

RESEARCH

Open Access

Selection of functional *EPHB2* genotypes from ENU mutated grass carp treated with GCRV



Meher un Nissa^{1,2,3†}, Zhu-Xiang Jiang^{1,2,3†}, Guo-Dong Zheng^{1,2,3*} and Shu-Ming Zou^{1,2,3*}

Abstract

Background: *N*-ethyl-*N*-nitrosourea (ENU) mutagenesis is a useful method for the genetic engineering of plants, and the production of functional mutants in animal models including mice and zebrafish. Grass carp reovirus (GCRV) is a haemorrhagic disease of grass carp which has caused noteworthy losses in fingerlings over the last few years. To overcome this problem, we used ENU mutant grass carp in an attempt to identify functional resistance genes for future hereditary rearing projects in grass carp.

Results: This study used ENU-mutated grass carp to identify genetic markers associated with resistance to the haemorrhagic disease caused by GCRV. Bulked segregant analysis (BSA) was performed on two homozygous gynogenetic ENU grass carp groups who were susceptible or resistant to GCRV. This analysis identified 466,162 SNPs and 197,644 InDels within the genomes of these mixed pools with a total of 170 genes annotated in the associated region, including 49 genes with non-synonymous mutations at SNP sites and 25 genes with frame shift mutations at InDel sites. Of these 170 mutated genes, 5 randomly selected immune-related genes were shown to be more strongly expressed in the resistant group as compared to the susceptible animals. In addition, we found that one immune-related gene, *EPHB2*, presented with two heterozygous SNP mutations which altered the animal's responded to GCRV disease. These SNPs were found in the intron region of *EPHB2* at positions 5859 (5859^{G>A}) and 5968 (5968^{G>A}) and were significantly ($p = 0.002, 0.003$) associated with resistance to GCRV. These SNP sites were also shown to correlate with the GCRV-resistant phenotype in these ENU grass carp. We also evaluated the mortality of the different ENU fish genotypes in response to GCRV and the SNPs in *EPHB2*. The outcomes of these evaluations will be useful in future selections of GCRV-resistant genes for genetic breeding in grass carp.

Conclusion: Our results provide a proof of concept for the application of BSA-sequence analysis in detecting genes responsible for specific functional genotypes and may help to develop better methods for marker-assisted selection, especially for disease resistance in response to GCRV.

Keywords: BSA, Mutant grass carp (ENU), GCRV, qPCR, SNP

* Correspondence: gdzheng@shou.edu.cn; smzou@shou.edu.cn

†Meher un Nissa and Zhu-Xiang Jiang contributed equally to this work.

¹Genetics and Breeding Center for Blunt Snout Bream, Ministry of Agriculture, Shanghai 201306, China

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Genetic breeding of aquaculture fish species primarily depends on the identification of naturally occurring mutants with high performance, and subsequent hybridisation or marker-assisted breeding to produce better-quality strains [1]. These methods have been used to produce more than 200 improved strains of aquaculture fish species in China, with 139 produced using genetic selection. Therefore, chemical mutagenesis is an effective way to make new mutants for upcoming genetic development in aquaculture species [2]. *N*-ethyl-*N*-nitrosourea (ENU) is a chemical mutagen that acts as an alkylating agent, exchanging its ethyl group to nucleophilic nitrogen or oxygen locales on the deoxyribonucleotides, leading to base inconsistencies during DNA replication [3, 4]. ENU mutagenesis has been demonstrated to be effective for inciting point mutations in the genome of the grass carp [5].

Grass carp reovirus (GCRV) belongs to the Reoviridae family genus Aquareovirus. The first instance of haemorrhagic disease in grass carp was reported at a fish farm in Hubei Province in 1972 [6]. GCRV was categorised as an aqua reovirus and in 1984 its genome was shown to include 11 sections of double stranded RNA. The Aquareovirus genus can be separated into groups A through G (AQRV-A to AQRV-G), with GCRV falling into the AQRV-C group [7]. GCRV infection results in haemorrhagic disease in grass carp and causes noteworthy losses in fingerlings, with recent outbreaks exhibiting a significant economic impact.

A genotyping-by-sequencing method, permits genome widespread association studies, bulked segregant analysis (BSA) and genomic selection and has previously been applied to various animal breeding programmes. Recent reductions in the cost of genome sequencing have opened up the opportunity for whole genome sequencing and resequencing in larger pools of individuals. BSA has been suggested as an effective tool for quickly identifying markers connected to specific characteristics of interest including those associated with disease

resistance [8]. This approach includes segregating the F2 population produced from a starting cross of two phenotypically different parents, which are then scored for the phenotype of interest. Bulk DNA or RNA tests are built from organisms with differentiating phenotypes. BSA has been primarily used in the development and identification of crop species as it facilitates the recognisable identification of large impact QTLs, such as disease resistance genes or for mapping subjective mutations [8–11].

The objective of the current study was to select the functional genotypes from ENU fish facilitating GCRV resistance using BSA sequence analysis. Our study also provides a useful method for carrying out marker-assisted selection, especially for disease resistance in response to GCRV.

Results

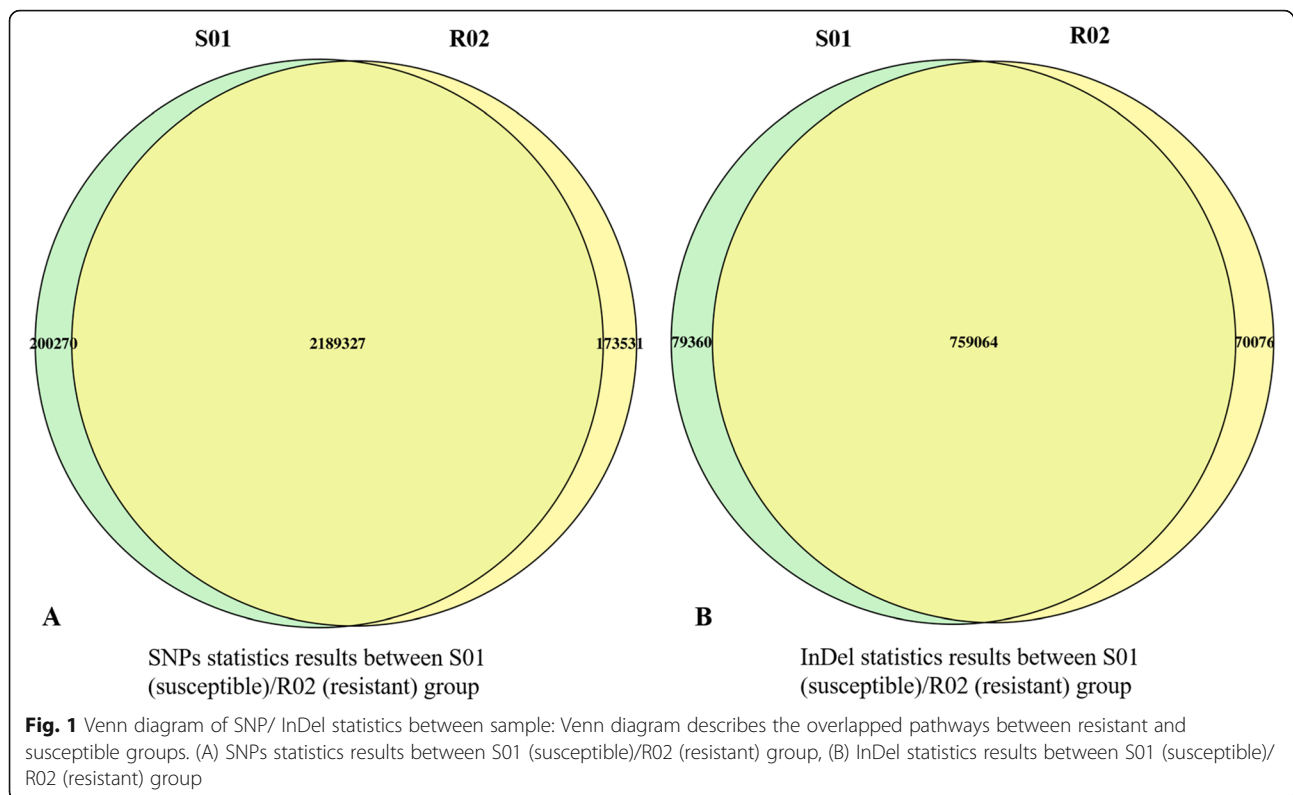
SNP and InDel annotation

After filtering our sequencing analysis produced 73.47 Gbp of clean read data with an average Q30 of 93.92%, and an average GC% of up to 38.11% (Table 1). The average genome coverage depth in each sample was 41.00X, and the genome coverage was about 99.11% (at least 1×) (Table 1), indicating the high quality of the sequencing data. The assembled genome size of grass carp is 900.50 Mb, GC content is 37.42% and the genome is currently annotated at Scaffold level [12]. The Venn diagram reveals that there are a different number of SNP and InDels in the S01(susceptible) and R02 (resistant) groups (Fig. 1). There were approximately 200,270 SNPs in the S01 and 173,531 SNPs in the R02 groups. The overlapping portion of the Venn diagram represented the total number of SNPs (2,189,327) and InDels (759,064), identified in these assays (Fig. 1).

We detected a total of 466,162, 197,644, 8492, and 3668 polymorphic sites across different regions of the genome using SNP, InDel, and SNP/InDel annotation (Table 2).

Table 1 Summary of Illumina sequencing data

Sample ID	Susceptible group (S01)	Resistant group (R02)	Average values
Clean_reads	125,577,407	117,707,723	73.47Gbp
Q30 (%)	93.52	94.31	93.92%
GC (%)	38.09	38.14	38.11%
Mapped ratio (%)	80.51	80.53	80.52%
Average depth	42	40	41.00X
Coverage_ratio_1× (%)	99.17	99.05	99.11%
Coverage_ratio_5× (%)	97.77	97.57	97.67
Coverage_ratio_10× (%)	96.26	95.97	96.12



These annotation were then used to demonstrate that the SNPs were randomly distributed throughout the genome (Table 2), while the InDel frameshift mutations were found in only 50 genes, 7 of which demonstrated an in codon_insertion and 6 of which had an in codon_deletion (Table 2). Synonymous and non-synonymous codons had a higher number of SNPs than other genes (7251, 9298) (Table 2).

Association analysis (SNP, InDel) and ED correlations

The SNPs and InDels were first filtered leaving 2,313,014 high-quality trusted SNP sites and 775,981 high-quality InDels which were then applied to the association analysis (Table 3). Through this, 1197 Scaffold/contig sequences with significantly enriched associated SNP sites were selected. Some results are shown in the table (Additional file 1). The same analysis method (Euclidean Distance (ED) method) as the SNP association analysis was used, and 85 Scaffold/contig sequences that were significantly enriched in association InDel sites were finally screened (Additional file 2). Take the intersection of the results obtained by these two association analysis methods (SNP & InDel along with ED), a total of 21 Scaffolds related to traits were obtained (Table 4).

Gene annotation and functional analysis

A total of 170 genes were identified within the candidate region, 49 of which had non-synonymous mutations and

25 of which presented with frameshift mutation within the mixed pool evaluations (Table 5, Additional file 3). A total of 55 genes in the genome were annotated and classified into biological processes, cellular components, and molecular functions (Fig. 2, Additional files 4, 5 & 6). An additional 11 genes were found to have non-synonymous mutations and 2 other genes were found to have frameshift mutations when evaluated for GO enrichment. The annotated KEGG databases showed that 29 SNPs were found in the top 20 pathways as shown in Fig. 3 and after multiple-testing corrections, we selected those pathways with Q values of ≤ 0.05 as significantly enriched for these genes (Fig. 4). Our data suggest that these genes may play an important role in the innate immune response to GCRV in ENU grass carp.

Changes in gene expression associated with grass carp reovirus (GCRV) resistance

The relative expression levels of the five genes significantly associated with grass carp reovirus (GCRV) resistance were evaluated using *qPCR* (Fig. 5). We selected five genes with genetic variations and examined their mRNA expression levels in liver, kidney and gill tissues. Each of the five genes were shown to be involved in the inflammatory response, cell proliferation, anti-apoptosis, tumour suppression and the immune response to viral infection pathways. Results showed that mRNA expression levels of SAMD9L,

Table 2 Annotation results statistics of SNPs and indels containing candidate region

Different region and types of Mutation	SNP annotation result statistics	InDel annotation result statistics	SNP annotation results Statistics in candidate regions	InDel annotation results statistics in the candidate regions
	S01vsR02	S01vsR02	S01vsR02	S01vsR02
INTERGENIC	220,527	89,734	4223	1797
INTRON	144,560	27	2538	1242
UPSTREAM	37,653	66,109	618	266
DOWNSTREAM	37,132	16,283	653	232
UTR_5_PRIME	3202	16,315	43	25
UTR_3_PRIME	4074	1945	60	26
SPLICE_SITE_ACCEPTOR	118	2169	1	1
SPLICE_SITE_DONOR	100	94	2	2
SPLICE_SITE_REGION	912	66	20	5
FRAME_SHIFT	–	–	–	50
CODON_INSERTION	–	–	–	7
CODON_DELETION	–	–	–	6
CODON_CHANGE_PLUS_CODON_INSERTION	–	–	–	3
CODON_CHANGE_PLUS_CODON_DELETION	–	–	–	3
START_GAINED	648	401	7	–
START_LOST	5	15	–	–
SYNONYMOUS_START	3	3272	1	–
NON_SYNONYMOUS_START	1	276	–	–
SYNONYMOUS_CODING	7251	271	139	–
NON_SYNONYMOUS_CODING	9298	150	180	–
SYNONYMOUS_STOP	71	277	–	–
STOP_GAINED	449	178	6	2
STOP_LOST	158	62	1	1
Other	0	0	0	0
Total	466,162	197,644	8492	3668

BNIP3L and *EPHB2* were significantly increased in the resistant group when compared to the susceptible group ($p < 0.01$), *APPL2* expression was higher in the liver and kidney when compared to the gill tissues of resistant GCRV resistant fish. Kidney-NLRP12 gene expression was significantly elevated in the resistant group when compared to the susceptible group ($p < 0.01$) (Fig. 5).

Verification of SNPs associated with grass carp reovirus (GCRV) resistance

After validation, we identified an SNP (CI01000190: 1067676 A > G) at the 1,067,676 position and another

SNP (CI01000190: 1067785 A > G) at 1067785 in chromosome CI01000190 (Table 6). Both SNPs were located in the intron region of *EPHB2* as shown in Fig. 6. SNPs in *EPHB2* were found in two positions as described above and were shown to encode a 5859^{G>A} and 5968^{G>A} mutation in the S01/R02 groups.

About 300 ENU fish were used for amplification and sequencing of SNP sites across the *EPHB2* gene. The confirmed *EPHB2* SNPs are summarised in Table 7 for both the resistant (R02) and susceptible (S01) groups. In the case of the first SNP, the allele frequencies of the A and G in surviving fish (resistant) were 24 and 76%,

Table 3 SNP and InDel filtering statistics

Filtering statistics	Total SNP	Total InDel	Multiple allele loci	Read support for sites less than 4	Loci with consistent pool	High quality number after filtration
SNP	2,563,128	–	2212	40,627	207,275	2,313,014
InDel	–	908,500	36,042	37,050	59,427	775,981

Table 4 Candidate Scaffold results

Scaffold	AllSNP	AssoSNP_ED	AssolnDel_ED
CI01000352	2484	306	85
CI01112186	5	3	4
CI01000136	2141	185	73
CI01000184	1334	121	51
CI01000190	2631	214	83
CI01000087	2767	286	84
CI01000240	274	50	18
CI01061811	5	2	3
CI01072320	5	7	3
CI01163490	5	1	3
CI01064712	6	2	3
CI01085134	23	10	5
CI01075334	3	4	2
CI01141415	3	3	2
CI01000257	280	30	16
CI01079816	4	2	2
CI01114229	4	2	2
CI01017852	12	4	3
CI01026461	5	3	2
CI01114140	5	3	2
CI01000262	112	33	8

respectively, while in the dead fish (susceptible), the allele frequencies of A and G were 66 and 34%, respectively. At the other SNP, the allele frequencies of A and G in the surviving fish were 18 and 82% whereas in the dead fish they were 62 and 38%, respectively (Table 7). The chi-squared test showed that the allele frequencies were significantly different between the dead and surviving animals ($p = 0.002$ and $p = 0.003$ for respective SNPs) (Table 7), suggesting that *EPHB2* is closely associated with GCRV resistance.

This SNP sequencing data was then used to place the ENU fish into four groups according to their genotype (Fig. 7): Group I 5859^G 5968^G, Group II 5859^A 5968^G,

Group III 5859^G 5968^A and Group IV 5859^A 5968^A (Fig. 7). We then evaluated the correlation between each genotype and GCRV resistance. The results show that resistance was related to genotype. The mortality rate in Groups II and III were 77.4 and 74.5%, which were lower than Group I. However, the mortality rate in Group IV was the lowest (64.3%) and significantly lower than that in Group I. Our previous research suggests that wild-type grass carp mortality is 81.68% (SNP screening showed that the genotype of the wild-type grass carp was 5859^G 5968^G) which indicates that there was a similar resistance in the wild-type and Group I grass carp subjects.

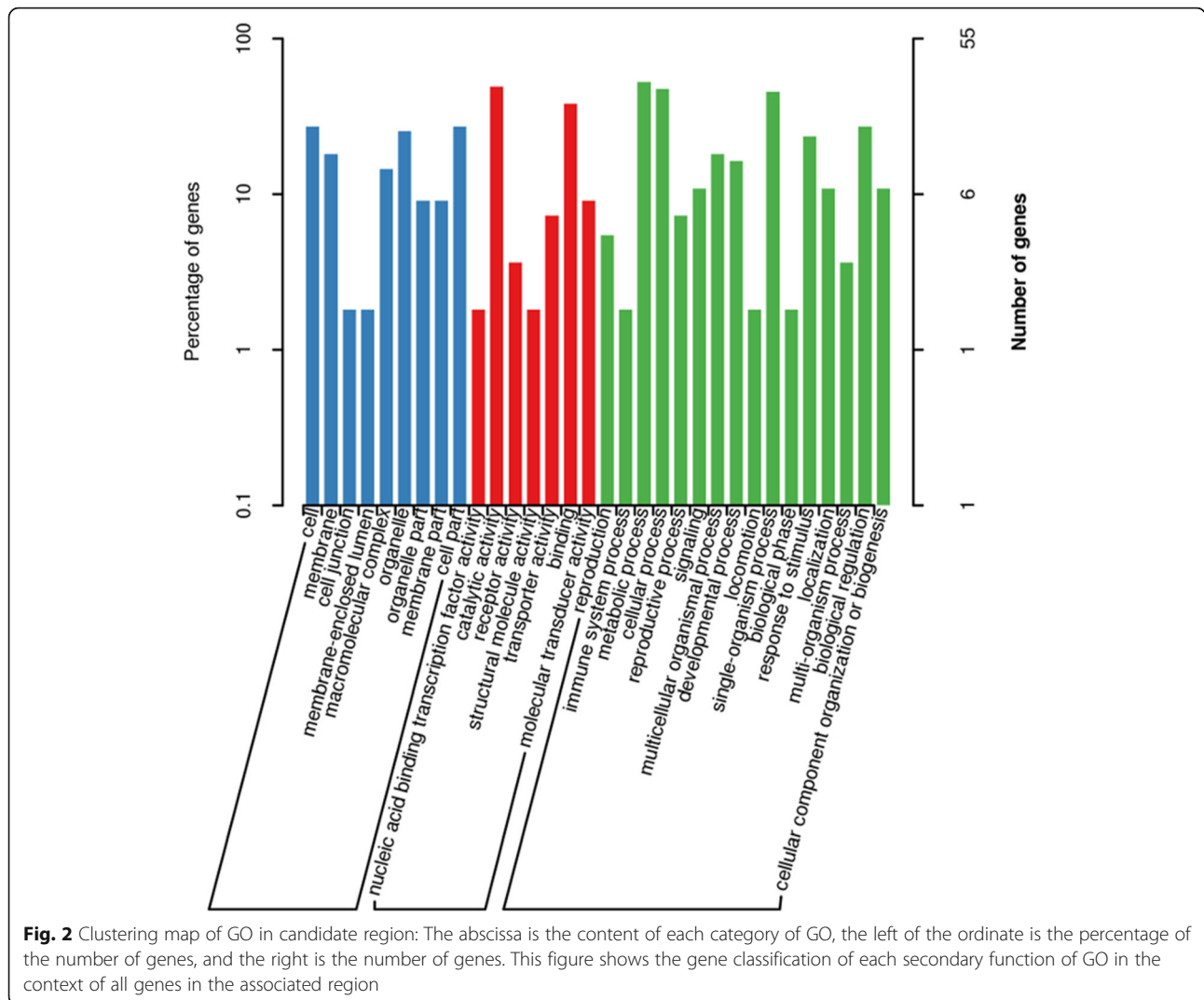
Discussion

Association analysis remains the key method for mining disease resistance SNPs in marine animals. Previous work has reported that SNP frequencies differ between different populations in numerous species [13–15]. However, to the best of our knowledge, there is limited information related to immune associated SNPs and their impact of GCRV resistance in grass carp. Therefore, the currently study focused on the selection of functional genes and their SNPs in relation to GCRV resistance in ENU grass carp, in the hope of improving our understanding of GCRV resistance and establishing a method to facilitate the selection of disease-resistant strains of ENU grass carp in culture.

Cultured grass carp are highly vulnerable to diseases which lead to yield reduction, and while antibiotics and drugs have been used to control the problem the regular use of these alternatives has various unwanted side effects on both the animals and the environment. Therefore, disease-resistant grass carp are highly desirable as they can greatly increase fish yield [5]. Due to the low occurrence of natural mutations in grass carp, chemical mutagenesis can be useful for increasing genetic mutation rates [2]. In this study, homozygous gynogenetic ENU grass carp, which have the advantage of strong disease resistance and rapid growth, were used to

Table 5 Statistics of gene function annotation results in SNP and InDel in candidate regions

Annotated_databases	Gene Num	Non_Syn Gene Num	FRAME_SHIFT Gene Num
NR	159	46	25
NT	170	49	25
trEMBL	170	49	25
SwissProt	104	30	19
GO	55	11	2
KEGG	69	13	7
COG	35	11	6
Total	170	49	25



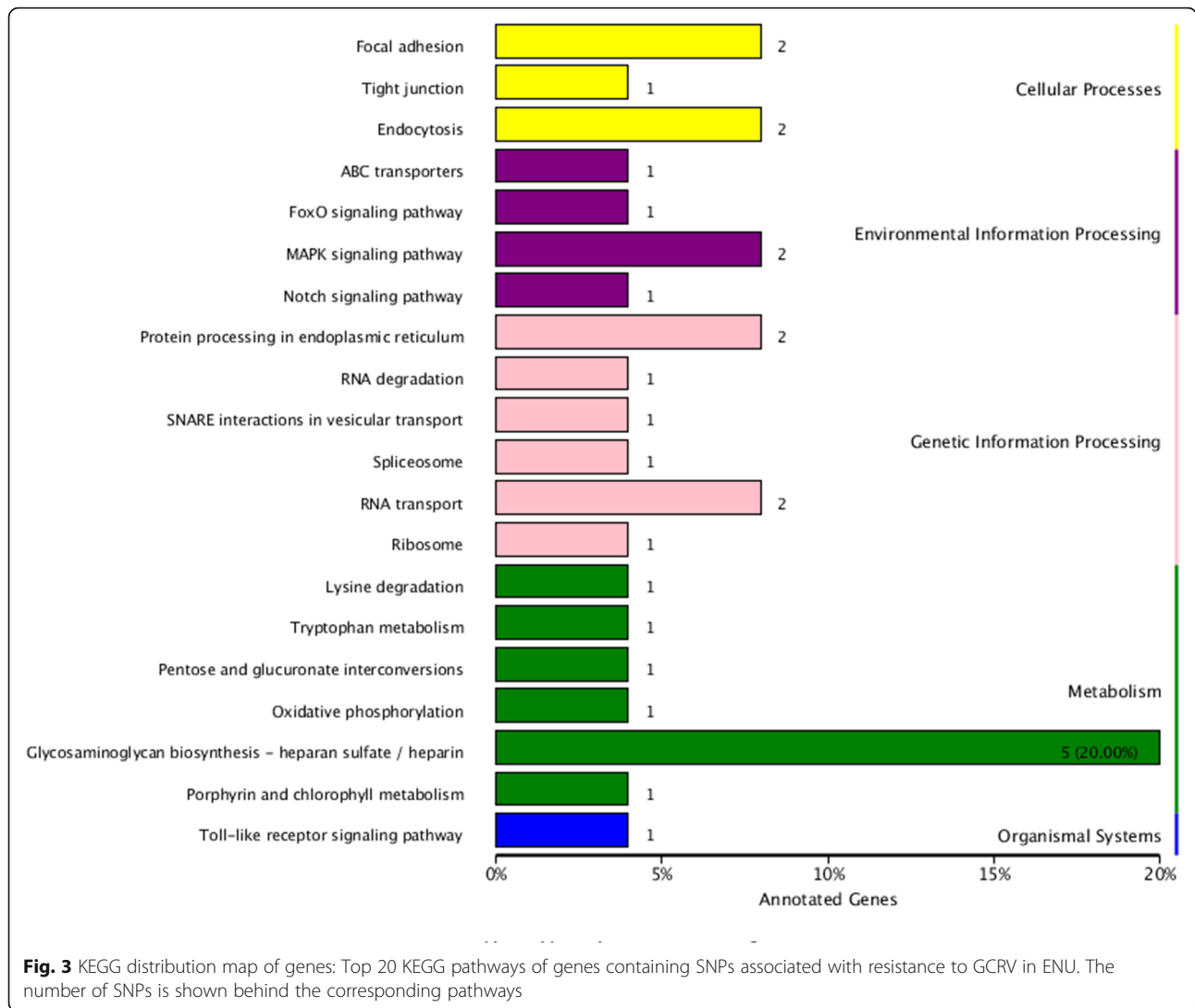
determine the functional genotypes associated with GCRV resistance using BSA sequence analysis.

We used BSA technology to determine SNP mutations within the genomic regions associated with GCRV resistance in homozygous gynogenetic ENU grass carp. Our findings verified the viability of using a BSA approach and the extensive information available for SNP typing for disease resistance traits as a fast and affordable method for marker development. Zhang et al (2019) also demonstrated that genetic analysis using BSA-sequencing is effective for accelerating the identification of disease resistance markers and will assist the selective breeding of turbot resistant to *Vibrio anguillarum* [16].

SNP and InDels are used for annotating mutations and predicting the effects of these mutations [17]. SNP mutation is a significant source of genetic diversity facilitating molecular evolution and disease resistance. Some researchers focus on non-synonymous coding SNPs, because those SNPs might directly influence protein

activity [18]. Wang et al (2015) reported the significant differences in both SNP and InDel rates in resistant and susceptible *C. idella* genomes. However, the SNPs and InDels associated with resistance/susceptibility to GCRV were not described in this omics sequencing project [12]. SNPs can provide innovative resources for genome sequence modification and facilitate the study of selective breeding [19]. This study identified a total of 466,162 SNPs and 197,644 InDel in the resistant/susceptible group, but only 9298 SNPs caused non-synonymous mutations.

In current study, there are indication of several resistance genes in the S01/R02 group of ENU grass carp. These were then evaluated by GO classification and KEGG pathway analysis to identify the molecular functions of the candidate genes in the resistant group. The correlation of differentially expressed sequences with the whole sequences of equivalent GO groups or KEGG pathways was observed as the key measure of

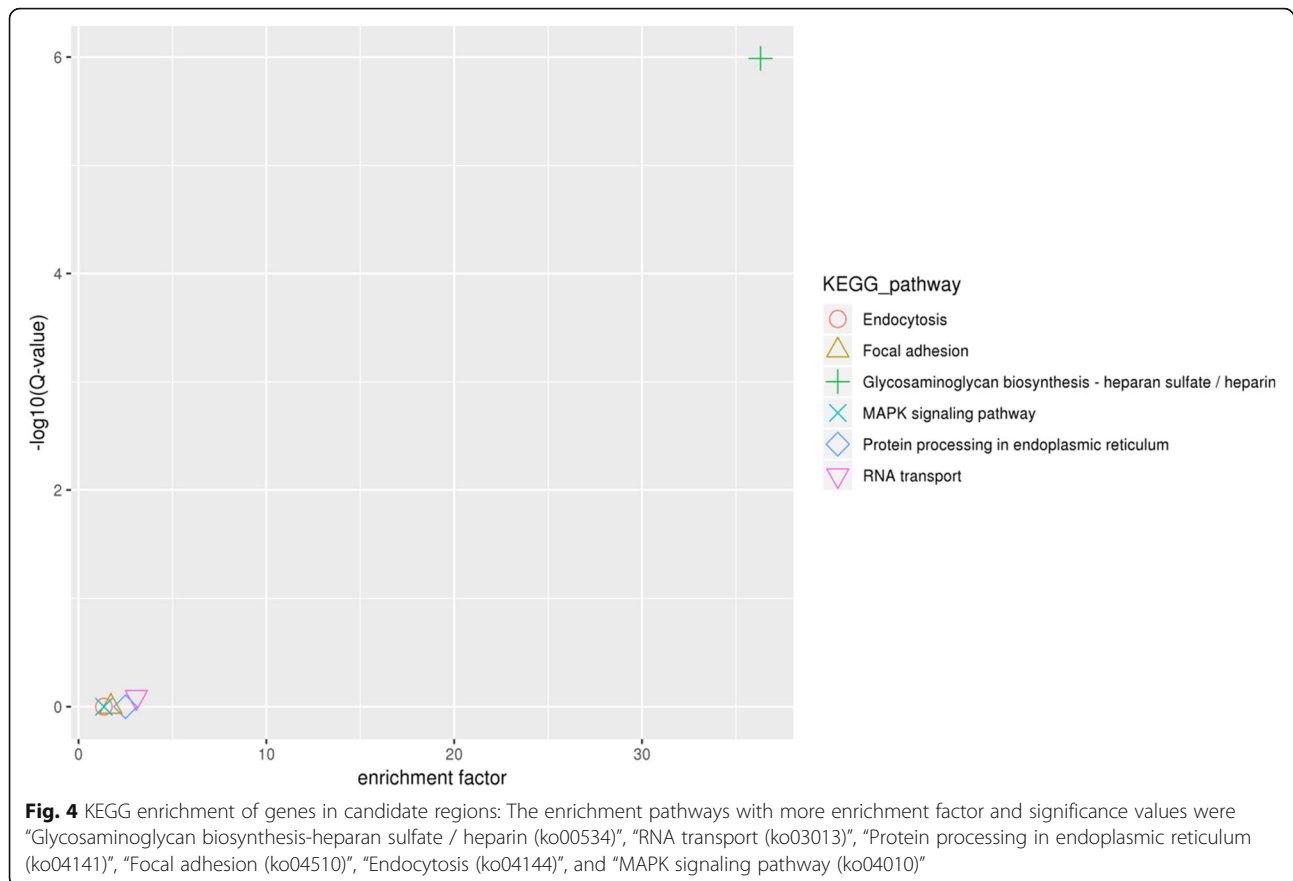


enrichment factor [20, 21]. Our results showed that the KEGG pathways with SNP enrichment ratio were binding in resistant groups. GO analysis also showed that some cellular components tended to be less polymorphic than others, whereas KEGG pathway analysis showed that some pathways tended to be more polymorphic than others. SNPs are inequitably distributed across the genome with many associated with higher contingency in functionally important positions. Similarly, SNPs can affect the components and their interactions [22]. These clarifications should be investigated in future studies.

Gene mutations may introduce phenotypic variation by influencing gene expression, including the possibility of hybrid vigour as useful traits that are oppressed in animal and crop breeding [23]. There has long been a tacit understanding that gene expression differences play a vital role in species differentiation and experiments on natural selection for gene expression can now be monitored in more effective ways [19, 24]. In our analysis, we

selected 5 genes with genetic variations and examined their mRNA expression in liver, kidney and gill tissues. The results of these analyses showed that mRNA expression levels of *SAMD9L* and *EPHB2* were increased in all three tissues from the resistant groups in response to GCRV, while *BNIP3L* and *APPL2* expression was increased in the liver and kidney tissues of resistant fish when compared to the gill tissues of the same animal. Kidney-*NLRP12* gene expression was higher in the resistant group as compared to susceptible group and the results of the expression analysis of all five of these genes suggest that each may be partially responsible for GCRV resistance.

Several more recent studies have found that SNPs in intron and intergenic regions also play a significant role in the adaptation of specific traits. It was reported that, SNPs in the third intron of the F-box and leucine rich repeat protein 17 (*FBXL17*) gene explains 58.4% of the phenotypic differences in sex reversal of Chinese tongue



sole *Cynoglossus semilaevis* [25]. Li, et al. (2016), also reported that heterozygous SNP variation can contribute to increase latex yield in these hybrids [26]. Here, we identified two heterozygous SNPs (at chromosomal position 5859G > A and 5968G > A) in the intron region of *EPHB2* which were shown to be significantly associated with GCRV responses in resistance/ susceptible ENU grass carp. Allele G was higher in the resistant group than in the susceptible group, indicating that these alleles could increase disease resistance. The survival rate of susceptible ENU after GCRV infection might be reduced by higher frequencies of allele A at these SNPs. Consequently, future breeding programmes for these ENU grass carp could emphasise selection for the G allele at both of these SNPs to improve GCRV disease resistance. This study is, to the best of our knowledge, the first to report SNPs located in a gene that is associated with GCRV disease resistance in ENU grass carp. Additionally, the SNPs obtained in this study provide the basis for genetic selection of GCRV disease resistance in ENU grass carp.

EPHB2 (Ephrin type-B receptor 2) influences the immune system in numerous ways, primarily through immune cell transfer and activation. Regulation of B and T lymphocyte as well as dendritic cell activation has all

been linked to *EPHB2* [27, 28]. The expression of *EPHB2* can also be regulated by certain inflammatory cytokines and pathogen-related molecular forms [29, 30]. Among the five resistant genes studied in this paper, we only found SNPs in *EPHB2* as explained above, and this gene was highly expressed in the resistant group. The modification and higher expression of this gene indicate that it is closely associated with GCRV resistance in these animals. This gene functions as a tumour suppressor and plays a role in immune cell enhancement during GCRV disease in ENU grass carp. Our results concluded that *EPHB2* is involved in the immune response and that it can repress viral replication and attenuate acute inflammatory responses to protect cells.

Conclusions

Our results demonstrate the utility of applying BSA-sequence analysis to the detection of genes responsible for interesting phenotypes and will help in developing new protocols for completing marker-assisted selection for GCRV disease resistance in farmed fish. *EPHB2* expression was higher in the kidney, liver and gills. Two SNPs found in the intron region of the *EPHB2* gene were significantly associated with GCRV resistance. Additionally, *EPHB2* is involved in immune response

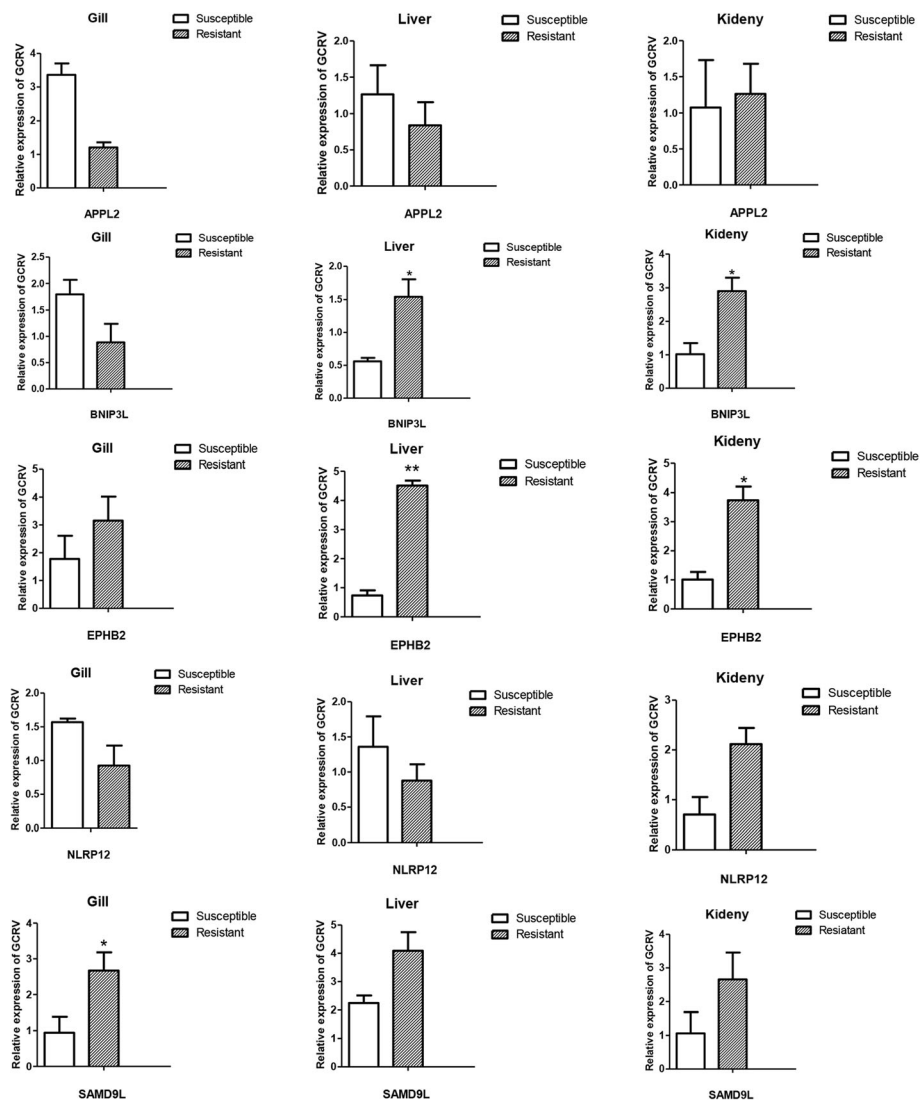


Fig. 5 genes expression analysis showed that mRNA expression level of SAMD9L, BNIP3L and EPHB2 was highly significant expressed in resistant group than infected group in response of GCRV ($p < 0.01$), APPL2 expression level was higher in liver and kidney as compared to gill tissue in case of resistant group. Kidney-NLRP12 gene expression level was significantly highly expressed in resistant group as compared to infected group after GCRV infection ($p < 0.01$). The statistical results (expressed as mean \pm standard deviation) were analyzed by one-way analysis of variance, followed by Dunnett's test for multiple comparisons using IBM SPSS Statistics 22 software. $p < 0.01$, $p < 0.05$ was considered to be statistically significant

and may suppress virus replication and reduce acute inflammatory responses protecting cells from apoptosis. Taken together these data suggest that EPHB2 may be an important gene for GCRV resistance. The SNPs associated with GCRV resistance could be applied to marker-assisted selection for breeding GCRV resistant grass carp.

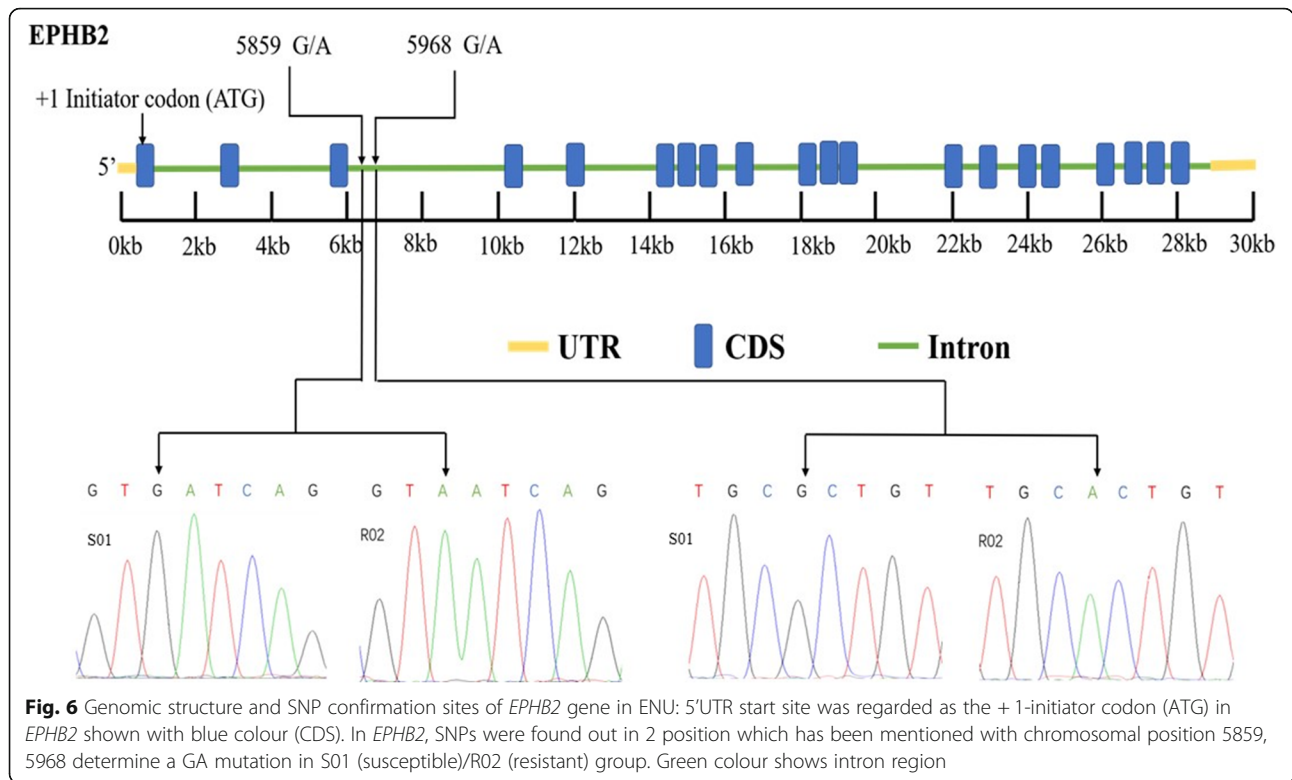
Materials and methods

Experimental design

ENU mutant grass carp (meiotic gynogenetic offspring induced by UV inactivated heterologous sperm from *Megalobrama amblycephala*) were obtained from the Bream Genetics and Breeding Center at Shanghai Ocean University, Shanghai, China. On arrival to the laboratory,

Table 6 Genes with validated SNP

Chromosome	Gene	SNP position	Ref	S01	R02	ED-value
CI01000190_01061817_01090001.mRNA	EPHB2	1,067,676	R	R	R	0.624991
		1,067,785	R	R	R	0.648181



fish were maintained at $28 \pm 0.5^\circ\text{C}$ for at least 7 days prior to experimental use and fed well to make them as healthy as possible. We used a total of 200 fish with an average weight of 4.4–6.0 g. The trials were conducted in aerated glass aquariums ($120 \times 40 \times 30$ cm) each containing 100 L of water. After acclimatisation, all fish were intraperitoneally inoculated with $20 \mu\text{l/g}$ of GCRV-873. Fish started to die 9 days post infection and 30 fish were

collected from this group (susceptible) and identified by their extreme GCRV symptoms. At 14 days post infection any surviving fish (30 animals) were collected and classified as resistant for further evaluation.

Fish sampling

Using the method described above we selected two groups of ENU fish with extremely different phenotypes.

Table 7 Number and allele frequency at two SNPs of *EPHB2* in susceptible and resistant ENU after GCRV infection

SNP1(CI01000190, 1,067,676) ($p = 0.002$)	Allele frequency in resistant group ($n = 50$)	Allele frequency in susceptible group ($n = 50$)
A	24%	66%
G	76%	34%
SNP2(CI01000190, 1,067,785) ($p = 0.003$)		
A	18%	62%
G	82%	38%
SNP1 genotype ($p = 0.004$)	Allele frequency in resistant group ($n = 50$)	Allele frequency in susceptible ($n = 50$)
AA	16%	34%
AG	48%	52%
GG	36%	14%
SNP2 genotype ($p = 0.003$)		
AA	12%	30%
AG	44%	52%
GG	44%	18%

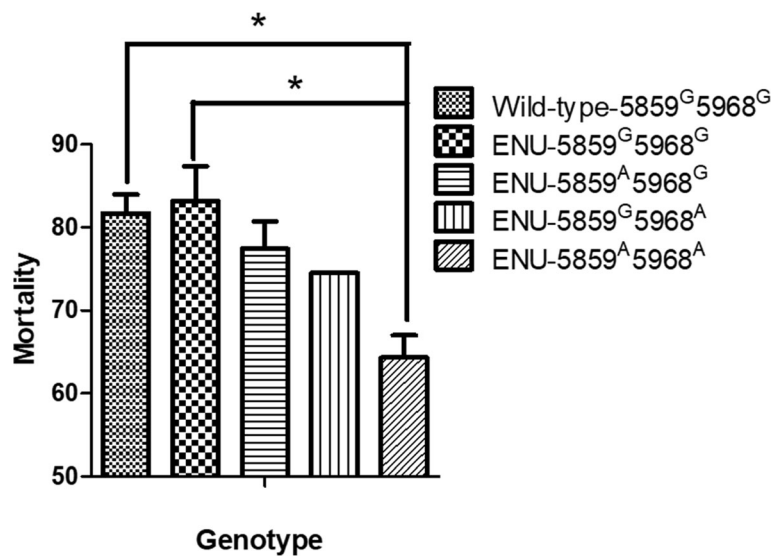


Fig. 7 Wild type and mutant grass verification on the basis of different SNP genotype and mortality rate

A total of 30 fish were selected as the susceptible/infected/morbid group (S01) and 30 fish were selected as the resistant group (R02). Infection was allowed to run for 14 days before we collected liver tissue samples from each group and applied these to the BSA analysis.

Sequencing analysis

Genomic DNA was isolated using a conventional phenol-chloroform extraction strategy in combination with RNase treatment and stored at -20°C until use. Two bulk isolations were produced by pooling an equal quantity of DNA from a susceptible (S01) and a resistant group (R02). DNA from each pool was then used to build paired-end (PE) sequencing libraries, which were sequenced on an Illumina HiSeq (Illumina Casava 1.8 version). The entire experimental procedure was completed as described in the protocol from Illumina, including sample testing, library construction, library-quality testing, and computer sequencing. The final sequence read length was 150 bp (Biomarker technology, Beijing, China).

After removing the connector sequences and low-quality reads, the clean reads were then re-evaluated for quality utilising FASTQC. High quality PE reads were mapped to the grass carp reference genome sequence (PRJEB5920) [12] using the BWA programme with default constraints [31]. The position of the clean reads on the reference genome was identified by comparing data such as the sequencing depth and genome coverage in each run and this was then used to map the mutation loci. Assessments of the sequencing output information, comparisons, depth of coverage, and the genome

coverage comparison at each depth for both the S01 and R02 groups are summarised in Table 1.

SNP and InDel detection

SnEff is a software package designed to facilitate the annotation of mutations (SNP, Small InDel) and predicting the effects of these mutations [17]. The detection of SNPs was mainly implemented using the GATK software toolkit [32] and Picard (<http://sourceforge.net/projects/picard/>) was used to complete the data preprocessing such as marking the duplicates. GATK was used for Local realignment and base recalibration to ensure detection.

InDel represents single base insertions and deletions, with the insertions detected using GATK. Small InDel variation is generally less frequent than SNP variation, and is also reflected in the differences between the sample and the reference genome, and InDels in the coding region will cause frameshift mutations, resulting in changes in gene function.

ED calculations

ED algorithm uses sequencing data to identify markers which demonstrate significant differences in occurrence between pools and is also often used to evaluate SNP/InDel associations [8]. Theoretically, two mixed pools constructed using the BSA system are likely to have differences in the target trait-related sites, but the other sites should be consistent, so the ED value of non-target sites should tend towards 0. The formula for the ED calculations is shown below. The larger the ED value, the greater the difference between the mark and the two mixing tanks.

Table 8 Selected primer and sequence

Gene	Primer name	Forward primer (5'-3')	Reverse primer (5'-3')	Temperature
CI01000184_00984445_00984988.mRNA	APPL2	CCGTGGATGACACCGCCTAC	CACACTCCACGGCCATGACA	59.7
CI01000087_01389941_01395452.mRNA	BNIP3L	TCTGAGTGTGTCCATCGAGT	GAGATCCAGCAGAACTGG	58.0
CI01000190_01446777_01465687.mRNA	NLRP12	TCTGAGTGTGTCCATCGAGT	GAGATCCAGCAGAACTGG	58.0
CI01000190_01061817_01090001.mRNA	EPHB2	CAGAGGCATTTCATCCCA	GAGAGAGGGCAGGGGAAAA	58.1
CI01000240_00049921_00050822.mRNA	SAMD9L	GCAGCAGAATGATGTGGTGAC	TGTGTTGGTGAATGCTTTGAAT	58.5

$$ED = \sqrt{(A_{mut}-A_{wt})^2 + (C_{mut}-C_{wt})^2 + (G_{mut}-G_{wt})^2 + (T_{mut}-T_{wt})^2}$$

where each letter (A, C, G, T) corresponds to the frequency of its corresponding nucleotide in the mutation and wild type pool or bulk preparation respectively.

Functional annotation of genes containing SNPs

Those genes with SNPs correlated to the resistance/susceptibility to GCRV were annotated using BLAST software against multiple databases including the NR [non-redundant protein database, NCBI], Swiss-Prot [<http://www.uniprot.org/>], GO [Gene Ontology, <http://www.geneontology.org/>], KEGG [<http://www.genome.jp/kegg/>] [33, 34], COG [<http://www.ncbi.nlm.nih.gov/COG/>] coding genes in the candidate interval. This in-depth annotation allowed for rapid screening of candidate genes for functional relevance.

Gene expression and SNP verification

qPCR was carried out on a CFX96 Touch™ Real-Time PCR System, using SYBR Premix Ex Taq kit (TaKaRa, Japan). All primers were designed using Primer Premier 5.0 software and are listed in Table 8. The comparative expression values of each of the designated genes versus 18 s rRNA (reference gene to normalise expression levels between samples) was calculated using the $2^{-\Delta\Delta Ct}$ method. SYBR Green reactions were performed in 20 μ L volumes containing 10 μ L of 2 \times SYBR® Green Realtime PCR Master Mix (Toyobo, Osaka, Japan), 1 μ L of each forward and reverse primer (10 μ M), 7 μ L of water, and 1 μ L of diluted cDNA (100 ng/ μ L). All experiments were performed in two groups. We identified five functional genes as disease-related mutations in response to GCRV in mutant grass carp.

For the SNP verification, we injected GCRV into 300 ENU grass carp and selected 50 dead mutant fish on 8-day and 50 alive mutant fish on day 14 for allelic frequency evaluations (Table 8). Both SNPs were located in the intron region of *EPHB2*. We further verified these SNPs in our previous data set in which the fish were divided into 4 group on the basis of their genotype and compared their mortality rate with the wild type grass carp after challenge with GCRV. We also evaluated the

statistical association between each genotype and resistance, using the *p*-values produced from a chi-square test. A subsequent *p*-value of 0.05 or less was determined to be statistically significant.

Statistical analysis

Statistical significance (expressed as mean \pm standard deviation) was evaluated using one-way analysis of variance, followed by a Dunnett's test for multiple comparisons using IBM SPSS Statistics 22 software. *p* < 0.01 and *p* < 0.05 were considered statistically significant. All experiments were repeated at least three times.

Abbreviations

BSA: Bulk segregant analysis; SNP: Single nucleotide polymorphism; InDel: Insertion/ deletion; ENU: Mutant grass carp; GCRV: Grass carp reovirus; GATK: Genomic analysis toolkit; SnpEff: SNP effect; ED: Euclidean distance; KEGG: Kyoto encyclopedia of genes and genomes; GO: Gene ontology; FDR: False discovery rate; APPL2: Adaptor protein, phosphotyrosine interacting with PH domain and leucine zipper 2; NLRP12: NLR family pyrin domain containing 12; BNIP3L: BCL2 interacting protein 3 like; *EPHB2*: EPH receptor B2; SAMD9L: Sterile alpha motif domain containing 9 like; MAPK: Mitogen-activated protein kinase; ROS: Reactive oxygen species; BWA: Burrows-wheeler aligner

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-021-07858-x>.

Additional file 1.

Additional file 2.

Additional file 3.

Additional file 4.

Additional file 5.

Additional file 6.

Acknowledgements

We are thankful to Prof. Yaping Wang of Institute of Hydrobiology, Chinese Academy of Sciences for providing the GCRV virus. BioMarker (qingdao) support early step of BSA sequence analysis and computer resources availability, respectively.

Authors' contributions

MN, ZJ, GZ and SZ conceived the project and designed scientific objectives. MN, ZJ and GZ carried out the transcriptome data processing and computational analyses, ZJ and MN performed the RNA isolation, Sanger sequencing and related analyses. MN and JZ wrote the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by grants from the National Key Research and Development “Blue Granary Technology Innovation” key project (No. 2020YFD0900400 to S.M. Zou), The National Natural Science Foundation of China (32002381 to G.D. Zheng).

Availability of data and materials

Raw genomic-seq reads data supporting the results of this article are available in the NCBI Sequence Read Archive (SRA) database (Accession number: PRJNA716293).

Declarations

Ethics approval and consent to participate

The study was carried out in compliance with the ARRIVE guidelines. All experiments were approved by the Shanghai Ocean University and conducted following the guidelines approved by the Shanghai Ocean University Committee on the Use and Care of Animals (Permit Number: SHOU-DW-2020-033).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Genetics and Breeding Center for Blunt Snout Bream, Ministry of Agriculture, Shanghai 201306, China. ²Key Laboratory of Freshwater Aquatic Genetic Resources, Ministry of Agriculture, Shanghai 201306, China. ³National Demonstration Center for Experimental Fisheries Science Education, Shanghai Ocean University, Shanghai 201306, China.

Received: 17 February 2021 Accepted: 25 June 2021

Published online: 07 July 2021

References

- Ozaki A, Okamoto H, Yamada T, Matuyama T, Sakai T, Fuji K, et al. Linkage analysis of resistance to *Streptococcus iniae* infection in Japanese flounder (*Paralichthys olivaceus*). *Aquaculture*. 2010;308(SUPPL.1):S62–7. <https://doi.org/10.1016/j.aquaculture.2010.07.039>.
- Jiang XY, Sun CF, Zhang QG, Zou SM.ENU-induced mutagenesis in grass carp (*Ctenopharyngodon idellus*) by treating mature sperm. *PLoS One*. 2011; 6:1–8.
- van Eeden FJM, Granato M, Odenthal J, Haffter P. Chapter 2 developmental mutant screens in the zebrafish. *Methods Cell Biol*. 1998;60(C):21–41. [https://doi.org/10.1016/S0091-679X\(08\)61892-0](https://doi.org/10.1016/S0091-679X(08)61892-0).
- Knapik EW.ENU mutagenesis in zebrafish - from genes to complex diseases. *Mamm Genome*. 2000;11(7):511–9. <https://doi.org/10.1007/s003350010098>.
- Zhang H, Liu SJ, Zhang C, Tao M, Peng LY, You CP, et al. Induced gynogenesis in grass carp (*Ctenopharyngodon idellus*) using irradiated sperm of Allotetraploid hybrids. *Mar Biotechnol*. 2011;13(5):1017–26. <https://doi.org/10.1007/s10126-011-9365-8>.
- Jiang Y. Hemorrhagic disease of grass carp - disease card. *Isr J Aquac*. 2009; 61:188–97.
- Wang Q, Zeng W, Liu C, Zhang C, Wang Y, Shi C, et al. Complete genome sequence of a Reovirus isolated from grass carp, indicating different genotypes of GCRV in China. *J Virol*. 2012;86(22):–12466. <https://doi.org/10.1128/JVI.02333-12>.
- Liu S, Yeh CT, Tang HM, Nettleton D, Schnable PS. Gene mapping via bulked segregant RNA-Seq (BSR-Seq). *PLoS One*. 2012;7:1–8.
- Hyten DL, Smith JR, Frederick RD, Tucker ML, Song Q, Cregan PB. Bulk segregant analysis using the goldengate assay to locate the Rpp3 locus that confers resistance to soybean rust in soybean. *Crop Sci*. 2009;49(1):265–71. <https://doi.org/10.2135/cropsci2008.08.0511>.
- Venuprasad R, Dalid CO, Del Valle M, Zhao D, Espiritu M, Sta Cruz MT, et al. Identification and characterization of large-effect quantitative trait loci for grain yield under lowland drought stress in rice using bulk-segregant analysis. *Theor Appl Genet*. 2009;120(1):177–90. <https://doi.org/10.1007/s00122-009-1168-1>.
- Lorenz AJ, Coors JG. What can be learned from silage breeding programs? *Appl Biochem Biotechnol*. 2008;148(1-3):261–70. <https://doi.org/10.1007/s12010-007-8116-9>.
- Wang Y, Lu Y, Zhang Y, Ning Z, Li Y, Zhao Q, et al. The draft genome of the grass carp (*Ctenopharyngodon idellus*) provides insights into its evolution and vegetarian adaptation. *Nat Genet*. 2015;47(6):625–31. <https://doi.org/10.1038/ng.3280>.
- Geng X, Sha J, Liu S, Bao L, Zhang J, Wang R, 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*. 2015;16:1–12.
- Wang L, Liu P, Huang S, Ye B, Chua E, Wan ZY, et al. Genome-wide association study identifies loci associated with resistance to viral nervous necrosis disease in Asian seabass. *Mar Biotechnol*. 2017;19(3):255–65. <https://doi.org/10.1007/s10126-017-9747-7>.
- Correa K, Lhorente JP, López ME, Bassini L, Naswa S, Deeb N, et al. Genome-wide association analysis reveals loci associated with resistance against *Piscirickettsia salmonis* in two Atlantic salmon (*Salmo salar* L.) chromosomes. *BMC Genomics*. 2015;16:1–9.
- Zhang K, Han M, Liu Y, Lin X, Liu X, Zhu H, et al. Whole-genome resequencing from bulked-segregant analysis reveals gene set based association analyses for the *Vibrio anguillarum* resistance of turbot (*Scophthalmus maximus*). *Fish Shellfish Immunol*. 2019;88:76–83. <https://doi.org/10.1016/j.fsi.2019.02.041>.
- Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*. 2012;6(2):80–92. <https://doi.org/10.4161/fly.19695>.
- Chen R, Davydov EV, Sirota M, Butte AJ. Non-synonymous and synonymous coding SNPs show similar likelihood and effect size of human disease association. *PLoS One*. 2010;5:1–6.
- Liao Z, Wan Q, Shang X, Su J. Large-scale SNP screenings identify markers linked with GCRV resistant traits through transcriptomes of individuals and cell lines in *Ctenopharyngodon idella*. *Sci Rep*. 2017;7:1–12.
- Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. 2005;21(18):3674–6. <https://doi.org/10.1093/bioinformatics/bti610>.
- Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, et al. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res*. 2008;36(10):3420–35. <https://doi.org/10.1093/nar/gkn176>.
- Zhong Q, Simonis N, Li QR, Charloreaux B, Heuze F, Klitgord N, et al. Edgetic perturbation models of human inherited disorders. *Mol Syst Biol*. 2009;5(1): 321. <https://doi.org/10.1038/msb.2009.80>.
- Huang X, Yang S, Gong J, Zhao Y, Feng Q, Gong H, et al. Genomic analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis. *Nat Commun*. 2015;6:1–9.
- López-Maury L, Marguerat S, Bähler J. Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nat Rev Genet*. 2008;9(8):583–93. <https://doi.org/10.1038/nrg2398>.
- Jiang L, Li H. Single locus maintains large variation of sex reversal in half-smooth tongue sole (*Cynoglossus semilaevis*). *G3 genes, genomes. Genet*. 2017;7:583–9.
- Li D, Zeng R, Li Y, Zhao M, Chao J, Li Y, et al. Gene expression analysis and SNP/InDel discovery to investigate yield heterosis of two rubber tree F1 hybrids. *Sci Rep*. 2016;6 August 2015;6:1–12.
- Darling TK, Lamb TJ. Emerging roles for Eph receptors and ephrin ligands in immunity. *Front Immunol*. 2019;10:1–15.
- Coulthard MG, Morgan M, Woodruff TM, Arumugam TV, Taylor SM, Carpenter TC, et al. Eph/ephrin signaling in injury and inflammation. *Am J Pathol*. 2012;181(5):1493–503. <https://doi.org/10.1016/j.ajpath.2012.06.043>.
- Ivanov A, Romanovsky A. Putative dual role of ephrin-Eph receptor interactions in inflammation. *IUBMB Life*. 2006;58(7):389–94. <https://doi.org/10.1080/15216540600756004>.
- Wohlfahrt JG, Karagiannidis C, Kunzmann S, Epstein MM, Kempf W, Blaser K, et al. Ephrin-A1 suppresses Th2 cell activation and provides a regulatory link to lung epithelial cells. *J Immunol*. 2004;172(2):843–50. <https://doi.org/10.4049/jimmunol.172.2.843>.
- Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*. 2009;25(14):1754–60. <https://doi.org/10.1093/bioinformatics/btp324>.

32. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20(9):1297–303. <https://doi.org/10.1101/gr.107524.110>.
33. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 2016;44(D1):D457–62. <https://doi.org/10.1093/nar/gkv1070>.
34. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28(1):27–30. <https://doi.org/10.1093/nar/28.1.27>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

