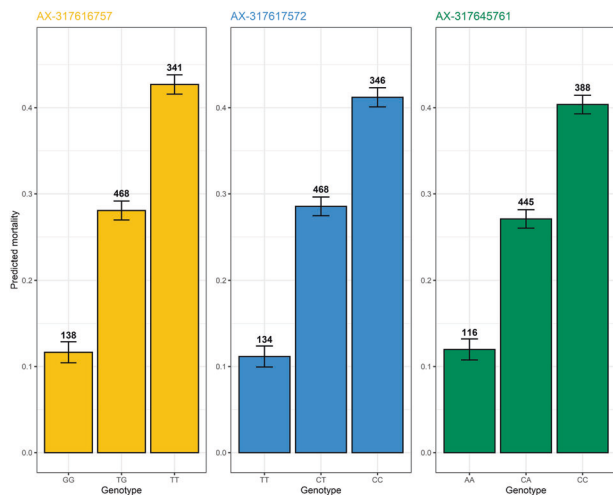
were not in LD with SNPs included on the SNP array.

The presence of a major QTL underlying a quantitative trait is scarce in animal breeding populations. To date, most of the production traits of commercial interest in aquaculture have been found to be controlled by several loci with minor effects (Fraslin et al. 2020; Houston et al. 2020). However, there are some exceptions, and interestingly, with exception of age at maturity (Sinclair-Waters et al. 2020), these examples are primarily associated with host resistance to viral diseases in Atlantic salmon. For example, Gonen et al. (2015) found a major QTL, located on Atlantic salmon chromosome 3, for host resistance to salmonid alphavirus, using data collected from an experimental challenge. Boison et al. (2019) identified a major QTL for the piscine myocarditis virus located on Atlantic salmon chromosome 27 using data from a field outbreak. However, the most remarkable example is the discovery of a major QTL for resistance to IPN virus using data collected from a natural field outbreak (Houston et al. 2008) in post-smolts and then confirmed through an experimental challenge (Moen et al. 2009) in fry. One possible explanation for the existence of several major QTL affecting disease resistance in farmed fish populations is their recent domestication, such that these disease pressures are new, and standing genetic variation associated with resistance has not yet been subject to natural or domestication selection (Houston et al. 2020).

Given the large effect of the QTL mapped in the current study, it is likely that MAS for the favorable allele at the *Oni22* QTL will be of substantial utility for breeding tilapia with improved innate resistance to TiLV. However, due to variation in allele frequencies and genetic background, it would be beneficial to confirm these findings in other Nile tilapia breeding populations prior to implementation in specific populations. Ideally, an independent population should be used to evaluate if the same QTLs proposed here, or if new associations are detected. For example, previous reports suggest an impact of genetic background on resistance to TiLV, with the Chitralada strain being more susceptible than GMT and GIFT strains of Nile tilapia (Ferguson et al. 2014; Kabuuusu et al. 2018). Another important future goal is to assess whether the QTL detected here is found under experimental conditions. If intraperitoneal injection is used as the challenge model, the innate immune response of the mucosal surfaces is largely bypassed, and therefore the host resistance trait may differ. This is important for facilitating future routine challenge testing of tilapia breeding populations for selective breeding purposes, and also for allowing further study of the underlying mechanisms of the QTL.

#### Candidate genes

Due to the relatively recent discovery of TiLV, there is a relative paucity of information about the entry, dissemination, and replication of the virus, and the host immune response. In the



**Fig. 2** Predicted mortality values for host resistance to Tilapia lake virus in a Nile tilapia breeding population. Host resistance as binary survival predictive values for each genotype of the SNPs with the stronger genome-wide association. The bars in yellow, light blue, and green show the predicted values for the top three SNPs located on *Oni22*. The bars show the standard error. Numbers above the bars indicate the number of fish with the specific genotype.

**Table 4.** Genes flanking the most important genome-wide associated SNPs within each chromosome for TiLV resistance.

<i>Oni</i> <sup>a</sup>	Trait	QTL region <sup>b</sup>		Gene names
		Left position	Right position	
03	BS	71,347,333	72,347,333	<b>zbed1</b> <sup>c</sup> , <i>kcnab1</i> , <i>trappc1</i> , <i>nlrc3</i> , <b>psmb6</b> , <i>pigr</i> , <b>nlrc3</b> , <b>mrc1</b> , <i>ephb4</i> , <b>fcgr2b</b> , <i>agp4</i> , <b>btnl2</b> , <b>btnl10</b>
22	BS and TD	1	755,104	<i>zhx1</i> , <b>lgals17</b> , <b>vps52</b> , <i>hmcn1</i> , <i>senp1</i> , <b>muc5ac</b> , <i>ha1f</i> , <b>rnf2</b> , <i>atad2</i> , <i>rps18</i> , <i>ptk7</i>
22	BS	1,439,192	2,439,192	<b>zbed1</b> , <i>zscan2</i> , <i>tmem65</i> , <i>sema6d</i> , <i>scgn</i> , <i>tatdn1</i> , <i>pomc</i> , <i>NADH</i> , <i>mtss1</i> , <i>grik5</i> , <i>fibcd1</i> , <i>carmil1</i> , <i>rnf139</i> , <i>ceacam5</i> , <b>atx2</b>
22	BS	1	739,073	<i>zhx1</i> , <b>lgals17</b> , <b>vps52</b> , <i>hmcn1</i> , <i>senp1</i> , <b>muc5ac</b> , <i>ha1f</i> , <b>rnf2</b> , <i>atad2</i> , <i>rps18</i> , <i>ptk7</i>
22	BS and TD	4,879,664	5,879,664	<i>cacgn7</i> , <i>cacgn6</i> , <i>cacgn4</i> , <i>vamp2</i> , <i>styk1</i> , <b>trim29</b> , <i>tpi1</i> , <b>glut1</b> , <i>phc1</i> , <i>iffo2</i> , <b>gnb3</b> , <i>gapdh</i> , <i>gtan</i> , <i>eno2</i> , <i>g2e3</i> , <b>trim21</b> , <i>cox6b1</i> , <i>chd4</i> , <i>clcn1</i> , <b>cdc42</b> , <i>cd209e</i> , <b>clec6a</b>

<sup>a</sup>Number of chromosome on the *Oreochromis niloticus* genome.

<sup>b</sup>QTL region size was defined as 500 kb upstream and downstream from the SNP location.

<sup>c</sup>In bold the name of the genes with a role known to be involved in a viral infection process.

current study, the SNP with the lowest *P* value for BS and TD (AX-317616757) is located at a position of 0.25 Mb on *Oni22*, specifically within the second intron of the *Igals17* gene. This gene encodes one of a family of proteins with functions as pattern-recognition receptors (PRRs) which are important in viral infections due to their role in activating the innate and adaptive immune response (Sato and Nieminen 2002; Vasta 2009). Galectins have shown increased expression and potential antiviral activity after exposure of several fish species to single-stranded RNA viruses, such as infectious salmon anemia virus (Jørgensen et al. 2008), viral hemorrhagic septicemia virus (VHSV) (*Halotis discus*) (O'Farrell et al. 2002; Sandamalika and Lee 2020) and sea bass nervous necrosis virus (SBNNV) (Poisa-Beiro et al. 2009). Considering the homology of the largest genomic segment of TiLV with the PB1 subunit of the Influenza C virus polymerase (Bacharach et al. 2016), it is promising that *Igals1* has been found to reduce influenza virus load and that knockout mice have higher susceptibility than their wild-type counterparts (Yang et al. 2011).

Another gene of interest proximal to the SNP with the lowest *P* value is the *vps52* subunit of the *GARP complex* gene. This gene is required for the replication and extracellular formation of two DNA viruses; vaccinia virus and monkeypox virus, and plays a crucial role in virus egress and cell-to-cell spread (Harrison et al. 2016; Realegeno et al. 2017). In addition, the *mf2* gene is located within the QTL region and has been linked to inhibition of *ifn1*, which plays a crucial role in the innate response by inducing multiple antiviral genes in mice (Liu et al. 2018).

The *cdc42* gene is also located in the major QTL region, and several RNA viruses are known to hijack the actin-regulating pathways to internalize into human cells (Swaine and Dittmar 2015). Furthermore, the role of *cdc42* in the host response to viral infection has been observed in some aquaculture species. For example, Xu et al. (2017) and Levican et al. (2017) showed the ability of *cdc42* to inhibit the replication of the white spot syndrome virus and IPNV, respectively. Studies in Atlantic salmon populations revealed that this gene mapped to a QTL associated with host resistance to cardiomyopathy syndrome (Boison et al. 2019), and was also to be up-regulated in salmon with high levels of genetic resistance to amoebic gill disease (Robledo et al. 2020). The genes *trim29* and *trim21* were also located within the significant QTL. The former negatively regulates the IFN production in response to an infection mediated by RNA virus (Xing et al. 2018), while the latter is mainly associated with neutralization of the virus, operating during both the early and the late stages of the infection (Mallery et al. 2010; Vaysburd et al. 2013; Foss et al. 2019). Furthermore, it has been shown that *trim29*<sup>-/-</sup> mice have higher survival and body weight than their wild-type counterparts following the influenza challenge (Xing et al. 2016).

A recent study performed by Wang et al. (2020), examined the host response to TiLV infection in an experimental challenge of Nile tilapia. They identified several genes and pathways upregulated after infection and identified enrichment of several pathways such as PPAR signaling, sucrose metabolism, phagosome, and cytokine-cytokine receptor interaction. None of the genes mapping to the identified QTLs in the current study were identified as being upregulated in response to infection in the study of Wang et al. These discrepancies could be due to a number of reasons, including the different nature of the infection. Whereas our data was obtained through a natural field outbreak, Wang et al. (2020) performed an IP experimental challenge. The latter could have bypassed the underlying factors related to the innate immune function of the mucosal layers and may have affected the host's ability to mount a complete immune response to TiLV. Secondly, it is plausible that the causative gene underlying the QTL identified herein may not itself be differentially expressed, but may act via changes to the structure and function of key proteins involved in the host response. Further characterization of the genes and polymorphisms in the QTL region using whole-

genome resequencing will help to identify potential underlying mutations for further study.

### Applications to tackle TiLV outbreaks

The presence of a large QTL underlying an important production trait is a relatively rare situation in farmed animal breeding. However, its usefulness has been effectively exploited in the aquaculture industry, by reducing the IPN-related mortalities in Norwegian Atlantic salmon populations to near zero (Hjeltnes 2014; Norris 2017). As such, the use of QTL in breeding programs has highlighted the power of genetic approaches to tackle infectious diseases by improving the disease resistance of aquaculture species. While the size of the effect of this QTL is not as large as for the IPN case, our results highlight a promising avenue to reduce the TiLV-related mortalities through the implementation of selection based on genetic markers in the QTL region. MAS for TiLV resistance could be performed without additional data recording, or in combination with TiLV challenge data on relatives of broodstock. It has been shown that the use of genetic markers together with family selection can achieve higher genetic gain (Spelman and Bovenhuis 1998) compared with the classic pedigree-based approach alone. In addition, future studies to elucidate the underlying causative gene and mutation would provide a putative target for CRISPR-Cas9 genome editing, which has significant potential to expedite the genetic improvement of disease resistance in aquaculture species (Gratacap et al. 2019).

### DATA AVAILABILITY

Details of the SNPs used in the genome-wide association study are available in Peñaloza et al. (2020). The phenotype and genotype data supporting this study belongs to a breeding program managed and owned by WorldFish and is available upon reasonable request. Table S1 contains the complete list of candidate genes flanking the significant SNPs not shown on the main text. Figure S1 shows the Manhattan plot for TD, whereas the QQ plot for BS and TD is shown in Fig. S2.

### REFERENCES

- Aljanabi SM, Martinez I (1997). Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Res* 25
- Aslam ML, Boison SA, Lillehammer M, Norris A, Gjerde B (2020) Genome-wide association mapping and accuracy of predictions for amoebic gill disease in Atlantic salmon (*Salmo salar*). *Sci Rep* 10:1–9
- Bacharach E, Mishra N, Briese T, Zody MC, Kembou Tsofack JE, Zamostiano R et al. (2016) Characterization of a novel orthomyxo-like virus causing mass die-offs of Tilapia. *Am Soc Microbiol* 7:1–7
- Bangera R, Ødegård J, Mikkelsen H, Nielsen HM, Seppola M, Puvanendran V et al. (2014) Genetic analysis of francisellosis field outbreak in Atlantic cod (*Gadus morhua* L.) using an ordinal threshold model. *Aquaculture* 420–421:550–556
- Barria A, Christensen KA, Yoshida GM, Correa K, Jedlicki A, Lhorente JP et al. (2018) Genomic predictions and genome-wide association study of resistance against piscirickettsia salmonis in Coho Salmon (*Oncorhynchus kisutch*) using ddRAD sequencing. *G3 Genes Genomes Genet* 4231.g3.200053.2018
- Barria A, Trinh TQ, Mahmuddin M, Benzie JAH, Chadag VM, Houston RD (2020). Genetic parameters for resistance to Tilapia lake virus (TiLV) in Nile tilapia (*Oreochromis niloticus*). *Aquaculture* 522
- Bassini LN, Lhorente JP, Oyarzún M, Bangera R, Yáñez JM, Neira R (2019) Genetic parameters for Piscirickettsia salmonis resistance, sea lice (*Caligus rogercresseyi*) susceptibility and harvest weight in rainbow trout (*Oncorhynchus mykiss*). *Aquaculture* 510:276–282
- Van den Berg S, Vandenplas J, van Eeuwijk FA, Lopes MS, Veerkamp RF (2019) Significance testing and genomic inflation factor using high-density genotypes or whole-genome sequence data. *J Anim Breed Genet* 136:418–429
- Boison S, Ding J, Leder E, Gjerde B, Bergtun PH, Norris A et al. (2019) QTLs associated with resistance to cardiomyopathy syndrome in Atlantic Salmon. *J Hered* 110:727–737
- Conte MA, Joshi R, Moore EC, Nandamuri SP, Gammerding WJ, Roberts RB et al. (2019) Chromosome-scale assemblies reveal the structural evolution of African cichlid genomes. *Gigascience* 8:1–20
- Core TR (2017). R: A language and environment for statistical computing. R Foundation for Statistical

- Dong HT, Siriroob S, Meemetta W, Santimanawong W, Gangnonngiw W, Pirarat N et al. (2017) Emergence of tilapia lake virus in Thailand and an alternative semi-nested RT-PCR for detection. *Aquaculture* 476:111–118
- Dong HT, Senapin S, Gangnonngiw W, Nguyen VV, Rodkhum C, Debnath PP et al. (2020) Experimental infection reveals transmission of tilapia lake virus (TiLV) from tilapia broodstock to their reproductive organs and fertilized eggs. *Aquaculture* 515:734541
- Eyngor M, Zamostiano R, Tsofack JEK, Berkowitz A, Bercovier H, Tinman S et al. (2014) Identification of a novel RNA virus lethal to tilapia. *J Clin Microbiol* 52:4137–4146
- FAO (2020). The state of world fisheries and aquaculture.
- Fathi M, Dickson C, Dickson M, Leschen W, Baily J, Muir F et al. (2017) Identification of Tilapia lake virus in Egypt in Nile tilapia affected by 'summer mortality' syndrome. *Aquaculture* 473:430–432
- Ferguson HW, Kabuusu R, Beltran S, Reyes E, Lince JA, del Pozo J (2014) Syncytial hepatitis of farmed tilapia, *Oreochromis niloticus* (L.): a case report. *J Fish Dis* 37:583–589
- Foss S, Bottermann M, Jonsson A, Sandlie I, James LC, Andersen JT (2019) TRIM21— from intracellular immunity to therapy. *Front Immunol* 10:1–14
- Fraslin C, Quillet E, Rochat T, Dechamp N, Bernardet J-F, Collet B et al. (2020) Combining multiple approaches and models to dissect the genetic architecture of resistance to infections in fish. *Front Genet* 11:677
- Gilmour AR, Gogel BJ, Cullis BR, Welham SJ, Thompson R (2015). ASReml user guide
- Gjedrem T (2015) Disease resistant fish and shellfish are within reach: a review. *J Mar Sci Eng* 3:146–153
- Gonen S, Baranski M, Thorland I, Norris A, Grove H, Arnesen P et al. (2015) Mapping and validation of a major QTL affecting resistance to pancreas disease (salmonid alphavirus) in Atlantic salmon (*Salmo salar*). *Heredity* 115:405–414
- Gratacap RL, Wargelius A, Edvardsen RB, Houston RD (2019) Potential of genome editing to improve aquaculture breeding and production. *Trends Genet* 35:672–684
- Gutierrez AP, Turner F, Gharbi K, Talbot R, Lowe NR, Peñaloza C et al. (2017) Development of a Medium density combined-species SNP array for Pacific and European oysters (*Crassostrea gigas* and *Ostrea edulis*). *G3 Genes Genomes Genet* 7:2209–2218
- Gutierrez AP, Bean TP, Hooper C, Stenton CA, Sanders MB, Paley RK et al. (2018) A genome-wide association study for host resistance to ostreid herpesvirus in Pacific oysters (*Crassostrea gigas*). *G3 Genes Genomes Genet* 8:1273–1280
- Harrison K, Haga IR, Pechenick Jowers T, Jasim S, Cintrat J-C, Gillet D et al. (2016) Vaccinia virus uses retromer-independent cellular retrograde transport pathways to facilitate the wrapping of intracellular mature virions during virus morphogenesis. *J Virol* 90:10120–10132
- Hjeltnes B (ed) (2014). The health situation in Norwegian aquaculture 2013. *Nor Vet Inst* 42
- Houston RD (2017) Future directions in breeding for disease resistance in aquaculture species. *Rev Bras Zootec* 46:545–551
- Houston RD, Haley CS, Hamilton A, Guy DR, Tinch AE, Taggart JB et al. (2008) Major quantitative trait loci affect resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*). *Genetics* 178:1109–1115
- Houston RD, Haley CS, Hamilton A, Guy DR, Mota-Velasco JC, Gheyas AA et al. (2010) The susceptibility of Atlantic salmon fry to freshwater infectious pancreatic necrosis is largely explained by a major QTL. *Heredity* 105:318–27
- Houston RD, Taggart JB, Cézard T, Bekaert M, Lowe NR, Downing A et al. (2014) Development and validation of a high density SNP genotyping array for Atlantic salmon (*Salmo salar*). *BMC Genomics* 15:90
- Houston RD, Bean TP, Macqueen DJ, Gundappa MK, Jin YH, Jenkins TL et al. (2020) Harnessing genomics to fast-track genetic improvement in aquaculture. *Nat Rev Genet* 21:389–409
- ICTV (2018). Virus taxonomy: 2018 Release. <https://talk.ictvonline.org/taxonomy/>.
- Jaemwimol P, Rawiwan P, Tattiyapong P, Saengnual P, Kamlangdee A, Surachetpong W (2018) Susceptibility of important warm water fish species to tilapia lake virus (TiLV) infection. *Aquaculture* 497:462–468
- Johnson RC, Nelson GW, Troyer JL, Lautenberger JA, Kessing BD, Winkler CA et al. (2010) Accounting for multiple comparisons in a genome-wide association study (GWAS). *BMC Genom* 11:724
- Jørgensen SM, Afanasyev S, Krasnov A (2008) Gene expression analyses in Atlantic salmon challenged with infectious salmon anemia virus reveal differences between individuals with early, intermediate and late mortality. *BMC Genom* 9:1–16
- Joshi R, Skaard A, de Vera M, Alvarez AT, Ødegård J (2020) Genomic prediction for commercial traits using univariate and multivariate approaches in Nile tilapia (*Oreochromis niloticus*). *Aquaculture* 516:734641
- Joshi R, Árnýasi M, Lien S, Gjoen HM, Alvarez AT, Kent M (2018) Development and validation of 58K SNP-array and high-density linkage map in Nile tilapia (*O. niloticus*). *Front Genet* 9:1–15
- Joshi R, Skaard A, Alvarez AT, Moen T, Ødegård J (2020). Bayesian genomic models boost prediction accuracy for resistance against *Streptococcus agalactiae* in Nile tilapia (*Oreochromis niloticus*). *bioRxiv*: 1–27
- Kabuusu RM, Aire AT, Stroup DF, Macpherson CNL, Ferguson HW (2018) Production-level risk factors for syncytial hepatitis in farmed tilapia (*Oreochromis niloticus* L.). *J Fish Dis* 41:61–66
- LaFrentz BR, Lozano CA, Shoemaker CA, García JC, Xu DH, Løvoll M et al. (2016) Controlled challenge experiment demonstrates substantial additive genetic variation in resistance of Nile tilapia (*Oreochromis niloticus*) to *Streptococcus iniae*. *Aquaculture* 458:134–139
- Levicán J, Miranda-Cárdenas C, Soto-Rifo R, Aguayo F, Gaggero A, León O (2017) Infectious pancreatic necrosis virus enters CHSE-214 cells via macropinocytosis. *Sci Rep* 7:1–12
- Lien S, Koop BF, Sandve SR, Miller JR, Matthew P, Leong JS et al. (2016) The Atlantic salmon genome provides insights into rediploidization. *Nature* 533:200–205
- Lillehammer M, Odegård J, Madsen P, Gjerde B, Refstie T, Rye M (2013) Survival, growth and sexual maturation in Atlantic salmon exposed to infectious pancreatic necrosis: a multi-variate mixture model approach. *Genet Sel Evol* 45:8
- Liu S, Jiang M, Wang W, Liu W, Song X, Ma Z et al. (2018) Nuclear RNF2 inhibits interferon function by promoting K33-linked STAT1 disassociation from DNA article. *Nat Immunol* 19:41–50
- Mallery DL, Mcewan WA, Bidgood SR, Towers GJ, Johnson CM, James LC (2010) Antibodies mediate intracellular immunity through tripartite motif-containing 21 (TRIM21). *Proc Natl Acad Sci USA* 107:2–7
- Misztal I, Tsuruta S, Lourenco D, Aguilar I, Legarra A, Vitezica Z (2015). Manual for BLUPF90 family of programs 125
- Moen T, Baranski M, Sonesson AK, Kjøglum S (2009) Confirmation and fine-mapping of a major QTL for resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*): population-level associations between markers and trait. *BMC Genom* 10:368
- Mohamed AR, Verbyla KL, Al-Mamun HA, McWilliam S, Evans B, King H et al. (2019) Polygenic and sex specific architecture for two maturation traits in farmed Atlantic salmon. *BMC Genom* 20:1–13
- Norris A (2017) Application of genomics in salmon aquaculture breeding programs by Ashie Norris: who knows where the genomic revolution will lead us? *Mar Genom* 36:13–15
- Nugent CM, Leong JS, Christensen KA, Rondeau EB, Brachmann MK, Easton AA et al. (2019) Design and characterization of an 87k SNP genotyping array for Arctic charr (*Salvelinus alpinus*). *PLoS ONE* 14:1–17
- O'Farrell C, Vaghefi N, Cantonnet M, Buteau B, Boudinot P, Benmansour A (2002) Survey of transcript expression in rainbow trout leukocytes reveals a major contribution of interferon-responsive genes in the early response to a rhabdovirus infection. *J Virol* 76:8040–8049
- Ødegård J, Baranski M, Gjerde B, Gjedrem T (2011) Methodology for genetic evaluation of disease resistance in aquaculture species: challenges and future prospects. *Aquac Res* 42:103–114
- Palaiokostas C, Cariou S, Bestin A, Bruant JS, Haffray P, Morin T et al. (2018) Genome-wide association and genomic prediction of resistance to viral nervous necrosis in European sea bass (*Dicentrarchus labrax*) using RAD sequencing. *Genet Sel Evol* 50:1–11
- Palaiokostas C, Robledo D, Vesely T, Prchal M, Pokorova D, Piackova V et al. (2018) Mapping and sequencing of a significant quantitative trait locus affecting resistance to koi herpesvirus in common carp. *G3 Genes Genomes Genet* 8:3507–3513
- Palti Y, Vallejo RL, Gao G, Liu S, Hernandez AG, Rexroad CE et al. (2015) Detection and validation of QTL affecting bacterial cold water disease resistance in rainbow trout using restriction-site associated DNA sequencing. *PLoS ONE* 10:e0138435
- Palti Y, Gao G, Liu S, Kent MP, Lien S, Miller MR et al. (2015) The development and characterization of a 57K single nucleotide polymorphism array for rainbow trout. *Mol Ecol Resour* 15:662–672
- Peñaloza C, Robledo D, Barria A, Trinh TQ, Mahmuddin M, Wiener P, et al (2020). Development and validation of an open access SNP array for Nile tilapia (*Oreochromis niloticus*). *G3 Genes Genomes Genet* 1–37
- Poisa-Beiro L, Dios S, Ahmed H, Vasta GR, Martínez-López A, Estepa A et al. (2009) Nodavirus infection of sea bass (*Dicentrarchus labrax*) induces up-regulation of galectin-1 expression with potential anti-inflammatory activity. *J Immunol* 183:6600–6611
- Purcell S, Neale B, Todd-brown K, Thomas L, Ferreira MAR, Bender D et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575
- Realegeno S, Puschnik AS, Kumar A, Goldsmith C, Burgado J, Sambhara S et al. (2017) Monkeypox virus host factor screen using haploid cells identifies essential virus formation. *J Virol* 91:1–16



- Robledo D, Hamilton A, Gutiérrez AP, Bron JE, Houston RD (2020). Characterising the mechanisms underlying genetic resistance to amoebic gill disease in Atlantic salmon using RNA sequencing. *BMC Genom* 21
- Rodríguez FH, Flores-mara R, Yoshida GM, Barria A (2019) Genome-wide association analysis for resistance to infectious pancreatic necrosis virus identifies candidate genes involved in viral replication and immune response in rainbow trout (*Oncorhynchus mykiss*). *G3 Genes Genomes Genet* 9:2897–2904
- Sandamalika G, Lee J (2020). Quadruple domain-containing galectin from marine invertebrate disk abalone (*Haliotis discus discus*): molecular perspectives in early development, immune expression, and potent antiviral responses. *Fish Shellfish Immunol* 135:577
- Sato S, Nieminen J (2002) Seeing strangers or announcing 'danger': Galectin-3 in two models of innate immunity. *Glycoconj J* 19:583–591
- Shoemaker CA, Lozano CA, Lafrentz BR, García JC, Soto E, Xu D et al. (2017) Additive genetic variation in resistance of Nile tilapia (*Oreochromis niloticus*) to *Streptococcus iniae* and *S. agalactiae* capsular type Ib: Is genetic resistance correlated? *Aquaculture* 468:193–198
- Sinclair-Waters M, Ødegård J, Korsvoll SA, Moen T, Lien S, Primmer CR et al. (2020) Beyond large-effect loci: large-scale GWAS reveals a mixed large-effect and polygenic architecture for age at maturity of Atlantic salmon. *Genet Sel Evol* 52:1–11
- Spelman R, Bovenhuis H (1998) Genetic response from marker assisted selection in an outbred population for differing marker bracket sizes and with two identified quantitative trait loci. *Genetics* 148:1389–1396
- Suebsong W, Poopuang S, Rye M, Srisapoom P (2019). Selection response for *Streptococcus agalactiae* resistance in Nile tilapia *Oreochromis niloticus*. *J Fish Dis* 1553–1562
- Sukhavachana S, Poopuang S, Onming S, Luengnaruemitchai A (2019) Heritability estimates and selection response for resistance to *Streptococcus agalactiae* in red tilapia *Oreochromis* spp. *Aquaculture* 502:384–390
- Swaine T, Dittmar MT (2015) CDC42 Use in viral cell entry processes by RNA viruses. *Viruses* 7:6526–6536
- Taslima K, Davie A, McAndrew BJ, Penman DJ (2016) DNA sampling from mucus in the Nile tilapia, *Oreochromis niloticus*: minimally invasive sampling for aquaculture-related genetics research. *Aquac Res* 47:4032–4037
- Tattiyapong P, Sirikanachana K, Surachetpong W (2018) Development and validation of a reverse transcription quantitative polymerase chain reaction for tilapia lake virus detection in clinical samples and experimentally challenged fish. *J Fish Dis* 41:255–261
- Thammatorn W, Rawiwan P, Surachetpong W (2019) Minimal risk of tilapia lake virus transmission via frozen tilapia filets. *J Fish Dis* 42:3–9
- Tsai HY, Hamilton A, Tinch AE, Guy DR, Gharbi K, Stear MJ et al. (2015) Genome wide association and genomic prediction for growth traits in juvenile farmed Atlantic salmon using a high density SNP array. *BMC Genom*. 16:1–9
- Vallejo RL, Cheng H, Fragomeni BO, Shewbridge KL, Gao G, Macmillan JR et al. (2019) Genome-wide association analysis and accuracy of genome-enabled breeding value predictions for resistance to infectious hematopoietic necrosis virus in a commercial rainbow trout breeding population. *Genet Sel Evol* 51:1–14
- Vasta GR (2009) Roles of galectins in infection. *Nat Rev Microbiol* 7:424–438
- Vaysburd M, Watkinson RE, Cooper H, Reed M, Connell KO, Smith J, et al. (2013). Intracellular antibody receptor TRIM21 prevents fatal viral infection. *Proc Natl Acad Sci USA* 110
- Wang Y, Wang Q, Li Y, Yin J, Ren Y, Shi C et al. (2020) Integrated analysis of mRNA-miRNA expression in Tilapia infected with Tilapia lake virus (TiLV) and identifies primarily immuneresponse genes. *Fish Shellfish Immunol* 99:208–226
- Wonmongkol P, Sukhavachana S, Ampolsak K, Srisapoom P, Suwanasopee T, Poopuang S (2017) Genetic parameters for resistance against *Flavobacterium columnare* in Nile tilapia *Oreochromis niloticus* (Linnaeus, 1758). *J Fish Dis* 41:321–328
- Xing J, Zhang A, Minze LJ, Li XC, Zhang Z (2018) TRIM29 negatively regulates the type I interferon production in response to RNA virus. *J Immunol* 201:183–192
- Xing J, Weng L, Yuan B, Wang Z, Jia L, Jin R et al. (2016) Identification of a role for TRIM29 in the control of innate immunity in the respiratory tract. *Nat Immunol* 17:1373–1380
- Xu J-D, Jiang H-S, Wei T-D, Zhang K-Y, Wang X-W, Zhao X-F et al. (2017) Interaction of the small GTPase Cdc42 with arginine kinase restricts white spot syndrome virus in shrimp. *J Virol* 91:1–16
- Yañez JM, Houston RD, Newman S (2014) Genetics and genomics of disease resistance in salmonid species. *Front Genet* 5:1–13
- Yañez JM, Naswa S, López ME, Bassini L, Correa K, Gilbey J et al. (2016) Genomewide single nucleotide polymorphism discovery in Atlantic salmon (*Salmo salar*): validation in wild and farmed American and European populations. *Mol Ecol Resour* 16:1002–1011
- Yañez JM, Yoshida G, Barria A, Palma-Véjares R, Travisany D, Díaz D, et al (2020). High-throughput single nucleotide polymorphism (SNP) discovery and validation through whole-genome resequencing in Nile Tilapia (*Oreochromis niloticus*). *Mar Biotechnol*
- Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 88:76–82
- Yang M-L, Chen Y-H, Wang S-W, Huang Y-J, Leu C-H, Yeh N-C et al. (2011) Galectin-1 binds to influenza virus and ameliorates influenza virus pathogenesis. *J Virol* 85:10010–10020
- Yin L (2020). CMplot: circle manhattan plot. <https://cran.r-project.org/web/packages/CMplot/CMplot.pdf>.
- Yoshida G, Barria A, Caceres G, Correa K, Jedlicki A, Cadiz MI et al. (2019) Genome-wide patterns of population structure and linkage disequilibrium in farmed Nile tilapia (*Oreochromis niloticus*). *Front Genet* 10:745
- Yoshida GM, Lhorente JP, Correa K, Soto J, Salas D, Yañez JM (2019) Genome-wide association study and cost-efficient genomic predictions for growth and fillet yield in Nile tilapia (*Oreochromis niloticus*). *G3 Genes Genomes Genet* 9:2597–2607
- Yu Y, Luo Z, Wang Q, Zhang Q, Zhang X (2020) Development of high throughput SNP genotyping approach using target sequencing in Pacific white shrimp and its application for genetic study. *Aquaculture* 528:735549
- Zeng Q, Fu Q, Li Y, Waldbieser G, Bosworth B, Liu S et al. (2017) Development of a 690 K SNP array in catfish and its application for genetic mapping and validation of the reference genome sequence. *Sci Rep* 7:1–14
- Zenger KR, Khatkar MS, Jones DB, Khalilislamani N, Jerry DR, Raadsma HW (2019). Genomic selection in aquaculture: Application, limitations and opportunities with special reference to marine shrimp and pearl oysters. *Front Genet* 10
- Zhou Q, Chen Y, Lu S, Liu Y, Xu W, Li Y, et al (2020). Development of a 50K SNP array for Japanese flounder and its application in genomic selection for disease resistance. *Engineering*

## ACKNOWLEDGEMENTS

We thank the Department of Fisheries Malaysia at Jitra and the WorldFish breeding team.

## AUTHOR CONTRIBUTIONS

AB performed the genomic analyses and wrote the first draft of the manuscript, TT and MM led the acquisition and provision of data and samples from throughout the TiLV outbreak, CP contributed to samples management, quality controls and analyses, AP and OG performed DNA extractions, MC collaborated on the experimental design and in the veterinary diagnosis, JB and RH led the experimental design and project management. All authors read and approved the manuscript.

## FUNDING

This work was undertaken as part of and funded by, the CGIAR Research Program on Fish Agri-Food Systems (FISH) led by WorldFish. The program is supported by contributors to the CGIAR Trust Fund. The authors acknowledge funding from the Institute Strategic Program funding from BBSRC to The Roslin Institute (BB/P013759/1 and BB/P013740/1).

## COMPETING INTERESTS

The authors declare no competing interests.

## ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Data collection and sampling were performed as part of a non-profit selective breeding program run by WorldFish. The animals from this breeding population are managed in accordance with the Guiding Principles of the Animal Care, Welfare, and Ethics Policy of WorldFish Center.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41437-021-00447-4>.

**Correspondence** and requests for materials should be addressed to R.D.H.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative

Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021