Selection of functional EPHB2 genotype in ENU mutated grass carp response to GCRV by the method of BSA Sequence Analysis

Meher un Nissa
Shanghai Ocean University
ZhuXiang Jiang
Shanghai Ocean University
GuoDong Zheng
Shanghai Ocean University
ShuMing Zou (smzou@shou.edu.cn)
Shanghai Ocean University

Research Article

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

Posted Date: April 1st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-250461/v1

Version of Record: A version of this preprint was published at BMC Genomics on July 7th, 2021. See the published version at https://doi.org/10.1186/s12864-021-07858-x.

Abstract

Background *N*-ethyl-*N*-nitrosourea (ENU) mutagenesis is a useful method for genetic development of plants, as well as for inducing functional mutants in animal models including mice and zebrafish. GCRV (Grass carp Reovirus) is the hemorrhage disease of grass carp and causes noteworthy loss of fingerlings in the last few years. To overcome this problem, we have used ENU mutant grass carp might be a helpful to find out some resistant functional gene for future hereditary rearing projects in grass carp.

Results In the present study, ENU-mutated grass carp were used to identify genetic markers correlated with anti hemorrhagic disease caused by grass carp reovirus (GCRV). The BSA (Bulked segregant analysis) technology were performed on two homozygous gynogenetic ENU grass carp groups with susceptible or resistant in response of GCRV. Our results showed that 466,162 SNPs and 197644 InDel were found out in the mixed pool. A total of 170 genes were 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. Among these 170 mutated genes, 5 randomly selected immune-related genes expression was higher in resistant group as compared to susceptible. Furthermore, we found one immune-related gene EPHB2 has confirmed two heterozygous SNP mutations which responded to GCRV disease. The SNPs within gene EPHB2 were found out in the intron region with position 5859 ($5859^{G>A}$) and 5968 ($5968^{G>A}$) were significantly (p = 0.002, 0.003) associated with the resistance against GCRV. These SNP sites were confirmed to correlate with the GCRV-resistant phenotype of ENU grass carp. The SNPs associated with GCRV resistance are useful in selecting GCRV-resistant genes for genetic breeding in grass carp.

Conclusion Our results proved the possibility of BSA-sequence analysis in detecting genes responsible for the exciting phenotypes and will help in carrying out the marker-assisted selection, especially for disease resistance in the response of GCRV.

Background

Genetic breeding of aquaculturally fish species primarily depends on the conclusion of natural mutants with high performance, and then better-quality strains can be formed by hybridization, genetic selection or marker-assisted breeding methods [1]. Thus far, more than 200 improved strains of aquaculturally fish species have been shaped in China, in which 139 classes were produced through 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 operator, exchanging its ethyl gather to nucleophilic nitrogen or oxygen locales on deoxyribonucleotides, directing to base inconsistency while DNA replication [3][4]. ENU mutagenesis has been demonstrated to be effective for inciting point mutation in the genome of the grass carp [5].

Grass carp reovirus (GCRV) belongs to the family Reoviridae and it has genus Aquareovirus. In 1972, Hemorrhagic disease of grass carp was first time revealed in a fish farm of Hubei Province [6]. GCRV was categorized as an aqua reovirus and in 1984, which contains 11 parts of double stranded RNA. The genus Aquareovirus is assigned A to G (AQRV-A to AQRV-G), and GCRV be appropriate to AQRV-C [7]. GCRV is the hemorrhage disease of grass carp and causes noteworthy loss of fingerlings in the last few years.

A genotyping-by-sequencing method, it permits the genome widespread association studies, bulked segregant analysis (BSA) and genomic selection, has previously been familiarized to breeding programs. In current times, because of reducing the prices of genome sequencing have loosened the opportunity for the whole genome sequencing and resequencing the higher pools of individuals. BSA was proposed to quickly distinguish markers connected to specific characteristics of intrigued such as disease resistance [8]. Their approach includes a segregating F2 population produced from a starting cross between two phenotypically different parents, which is at that point scored for a phenotype of intrigued. Bulked DNA or RNA tests are built from organism that appears differentiating phenotypes. BSA has been primarily utilized in crop species either for the recognizable proof of huge impact QTL, such as disease resistance genes or for mapping subjective mutation [9][10][8] [11].

The objective of the current study was to select the functional genotype in ENU fish response to GCRV by the method of BSA sequence analysis. Thus, it will be helpful in carrying out the marker-assisted selection, especially for disease resistance in the response of GCRV.

Results

SNPs and indels annotation results

After filteration, the Clean Reads data was generated up to 73.47 Gbp, average Q30 reached 93.92%, and the average GC% was up to 38.11% **(Table 1)**. The average genome coverage depth of 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%, the genome is Scaffold level[12]. Venn diagram showed that different number of SNP and inDel between sample S01(suspected) and R02

(resistant) like in **Fig.1**, 200,270 SNPs were specified in the S01 and 173,531 SNPs were specifically found out in R02. The center overlapping portion of the Venn diagram represented the total number of identified SNPs 2,189,327 and inDel 759,064 results **(Fig. 1)**.

We detected a total number of 466,162, 197,644, 8,492, and 3,668 polymorphic sites in different regions like SNP annotation results, InDel annotation results, SNP and inDel annotation results in candidate region respectively **(Table 2)**. Based on the annotation, the SNPs were randomly distributed throughout the genome **(Table 2)**. InDel frameshift mutation type was found out in 50 genes, 7 genes figured out in codon_insertion and 6 genes found out in codon_deletion **(Table 2).** Synonymous and non-synonymous codon had a higher number of genes that were found out in the SNP annotated region (7251, 9298) **(Table 2)**.

Association Analysis (SNP, InDel) and ED method correlation results

Before the association analysis, first SNP was filtered, finally obtained 2,313,014 high-quality trusted SNP sites. Before using InDel for correlation analysis, same like SNP, first InDel was filtered, and finally got 775,981 high-quality trusted InDel sites **(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 (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 results

A total of 170 genes were annotated in the candidate region, among which 49 were non-synonymous mutant genes and 25 frameshift mutation genes were annotated in the mixed pool **(Table 5, Additional file 3)**. Total 55 genes in the genome were annotated and classified into biological processes, cellular components, and molecular functions **(Fig.2, Additional file 4, 5 & 6).** 11 genes were found out with non-synonymous mutation and 2 genes were found out with frameshift mutation in GO enrichment analysis. The annotated KEGG databases showed that 29 SNP were found among top 20 pathways as shown in **Fig.3.** After multiple-testing corrections, we selected the pathways with Q values ≤ 0.05 as significantly enriched among the genes **(Fig.4)**. Our data suggest that these genes may play an important in innate immune response to GCRV in ENU grass crap.

Gene expression results associated with grass carp reovirus (GCRV) resistance

The relative expression levels of the five genes significantly associated with grass carp reovirus (GCRV) resistance were detected through *q*PCR **(Fig.5)**. We selected 5 genes with genetic variation to examine their mRNA expression levels in liver, kidney and gill tissue. All these five genes involved in the inflammatory response, cell proliferation, anti-apoptosis, tumor suppressor and immune response to viral infection pathways. Results showed that mRNA expression level of SAMD9L, BNIP3L and EPHB2 was highly significant expressed in resistant group than susceptible 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 ($p<0.01$) **(Fig. 5)**.

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

After validation, we identified an SNP (CI01000190: 1067676 A > G) at the position of 1067676 and another SNP (CI01000190: 1067785 A > G) at the position of 1067785 in chromosome CI01000190, respectively **(Table 6).** Both SNPs were located in the intron region of EPHB2 as shown in **Fig.6**. SNPs in EPHB2 were found out in 2 position which has been mentioned with chromosomal position 5859 ($5859^{G>A}$), 5968 ($5968^{G>A}$) determine G to A mutation in S01/R02.

EPHB2 gene with confirmed SNPs has been showed in **Table 7** in case of resistant (S01) and susceptible (R02) group in response of GCRV. In the first SNP, the allele frequencies of A and G in survived fish (resistant) were 24% and 76%, correspondingly. For dead fish (susceptible), the allele frequencies of A and G f were 66% and 34%, respectively. In the other SNP, the allele frequencies A and G in survived fish were 18% and 82%whereas in dead fish were 62% and 38%, respectively **(Table 7)**. The chi-squared test showed that allele frequencies were significantly different between dead and survived individuals ($p = 0.002$ and $p = 0.003$ for respective SNPs) **(Table 7)**, signifying that EPHB2 was significantly associated with GCRV disease resistance.

Discussion

Up to the present time, association analysis is still the key method for the mining of SNPs for disease resistance genes in marine animals. Previous work has been reported that SNP frequencies diverse among different populations in numerous species[13][14][15]. However, according to our knowledge, there is limited information on immune related SNP in relation to the response of GCRV in grass carp. Therefore,

the currently study focused on the selection of functional genes and its SNPs related to GCRV resistance of ENU grass carp, this could help to improve our understanding on GCRV resistance and might help to select disease-resistant strains of ENU grass carp culture.

Cultured grass carp are highly vulnerable to diseases which lead to yield reduction, antibiotics and drugs have been used to control the problem but the regular use of these alternatives showed side effects to animal and environment. Therefore, disease-resistant grass carp are highly desirable as they can greatly increase fish yields [5]. Due to low occurrence of natural mutation in grass carp, chemical mutagenesis can be useful for increasing genetic mutations [2]. In this study, homozygous gynogenetic ENU grass carp which has the advantages of strong disease resistance and rapid growth was used and its functional genotype was selected by using BSA sequence analysis after challenged to GCRV.

In this study, we used BSA technology to determine SNP mutation within the genic region associated with GCRV resistance in homozygous gynogenetic ENU grass carp. Our findings verified the viability of BSA approach and the extensive information of SNP gene for disease resistance traits as a fast and affordable method for marker development. Zhang *et al* (2019) also 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 *Vibrio anguillarum* resistance [16].

SNP and InDel are used for annotating mutations and predicting the effects of mutations[17]. SNP mutation is significant source for genetic diversity, molecular evolution and disease resistance. Some researchers primarily focus on non-synonymous coding SNPs, because those SNPs might influence the protein activity directly[18]. Wang *et al* (2015) reported the significant differences of SNPs and InDel in resistant and susceptible group in the *C. idella* genome; However, the SNPs and InDel associated with resistance/susceptibility to GCRV were not described by omics sequencing [12]. SNPs can provide innovative resources for genome sequence modification and facilitate the studies on selective breeding [19]. In current study, a total of 466,162 SNPs and 197,644 InDel were found in the resistant/susceptible group, but only 9,298 SNPs caused non-synonymous mutations.

In current study, we determined several resistance genes in S01/R02 group of ENU grass carp. Followed by the performance of GO classification, KEGG pathway analyses to well comprehend the molecular functions of candidate genes in resistance group. The correlation of differentially expressed sequences to the whole sequences of equivalent GO groups or KEGG pathways were observed as the key measures for enrichment factor [20][21]. Our results showed that KEGG pathways with SNP enrichment ratio were binding in resistant groups. GO analysis also showed that some cellular components tended to be less polymorphic. Whereas KEGG pathway analysis showed that some pathways tended to be more polymorphic. Since disease correlated SNP are inequitably diffused by gene sequences, taking a higher contingency in functionally important positions. Similarly, SNP can affect the components and their interactions[22]. These clarifications should be investigated in future studies.

Gene mutation may bring phenotypic variation by influencing the gene expression, including the possibility of hybrid vigor as useful traits that are oppressed in animal and crop breeding[23]. There has long been a comprehension that gene expression differences play vital roles in species differentiation and experiments on natural selection for gene expression which can now be monitored in a more effortless way than that in the past[24][19]. In our analysis, we selected 5 genes with genetic variation to examine their mRNA expression levels in liver, kidney and gill tissue. Results showed that mRNA expression level of SAMD9L and EPHB2 in all tissues was higher in resistant group than susceptible group in response of GCRV, while BNIP3L and APPL2 expression level was higher in liver and kidney compared to gill tissue of resistant group. Kidney-NLRP12 gene expression level was higher in resistant group as compared to susceptible group after GCRV infection. The results of expression analysis of five genes in resistant and susceptible group pinpoint that these gene may be partially responsible for GCRV resistance.

In recent years, studies have found out that SNPs in intron and intergenic also play a great role in the changes of traits. It was reported that, SNP in the third intron of the F-box and leucine rich repeat protein 17 (FBXL17) gene explains 58.4% of the phenotypic difference 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 the hybrid [26]. In present study, we identified two heterozygous SNP (in chromosomal position 5859G>A and 5968G>A) in the intron region of EPHB2 gene after validation which significantly associated with GCRV response in resistance/ susceptible ENU grass carp. Allele G in resistance group was higher than in susceptible group, indicating that these alleles could increase disease resistance. The survival rate of susceptible ENU after GCRV infection might be reduced by higher frequency of allele A in SNP. Consequently, future breeding program of the ENU grass carp can emphasize on allele G in the two SNPs for GCRV disease resistance. Based on our knowledge, this study is 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 base for genetic selection of GCRV disease resistance in ENU grass carp.

EPHB2 (Ephrin type-B receptor 2) influence the immune system in numerous ways, primarily through immune cell transferring and immune cell activation. B and T lymphocyte as well as dendritic cell activation has been revealed to be controlled through 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 study, we only found SNPs in EPHB2 gene as explained above, and this gene was highly expressed in resistance group than in susceptible group. The modification and higher expression of this gene indicate that this it related to the resistance of GCRV. This gene

functioned as a tumor suppressor and also had a role in immune cell enhancement against GCRV disease in ENU grass carp. Our results prove that EPHB2 is involved in immune response that can repress virus replication and attenuates acute inflammatory responses to protect cells.

Conclusions

Our results proved the possibility of BSA-sequence analysis in detecting genes responsible for the exciting phenotypes and will help in carrying out the marker-assisted selection for GCRV disease resistance. EPHB2 expression was higher in 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 that may suppress virus replication and reduces acute inflammatory responses to protect cells. All 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 GCRV resistance in mutant grass carp.

Materials And Methods

Experimental Design:

ENU mutant grass carp (meiotic gynogenetic offsprings induced by UV inactivated heterologous sperm from *Megalobrama amblycephala*) were obtained from the Bream Genetics and Breeding Center of Shanghai Ocean University, Shanghai, China. After on arrival in the laboratory, fish were maintained in the laboratory at 28 ± 0.5 °C for at least 7 days prior to experimental use and feed them well to make healthier. A total of 60 fishes were used. The average weight of the fish was 4.4-6.0g. The trials were conducted in aerated glass aquariums (120 x 40 x 30 cm) each containing 100 L of water. After the acclimatization period, all the fishes were intraperitoneally inoculated with 20 µl/g of GCRV-873 strain. After 14 days of the challenge, all survived fish were collected as a safe fish from which 30 fishes were chosen arbitrarily as a resistant group.

Fish sampling:

In this way, mutant grass carp were divided into two group. 30 fishes were selected as susceptible/ infected/morbid group (S01) and leftover 30 fishes which were selected as resistant group (R02) to GCRV. About fourteen days of the injection later, liver tissue collected as sample and transferred it for BSA analysis.

Sequencing analysis:

Genomic DNA was taken out utilizing conventional phenol-chloroform extraction strategy in combination with RNase treatment and put away at −20°C until utilize. Two bulks were produced by pooling an equal quantity of DNA from a susceptible (S01) and a resistant group (R02). DNA from each bulk was utilized to build paired-end (PE) sequencing libraries, which were sequenced on an Illumina HiSeq (Illumina Casava 1.8 version). Whole experimental procedure is prepared according to the standard convention was given by Illumina, including sample testing, library construction, library-quality testing, and computer sequencing and Sequence read length was 150 bp (Biomarker technology, Beijing, China).

After evacuating connector and low-quality reads, the clean reads were advance rechecked for quality utilizing FASTQC. High quality paired-end reads were mapped to the grass carp reference genome sequence (PRJEB5920) [12] using the BWA program with default constraints [31]. Position of Clean Reads on the reference genome was found out by comparing the data such as the sequencing depth and genome coverage of each test, and then mutation location was achieved. The assessment results of the sequencing output information of each sample, the comparison results of the samples, the average coverage depth of each sample, and the genome coverage proportion comparison to each depth among S01 and R02 group can be seen in the following **Table 1.**

SNP and InDel detection:

SnpEff is a software for annotating mutations (SNP, Small InDel) and predicting the effects of mutations[32]. The detection of SNP is mainly implemented using GATK software toolkit [33]. According to the positioning results of Clean Reads in the reference genome, Picard (http://sourceforge.net/projects/picard/) was used to perform preprocessing such as Mark duplicates. GATK was used for Local realignment and base recalibration to ensure detection.

InDel represents single base insertion and deletion. The insertion loss of the sample was detected using GATK. Small InDel variation is generally less than SNP variation, which also reflects the difference between the sample and the reference genome, and InDel in the coding region will cause frameshift mutations, resulting in changes in gene function.

Euclidean distance calculation:

The Euclidean Distance (ED) algorithm is a method that uses sequencing data to find markers that have significant differences between pools and also evaluated the SNP, inDel association analysis [8]. Theoretically, two mixed pools constructed by the BSA project have differences in the target trait-related sites, and other sites tend to be consistent, so the ED value of non-target sites should tend to 0. The calculation formula of the ED method is shown below. The larger the ED value, the greater the difference between the mark and the two mixing tanks.

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

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

Functional annotation of genes containing SNPs:

The genes having SNPs correlated with resistance/susceptibility to GCRV were annotated using NCBI non-redundant database by BLAST software to perform in-depth annotation of multiple databases such as (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/], COG [http://www.ncbi.nlm.nih.gov/COG/]) coding genes in the candidate interval. Quickly screening the candidate genes were through detailed annotations.

Gene expression and SNP verification analysis:

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

For SNP verification, again we injected the virus (GCRV) in 100 ENU grass carp and divided the fishes into 50 resistant and 50 susceptible groups to find out their allelic frequency as we can see in the Table 8. Both SNPs were located in the intron region of EPHB2. Using these two SNPs, we investigated the genotype number in 50 dead (susceptible and 50 survived(resistant) fish after the GCRV challenge to detect whether there was a statistically association between genotype and resistance, using p-values of chi-square test. A subsequent p-value 0.05 0r less was measured statistically significantly.

Statistical Analysis:

The statistical results (expressed as mean ± 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. 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.

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.

Author's 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.

Availability of Supporting Data

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

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 Science Foundation of China (32002381 to G.D. Zheng).

Competing interest

The authors declare that they have no competing interests.

Acknowledgements

We are thankful to: Prof. Dr. Shuming Zou for providing mutant grass carp (ENU) and helped during the last stages of manuscript; Prof. Dr. Libo He for providing GCRV virus; Dr. Guadong Zheng for funding and guidance. BioMarker (qingdao) support early step of BSA sequence analysis and computer resources availability, respectively.

References

1. 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.
2. 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.
3. 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.
4. Knapik EW. ENU mutagenesis in zebrafish - From genes to complex diseases. Mamm Genome. 2000;11:511–9.
5. 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;13:1017–26.
6. Jiang Y. Hemorrhagic disease of grass carp - Disease card. Isr J Aquac. 2009;61:188–97.
7. 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;86:12466–12466.
8. 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.
9. Hyten DL, Smith JR, Frederick RD, Tucker ML, Song Q, Cregan PB. Bulked segregant analysis using the goldengate assay to locate the Rpp3 locus that confers resistance to soybean rust in soybean. Crop Sci. 2009;49:265–71.
10. 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; 120:177–190
11. Lorenz AJ, Coors JG. What can be learned from silage breeding programs? Appl Biochem Biotechnol. 2008;148:261–70.
12. 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: 625–631
13. 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.
14. 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;19:255–65.
15. 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.

16. 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.
17. 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;6:80–92.
18. Chen R, Davydov E V., Sirsota 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.
19. 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.
20. 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;21:3674–6.
21. 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;36:3420–35.
22. Zhong Q, Simonis N, Li QR, Charloteaux B, Heuze F, Klitgord N, et al. Edgetic perturbation models of human inherited disorders. Mol Syst Biol. 2009; 5: 321
23. 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.
24. López-Maury L, Marguerat S, Bähler J. Tuning gene expression to changing environments: From rapid responses to evolutionary adaptation. Nat Rev Genet. 2008;9:583–93.
25. 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.
26. 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.
27. Darling TK, Lamb TJ. Emerging roles for Eph receptors and ephrin ligands in immunity. Front Immunol. 2019; 10:1–15.
28. Coulthard MG, Morgan M, Woodruff TM, Arumugam T V., Taylor SM, Carpenter TC, et al. Eph/ephrin signaling in injury and inflammation. Am J Pathol. 2012; 181:1493–503.
29. Ivanov A, Romanovsky A. Putative dual role of ephrin-Eph receptor interactions in inflammation. IUBMB Life. 2006;58:389–94.
30. 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;172:843–50.
31. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60.
32. 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;6:80–92.
33. 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;20:1297–303.

Tables

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

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	220527	89734	4,223	1797
INTRON	144560	27	2,538	1242
UPSTREAM	37653	66109	618	266
DOWNSTREAM	37132	16283	653	232
UTR_5_PRIME	3202	16315	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	466162	197644	8492	3668

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 filteration
SNP	2,563,128	—	2,212	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	AssoInDel_ED
Cl01000352	2484	306	85
Cl01112186	5	3	4
Cl01000136	2141	185	73
Cl01000184	1334	121	51
Cl01000190	2631	214	83
Cl01000087	2767	286	84
Cl01000240	274	50	18
Cl01061811	5	2	3
Cl01072320	5	7	3
Cl01163490	5	1	3
Cl01064712	6	2	3
Cl01085134	23	10	5
Cl01075334	3	4	2
Cl01141415	3	3	2
Cl01000257	280	30	16
Cl01079816	4	2	2
Cl01114229	4	2	2
Cl01017852	12	4	3
Cl01026461	5	3	2
Cl01114140	5	3	2
Cl01000262	112	33	8

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

Table. 6 Genes with validated SNP

Chromosome	Gene	SNP position	Ref	S01	R02	ED-value
Cl01000190_01061817_01090001.mRNA	EPHB2	1067676	R	R	R	0.624991
		1067785	R	R	R	0.648181

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

SNP1(CI01000190, 1067676) (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, 1067785) (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%

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	GAGATCCAGCAGAACACTGG	58.0
CI01000190_01446777_01465687.mRNA	NLRP12	TCTGAGTGTGTCCATCGAGT	GAGATCCAGCAGAACACTGG	58.0
CI01000190_01061817_01090001.mRNA	EPHB2	CAGAGGCATTCATCTTCCCA	GAGAGAGGGCAGGGGAAAA	58.1
CI01000240_00049921_00050822.mRNA	SAMD9L	GCAGCAGAATGATGTGGTGAC	TGTGTTGGTGAATGCTTTGAAT	58.5

Figures

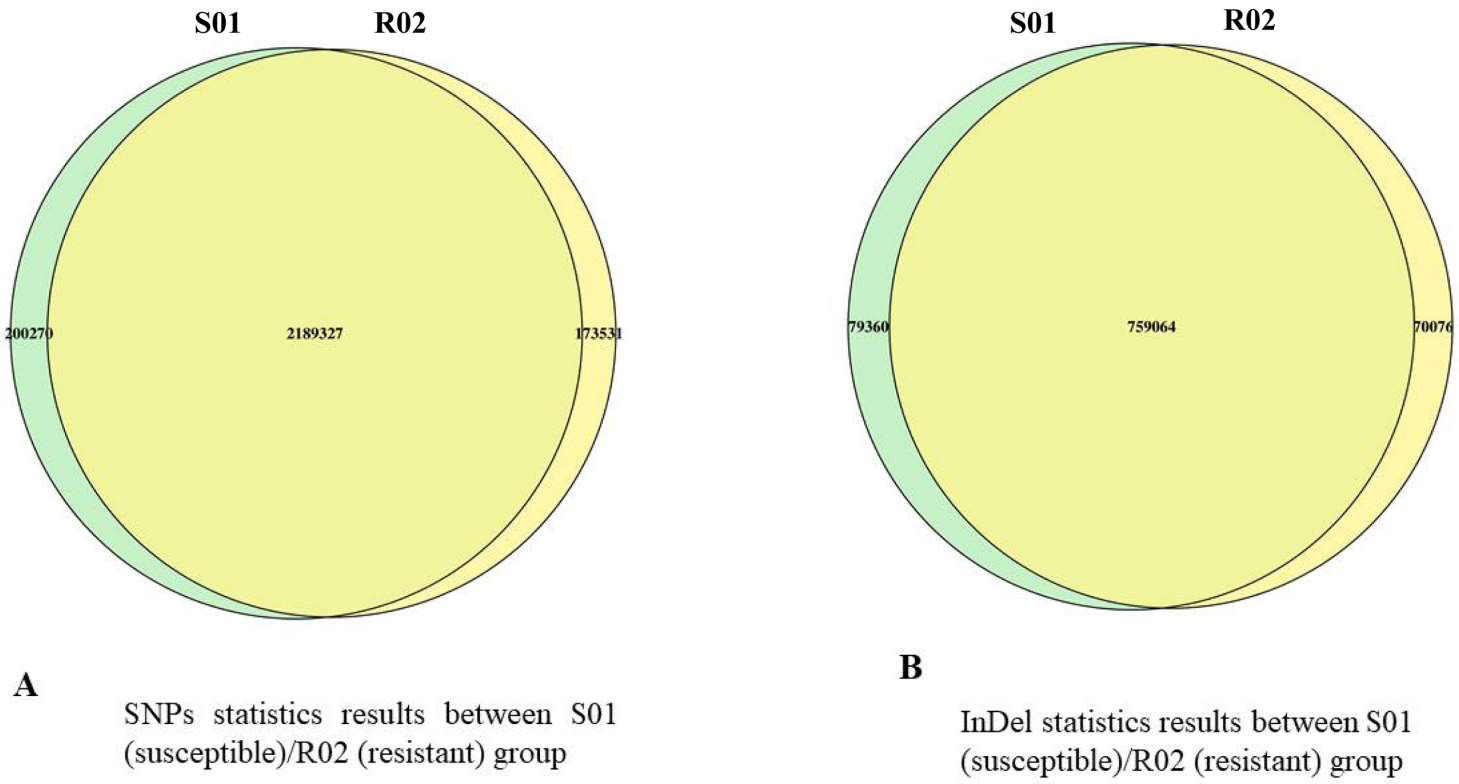

Figure 1

Venn diagram of SNP/ InDel statistics between sample: Venn diagram describes the overlapped pathways between resistant and susceptible groups. (A) SNPs statistics results between S01 (susceptible)/R02 (resistant) group, (B) InDel statistics results between S01 (susceptible)/R02 (resistant) group

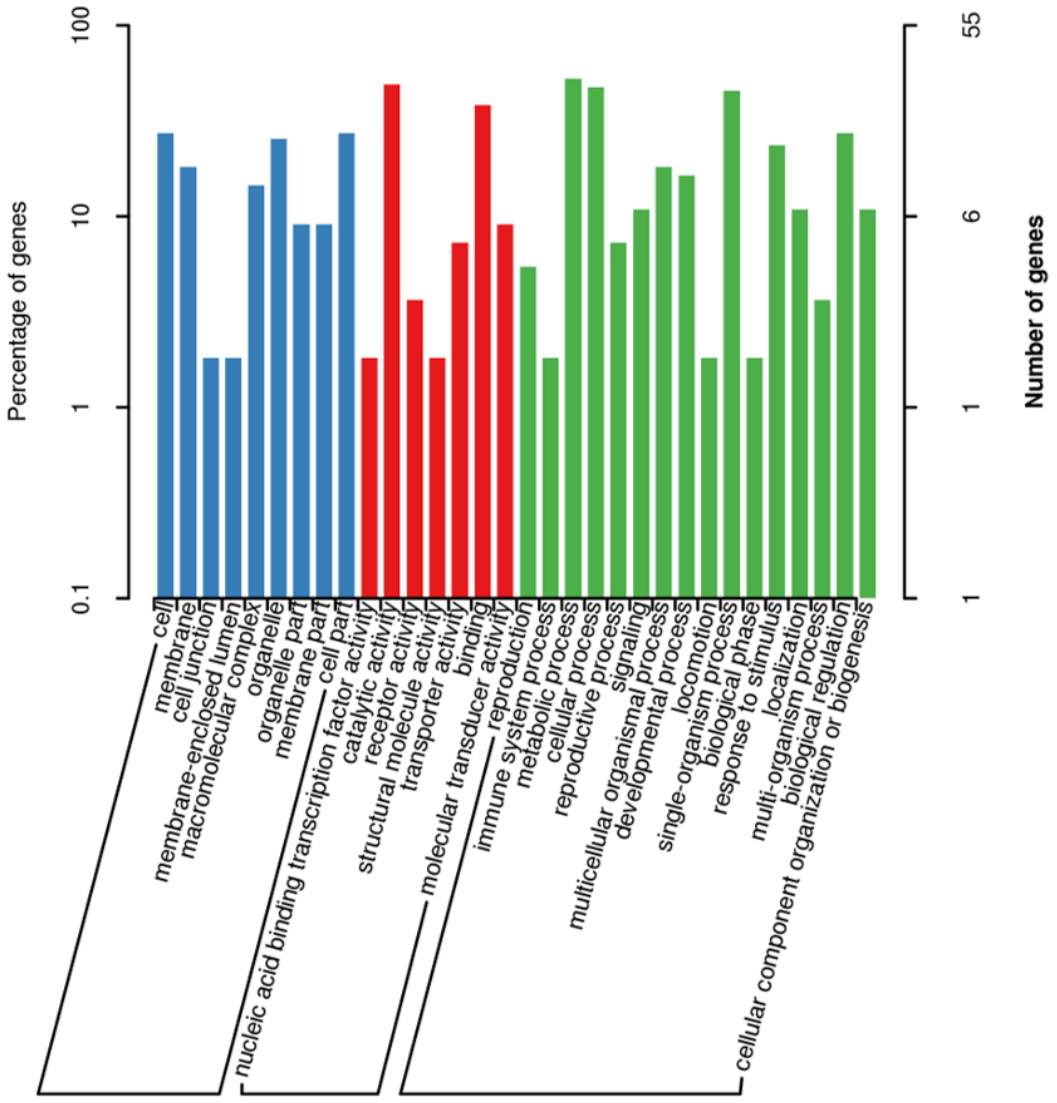

Figure 2

Clustering map of GO in candidate region: The abscissa is the content of each category of GO, the left of the ordinate is the percentage of the number of genes, and the right is the number of genes. This figure shows the gene classification of each secondary function of GO in the context of all genes in the associated region.

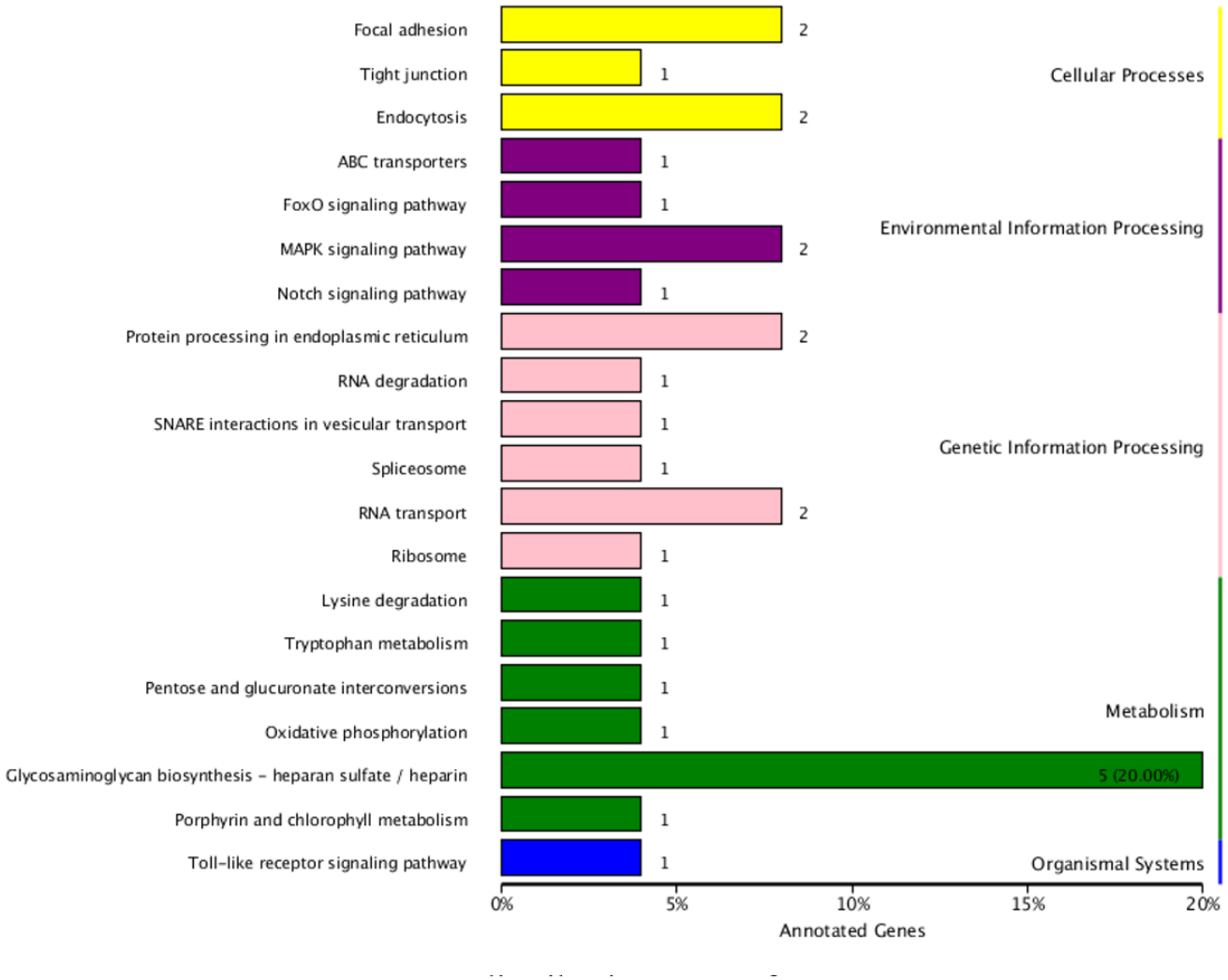

Figure 3

KEGG distribution map of genes: Top 20 KEGG pathways of genes containing SNPs associated with resistance to GCRV in ENU. The number of SNPs is shown behind the corresponding pathways.

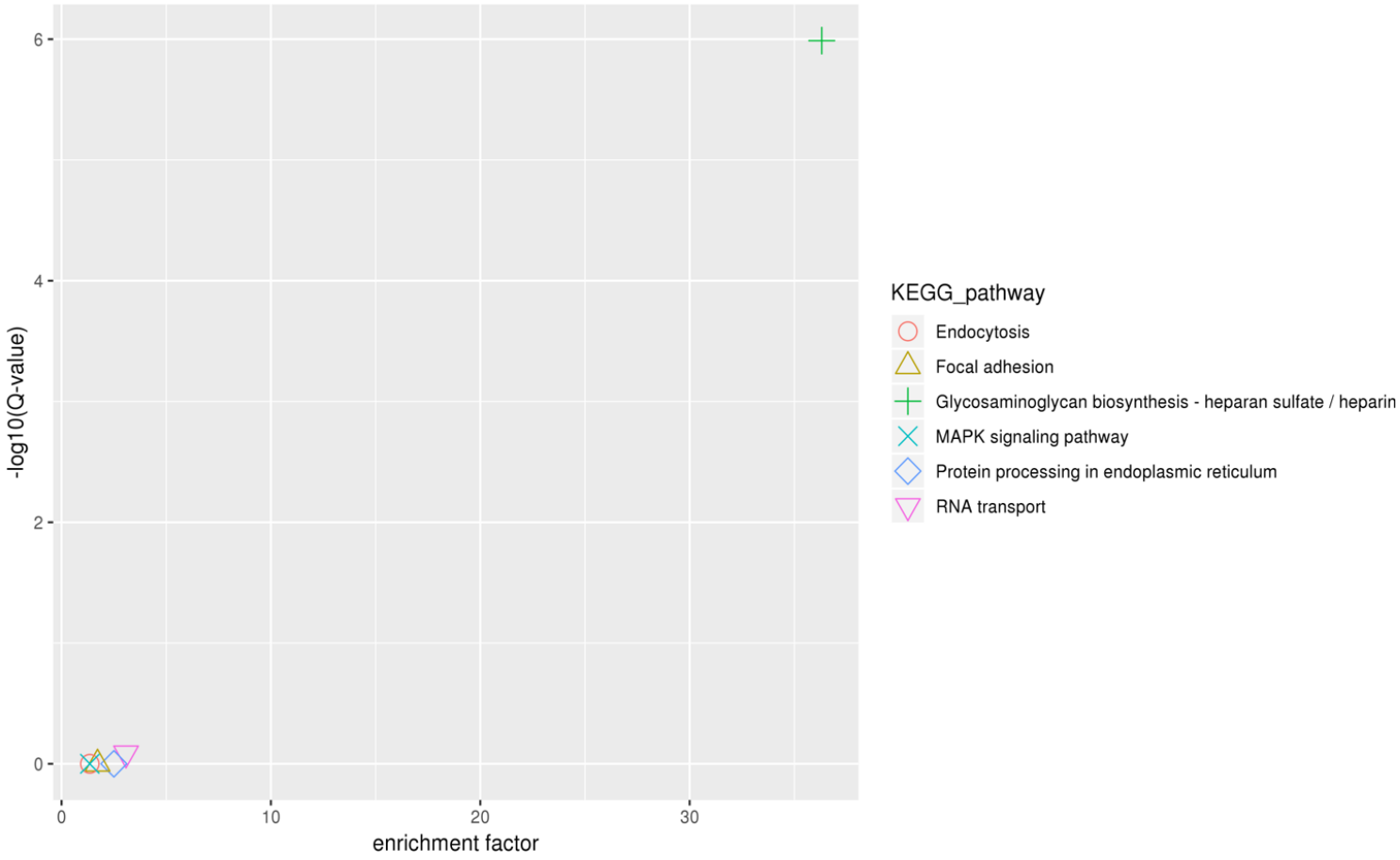

Figure 4

KEGG enrichment of genes in candidate regions: The enrichment pathways with more enrichment factor and significance values were “Glycosaminoglycan biosynthesis-heparan sulfate / heparin (ko00534)”, “RNA transport (ko03013)”, “Protein processing in endoplasmic reticulum (ko04141)”, “Focal adhesion (ko04510)”, “Endocytosis (ko04144)”, and “MAPK signaling pathway (ko04010)”.

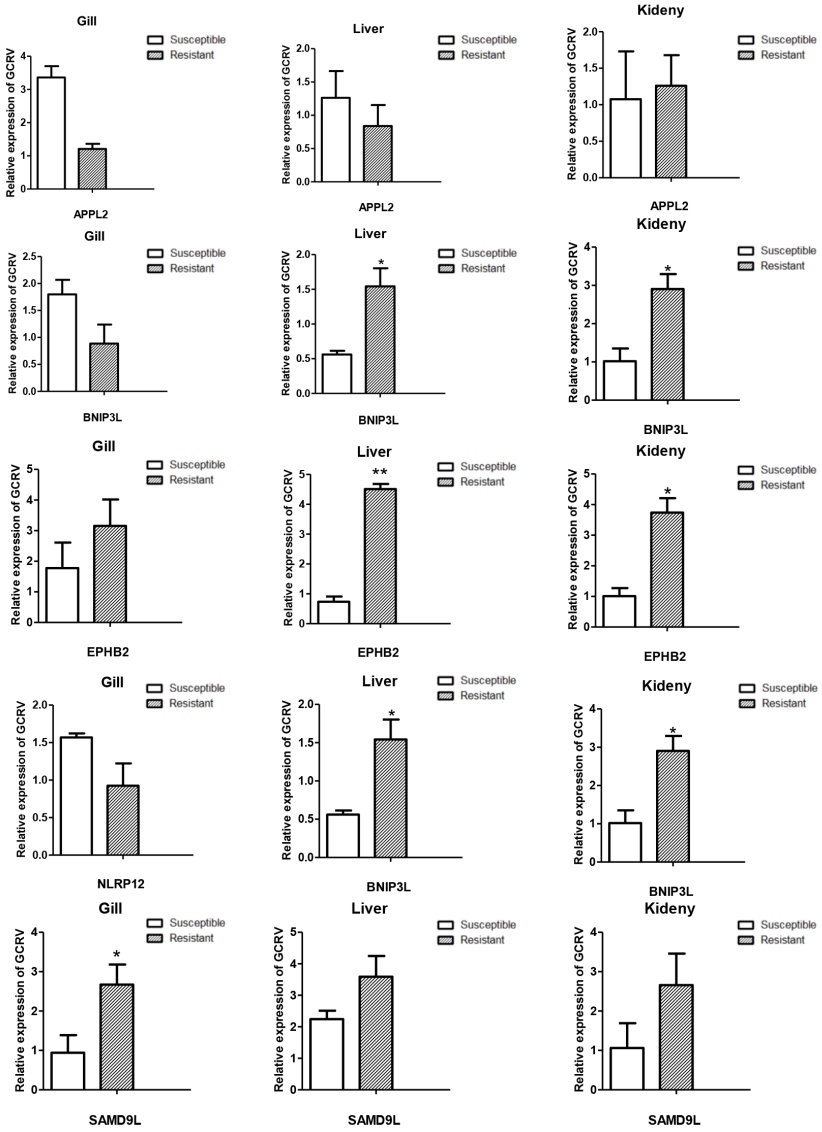

Figure 5

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 ± 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.

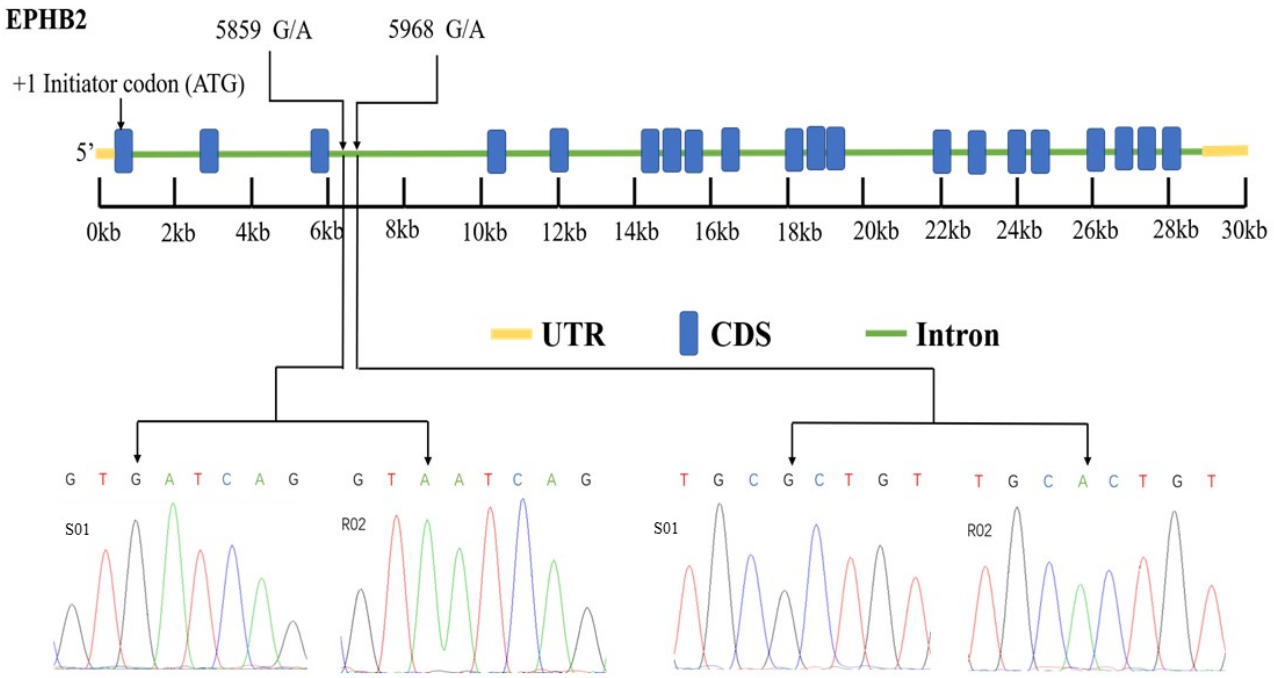

Figure 6

Genomic structure and SNP confirmation sites of EPHB2 gene in ENU: 5'UTR start site was regarded as the +1-initiator codon (ATG) in EPHB2 shown with blue colour (CDS). In EPHB2, SNPs were found out in 2 position which has been mentioned with chromosomal position 5859, 5968 determine a GA mutation in S01 (susceptible)/R02 (resistant) group. Green colour shows intron region.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementaryfile1.xls
- supplementaryfile2.xls
- supplementaryfile3.xls
- supplementaryfile4.xls
- supplementaryfile5.xls
- supplementaryfile6.xls