Aquaculture 548 (2022) 737592

Contents lists available at ScienceDirect

Aquaculture

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

Genome-wide association study toward efficient selection breeding of resistance to *Vibrio alginolyticus* in Pacific oyster, *Crassostrea gigas*

Ben Yang [a], Shangyu Zhai [a], Fuqiang Zhang [a], Hebing Wang [a], Liting Ren [a], Yongjing Li [a], Qi Li [a,b], Shikai Liu [a,b,*]

[a] Key Laboratory of Mariculture (Ocean University of China), Ministry of Education, and College of Fisheries, Ocean University of China, Qingdao 266003, China
[b] Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266237, China

ARTICLE INFO

Keywords:
Crassostrea gigas
Resistance
Vibrio
Genome wide association study

ABSTRACT

The oysters are facing with risk of pathogen infection due to complex and variable intertidal habitats, which frequently causes serious disease problem to oyster industry globally. Genetic improvement of disease resistance can be an effective solution to resolve this problem, while dissection of genetic basis underlying disease resistance is critical toward efficient selection breeding. In this work, we performed a genome wide association study (GWAS) in the Pacific oyster (*Crassostrea gigas*) to identify genetic loci underlying resistance to bacterial *Vibrios*, a major group of pathogens in the ocean. We performed artificial infection of 1402 oysters collected from 52 families using the previously isolated pathogenic *Vibrio alginolyticus*. Selective genotyping of 295 resistant and susceptible individuals were conducted using ddRAD sequencing approach. A total of 48,099 high quality SNPs were detected across whole genome of *C. gigas*. Linkage disequilibrium analysis revealed a rapid decay of linkage across the genome. GWAS revealed polygenic architecture of resistance traits, and identified a total of 18 SNP loci and three QTL underlying the trait of resistance to *Vibrio*, allowing for identification of 178 genes involved in various processes of immune response, including pathogen recognition, immune signaling, and immune defense. This work provides insights into genetic basis of resistance to *V. alginolyticus*, and valuable genetic resources toward genomic selection breeding of strains resistant to *Vibrio* in the *C. gigas*.

1. Introduction

As a fast-growing industry for food production around the world, aquaculture supplies over 50% of aquatic protein sources to fulfill human needs (Little et al., 2016; Norman et al., 2019). However, pathogenic infections by bacteria, viruses, and parasites cause serious disease outbreaks leading to great economic loss, which hamper sustainable development of aquaculture industry globally (Bondad-Reantaso et al., 2005; Ghittino et al., 2003; Stentiford et al., 2012). Especially, with global climate change, aquatic animals are subject to greater environmental stress along with accelerated proliferation of pathogens in natural water, resulting in frequent outbreaks of mass mortality events (Lauringson et al., 2021; Sae-Lim et al., 2017). Intensive recirculating aquaculture system can be applied to some aquaculture species to keep water quality and control outbreaks of diseases (Bartelme et al., 2019; Patil et al., 2020). However, this is not applicable to the species that are cultured in the open ocean, such as marine bivalves.

The Pacific oyster, *C. gigas*, is an economic marine bivalve naturally distributed along the coasts of Northeast Asian countries (Zhai et al., 2021; Zhang et al., 2019b). It has been introduced to various countries for aquaculture due to its strong environmental adaptability, high nutrition and fast growth (Meng et al., 2019; Zhang et al., 2019b). As in other marine bivalves, frequent exposure of environmental stressors and pathogens cause disease outbreaks and adversely affect the oyster industry. In the past two decades, the mortality events of oysters caused by pathogenic infections have been reported in many countries, leading to great economic losses (Dégremont et al., 2015a; Pernet et al., 2016). In recent years, high mortality events of oysters have also been frequently observed in China, from which the causative pathogens were investigated, leading to identification of pathogenic *Vibrios* including *V. alginolyticus* (Zhai et al., 2021). Oysters lack acquired immunity, therefore, their resistance to pathogens can only be enhanced by genetic approach to boost host innate immunity rather than use of vaccines

* Corresponding author at: Key Laboratory of Mariculture (Ocean University of China), Ministry of Education, and College of Fisheries, Ocean University of China, China.
E-mail address: liushk@ouc.edu.cn (S. Liu).

https://doi.org/10.1016/j.aquaculture.2021.737592
Received 28 August 2021; Received in revised form 9 October 2021; Accepted 11 October 2021
Available online 14 October 2021

(Wang et al., 2018).

Understanding genetic basis of disease resistance is critical for its improvement through genetic breeding. Recent development of genomic technologies accelerates dissection of genetic basis of complex traits, thus promoting the progress of genetic improvement (Goddard et al., 2016; Zhu et al., 2016). Genome wide association study (GWAS) is an effective method for analyzing complex traits based on genome-wide genetic variations (Visscher et al., 2012), which has been carried out in numerous aquatic species to dissect genetic basis of various economic traits such as disease resistance, growth, amino acid content, and heat tolerance (Holborn et al., 2020; Meng et al., 2019; Ning et al., 2019; Yu et al., 2021). With fine-scale genetic analyses of traits by GWAS, great progresses have been made in genetic improvement of various economic traits in various aquaculture species such as tilapia, salmon and yellow croaker (Xu et al., 2020; Zhao et al., 2021).

In this study, toward genetic breeding of the Pacific oyster strains resistant to pathogenic *Vibrios*, we performed GWAS to investigate the genetic basis of resistance to *Vibrio* infection in *C. gigas*. The ddRAD-seq (double digest restriction-site associated sequencing) was used to genotype resistant and susceptible oysters after artificial infection with the previously isolated pathogenic *V. alginolyticus* (Yang et al., 2021). This work provides insights into genetic basis of resistance to *Vibrio* in *C. gigas*, which will be valuable information for selective breeding of resistant strains as well as revealing molecular mechanism of resistance to bacterial infection.

2. Materials and methods

2.1. Experiment oysters

The *C. gigas* used in this experiment was bred and cultured as described in the previous study (Zhai et al., 2021). Briefly, a total of 52 families of *C. gigas* were constructed in the breeding season of 2019. The broodstocks (80.6 mm in mean shell height, 40.9 mm in mean shell length) were transported from oyster farm to the hatchery in Laizhou, Shandong Province in March. All oysters were cleaned and placed in baskets and cultured in a cement pond containing 30 m^3 filtered seawater. In order to promote gonadal development, oysters were cultured at 24 °C, and fed with the algae (*Nitzschia closterium* and *Isochrysis galbana*) ad libitum. The water was changed 50% every day to maintain water quality (Li et al., 2011). Artificial fertilization was performed when the gonad develops well. The gonads of oysters were dissected to collect sperm and egg, which were mixed for fertilization in a beaker followed by hatchery in 100 L plastic bucket. The larvae were cultured in the100 L plastic bucket with regular culture practice as described previously (Chi et al., 2021). When the spats grew to 2–3 mm in shell height, they were transferred to the oyster farm in Rongcheng (Shandong, China) for culture in the ocean.

2.2. Bacterial challenge experiment

The bacterial challenge experiment was performed as reported in our previous study (Zhai et al., 2021), using the highly pathogenic *V. alginolyticus* strain Cg5 that was isolated from diseased oysters (Yang et al., 2021). The bacteria were cultured on agar medium to recover the virulence, and then cultured in flasks containing 1 L sterile 2216E broth (121 °C, 30 min for sterilization) for 12 h in constant temperature shaker at 28 °C. The bacteria were harvested by centrifugation (4 °C, 8000 *g*, 5 min) to prepare the suspension for use (Yang et al., 2021). Oysters randomly collected from the 52 families were used for the challenge experiment. These oysters were identified as healthy oysters before experiment by random PCR detection and hemolymph bacterial load analysis (Yang et al., 2021). Before the challenge experiment, oysters were acclimated to the laboratory environment and no oyster died during acclimation. After two-week acclimation, a total of 1402 oysters were infected via intramuscular injection with the same dose of suspension (5×10^7 CFU), then assigned to 52 baskets (one family per basket) that were randomly arranged in four water tanks. The water temperature during the experiment was maintained at 22 ± 1 °C. To keep good water quality, 50% of the seawater in the tanks was changed and the water quality was monitored every day. The death of oysters was observed every two hours and the survival time was recorded. The adductor muscle of dead oysters (denoted as susceptible) was collected and deposit into 1.5 mL DNA store (Sangon Biotech Co., Ltd., China, CAT. NO. B644771) at room temperature. The challenge experiment last for 12 days post infection, and the survived oysters at the end of the experiment was denoted as resistant. The adductor muscle of these survivors was collected and deposit into 1.5 mL DNA store (Sangon Biotech Co., Ltd., China, CAT. NO. B644771) at room temperature.

2.3. SNP genotyping with ddRAD

A total of 131 susceptible individuals (dead oysters) and 164 resistant individuals (survivors) were selected for SNP genotyping using ddRAD sequencing. Genomic DNA was extracted according to the standard phenol chloroform method. The RAD library for sequencing was constructed as described in the previous study (Peterson et al., 2012). Briefly, the DNA products digested by both *Eco*RI and *Msp*I (20 units of each enzyme) were ligated to adaptors with unique barcodes using T4 ligase. The ligation products of 72 individuals were pooled into one library and separated on 2% agarose gel. Fragments of size with 280-430 bp were re-isolated using *E*-Gel system (Zhou et al., 2019). The RAD library was amplified by PCR and the PCR product was resolved on 2% agarose gel. The high throughput sequencing was performed with Illumina HiSeq 2500 for 150 bp paired-end reads. The Stacks program developed for ddRAD was used for SNP identification following the standard procedure (Rochette et al., 2019). The filtering of SNPs with high quality was carried out according to the following parameters: the MAF (minor allele frequency) was >0.05, call rate > 0.9, the individual missing <0.9.

2.4. Linkage disequilibrium analysis

The linkage disequilibrium (LD) was analyzed to determine the LD decay within the genome of *C. gigas*. PopLDdecay was used to calculate the LD value (r^2) at pairwise loci, then the average r^2 value was calculated within 5 bp interval. The LD decay map was drawn with the average r^2 value along the physical distance (Zhang et al., 2019a). The half of the maximum r^2 value was defined as decay distance (Li et al., 2019).

2.5. Genome wide association study

The phenotype of resistance to bacterial infection was defined as survival status (death versus survival) and survival time (the time last from the beginning of infection to the death) (Yi and Xu, 1999). Genome wide association analysis was performed with R-package GAPIT (Tang et al., 2016), using the next generation GWAS model of BLINK (Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway) to estimate *P* value of each locus in the association test between SNP and resistance trait (Huang et al., 2019). The model is described as:

$$y = \mu + Zu + QTNs + Q + e$$

where y is the vector of phenotypes; μ is the intercept; u is vector of additive genetic effects for individuals; Z are the known design matrices; the detected pseudo QTNs (quantitative trait nucleotides) were used as cofactors to control false positives; Q is population structure and e is the vector of residuals. In order to correct the effects of phenotypic differences among families, the family effect was included as a covariate in GWAS analysis.

BLINK first estimated QTNs with GLM model, then used QTNs as

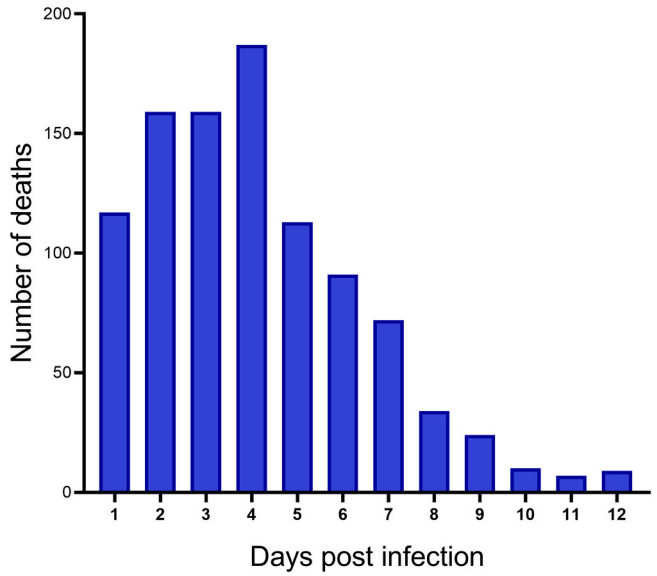

Fig. 1. Daily mortality of *C. gigas* after infection with *Vibrio alginolyticus* during challenge experiment.

covariates for marker association analysis. Then, Bayesian information criterion strategy was used to detect the accuracy of QTNs, retaining the real QTNs, and then repeated iterations until all QTNs were detected. In GWAS, the first three PCA (Principal Component Analysis) were used to correct the population structure. This method has been proved to be equivalent to Q matrix and can effectively correct the population stratification (Zhao et al., 2007). The validity of population structure correction results was verified by Q-Q plot. To avoid false positives, Bonferroni method was used to correct the significance of multiple comparisons, and Bonferroni's threshold was 1 / N (N is the number of SNPs used for analysis), representing suggestive threshold at genome level (Lander and Kruglyak, 1995). Besides the whole genome level threshold, the suggestive threshold on chromosome level was set as 1 / n (n is the number of SNPs on chromosome) (Yu et al., 2021; Zhong et al., 2017). Manhattan plot and Q-Q plot were generated according to GWAS results using R package CMplot.

Furthermore, Bayesian window analysis was performed, in which the ratio of the of additive genetic variance estimated by each interval to the total genetic variance was used to determine the potential QTL (quantitative trait locus) regions. In order to determine the resistance QTL, 300000 iterations were constructed in BayesB model to calculate the variance of genetic variation of each 1 Mb non overlapping window, and the interval with more than 1% variance explanation was defined as QTL (Peters et al., 2013).

2.6. Gene annotation and enrichment analysis

The SnpEff program was used to construct the genome annotation database of *C. gigas* using genome annotation files and reference genome sequences, followed by annotation of identified SNPs from GWAS (Cingolani et al., 2012). Considering the genome level linkage disequilibrium in *C. gigas*, genomic regions flanking the detected SNPs (2 kb up- and downstream) were identified for candidate genes (Gutierrez et al., 2018). The genes were annotated by homology search in Swissprot and nr database, followed by Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) enrichment analysis using R package clusterProfiler.

3. Results

3.1. Survival of oysters post infection

As shown in Fig. 1, the number of daily deaths was recorded during the 12-day challenge experiment. After infection, oysters began to die on the first day, and reached peak of deaths on the fourth day, then gradually decreased (Fig. 1). Molecular characterization and morphological observation were carried out by re-isolating bacteria from dead oysters and confirmed that the deaths were caused by the infection of *V. alginolyticus*.

3.2. SNP identification and population analysis

A total of 523,313,588 clean reads were generated from ddRAD sequencing, of which 90.17% were mapped to the reference genome (NCBI accession: GCA_902806645.1), resulting in the average sequencing depth of 17.5× and genome coverage of 3.27%. We identified a total of 3.3 million SNPs, of which 48,099 high-quality SNPs were obtained from 128 susceptible and 149 resistant individuals after stringent filtering. The high-quality SNPs were evenly distributed across the ten chromosomes with relatively low density in chromosome 2 and chromosome 5 (Fig. 2A). Analysis of population structure showed that significant genetic differentiation existed among challenged *C. gigas*, while the population structure between resistant and susceptible oysters did not show significant difference. (Fig. 2B). Linkage disequilibrium analysis revealed the extremely fast decay at genome level (Fig. 2C). The maximum value of r^2 was 0.46 and quickly decreased to half of the maximum value to 0.23 at 48 bp. Notably, LD decay in resistant and susceptible oysters showed similar pattern (Supplementary Table 1).

3.3. GWAS of resistance to Vibrio infection

GWAS of resistance to *Vibrio* infection in terms of survival status and survival time were carried out respectively, resulting in identification of 18 suggestive SNP loci associated with resistance at chromosome level ($P = 2.07 \times 10^{-4}$, Table 1). Of which, 12 SNPs were detected when the resistance trait was defined as survival status, while 15 SNPs were detected when the resistance trait was defined as survival time (Fig. 3A and B). The phenotypic variation explained by the suggestive loci ranged from 4.99% to 6.32% for survival status and from 4.62% to 6.26% for survival time, indicating the polygenic architecture of resistance to *Vibrio*. The Q-Q plot analyses showed stringent corrections of population stratification in association analysis (Fig. 3C and D). Further Bayesian window analysis revealed three windows with more than 1% of explained additive genetic variance that were identified as QTL for survival status, and only one QTL was detected for the survival time (Table 2). These QTL together explained 3.83% of genetic variance. Notably, seven of 18 SNPs with suggestive significance were located in these three QTL, including the two SNP loci, SNP26634 and SNP26635, with genome wide suggestive significance (Table 1).

3.4. Functional annotation of genes associated with SNPs for Vibrio resistance

A total of 10 genes were annotated from the total 18 SNP loci associated with *Vibrio* resistance (Table 3). Furthermore, a total of 171 genes were annotated from the three QTL for resistance (Supplementary Table 2). Many of these genes were well recognized to be involved in immune related biological processes including pathogen recognition, immune signaling transduction and immune defense response. For instance, Toll like receptor 6 and scavenger receptor class F member 1 are two well-known pattern recognition receptors involved in pathogen sensing. The ubiquitin system includes E3 ubiquitin-protein ligase Mdm2, E3 ubiquitin-protein ligase HACE1 and ubiquitin carboxyl-terminal hydrolase CYLD, which transduce immune signals through

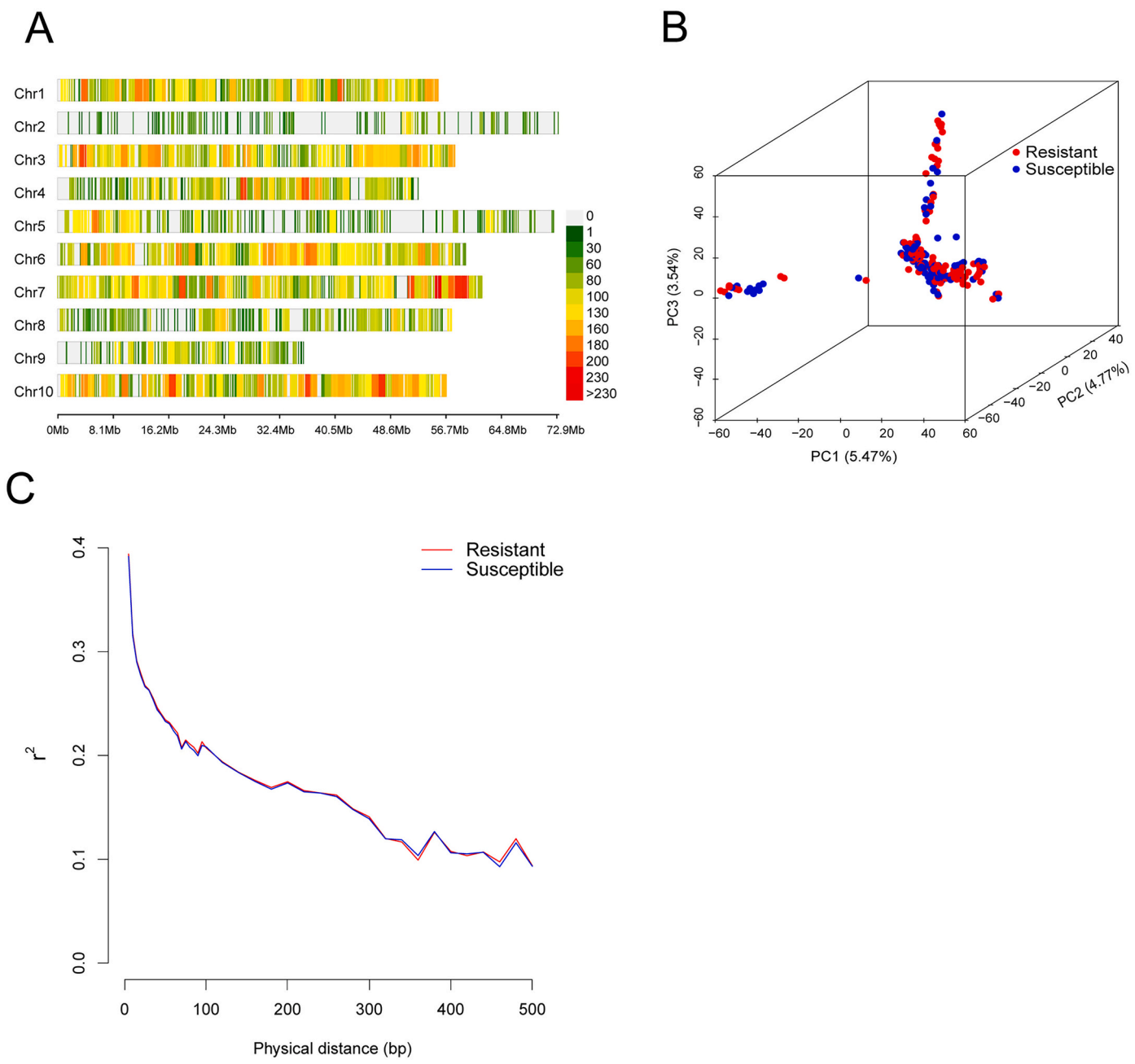

Fig. 2. (A) Identification of high-quality SNPs and their distribution across 10 chromosomes of *C. gigs*. The gradient colors from green to red denote the increase of SNP density within 1 Mb interval. (B) Population structure of *C. gigas* as revealed by PCA analysis. The percentages of variance explained by the three PCA were 5.47%, 4.77% and 3.54%. The two subpopulations were classified according to phenotypic traits. The subgroup of resistant was the individuals who survived at the end of the experiment, and the subgroup of susceptible was individuals died of *V. alginolyticus* infection in challenge trial. (C) Linkage disequilibrium decay of resistant and susceptible groups of oysters. The X-axis is the physical location, and the Y-axis is the linkage disequilibrium value (r^2). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

ubiquitin degradation of proteins. Laccase-19 is a laccase type phenoloxidase that directly kills bacteria as an immune effector, and cytochrome *b*-245 can produce superoxide to kill invading pathogens. Notably, a set of genes related to cytoskeleton regulation were also detected, including ras-like GTP-binding protein rhoA, GTP-binding protein REM and ras-like and GTP-binding protein Rho1.

Enrichment analysis of GO terms and KEGG pathways related to these genes provided insights into biological processes associated with the resistance to *Vibrio* in *C. gigas* (Fig. 4). For instance, enriched GO terms of T cell lineage commitment and T cell migration were observed (Fig. 4A), suggesting the activation and defense response of lymphocytes involved in innate immunity. Similarly, KEGG pathways of C-type lectin and T cell receptor as pattern recognition receptor were enriched, and key immune signal transduction pathways include MAPK signaling pathway and NOD-like receptor signaling pathway were enriched (Fig. 4B). Additionally, GO terms of mitotic cleavage furrow formation and regulation of Rho protein signal transduction, and gene pathways including regulation of actin cytoskeleton, mTOR signaling pathway and Rap1 signaling pathway related to cytoskeleton regulation were significantly enriched. Notably, the signaling pathways of Yersinia infection and shigellosis also showed the role of host cytoskeleton regulation in response to *Vibrio* infection (Fig. 4B).

Table 1
Identification of SNPs with suggestive significance for resistance to *V. alginolyticus* in *C. gigas* by GWAS.

SNP	Chromosome	Position	*Survival status*		*Survival time*	
			P value	PVE	*P* value	PVE
SNP9517	Chr 3	4,894,061	3.26E-05	5.87%	1.92E-04	4.62%
SNP9919	Chr 3	8,491,408	NS		7.95E-05	5.41%
SNP21002	Chr 5	19,139,714	1.33E-04	4.99%	NS	
SNP23277	Chr 6	6,022,228	7.01E-05	5.71%	NS	
SNP23280	Chr 6	6,022,417	1.08E-04	5.42%	1.79E-04	5.36%
SNP23281	Chr 6	6,022,432	1.55E-04	5.17%	NS	
SNP23282	Chr 6	6,022,433	1.08E-04	5.42%	1.79E-04	5.36%
SNP26634*	Chr 6	36,346,559	9.75E-06	6.49%	5.34E-05	5.31%
SNP26635*	Chr 6	36,346,566	9.75E-06	6.49%	5.34E-05	5.31%
SNP28770	Chr 6	54,458,946	8.10E-05	5.47%	1.62E-04	5.11%
SNP40901	Chr 9	27,514,812	7.90E-05	5.61%	2.81E-05	6.26%
SNP40907	Chr 9	27,514,868	NS		7.35E-05	5.74%
SNP44591	Chr 10	30,456,003	NS		9.79E-05	4.88%
SNP44592	Chr 10	30,456,014	NS		9.79E-05	4.88%
SNP44594	Chr 10	30,456,027	NS		4.75E-05	5.49%
SNP44598	Chr 10	30,456,147	NS		4.75E-05	5.49%
SNP45726	Chr 10	40,049,210	4.51E-05	5.83%	7.46E-05	5.55%
SNP47547	Chr 10	52,397,980	2.92E-05	6.32%	7.49E-05	6.14%

* indicted SNP loci with genome level suggestive significance with *P* value of 2.07×10^{-5}. PVE, phenotypic variation explained by SNPs. NS, not significant.

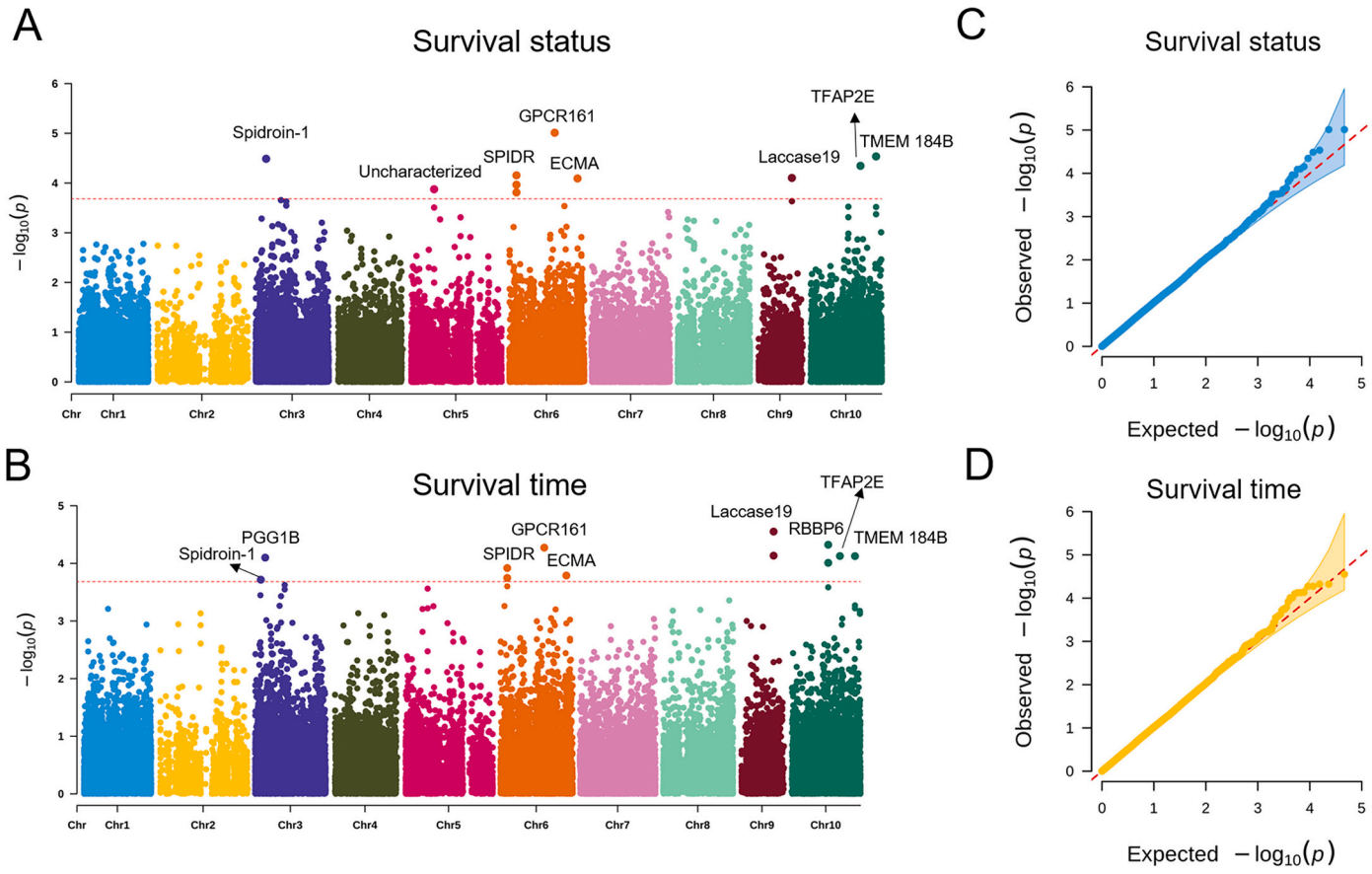

Fig. 3. Genome wide association analysis of resistance to *V. alginolyticus* in *C. gigas*. (A) Manhattan plot for GWAS of survival status, and (B) Manhattan plot for GWAS of survival time. The red line indicated chromosome-level suggestive significance threshold *P* value of 2.07×10^{-4}. (C) Q-Q plot of survival status, and (D) Q-Q plot of survival time. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Discussion

As an important aquaculture species, the Pacific oyster has been under great threat from outbreaks of disease over decades (Alfaro et al., 2019; Nguyen et al., 2019; Petton et al., 2015). Bacterial pathogens, especially *Vibrio*, have become the major risk to the oyster industry (Petton et al., 2015). Lack of acquired immunity makes it difficult to use vaccines against pathogens in oysters. Therefore, improvement through genetic breeding is the only effective way for healthy and sustainable development of oyster industry (Dégremont et al., 2015a). In *C. gigas*, studies on the disease resistance has mainly focused on the OsHV-1 virus, while resistance to *Vibrio* drawn less attention (Dégremont et al., 2015b). In this work, we first investigated the resistance to *Vibrio* in *C. gigas* using genome wide association study approach, which will lay

Table 2
Identification of QTLs for resistance to *V. alginolyticus* in *C. gigas* by GWAS based on Bayesian window analysis.

Traits	Chr	Start (Mb)	End (Mb)	Number of SNPs	Explained additive genetic variance
Survival status	Chr6	6	7	158	1.38%
Survival status	Chr5	19	20	40	1.24%
Survival status	Chr6	36	37	185	1.21%
Survival time	Chr6	6	7	158	1.41%

Table 3
Functional annotation of genes associated with SNPs for *Vibrio* resistance.

SNP	Gene ID	Gene name	SNP location
SNP9517	LOC117688787	Spidroin-1-like (Spidroin-1)	exon
SNP9919	LOC105339449	geranylgeranyl transferase type-1 subunit beta (PGG1B)	intergenic region
SNP21002	LOC105348032	uncharacterized LOC105348032	intron
SNP23277	LOC105321090	DNA repair-scaffolding protein (SPIDR)	3′ UTR
SNP23280	LOC105321090	DNA repair-scaffolding protein (SPIDR)	3′ UTR
SNP23281	LOC105321090	DNA repair-scaffolding protein (SPIDR)	3′ UTR
SNP23282	LOC105321090	DNA repair-scaffolding protein (SPIDR)	3′ UTR
SNP26634	LOC117692375	G protein-coupled receptor 161-like (GPCR161)	intron
SNP26635	LOC117692375	G protein-coupled receptor 161-like (GPCR161)	intron
SNP28770	LOC105328113	extracellular matrix protein A (ECMA)	intron
SNP40901	LOC105334004	putative laccase-19 (Laccase 19)	intron
SNP40907	LOC105334004	putative laccase-19 (Laccase 19)	intron
SNP44591	LOC105332786	E3 ubiquitin-protein ligase RBBP6 (RBBP6)	intron
SNP44592	LOC105332786	E3 ubiquitin-protein ligase RBBP6 (RBBP6)	intron
SNP44594	LOC105332786	E3 ubiquitin-protein ligase RBBP6 (RBBP6)	intron
SNP44598	LOC105332786	E3 ubiquitin-protein ligase RBBP6 (RBBP6)	intron
SNP45726	LOC105325417	transcription factor AP-2-epsilon (TFAP2E)	intron
SNP47547	LOC117684183	transmembrane protein 184B-like (TMEM 184B)	intron

a solid foundation for efficient selection breeding of *C. gigas* strains resistant to *Vibrio* bacteria, and for research to unravel molecular mechanism underlying resistance.

The availability of genome resources and advancement of genotyping technologies based on high throughput sequencing enabled genome-wide association studies of disease resistance. With the previously published oyster genome sequences (Zhang et al., 2012a), the recently updated chromosome level assembly provided valuable resources for genome research in *C. gigas* (Peñaloza et al., 2021; Qi et al., 2021). Considering the cost of large-scale genotyping (Weymann et al., 2017), various reduced representation genome resequencing methods have been developed for large-scale genotyping with low cost (Scheben et al., 2017). The ddRAD is one of the widely used methods by constructing a reduced representation genome sequencing library using double restriction enzyme digestion followed by genotyping with next generation high-throughput sequencing (Peterson et al., 2012). Compared with SNP chip, it does not require prior genome information, and can provide flexibility in marker density in the whole genome according to the selection of restriction sites. Compared with whole genome sequencing, the cost of this method can be greatly reduced. In this study, we obtained a total of 48,099 high quality SNPs using ddRAD approach. Overall, the SNPs were well-distributed across 10 chromosomes with relatively low density in chromosome 2 and 5. This could be due to the uneven distribution of restriction enzyme sites that were used.

LD at the whole genome level affects the accuracy of association analysis and the density of the number of markers needed (Siol et al., 2017). In other species, the decay distance is tens to hundreds of kb (Gaut and Long, 2003; Siol et al., 2017), but it was less than 100 bp in *C. gigas* (Meng et al., 2019), which implies a high recombination rate at the genome level. Consistently, the genome-wide LD estimated in this work showed rapid decay. The pattern of LD decay has a great effect on association analysis by impacting on genotype imputation (Kabisch et al., 2017). Rapid LD decay requires more SNPs to ensure the validity of association analysis (Flint-Garcia et al., 2003). Furthermore, considering the rapid LD decay, we annotated genes within flanking regions of 2 kb upstream and downstream of the identified SNPs with suggestive significance, instead of the commonly used 100 kb flanking regions for gene identification. Population structure is another important factor to affect LD (Park, 2019; Vos et al., 2017). In this work, LD analysis in two subgroups with different phenotype showed the same decay distance, indicating the consistent LD decay between the two subgroups and similar pattern of recombination events within the population.

Further annotation and functional enrichment of genes identified in genomic regions with suggestive SNPs allowed identification of genes and pathways associated with *Vibrio* resistance in *C. gigas* (Fig. 5).

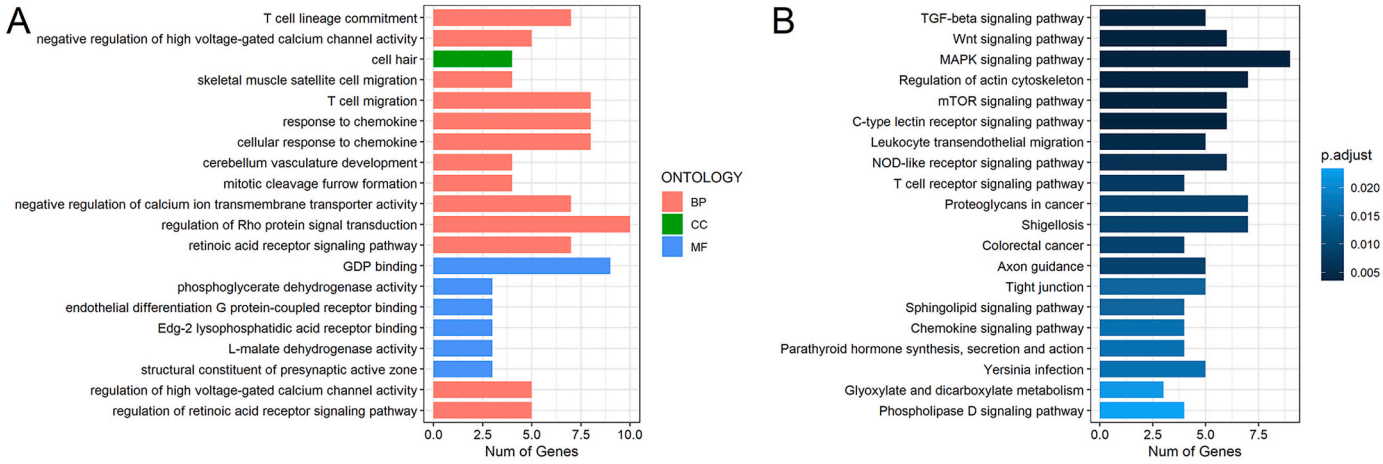

Fig. 4. Enrichment analysis of genes identified from three QTL detected by GWAS based on Bayesian window analysis. (A) The top 20 terms of GO enrichment analysis. (B) The top 20 pathways of KEGG enrichment analysis. The GO terms and pathways were ranked according to the statistical significance.

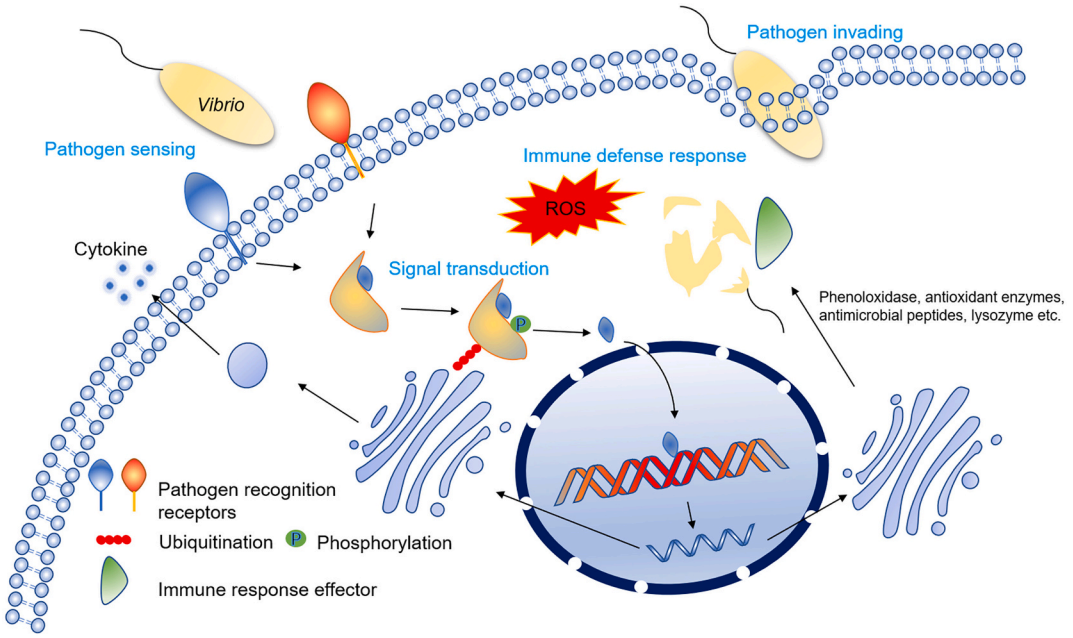

Fig. 5. Schematic diagram of putative molecular basis underlying *Vibrio* resistance in *C. gigas*.

Various pattern recognition receptors of the host are involved in sense infection from pathogens (Takeuchi and Akira, 2010). Furthermore, the signals can be transduced to activate immune response through MAPK signaling pathway (Arthur and Ley, 2013). Meanwhile, ubiquitination-mediated degradation of the proteins that suppress immune signaling molecules promotes transcription factors to enter the nucleus for regulating expression of various immune genes (Sun, 2017; Takeda and Akira, 2004). The resulting cytokines activated lymphocytes to phagocytize and eliminate pathogens (Commins et al., 2010). Upon immune activation, the direct immune effector proteins such as phenoloxidase and lysozyme directly act on the bacteria (Luna-Acosta et al., 2017), or through ROS and antioxidant enzymes system to defend against bacteria (Sies and Jones, 2020; Yang et al., 2013). The pathogen could regulate cytoskeleton through its secretory system to invade host cells for immune evasion and proliferation (Ribet and Cossart, 2015; Skirpstunas and Baldwin, 2002; Zhang et al., 2012b). Host cells invaded by bacteria can eliminate pathogens through autophagy or apoptosis (Campisi et al., 2014; Cui et al., 2019; Navratil et al., 2004). Together, the resistance or susceptibility of an organism relies on the interaction between host immune system and pathogen invasion. When pathogenic microorganisms burst to cause disease, genetic differences could lead to differences in the expression of immune genes among individuals, affecting the survival of oysters during disease outbreaks. Previously reported gene expression analysis of resistant and susceptible oysters in naturally occurring diseases revealed that differences in disease resistance lied in differential expression of a set of immune related genes (de Lorgeril et al., 2018; de Lorgeril et al., 2020), which included the immune molecules identified in our study such as toll like receptors, scavenger receptors, tripartite motif containing proteins, ubiquitin ligases, superoxide dismutase and helicases. These reported immune genes support our proposed molecular mechanism related to *Vibrio* disease resistance.

Selective breeding is effective to genetically improve monogenic traits or the traits controlled by major QTL, in which a few markers can be sufficient for marker-assisted selection breeding (Fu et al., 2017; Lammerts van Bueren et al., 2010). However, for complex traits with polygenic architecture, traditional selection based on small numbers of marker is generally ineffective. Most of traits for resistance to pathogens reported in a variety of economic aquatic animals are with polygenic architecture (Fraslin et al., 2020; Holborn et al., 2018). For instance, a recently published work reported the polygenic trait of resistance to OsHV-1 infection in *C. gigas* (Gutierrez et al., 2018). In our work, we showed that resistance to *V. alginolyticus* infection in *C. gigas* is also a trait with polygenic architecture, which is consistent with our previous estimate of the low to moderate heritability of this trait (Zhai et al., 2021). For the trait controlled by many genes with minor effects, reliable prediction of genomic breeding values requires large numbers of SNPs at genome scale. Therefore, genomic selection breeding will be an efficient approach for genetic improvement (Hollenbeck and Johnston, 2018; Zenger et al., 2018).

5. Conclusion

In this work, we, for the first time, dissected genetic basis of resistance to *V. alginolyticus* in the *C. gigas*. Genome-wide association study revealed a total of 18 loci and three QTL underlying the trait of resistance to *Vibrio*, allowing for identification of critical genes and pathways involved in various processes of immune response, including pathogen recognition (C-type lectin receptor and T-cell receptor signaling), immune signaling (MAPK signaling and ubiquitination process), and immune defense (phenoloxidase and superoxide). This work provides insights into genetic basis of resistance to *V. alginolyticus*, and valuable genetic resources toward genomic selection breeding of strains resistant to *Vibrio* in the *C. gigas*.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aquaculture.2021.737592.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgements

This work was supported by the grants from National Natural Science Foundation of China (No. 31802293 and No. 41976098), and the Young Talent Program of Ocean University of China (No. 201812013), and the Fundamental Research Funds for the Central Universities (No. 202042011).

References

Alfaro, A.C., Nguyen, T.V., Merien, F., 2019. The complex interactions of Ostreid herpesvirus 1, *Vibrio* bacteria, environment and host factors in mass mortality outbreaks of *Crassostrea gigas*. Rev. Aquac. 11, 1148–1168.

Arthur, J.S., Ley, S.C., 2013. Mitogen-activated protein kinases in innate immunity. Nat. Rev. Immunol. 13, 679–692.

Bartelme, R.P., Smith, M.C., Sepulveda-Villet, O.J., Newton, R.J., 2019. Component microenvironments and system biogeography structure microorganism distributions in recirculating aquaculture and aquaponic systems. mSphere 4 e00143-19.

Bondad-Reantaso, M.G., Subasinghe, R.P., Arthur, J.R., Ogawa, K., Chinabut, S., Adlard, R., Shariff, M., 2005. Disease and health management in Asian aquaculture. Vet. Parasitol. 132, 249–272.

Campisi, L., Cummings, R.J., Blander, J.M., 2014. Death-defining immune responses after apoptosis. Am. J. Transplant. 14, 1488–1498.

Chi, Y., Li, Q., Liu, S., Kong, L., 2021. Genetic parameters of growth and survival in the Pacific oyster *Crassostrea gigas*. Aquac. Res. 52, 282–290.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Ruden, D.M., 2012. 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 6, 80–92.

Commins, S.P., Borish, L., Steinke, J.W., 2010. Immunologic messenger molecules: cytokines, interferons, and chemokines. J. Allergy Clin. Immunol. 125, S53–S72.

Cui, B., Lin, H., Yu, J., Yu, J., Hu, Z., 2019. Autophagy and the immune response. Adv. Exp. Med. Biol. 1206, 595–634.

de Lorgeril, J., Escoubas, J.-M., Loubiere, V., Pernet, F., Le Gall, P., Vergnes, A., Bachère, E., 2018. Inefficient immune response is associated with microbial permissiveness in juvenile oysters affected by mass mortalities on field. Fish Shellfish Immunol. 77, 156–163.

de Lorgeril, J., Petton, B., Lucasson, A., Perez, V., Stenger, P.-L., Dégremont, L., Mitta, G., 2020. Differential basal expression of immune genes confers *Crassostrea gigas* resistance to Pacific oyster mortality syndrome. BMC Genomics 21, 63.

Dégremont, L., Garcia, C., Allen Jr., S.K., 2015a. Genetic improvement for disease resistance in oysters: a review. J. Invertebr. Pathol. 131, 226–241.

Dégremont, L., Lamy, J.B., Pépin, J.F., Travers, M.A., Renault, T., 2015b. New insight for the genetic evaluation of resistance to Ostreid herpesvirus infection, a worldwide disease, in *Crassostrea gigas*. PLoS One 10, e0127917.

Flint-Garcia, S.A., Thornsberry, J.M., Buckler, E.S., 2003. Structure of linkage disequilibrium in plants. Annu. Rev. Plant Biol. 54, 357–374.

Fraslin, C., Quillet, E., Rochat, T., Dechamp, N., Bernardet, J.F., Collet, B., Boudinot, P., 2020. Combining multiple approaches and models to dissect the genetic architecture of resistance to infections in fish. Front. Genet. 11, 677.

Fu, Y.-B., Yang, M.-H., Zeng, F., Biligetu, B., 2017. Searching for an accurate marker-based prediction of an individual quantitative trait in molecular plant breeding. Front. Plant Sci. 8, 1182.

Gaut, B.S., Long, A.D., 2003. The lowdown on linkage disequilibrium. Plant Cell 15, 1502.

Ghittino, C., Latini, M., Agnetti, F., Panzieri, C., Lauro, L., Ciappelloni, R., Petracca, G., 2003. Emerging pathologies in aquaculture: effects on production and food safety. Vet. Res. Commun. 27 (Suppl. 1), 471–479.

Goddard, M.E., Kemper, K.E., MacLeod, I.M., Chamberlain, A.J., Hayes, B.J., 2016. Genetics of complex traits: prediction of phenotype, identification of causal polymorphisms and genetic architecture. Proc. R. Soc. B 283, 1835.

Gutierrez, A.P., Bean, T.P., Hooper, C., Stenton, C.A., Sanders, M.B., Paley, R.K., Houston, R.D., 2018. A genome-wide association study for host resistance to Ostreid herpesvirus in Pacific Oysters (*Crassostrea gigas*). G3-Genes Genom. Genet. 8, 1273–1280.

Holborn, M.K., Ang, K.P., Elliott, J.A.K., Powell, F., Boulding, E.G., 2018. Genome wide association analysis for bacterial kidney disease resistance in a commercial North American Atlantic salmon (*Salmo salar*) population using a 50 K SNP panel. Aquaculture 495, 465–471.

Holborn, M.K., Ang, K.P., Elliott, J.A.K., Powell, F., Boulding, E.G., 2020. Genome wide analysis of infectious salmon anemia resistance in commercial Saint John River Atlantic salmon. Aquaculture 514, 734514.

Hollenbeck, C.M., Johnston, I.A., 2018. Genomic tools and selective breeding in molluscs. Front. Genet. 9, 253.

Huang, M., Liu, X., Zhou, Y., Summers, R.M., Zhang, Z., 2019. BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. Gigascience 8, 1–12.

Kabisch, M., Hamann, U., Lorenzo Bermejo, J., 2017. Imputation of missing genotypes within LD-blocks relying on the basic coalescent and beyond: consideration of population growth and structure. BMC Genomics 18, 798.

Lammerts van Bueren, E.T., Backes, G., de Vriend, H., Østergård, H., 2010. The role of molecular markers and marker assisted selection in breeding for organic agriculture. Euphytica 175, 51–64.

Lander, E., Kruglyak, L., 1995. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat. Genet. 11, 241–247.

Lauringson, M., Nousiainen, I., Kahar, S., Burimski, O., Gross, R., Kaart, T., Vasemägi, A., 2021. Climate change-driven disease in sympatric hosts: temporal dynamics of parasite burden and proliferative kidney disease in wild brown trout and Atlantic salmon. J. Fish Dis. 44, 689–699.

Li, Q., Wang, Q., Liu, S., Kong, L., 2011. Selection response and realized heritability for growth in three stocks of the Pacific oyster *Crassostrea gigas*. Fish. Sci. 77, 643–648.

Li, N., Zheng, H., Cui, J., Wang, J., Liu, H., Sun, J., Zou, D., 2019. Genome-wide association study and candidate gene analysis of alkalinity tolerance in japonica rice germplasm at the seedling stage. Rice 12, 24.

Little, D.C., Newton, R.W., Beveridge, M.C., 2016. Aquaculture: a rapidly growing and significant source of sustainable food? Status, transitions and potential. Proc. Nutr. Soc. 75, 274–286.

Luna-Acosta, A., Breitwieser, M., Renault, T., Thomas-Guyon, H., 2017. Recent findings on phenoloxidases in bivalves. Mar. Pollut. Bull. 122, 5–16.

Meng, J., Song, K., Li, C., Liu, S., Shi, R., Li, B., Zhang, G., 2019. Genome-wide association analysis of nutrient traits in the oyster *Crassostrea gigas*: genetic effect and interaction network. BMC Genomics 20, 625.

Navratil, J.S., Sabatine, J.M., Ahearn, J.M., 2004. Apoptosis and immune responses to self. Rheum. Dis. Clin. N. Am. 30, 193–212.

Nguyen, T.V., Alfaro, A.C., Merien, F., 2019. Omics approaches to investigate host–pathogen interactions in mass mortality outbreaks of *Crassostrea gigas*. Rev. Aquac. 11, 1308–1324.

Ning, X., Li, X., Wang, J., Zhang, X., Kong, L., Meng, D., Bao, Z., 2019. Genome-wide association study reveals E2F3 as the candidate gene for scallop growth. Aquaculture 511, 734216.

Norman, R.A., Crumlish, M., Stetkiewicz, S., 2019. The importance of fisheries and aquaculture production for nutrition and food security. Rev. Sci. Tech. 38, 395–407.

Park, L., 2019. Population-specific long-range linkage disequilibrium in the human genome and its influence on identifying common disease variants. Sci. Rep. 9, 11380.

Patil, H.J., Gatica, J., Zolti, A., Benet-Perelberg, A., Naor, A., Dror, B., Cytryn, E., 2020. Temporal resistome and microbial community dynamics in an intensive aquaculture facility with prophylactic antimicrobial treatment. Microorganisms 8, 1984.

Peñaloza, C., Gutierrez, A.P., Eöry, L., Wang, S., Guo, X., Archibald, A.L., Houston, R.D., 2021. A chromosome-level genome assembly for the Pacific oyster *Crassostrea gigas*. GigaScience 10, 1–9.

Pernet, F., Lupo, C., Bacher, C., Whittington, R.J., 2016. Infectious diseases in oyster aquaculture require a new integrated approach. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 371, 20150213.

Peters, S.O., Kizilkaya, K., Garrick, D.J., Fernando, R.L., Reecy, J.M., Weaber, R.L., Thomas, M.G., 2013. Heritability and Bayesian genome-wide association study of first service conception and pregnancy in Brangus heifers1. J. Anim. Sci. 91, 605–612.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E., 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One 7, e37135.

Petton, B., Bruto, M., James, A., Labreuche, Y., Alunno-Bruscia, M., Le Roux, F., 2015. *Crassostrea gigas* mortality in France: the usual suspect, a herpes virus, may not be the killer in this polymicrobial opportunistic disease. Front. Microbiol. 6, 686.

Qi, H., Li, L., Zhang, G., 2021. Construction of a chromosome-level genome and variation map for the Pacific oyster *Crassostrea gigas*. Mol. Ecol. Resour. 21, 1670–1685.

Ribet, D., Cossart, P., 2015. How bacterial pathogens colonize their hosts and invade deeper tissues. Microbes Infect. 17, 173–183.

Rochette, N.C., Rivera-Colón, A.G., Catchen, J.M., 2019. Stacks 2: analytical methods for paired-end sequencing improve RADseq-based population genomics. Mol. Ecol. 28, 4737–4754.

Sae-Lim, P., Kause, A., Mulder, H.A., Olesen, I., 2017. Breeding and genetics symposium: climate change and selective breeding in aquaculture. J. Anim. Sci. 95, 1801–1812.

Scheben, A., Batley, J., Edwards, D., 2017. Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. Plant Biotechnol. J. 15, 149–161.

Sies, H., Jones, D.P., 2020. Reactive oxygen species (ROS) as pleiotropic physiological signalling agents. Nat. Rev. Mol. Cell Biol. 21, 363–383.

Siol, M., Jacquin, F., Chabert-Martinello, M., Smýkal, P., Le Paslier, M.C., Aubert, G., Burstin, J., 2017. Patterns of genetic structure and linkage disequilibrium in a large collection of Pea Germplasm. G3-Genes Genom. Genet. 7, 2461–2471.

Skirpstunas, R.T., Baldwin, T.J., 2002. *Edwardsiella ictaluri* invasion of IEC-6, Henle 407, fathead minnow and channel catfish enteric epithelial cells. Dis. Aquat. Org. 51, 161–167.

Stentiford, G.D., Neil, D.M., Peeler, E.J., Shields, J.D., Small, H.J., Flegel, T.W., Lightner, D.V., 2012. Disease will limit future food supply from the global crustacean fishery and aquaculture sectors. J. Invertebr. Pathol. 110, 141–157.

Sun, S.C., 2017. The non-canonical NF-κB pathway in immunity and inflammation. Nat. Rev. Immunol. 17, 545–558.

Takeda, K., Akira, S., 2004. TLR signaling pathways. Semin. Immunol. 16, 3–9.

Takeuchi, O., Akira, S., 2010. Pattern recognition receptors and inflammation. Cell 140, 805–820.

Tang, Y., Liu, X., Wang, J., Li, M., Wang, Q., Tian, F., Zhang, Z., 2016. GAPIT version 2: an enhanced integrated tool for genomic association and prediction. Plant Genom. 9, 1–9.

Visscher, P.M., Brown, M.A., McCarthy, M.I., Yang, J., 2012. Five years of GWAS discovery. Am. J. Hum. Genet. 90, 7–24.

Vos, P.G., Paulo, M.J., Voorrips, R.E., Visser, R.G.F., van Eck, H.J., van Eeuwijk, F.A., 2017. Evaluation of LD decay and various LD-decay estimators in simulated and SNP-array data of tetraploid potato. Theor. Appl. Genet. 130, 123–135.

Wang, L., Song, X., Song, L., 2018. The oyster immunity. Dev. Comp. Immunol. 80, 99–118.

Weymann, D., Laskin, J., Roscoe, R., Schrader, K.A., Chia, S., Yip, S., Regier, D.A., 2017. The cost and cost trajectory of whole-genome analysis guiding treatment of patients with advanced cancers. Mol. Genet. Genom. Med. 5, 251–260.

Xu, P., David, L., Martínez, P., Yue, G.H., 2020. Editorial: genetic dissection of important traits in aquaculture: genome-scale tools development, trait localization and regulatory mechanism exploration. Front. Genet. 11, 642.

Yang, Y., Bazhin, A.V., Werner, J., Karakhanova, S., 2013. Reactive oxygen species in the immune system. Int. Rev. Immunol. 32, 249–270.

Yang, B., Zhai, S., Li, X., Tian, J., Li, Q., Shan, H., Liu, S., 2021. Identification of *Vibrio alginolyticus* as a causative pathogen associated with mass summer mortality of the Pacific Oyster (*Crassostrea gigas*) in China. Aquaculture 535, 736363.
Yi, N., Xu, S., 1999. Mapping quantitative trait loci for complex binary traits in outbred populations. Heredity 82, 668–676.
Yu, F., Peng, W., Tang, B., Zhang, Y., Wang, Y., Gan, Y., Ke, C., 2021. A genome-wide association study of heat tolerance in Pacific abalone based on genome resequencing. Aquaculture 536, 736436.
Zenger, K.R., Khatkar, M.S., Jones, D.B., Khalilisamani, N., Jerry, D.R., Raadsma, H.W., 2018. Genomic selection in aquaculture: application, limitations and opportunities with special reference to marine shrimp and pearl oysters. Front. Genet. 9, 693.
Zhai, S., Yang, B., Zhang, F., Li, Q., Liu, S., 2021. Estimation of genetic parameters for resistance to *Vibrio alginolyticus* infection in the Pacific oyster (*Crassostrea gigas*). Aquaculture 736545.
Zhang, G., Fang, X., Guo, X., Li, L., Luo, R., Xu, F., Wang, J., 2012a. The oyster genome reveals stress adaptation and complexity of shell formation. Nature 490, 49–54.
Zhang, L., Krachler, Anne M., Broberg, Christopher A., Li, Y., Mirzaei, H., Gilpin, Christopher J., Orth, K., 2012b. Type III effector Vopc mediates invasion for *Vibrio* species. Cell Rep. 1, 453–460.
Zhang, C., Dong, S.S., Xu, J.Y., He, W.M., Yang, T.L., 2019a. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. Bioinformatics 35, 1786–1788.
Zhang, F., Hu, B., Fu, H., Jiao, Z., Li, Q., Liu, S., 2019b. Comparative transcriptome analysis reveals molecular basis underlying fast growth of the selectively bred Pacific oyster, *Crassostrea gigas*. Front. Genet. 10, 610.
Zhao, K., Aranzana, M.J., Kim, S., Lister, C., Shindo, C., Tang, C., Nordborg, M., 2007. An Arabidopsis example of association mapping in structured samples. PLoS Genet. 3, e4.
Zhao, J., Bai, H., Ke, Q., Li, B., Zhou, Z., Wang, H., Xu, P., 2021. Genomic selection for parasitic ciliate *Cryptocaryon irritans* resistance in large yellow croaker. Aquaculture 531, 735786.
Zhong, X., Wang, X., Zhou, T., Jin, Y., Tan, S., Jiang, C., Liu, Z., 2017. Genome-wide association study reveals multiple novel QTL associated with low oxygen tolerance in hybrid catfish. Mar. Biotechnol. 19, 379–390.
Zhou, Z., Han, K., Wu, Y., Bai, H., Ke, Q., Pu, F., Xu, P., 2019. Genome-wide association study of growth and body-shape-related traits in large yellow croaker (*Larimichthys crocea*) using ddRAD sequencing. Mar. Biotechnol. 21, 655–670.
Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M.R., Powell, J.E., Yang, J., 2016. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat. Genet. 48, 481–487.