Fish and Shellfish Immunology 88 (2019) 76–83

Contents lists available at ScienceDirect

Fish and Shellfish Immunology

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

Full length article

Whole-genome resequencing from bulked-segregant analysis reveals gene set based association analyses for the *Vibrio anguillarum* resistance of turbot (*Scophthalmus maximus*)

Kai Zhang[a], Miao Han[a], Yuxiang Liu[a], Xiaohan Lin[a], Xiumei Liu[a,c], He Zhu[a], Yan He[a,b], Quanqi Zhang[a,b], Jinxiang Liu[a,b,*]

[a] Key Laboratory of Marine Genetics and Breeding, Ministry of Education, Ocean University of China, Qingdao, Shandong, 266003, China
[b] Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology, Qingdao, Shandong, 266237, China
[c] College of Life Sciences, Yantai University, Yantai, 264005, China

ARTICLE INFO

Keywords:
BSA-seq
Turbot
Vibrio anguillarum
SNP
Association analysis

ABSTRACT

Many achievements have been made to develop quantitative trait loci (QTLs) and gene-associated single nucleotide polymorphisms (SNPs) to facilitate practical marker-assisted selection (MAS) in aquatic animals. However, the systematic studies of SNPs associated with extreme threshold traits were poor in populations lacking of parental genomic information. Coupling next generation sequencing with bulked segregant analysis (BSA) should allow identification of numerous associated SNPs with extreme phenotypes. In the present study, using combination of SNP frequency difference and Euclidean distance, we conducted linkage analysis of SNPs located in genes involved in immune responses, and identified markers associated with *Vibrio anguillarum* resistance in turbot (*Scophthalmus maximus*). A total of 221 SNPs was found as candidate SNPs between resistant and susceptible individuals. Among these SNPs, 35 loci located in immune related genes were genotyped in verification population and 7 of them showed significant association with *V. anguillarum* resistance in both alleles and genotypes ($P < 0.05$). Among these 7 genes, PIK3CA-like, CYLD, VCAM1, RhoB and RhoGEF are involved in PI3K/Akt/mTOR pathway and NF-κB pathway, which influence the efficiency of bacteria entering the host and inflammation. SNP-SNP interaction analysis was performed by generalized multifactor dimensionality reduction (GMDR). The combination of SNP loci in RhoB, PIK3CA-like and ADCY3 showed a significant effect on *V. anguillarum* resistance with the verification rate in the sequencing population up to 70.8%. Taken all, our findings demonstrated the feasibility of BSA-seq approach in identifying genes responsible for the extreme phenotypes and will aid in performing MAS in turbot.

1. Introduction

In recent decades, aquiculture breeding has been greatly accelerated by marker-assisted selection (MAS), especially for disease resistance. Association analysis, including pedigree-based QTL mapping and genome-wide association studies (GWAS), was regarded as a powerful tool for mapping both qualitative and quantitative traits [1–3]. GWAS has been widely applied for mapping complex diseases traits in human beings, economic crops and marine animals [4–6]. However, the candidate gene strategy was more widely applied due to its advantages in specificity and accuracy for association analysis, especially in species with no genome sequences and experimental individuals with low genetic diversity [7].

Different kinds of techniques have been used to anatomize the genes affecting the production and performance traits in aquaculture species. For example, traditional QTL mapping was applied to localize the region associated with resistance to *Vibrio*-resistant in Japanese flounder [8,9]. With the continuous development of next-generation sequencing (NGS), BSA-seq, a NGS-based QTL mapping strategy analysis in which two phenotypically distinct pool bulks of recombinant progeny are set up and genotyped, has successfully applied for mapping genes in various of plants and marine animals [10,11].

With the development of SNPs genotyping technologies, large numbers of SNPs can be obtained simultaneously with lower cost.

* Corresponding author. Key Laboratory of Marine Genetics and Breeding, Ministry of Education, Ocean University of China, No.5 Yushan Road, Qingdao, 266003, China.
E-mail address: liujinxiang@ouc.edu.cn (J. Liu).

https://doi.org/10.1016/j.fsi.2019.02.041
Received 6 January 2019; Received in revised form 18 February 2019; Accepted 19 February 2019
Available online 23 February 2019

Recently, the gene sets or pathway-based association analysis was widely used based on GWAS database [12–14]. This method eliminate redundant information, improve breeding efficiency and combine the subtle effects of multiple SNPs which would boost the joint signal in a gene set potentially. Gene set based association analyses have been applied in case-control studies of human diseases and economic trait studies of animals [15,16].

Turbot, *Scophthalmus maximus,* is a flatfish with increasing commercial demand world-wide. After first introduced in 1992, it has become one of the dominant marine aquaculture species in China with an annual production of about 40,000 tons since 2010 [17]. However, increased industrial large scale cultivation and the lack of disease-resistant strains have resulted in its susceptibility to various pathogens, including *Edwardsiella tarda*, *Streptococcus iniae*, and different species of *Vibrio*. *Vibrio anguillarum*, a widely reported gram-negative bacterial which can infect a variety of fish species through skin, gill, lateral line and intestine and cause systemic histopathologic changes, has brought high mortality and severe economic losses in fish aquaculture [8,18,19]. Selection for *V. anguillarum*-resistant strains should be the most effective way to solve this disease problem.

In the present study, we performed *V. anguillarum*-resistant association analysis based on whole genome resequencing of two DNA bulks from three populations (each bulks with 20 individuals) showing extreme phenotypic values by NGS technology. SNPs in the genes participated in immune related signaling pathways were selected and genotyped in validated families. The GMDR method was applied to further explore the influence of gene-by-gene interactions on the *V. anguillarum* resistance/susceptibility of Turbot. These results serve as a foundation for analyses of the genetic mechanisms underlying bacterial infection in turbot and for MAS for *V. anguillarum* resistant individuals of *S. maximus*.

2. Materials and methods

2.1. Ethics statement

All handling of fishes was conducted in accordance with the guidelines of the Institutional Animal Care and Use ACCEPTEDMANUSCRIPT Committee of the Ocean University of China (protocol number 11–06) and the China Government Principles for the Utilization and Care of Vertebrate Animals Used in Testing, Research, and Training (State Science and Technology Commission of the People's Republic of China for No. 2, October 31, 1988. http://www.gov.cn/gongbao/content/2011/content_1860757.htm).

2.2. Fish and bacteria

The fish used in this study was turbot *S. maximus.* In April 2016, 300 one-year-old healthy individuals (average size 0.38 ± 0.02 kg) were purchased from three different hatcheries (Haiyang, Shandong, China) with 100 each, named stock A, B and C. All fish were maintained in the laboratory at 19 ± 0.5 °C for at least 3 days prior to experimental use and challenged in three 3000 L tanks (2600 L fresh sea water) with continuous aeration. The pathogenic *V. anguillarum* strain LMG 4437(T) used in this study was Norway. Before challenge, *V. anguillarum* were cultured overnight in 2216 E medium at 28 °C, and then centrifuged and re-suspended to a final concentration of 3×10^8 CFU (half lethal concentration) with Ringer's solution for marine teleosts.

In April 2017, 247 fish, from 30 half-sib families, raised under the same conditions as those in 2016 were randomly harvested for confirmation of our detected polymorphic SNPs associated with *V. anguillarum* resistance.

2.3. Challenge experiment

The fish were divided by stocks with control group containing 30 individuals (10 per stock) and treatment group containing 270 fish (90 per stock). Fish in the treatment group were injected intraperitoneally with *V. anguillarum* suspension by weight (0.1 ml/100 g), whereas the control group were injected with the same amount of Ringer's solution. Mortality was recorded every 3 h based on the occurrences of dead fish. During 24 h after challenge, the first 20 dying individuals with obvious *V. anguillarum* clinic infection symptom were collected as susceptible group. After seven days of the challenge, all survived fish were collected as resistant fish from which 20 individuals were selected randomly to build resistant group.

2.4. Sequencing, variation calling and association analysis

Muscle tissue was dissected from each individual. Genomic DNA was extracted using traditional phenol–chloroform extraction method in combination with RNase treatment and stored at −20 °C until use [20]. Two bulks were generated by pooling equal amounts of DNA from susceptible and resistant individuals from each stock. DNA from each bulks of three stocks were used to construct paired-end sequencing libraries, which were sequenced on an Illumina HiSeqTM X Ten platform (Novogene, Beijing, China).

After removing adapter and low quality reads, the clean reads were further rechecked for quality using FASTQC. High quality paired-end reads were mapped to the turbot reference genome sequence (PRJEB11743) [21] using the BWA software with default parameters [22]. SAMtools (Version: 0.1.18) [23] software was used to convert mapping results into the BAM format and to filter the unmapped and non-unique reads. Picard package (picard.sourceforge.net, Version: 1.87) was used to filter duplicated reads. SNP detection was performed using the Genome Analysis Toolkit (GATK, version 2.4–7-g5e89f01) [24] and SAMtools. Only the SNPs detected by both two methods were used in further analyses. The detailed processes could be identified in Zhou' study [25]. SNP annotation was performed using the package ANNOVAR [26] (Version: 2013-08-23). Based on the annotation, the SNPs were categorized in exonic regions, splicing sites, 5′UTRs (untranslated regions), 3′UTRs, intronic regions, upstream and downstream regions as well as intergenic regions. The SNPs in coding exons were further grouped into synonymous SNPs (sSNP, did not cause amino acid changes) or nonsynonymous SNPs (nsSNP, caused amino acid changes; mutations causing stop gain and stop loss were also classified into this group).

2.5. Euclidean distance calculation

For unknown information of parents, Euclidean distance association analysis [27] and SNP-index association analysis [28] were performed in this study. Euclidean distance (ED) is a metric that does not require parental strain information and resistant to noise.

The higher the ED value is, the closer the object site. The equation was followed as:

$$ED = \sqrt{(A_{Res} - A_{Sus})^2 + (C_{Res} - C_{Sus})^2 + (G_{Res} - G_{Sus})^2 + (T_{Res} - T_{Sus})^2} \quad (1)$$

where each letter (*A*, *C*, *G*, *T*) corresponds to the frequency of its corresponding DNA nucleotide in the resistant and susceptible bulk respectively. Generally, the SNP loci with more than two mutants are extremely rare, so two of the terms will be zero. We also raising the distance measurement to a power (ED^x) to decrease noise created by small variations in the allelic frequency estimations. A sliding window approach (100 kb windows sliding in 10 kb steps) was applied to quantify the polymorphism levels.

Table 1
Summary of sequencing and mapping.

Group	Read length (bp)	Raw reads	Raw Q30 (%)	Clean reads	Clean Q30 (%)	Clean reads (%)	Mapped reads ratio (%)	Depth(X)
AS	150	45,009,572	94.38	44,916,554	95.85	99.79	85.4	29.5
AR	150	42,574,882	93.85	39,400,455	95.79	92.54	86.2	25.8
BS	150	52,133,439	94.64	51,982,781	93.92	99.71	86.2	34.2
BR	150	42,494,776	94.50	42,439,977	94.53	99.87	85.4	27.8
CS	150	41,720,105	94.36	41,586,626	94.91	99.68	84.8	27.3
CR	150	53,459,592	93.72	48,933,708	93.66	91.53	85.6	32.1

2.6. SNP selection, genotyping and data analysis

To identify loci associated with *V. anguillarum* resistance, Euclidean distance of each SNP locus in susceptible and resistant bulks was calculated using high-quality SNPs with QD quality score ≥ 10 and $10 \leq$ read depth ≤ 100. The average ED^4 value was measured and plotted for all scaffolds of the turbot genome. The top 1% windows in each three stocks and SNPs located in exonic, 5′UTRs, 3′UTRs and upstream regions were chosen for further analysis. Only SNPs with ED^4 value higher than 0.1 were chosen for further analysis. According to annotations, all SNPs located in immune related pathways or genes were genotyped in *V. anguillarum* susceptible/resistant families (Family S and Family R) to verify its accuracy by Fluidigm EP1 KASP (The Kompetitive Allele Specific PCR genotyping) system. For KASP genotyping, the SNP must meet the requirement of at least a 100 bp-length flanking region at each side of the putative SNP position and there were no other putative SNPs in the flanking regions. Accordingly, 35 SNPs in 34 genes were genotyped. The name of the gene was used to represent the testing SNP loci in it.

In order to detect the potential interactions among SNPs in different genes, SNPs associated with *V. anguillarum* resistance were chosen for further gene-gene interaction analysis. GMDR could divided various combinations of genotypes into 'high' or 'low' risk genotypes with cross-validation strategy [29]. In this research, the best models for one, two and three factors were given. Individuals with high risk genotypes were identified as more susceptible to *V. anguillarum* than those with low risk genotypes. Models were considered to be significant at $P < 0.05$.

2.7. Validation of SNPs interaction model

Depending on the results of GMDR analysis, 192 individuals were selected from the sequencing group with resistant and susceptible individuals in equal proportions. After randomly disturb the individual order, muscle from each fish was taken to extract DNA for SNP genotyping. Genomic DNA extracted from each fish were used for SNPs genotyping by KASP system with Fluidigm EP1. The primers used for genotyping were shown in Table S1. Individuals predicted with best combinations were grouped as 'low' risk (lr) group and vice versa. The accuracy of best model was calculated by combining the experimental resistance of each individual.

2.8. Gene expression analysis

The individuals were classified into different groups based on the result of SNP genotyping. Three individuals were randomly chosen from each group and the liver tissues were used for RT-PCR. Quantitative real time PCR (qPCR) was implemented using 2 × SYBR Green qPCR Master Mix (US Everbright Inc.) and analyzed on Multicolor Real-Time PCR Detection System (Roche Lightcycler 480, Switzerland) with triplicates for each PCR reaction. The specific primers and annealing temperatures are listed in Table S1. The results of were calculated using the $2^{-\Delta\Delta Ct}$ comparative Ct method and the β-actin gene was set as the internal reference to normalize expression levels between samples.

2.9. Statistical analysis

The significant differences of genotype and allele frequencies between the susceptible and resistance groups, as well as between validated groups, were detected using Pearson's Chi-square test in SPSS 19.0. A one-way ANOVA with Tukey's test was used to analyze the mRNA expression, and $P < 0.05$ represented significantly different.

3. Results

3.1. Genome resequencing of three stocks

We sequenced six pooled samples of resistant and susceptible fish in each three stocks and obtained 89.4 G bp of sequence data with approximately 29-fold coverage (25.8–34.2 ×). About 85% of these reads were aligned to a unique position in the turbot genome by BWA software and these positions were used for single nucleotide polymorphism (SNP) calling (Table 1). Raw reads data were archived at the NCBI Sequence Read Archive (SRA) under Accession PRJNA512430.

We detected 370,502, 450,897 and 487,465 polymorphic sites from the three stocks, respectively. The number of SNPs per base pair was 1.40 times larger at synonymous sites than that at non-synonymous sites in coding regions (Table 2). The SNPs annotated in different regions were evenly distributed in each stock indicating that these three stocks repeat well.

3.2. Identification of selective loci and Euclidean distance association analysis

After blasting the related cDNA sequences of these unigenes to the assembled genome sequences, all 221 loci distributed in 131 unigenes with high quality were obtained for further SNPs genotyping analyses in susceptible and resistant groups. Among these loci, 73 SNPs were located in 34 immune related genes. All information was shown in Table S2. Although stock C identified more candidate SNPs and unigenes than the other two, no same unigene was screened out from all three stocks. Due to the assay capacity and the limitations of primer design for the genotyping analysis, a total of 35 SNPs were chosen for the genotyping based on the SNP annotation. However, 2 SNPs were not

Table 2
SNPs detected in different stocks.

Stock	A	B	C
SUM	370,502	450,897	487,465
Exonic	20,015	21,424	21,932
Intergenic	162,100	198,786	214,434
Intronic	160,212	196,521	212,830
Splicing	69	68	74
Upstream	12,949	15,797	17,193
Downstream	10,099	12,354	14,040
3′UTR	3870	4650	5615
5′UTR	1188	1297	1347
Nonsynonymous	8311	8999	9097
Synonymous	11,597	12,394	12,710
Syn/Nonsyn	1.40	1.38	1.40

genotyped successfully with technical reasons (Table S3).

3.3. Confirmation of genetic locus in families

The resistance-associated SNPs identified based on stocks A, B and C were genotyped in a total of 247 individuals from Family S (n = 118) and Family R (n = 129), respectively. After performing general statistical analysis and the Chi-square test, 13 SNPs were found significantly associated with *V. anguillarum* resistance in genotypes, while 10 SNPs showed significant association in alleles ($P < 0.05$). Further analysis found 7 SNPs to be significantly associated both in genotypes and alleles which were located in rho guanine nucleotide exchange factor 11 (RhoGEF), ras homolog family member B (RhoB), phosphatidylinositol 4, 5-bisphosphate 3-kinase catalytic subunit alpha isoform-like (PIK3CA-like), ubiquitin carboxyl-terminal hydrolase CYLD (CYLD), vascular cell adhesion protein 1-like (VCAM1), adenylate cyclase type 3 (ADCY3) and claudin-10 (CLDN10), respectively. Among these SNPs, one was nonsynonymous (VCAM1), two were synonymous (ADCY3 and PIK3CA-like), two were in 5′UTRs (CYLD and CLDN10), and the remaining two were in 3′UTRs (RhoGEF) and upstream region (RhoB). Therefore, these genes were considered as candidate genes for *V. anguillarum* resistance in turbot.

3.4. Gene expression in turbots with different genotypes

The relative expression levels of the seven genes in which SNPs with both genotypes and alleles significantly associated with *V. anguillarum* resistance were detect through qPCR. In ADCY3 and CLDN10 heterozygotes showed slightly higher expression than the two homozygotes, but no obvious difference was observed between any two genotypes. Contrarily, in CYLD, PIK3CA-like, RhoB, RhoGEF and VCAM1 the relative expression of the two homozygotes showed significant differences ($P < 0.05$), while the expression of the heterozygotes were in between the two homozygotes (Fig. 1).

The expressions of ADCY3, RhoB and PIK3CA-like mRNA in liver tissue of both susceptible and resistant groups after *V. anguillarum* challenge are shown in Fig. 2. The expressions of ADCY3 in resistant group showed significantly higher expression than in susceptible group that reached the peak at 48 h after infection ($P < 0.05$). Similar phenomena were observed in RhoB and PIK3CA-like in which mRNA expressions were obviously elevated from 24 h after infection ($P < 0.05$).

3.5. GMDR analysis for SNP-SNP interaction associated with V. anguillarum resistance

GMDR analysis was performed to detect the combined effects of all the 7 SNPs showing associations with *V. anguillarum* resistance in both genotypes and alleles. The best interaction models of single-loci, two-loci and three-loci were shown in Table 3. One-factor model consisted of locus from PIK3CA-like, two-factor model consisted of loci from PIK3CA-like and ADCY3, whereas three-factor model combined with PIK3CA-like, ADCY3 and RhoB. All of these models were presented significant association with *V. anguillarum* resistance/susceptibility in the confirmation families of turbot. The interactions of the three-factor SNPs are shown in Fig. 3. The risk of turbot to *V. anguillarum* infection caused by different combinations of SNPs are exhibited. These combinations were classified into high risk genotype and low risk genotype. Individuals with high risk genotypes were considered to have higher risk to *V. anguillarum* infection than those with low risk genotypes. As a result the combinations of GG-GG-CC, GG-AA-TT, GG-AA-CT, GG-AA-CC, GT-GG-CC, GT-AG-CT, GT-AG-CC, GT-AA-TT, GT-AA-CC, TT-GG-CT, TT-GG-CC, TT-AG-CC, TT-AA-CT and TT-AA-CC in ADCY3, RhoB and PIK3CA-like were regarded to be susceptible to *V. anguillarum*. The rest combinations were supposed to be *V. anguillarum* resistance.

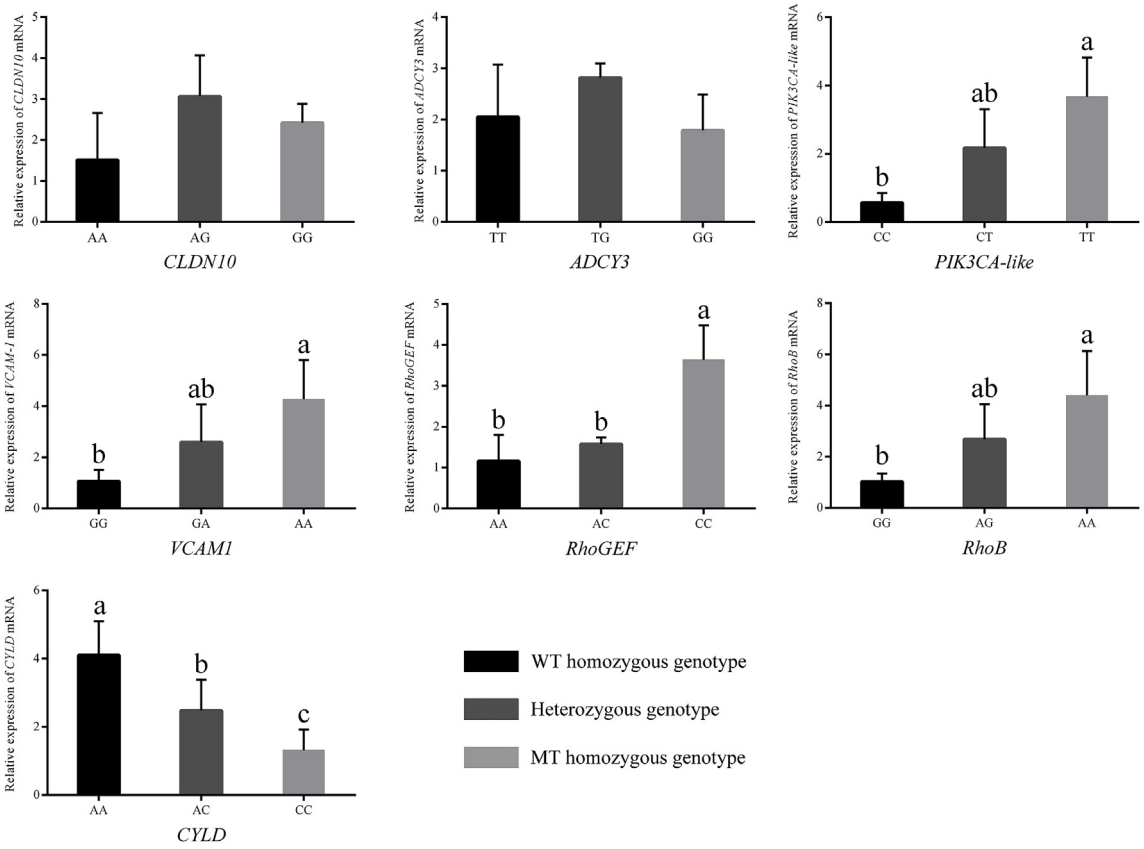

Fig. 1. Relative mRNA expression of seven genes measured by qPCR in the liver tissue of *S. maximus* after injection with *V. anguillarum*. Error bars represent standard deviation of triplicates. Different lowercase letters indicate significant differences ($P < 0.05$).

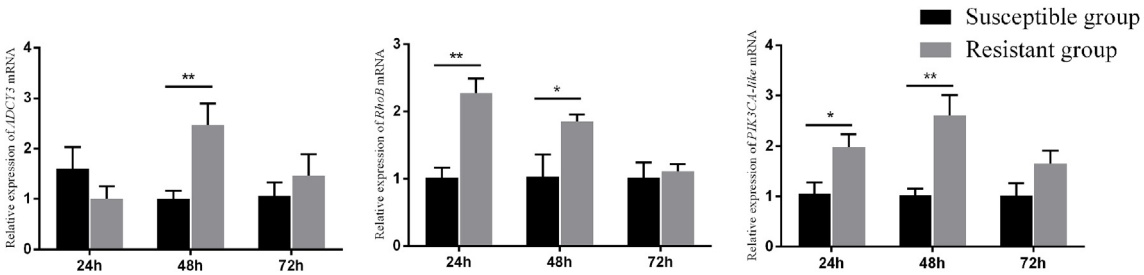

Fig. 2. Expression profiles of ADCY3, RhoB and PIK3CA-like in groups with different levels upon *V. anguillarum* challenge. Genes selected by the best three model were used to detected expression profiles. Error bars represent standard deviation of triplicates. The different asterisks letters indicate significant differences ($P < 0.05$).

Table 3
Best gene-gene interaction models of selected genes by GMDR.

The best model	Training Bal. Acc.	Testing Bal. Acc.	*P* value	CV Consistency
PIK3CA-like	0.6548	0.6580	0.0107*	10
PIK3CA-like & ADCY3	0.6890	0.6695	0.0010**	9
PIK3CA-like & ADCY3 & RhoB	0.7557	0.7227	0.0010**	10

Superscript (*) indicates significant difference at the $P < 0.05$ level. Superscript (**) indicates extremely significant difference at the $P < 0.01$ level.

3.6. Verification of GMDR analysis

A total of 192 individuals randomly collected from stock A, B and C were used for genotyping with SNPs in ADCY3, RhoB and PIK3CA-like to confirm the actual results of the predicted models in selective breeding. After genotyping in susceptible and resistant groups, the total identifications accuracy by three-factor model was about 70.8%, which was relatively higher than one-factor model (51.1%) and two-factor model (53.2%) (Table 4). In this model, 92 individuals were proposed to be susceptible to *V. anguillarum*, while 93 were resistant ones. Combined with the phenotype of injection experiment, 58 individuals were found to be consistent with the expected genotype in susceptible group with an accuracy of 63.0%. In the resistant group it had a higher identification rate of 78.5%.

4. Discussion

Linkage mapping and association analysis provide a tool to explain the basis of complex traits by genes. To date, association analysis is still the major approach for the mining of SNPs for disease resistance traits in marine animals. Previous studies reported that SNP frequencies varied among different populations in various species [1,6,30]. Nevertheless, to our knowledge, there is limited immune related SNP information available in relation with the treatment response to *V. anguillarum* in *S. maximus*. Therefore, studies conducted on the identification of genes and its SNPs related to disease resistance could help to improve our understanding of disease resistance in turbot and might be beneficial to select disease-resistant strains of aquaculture of turbot. In this study, we used BSA to discover SNP variation within the genic region associated with *V. anguillarum* resistance in a diverse collection of populations. This approach helps the development of a tightly linked-codominant genetic markers. Our results demonstrated the feasibility of NGS data with BSA approach and the extensive information of SNP/gene for disease resistance traits as a rapid and low-cost approach for marker development.

The use of phenotypic extremes, like the resistant and susceptible fish in this study, allowed the identification of significant SNPs between the pools. Recently, using contrasting pools of F2 populations or recombinant inbred lines (RILs), BSA coupled with NGS genotyping pipelines has been applied in plants for development of economic character linked markers [31,32] and identification of disease resistance genes [33]. However, in marine teleost the construction of segregation population usually takes several years for their slow sexual maturation, which also raises the difficulty of obtaining significant SNPs in disease

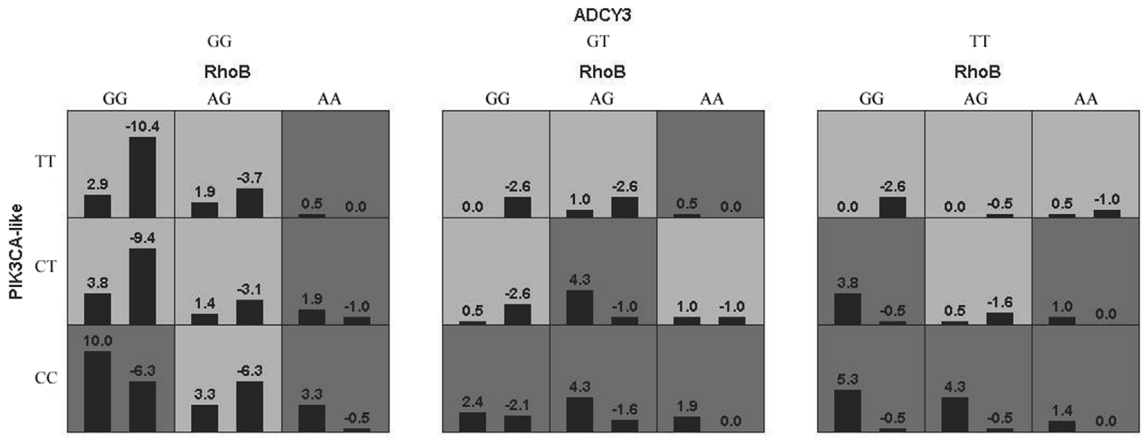

Fig. 3. GMDR analysis results of best three-factor model. Each cell is categorized as high risk or low risk when the sum of scores is above or below zero respectively. High risk cell is marked with dark gray and low risk with light gray. The left and right columns in each cell means individuals being susceptible and resistant, respectively. Genotypes of ADCY3 are GG, GT and TT, genotypes of RhoB are GG, AG and AA, and genotypes of PIK3CA-like are TT, CT and CC.

Table 4
Identifications of the best three-loci model in 192 turbots from stock A, B and C.

	groups	Expectation	Observation	Accuracy (%)	Total accuracy (%)
One-factor model	Susceptible group	93	47	50.5%	51.1%
	Resistant group	97	50	51.5%	
Two-factor model	Susceptible group	135	75	55.6%	53.2%
	Resistant group	55	26	47.3%	
Three-factor model	Susceptible group	92	58	63.0%	70.8%
	Resistant group	93	73	78.5%	

resistance owing to the mortality caused by long term cultivation. In this study, the use of populations without any parental strain or genotype information, together with the pooling of samples from phenotypic extremes, allowed analysis of significant SNPs between the resistant and susceptible fish. Euclidean distance equation was utilized to ignore the issues of parental orientation when calculate the allele frequencies. Our results also validated the feasibility of this method in natural stocks.

4.1. Identification and verification of SNPs for V. anguillarum resistance

SNPs are widely-distributed mutations in genomes. Based on the different positions in genomes, SNPs influence the gene translation or transcription through a variety of mechanisms [34]. Previous studies have primarily focused on nsSNPs as those SNPs might directly influence the protein structure and activity. However, Chen et al. [35] reported that sSNPs also have effects of human disease association similar to nsSNPs in GWAS. SNPs located in 3′ or 5′ UTR regions as well as 3′ or 5' flanking regions also feature in the changes in mRNA binding sites [36]. Therefore, we selected SNPs located in exons, UTRs and flanking regions for further association analysis. In recent years, increasing number of studies have found that SNPs in intron and intergenic also play a decision role in the changes of characters. A SNP in the third intron of the F-box and leucine rich repeat protein 17 (FBXL17) gene explains 58.4% of the phenotypic variation in sex reversal of Chinese tongue sole *Cynoglossus semilaevis* [37]. Hence, the *V. anguillarum* resistance association analysis of SNPs in intron and intergenic regions remain to be completed in the future.

In the present study, a total of 370,502, 450,897 and 487,465 SNPs were identified from WGS in three stocks respectively, and 73 SNPs located in the immune related genes were chosen as candidate SNPs for *V. anguillarum* resistance analyses. We found a locus on gene RhoB (5' flanking region) that was associated with the *V. anguillarum* resistance both in Euclidean distance model and *P*-value. Because that it lies in a region that is likely to affect promoter activity through the combination between mRNA and transcription factors, we hypothesized this SNP yields allelic differences in RhoB transcription [36]. LPS mediated activation of TRIF-GEFH1-RhoB pathway in dendritic cells could regulate the activity of $CD4^+$ T cells [38].

Compared to non-synonymous mutations, more sSNPs were recognized in immune effectors showing association with *V. anguillarum* resistance. For instance, individuals with allele C at locus 92,177 of PI3KCA-like showed resistance to *V. anguillarum*. Although synonymous mutation does not cause changes in amino acid sequences, it is more likely to have effects on linked non-synonymous mutation nearby with strong linkage disequilibrium which affects the expression or functions of immune genes. This kind of synonymous allele-specific differences might also directly affect the splicing, transcription and translation in mRNA folding [39]. Since the expression level of PIK3CA-like is higher in individuals with CC genotype (Fig. 1), it is plausible that the C allele of locus 92,177 is linked to another locus that might increases PIK3CA-like expression.

4.2. Mining of favorable SNP alleles and candidate genes for V. anguillarum resistance

Localization of the significant SNPs along the scaffolds should allow the identification of positional candidate genes responsible for the *V. anguillarum* resistance. A number of genes involved in the PI3K/Akt/mTOR and NF-kB pathways were found to be within the significantly associated genetic regions, suggesting the involvement of these two pathways in the resistance against *V. anguillarum*. Compared to one gene, mutations in gene pathways could explain more phenotypic variance for certain traits [16,40]. Of the significantly associated genes found in our analysis PIK3CA-like and CYLD were involved in PI3K/Akt/mTOR pathway while RhoB and VCAM-1 were in NF-kB pathway. In addition, many other genes involved in these two pathways were also identified in our SNP screening (Fig. 4), which further increased the likelihood that PI3K/Akt/mTOR pathway and NF-kB pathway may be the main participated pathways in the resistance of *V. anguillarum* infection.

PI3K/Akt/mTOR pathway was reported to be involved in immunity and resistance [41]. Previously studies have shown that PI3K/Akt/mTOR signaling pathway is take part in many cellular activities and biological processes such as apoptosis and autophagy. PI3K/Akt/mTOR signaling pathway also participate in the process of survival and reproduction of pathogens in cells by regulating the cytoskeleton dynamics, membrane permeability and pseudopod extension that could be used by some pathogenic bacteria to entry into host cell. Ireton et al. [42] elucidated that the protein InlB from *Listeria monocytogenes* will causes rapid increases in cellular amounts of PI3 kinases as an agonist of mammalian PI3K. It was also reported that PI3K could be activated by *Porphyromonas gingivalis*, which blocks phagocytosis and promotes inflammation [43]. In this study, a PI3K family subunit PI3KCA-like was identified as significant association gene involved in the immune response of *V. anguillarum* infection. Individuals with different genotypes showed significantly different expression levels ($P < 0.05$) (Fig. 1). The deubiquitinating enzyme CYLD was another significantly associated candidate gene found in this study that is involved in PI3K/

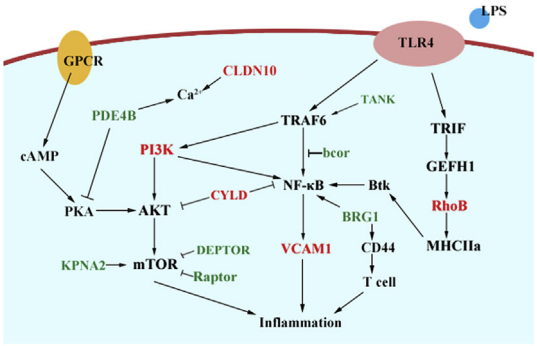

Fig. 4. Signal transduction pathways involving PI3Ks and the other candidate genes. The candidate genes significant in both genotypes and alleles in confirmation families are in red. The other candidate genes screen out from sequencing bulks are in green. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Akt/mTOR signaling pathway. It cleaves lysine 63-linked polyubiquitin chains from specific substrates, containing tumor necrosis factor receptor-associated factors (TRAF)-6, B cell lymphoma 3 (BCL3), nuclear factor kappa B essential modulator (NEMO) and so on, and negatively regulate PI3K, NF-κB and MAPKs pathways [44]. CYLD could inhibit the activation of NF-κB and Akt to suppress inflammation and the innate immune response [45]. With the deubiquitinase activity, CYLD was also demonstrated to function in TNF-induced cell apoptosis and necroptosis [46].

In leukocytes and vascular smooth muscle cells, VCAM1 is an important adhesion molecule that plays a pivotal role in mediating the adhesion and migration [47,48]. Mestre [49] reported that VCAM1 inhibition could reduce leukocyte transmigration. Previous evidence showed that PI3K, p38MAPK and NF-κB pathways regulate VCAM-1 through Platelet-derived growth factor BB (PDGFBB) [50].

Nuclear factor-κB plays a key role in activation of proinflammatory gene expression after diverse inflammatory stimuli and environmental stressors [51]. Researchers demonstrated that ubiquitination can regulate phosphorylation by activating NF-κB through a proteasome-independent mechanism [52]. Taken together, genes involved in PI3K pathway and NF-κB pathway, including PI3 kinases and NF-κB themselves, have great effects on *V. anguillarum* resistance. The most interesting finding in our study was that genes near the most significant SNPs were functionally related. Although positioned in different scaffolds, they were also considered as “functional hubs”. As we know, genes that perform the same function often have similar expression patterns [53]. Geng et al. [1] reported that genes both in neighbored and functionally related are much more likely to be co-expressed than expected by chance. In the current study, the four hub genes, PI3KCA-like, RhoB, RhoGEF and VCAM1, exhibited a significant expression elevation in individuals with resistant mutant (MT) homozygous alleles. The results observed in CYLD was opposite where the expression level of individuals with resistant allele was obviously lower due to its ubiquitination inhibition regulation. Our results were coincided with the previous reports.

4.3. Multigene interactions for MAS

Multifactor dimensionality reduction (MDR) was the leading method for studying gene or SNP interaction networks for moderate sample size data [54]. Advanced versions of MDR, such as OR-MDR, GMDR, MB-MDR, have been created for different purposes. Wang [13] elucidated that SNP-SNP combination from the IGF1/FoxO signal transduction pathway was associated with economical traits in pigs with GMDR approach. For the first time the GMDR method was applied to association analysis in turbot resistant against *V. anguillarum* infection. A three-factor model consisted of loci from ADCY3, RhoB and PIK3CA-like was proved to have significant association with vibrio resistance and higher precision rate in verification populations. According to our GMDR analysis results, the expected resistant and susceptible group could be divided based on the genotyping result of these three loci. After combined with the phenotype of these fish, 70.8% individuals could be identified with this three-factor model. To date, no practical application of such approach has been reported in marine animals. When we compare the expression level of ADCY3, RhoB and PIK3CA-like in susceptible group and resistant group, we found that these three genes showed higher expression levels in resistant group than in the susceptible group (Fig. 2). All these data demonstrated that the three-factor model could be used as a high-efficiency marker in future selective breeding of turbot.

5. Conclusions

This study presents the BSA approach that employs the high-throughput whole genome re-sequencing information for the study of *V. anguillarum* resistance in turbot. A total of 7 SNPs in these immune related genes were detected to be associated with resistance in both sequencing and verification populations. A three-factor model were identified and confirmed through SNP-SNP interaction analysis which showed a high accuracy in identification of disease-resistant individuals. All this fact demonstrated that the genetic analysis with the application of BSA-seq has a propelling effect in disease resistance and will assist the selective breeding of turbot with *V. anguillarum* resistance.

Acknowledgments

This study was supported by The Scientific and Technological Innovation Project from Laboratory for Marine Fisheries and Aquaculture, Qingdao National Laboratory for Marine Science and Technology (2015ASKJ02).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fsi.2019.02.041.

Conflicts of interest

The authors declare no conflict of interest.

References

[1] X. Geng, J. Sha, S. Liu, L. Bao, J. Zhang, R. Wang, et al., A genome-wide association study in catfish reveals the presence of functional hubs of related genes within QTLs for columnaris disease resistance, BMC Genomics 16 (2015) 196.
[2] X. Zhong, X. Wang, T. Zhou, Y. Jin, S. Tan, C. Jiang, et al., Genome-wide association study reveals multiple novel QTL associated with low oxygen tolerance in hybrid catfish, Mar. Biotechnol. 19 (4) (2017) 379–390.
[3] J. Ye, X. Wang, T. Hu, F. Zhang, B. Wang, C. Li, et al., An InDel in the promoter of Al-activated MALATE TRANSPORTER9 selected during tomato domestication determines fruit malate contents and aluminum tolerance, Plant Cell 29 (9) (2017) 2249–2268.
[4] J. Waage, M. Standl, J.A. Curtin, L.E. Jessen, J. Thorsen, C. Tian, et al., Genome-wide association and HLA fine-mapping studies identify risk loci and genetic pathways underlying allergic rhinitis, Nat. Genet. 50 (8) (2018) 1072–1080.
[5] A. Visioni, S. Gyawali, R. Selvakumar, O.P. Gangwar, P.S. Shekhawat, S.C. Bhardwaj, et al., Genome wide association mapping of seedling and adult plant resistance to barley stripe rust (*Puccinia striiformis f. sp. hordei*) in India, Front. Plant Sci. 9 (2018) 520.
[6] L. Wang, P. Liu, S. Huang, B. Ye, E. Chua, Z.Y. Wan, et al., Genome-wide association study identifies loci associated with resistance to viral nervous necrosis disease in asian seabass, Mar. Biotechnol. 19 (3) (2017) 255–265.
[7] R. Patnala, J. Clements, J. Batra, Candidate gene association studies: a comprehensive guide to useful in silico tools, BMC Genet. 14 (2013) 39-39.
[8] C. Shao, Y. Niu, P. Rastas, Y. Liu, Z. Xie, H. Li, et al., Genome-wide SNP identification for the construction of a high-resolution genetic map of Japanese flounder (*Paralichthys olivaceus*): applications to QTL mapping of Vibrio anguillarum disease resistance and comparative genomic analysis, DNA Res. 22 (2) (2015) 161–170.
[9] L. Wang, C. Fan, Y. Liu, Y. Zhang, S. Liu, D. Sun, et al., A genome scan for quantitative trait loci associated with Vibrio anguillarum infection resistance in Japanese flounder (*Paralichthys olivaceus*) by bulked segregant analysis, Mar. Biotechnol. 16 (5) (2014) 513–521.
[10] H. Takagi, A. Abe, K. Yoshida, S. Kosugi, S. Natsume, C. Mitsuoka, et al., QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations, Plant J. 74 (1) (2013) 174–183.
[11] R. Wang, L. Sun, L. Bao, J. Zhang, Y. Jiang, J. Yao, et al., Bulk segregant RNA-seq reveals expression and positional candidate genes and allele-specific expression for disease resistance against enteric septicemia of catfish, BMC Genomics 14 (2013) 929-929.
[12] R. Braun, K. Buetow, Pathways of distinction analysis: a new technique for multi-SNP analysis of GWAS data, PLoS Genet. 7 (6) (2011) e1002101.
[13] M. Wang, Q. Wang, Y. Pan, From QTL to QTN: candidate gene set approach and a case study in porcine IGF1-FoxO pathway, PLoS One 8 (1) (2013) e53452.
[14] T. He, P.S. Zhong, Y. Cui, A set-based association test identifies sex-specific gene sets associated with type 2 diabetes, Front. Genet. 5 (2014) 395.
[15] Y. Yu, Y. Pan, M. Jin, M. Zhang, S. Zhang, Q. Li, et al., Association of genetic variants in tachykinins pathway genes with colorectal cancer risk, Int. J. Colorectal Dis. 27 (11) (2012) 1429–1436.
[16] L.-A. Raven, B.G. Cocks, J.E. Pryce, J.J. Cottrell, B.J. Hayes, Genes of the RNASE5 pathway contain SNP associated with milk production traits in dairy cattle, Genet. Sel. Evol. : GSE 45 (1) (2013) 25-25.
[17] W. Wang, Y. Hu, Y. Ma, L. Xu, J. Guan, J. Kong, High-density genetic linkage

mapping in turbot (*Scophthalmus maximus L.*) based on SNP markers and major sex- and growth-related regions detection, PLoS One 10 (3) (2015) e0120410.
[18] X. Zhang, S. Wang, S. Chen, Y. Chen, Y. Liu, C. Shao, et al., Transcriptome analysis revealed changes of multiple genes involved in immunity in Cynoglossus semilaevis during Vibrio anguillarum infection, Fish Shellfish Immunol. 43 (1) (2015) 209–218.
[19] D. Tan, L. Gram, M. Middelboe, Vibriophages and their interactions with the fish pathogen Vibrio anguillarum, Appl. Environ. Microbiol. 80 (10) (2014) 3128–3140.
[20] T.M.J. Sambrook, E.F. Fritsch, Mol. Clon. Lab. Man. 33 (1) (1982).
[21] A. Figueras, D. Robledo, A. Corvelo, M. Hermida, P. Pereiro, J.A. Rubiolo, et al., Whole genome sequencing of turbot (*Scophthalmus maximus; Pleuronectiformes*): a fish adapted to demersal life, DNA Res. 23 (3) (2016) 181–192.
[22] H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform, Bioinformatics 25 (14) (2009) 1754–1760.
[23] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, et al., The sequence alignment/map format and SAMtools, Bioinformatics 25 (16) (2009) 2078–2079.
[24] A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, et al., The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data, Genome Res. 20 (9) (2010) 1297–1303.
[25] Z. Zhou, Y. Jiang, Z. Wang, Z. Gou, J. Lyu, W. Li, et al., Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean, Nat. Biotechnol. 33 (4) (2015) 408–414.
[26] K. Wang, M. Li, H. Hakonarson, ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data, Nucleic Acids Res. 38 (16) (2010) e164.
[27] M.M. Deza, E. Deza, Encyclopedia of Distances, (2009).
[28] A. Abe, S. Kosugi, K. Yoshida, S. Natsume, H. Takagi, H. Kanzaki, et al., Genome sequencing reveals agronomically important loci in rice using MutMap, Nat. Biotechnol. 30 (2) (2012) 174–178.
[29] X.Y. Lou, G.B. Chen, L. Yan, J.Z. Ma, J. Zhu, R.C. Elston, et al., A generalized combinatorial approach for detecting gene-by-gene and gene-by-environment interactions with application to nicotine dependence, Am. J. Hum. Genet. 80 (6) (2007) 1125–1137.
[30] K. Correa, J.P. Lhorente, M.E. Lopez, L. Bassini, S. Naswa, N. Deeb, et al., Genome-wide association analysis reveals loci associated with resistance against Piscirickettsia salmonis in two Atlantic salmon (*Salmo salar L.*) chromosomes, BMC Genomics 16 (2015) 854.
[31] J. Song, Z. Li, Z. Liu, Y. Guo, L.J. Qiu, Next-generation sequencing from bulked-segregant analysis accelerates the simultaneous identification of two qualitative genes in soybean, Front. Plant Sci. 8 (2017) 919.
[32] H. Huo, I.M. Henry, E.R. Coppoolse, M. Verhoef-Post, J.W. Schut, H. de Rooij, et al., Rapid identification of lettuce seed germination mutants by bulked segregant analysis and whole genome sequencing, Plant J. 88 (3) (2016) 345–360.
[33] H. Yang, J. Jian, X. Li, D. Renshaw, J. Clements, M.W. Sweetingham, et al., Application of whole genome re-sequencing data in the development of diagnostic DNA markers tightly linked to a disease-resistance locus for marker-assisted selection in lupin (*Lupinus angustifolius*), BMC Genomics 16 (1) (2015) 660.
[34] J. Ragoussis, Genotyping technologies for genetic research, Annu. Rev. Genom. Hum. Genet. 10 (1) (2009) 117–133.
[35] R. Chen, E.V. Davydov, M. Sirota, A.J. Butte, Non-Synonymous and Synonymous Coding SNPs Show Similar Likelihood and Effect Size of Human Disease Association, PLoS One 5 (10) (2010) e13574.
[36] Y. Miyamoto, A. Mabuchi, D. Shi, T. Kubo, Y. Takatori, S. Saito, et al., A functional polymorphism in the 5 ' UTR of GDF5 is associated with susceptibility to osteoarthritis, Nat. Genet. 39 (4) (2007) 529.
[37] L. Jiang, H.D. Li, Single locus maintains large variation of sex reversal in half-smooth tongue sole (*Cynoglossus semilaevis*), G3-Genes Genomes Genet. 7 (2) (2017) 583–589.
[38] H. Kamon, T. Kawabe, H. Kitamura, J. Lee, D. Kamimura, T. Kaisho, et al., TRIF-GEFH1-RhoB pathway is involved in MHCII expression on dendritic cells that is critical for CD4 T-cell activation, EMBO J. 25 (17) (2006) 4108–4119.
[39] A.A. Komar, SNPs, silent but not invisible, Science 315 (5811) (2007) 466–467.
[40] M. Peukert, S. Weise, M.S. Röder, I.E. Matthies, Development of SNP markers for genes of the phenylpropanoid pathway and their association to kernel and malting traits in barley, BMC Genet. 14 (2013) 97-97.
[41] S. Koyasu, The role of PI3K in immune cells, Nat. Immunol. 4 (2003) 313.
[42] K. Ireton, B. Payrastre, P. Cossart, The Listeria monocytogenes protein InlB is an agonist of mammalian phosphoinositide 3-kinase, J. Biol. Chem. 274 (24) (1999) 17025–17032.
[43] T. Maekawa, J.L. Krauss, T. Abe, R. Jotwani, M. Triantafilou, K. Triantafilou, et al., Porphyromonas gingivalis manipulates complement and TLR signaling to uncouple bacterial clearance from inflammation and promote dysbiosis, Cell Host Microbe 15 (6) (2014) 768–778.
[44] K. Wex, U. Schmid, S. Just, X. Wang, R. Wurm, M. Naumann, et al., Receptor-Interacting protein kinase-2 inhibition by CYLD impairs antibacterial immune responses in macrophages, 6 (2016) 650.
[45] S.C. Sun, CYLD: a tumor suppressor deubiquitinase regulating NF-κB activation and diverse biological processes, Cell Death Differ. 17 (1) (2010) 25.
[46] J.I. Hitomi, D.E. Christofferson, A. Ng, J.H. Yao, A. Degterev, R.J. Xavier, et al., Identification of a molecular signaling network that regulates a cellular necrotic cell death pathway, Cell 135 (7) (2008) 1311–1323.
[47] Y.M. Hyun, H.L. Chung, J.L. Mcgrath, R.E. Waugh, M.J.J.o.I. Kim, Activated integrin VLA-4 localizes to the lamellipodia and mediates T cell migration on, VCAM-1 183 (1) (2009) 359–369.
[48] X. Wang, G.Z. Feuerstein, J.L. Gu, P.G. Lysko, T.L. Yue, Interleukin-1 beta induces expression of adhesion molecules in human vascular smooth muscle cells and enhances adhesion of leukocytes to smooth muscle cells, Atherosclerosis 115 (1) (1995) 89–98.
[49] L. Mestre, P.M. Iñigo, M. Mecha, F.G. Correa, M. Hernangómez-Herrero, F. Loría, et al., Anandamide inhibits Theiler's virus induced VCAM-1 in brain endothelial cells and reduces leukocyte transmigration in a model of blood brain barrier by activation of CB(1) receptors, 8 (1) (2011) 102.
[50] Y. Hu, P. Cheng, J.C. Ma, Y.X. Xue, Y.H.J.O.R. Liu, Platelet-derived growth factor BB mediates the glioma-induced migration of bone marrow-derived mesenchymal stem cells by promoting the expression of vascular cell adhesion molecule-1 through the PI3K, P38 MAPK and NF-κB pathways, 30 (6) (2013) 2755–2764.
[51] B.K. De, B.W. Vanden, G.J.O. Haegeman, Cross-talk between nuclear receptors and nuclear factor kappaB, 25 (51) (2010) 6868.
[52] Z.J. Chen, Ubiquitin signalling in the NF-κB pathway, 7 (8) (2005) 758–765.
[53] E.J. Williams, D.J. Bowles, Coexpression of neighboring genes in the genome of Arabidopsis thaliana, Genome Res. 14 (6) (2004) 1060.
[54] M.D. Ritchie, L.W. Hahn, N. Roodi, L.R. Bailey, W.D. Dupont, F.F. Parl, et al., Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer, Am. J. Hum. Genet. 69 (1) (2001) 138–147.