

Article

Predict the Relationship between Gene and Large Yellow Croaker's Economic Traits

Xiangxiang Zeng ^{1,†}, Shuting Jin ^{1,†}, Jin Jiang ¹, Kunhuang Han ², Xiaoping Min ^{1,*}
and Xiangrong Liu ^{1,*} 

¹ School of Information Science and Technology, Xiamen University, Xiamen 361005, China; xzeng@xmu.edu.cn (X.Z.); stjia.xmu@gmail.com (S.J.); jingdefengye1314@126.com (J.J.)

² State Key Laboratory of Large Yellow Croaker Breeding, Ningde Fufa Fisheries Company Limited, Ningde 352000, China; hankunhuang@foxmail.com

* Correspondence: mxp@xmu.edu.cn (X.M.); xrlu@xmu.edu.cn (X.L.)

† X.Z. and S.J. contribute equally to the work.

Received: 21 October 2017; Accepted: 6 November 2017; Published: 16 November 2017

Abstract: The importance of a gene's impact on traits is well appreciated. Gene expression will affect the growth, immunity, reproduction and environmental resistance of some fish, and then affect the economic performance of fish-related business. Studying the connection between gene and character can help elucidate the growth of fishes. Thus far, a collected database containing large yellow croaker (*Larimichthys crocea*) genes does not exist. The gene having to do with the growth efficiency of fish will have a huge impact on research. For example, the protein encoded by the IFIH1 gene is associated with the function of viral infection in the immune system, which affects the survival rate of large yellow croakers. Thus, we collected data through the published literature and combined them with a biological genetic database related to the large yellow croaker. Based on the data, we can predict new gene–trait associations which have not yet been discovered. This work will contribute to research on the growth of large yellow croakers.

Keywords: large yellow croaker; KATZ; trait gene prediction

1. Introduction

The large yellow croaker (*Larimichthys crocea*), whose meat is delicious and has therapeutic effects, is a crucial source for the fish markets. Larger yellow croakers, belonging to Asian croakers, are economically important, however few studies focus on the genes related to economic traits of the large yellow croaker. With the development of molecular biology technology [1], such as genomics in recent years, research on important economic traits of aquatic animals has attracted much attention. Molecular theory on sex, growth, disease resistance, cold and hypoxia, and other traits of common aquatic organisms has made some achievements, and a large number of related functional genes have also been identified from the large yellow croaker. In 2016, Shen et al. [2] sequenced and characterized the melanoma differentiation-associated antigen 5 (LcMDA5), laboratory of genetics and physiology 2 (LcLGP2) and mitochondrial antiviral signaling protein (LcMAVS) from the large yellow croaker. In 2017, Zou et al. [3] characterized an IRAK4 orthologue from the large yellow croaker, named Lc-IRAK4. Liu et al. [4] identified and characterized the chemokine receptor genes CXCR2 (LycCXCR2), CXCR3 (LycCXCR3), and CXCR4 (LycCXCR4) of the large yellow croaker and so on. This study gives us more information on the gene expression in large yellow croakers. However, due to the limitations of the growth of research on large yellow croakers, previous studies only focused on individual genes' expression and, until now, research on the gene–trait accession of this fish are still relatively scarce [5].

In order to systematically study the relationship between various traits and genes of the large yellow croaker, two contributions are made in this work. First, we collected large yellow croaker

gene–trait associations that can affect economic activities related to the fish by searching the existing literature. Second, on the basis of these data, we predicted unknown gene–trait relationships. We collected papers related to the large yellow croaker through PubMed (<https://www.ncbi.nlm.nih.gov/pubmed/>) and the Baidu Encyclopedia (<https://baike.baidu.com/>). While collecting these papers, we filtered out the economic traits associated with the large yellow croaker. We used relevant genes to find data on the corresponding protein and protein sequence through Uniprot. The collected data contained the following information: reference, gene symbol/name, Uniprot Entry, protein sequence, trait, cause, pubmedID, chromosome location, and more.

KATZ [6] is a method based on network measurements which calculates the similarities between nodes in a heterogeneous network in order to solve the problem of association prediction. This method takes the number of walks and walk lengths between nodes in a graph into account as effective similarity measures. We translated the problem of gene–trait association measurement into the number of connections between trait and gene nodes in the heterogeneous network. With the collected experimental data, we constructed a large yellow croaker gene–trait association network and further used a new computational model of KATZ measurement for the large yellow croaker gene–trait association prediction (KATZ-YC), based on the assumption that functionally similar genes tend to have similar interaction and non-interaction patterns with trait, and vice versa [7]. To our best of knowledge, the KATZ-YC is the first tool created for the large yellow croaker gene–trait association prediction. The credible prediction performance could be owed to the use of KATZ measurements, and to the use of the Gaussian Interaction Profile (GIP) [8] kernel to obtain similarities for genes and traits. We propose a method that relies on the known gene–trait association network topology information, trait similarity network, and gene similarity network. The trait similarity network and gene similarity network were obtained by the GIP kernel [9]. LOOCV (Leave one out cross validation) and K-fold cross validation were performed in order to evaluate the validity of this novel computational model based on known gene–trait associations obtained from the data we collected. As a result, we obtained the LOOCV framework and 2-fold and 5-fold cross validation with average AUCs (Area under curve) of about 0.7766, 0.7183 and 0.7476, respectively. It is anticipated that the KATZ-YC could be used to obtain more novel genes associated with important large yellow croaker traits, thereby facilitating the study of the genes and traits of the fish. Based on the existing research literature, we collected large yellow croaker gene–trait associations, including 24 traits and 102 genes. Importantly, the prediction accuracy was not ideal due to the lack of data. With the development of molecular technology, the relevant research will increase remarkably, making more data available, thus improving predictive performance. Moreover, this predicting method is more convenient and effective because it is completely based on known gene–trait associations.

2. Results

In this article, we collected the genes related to the economic traits of large yellow croakers. The existing studies on the genes and traits of large yellow croakers are still relatively few, and the related data is inadequate. We know that similar genes may contribute to similar traits. Based on this phenomenon and on the data we collected, new methods of computing status with KATZ [10] were developed in order to predict probable existing associations between genes and traits. We proposed the large yellow croaker gene–trait association prediction. In order to obtain similarity between genes and traits for the prediction of potential associations, we used the GIP kernel [11]. We used the KATZ-YC to predict associations between genes and traits. To obtain experimental results, we used LOOCV and K-fold cross validation. The KATZ-YC achieved a prediction performance with average AUCs of 0.7766, 0.7183 and 0.7476 when using the LOOCV framework and 2-fold and 5-fold cross validations, respectively. The prediction new genes that affect traits was achieved by using known gene–trait relationships that can help promote the survival rates of large yellow croakers and that can improve economic efficiency.

2.1. Leave-One-Out Cross Validation

We evaluated the performance of the KATZ-YC on gene–trait link prediction with the LOOCV method, based on the data we collected from various papers published about biological information. Many gene–trait associations from these data were correct. In the validation framework of LOOCV [12], each known gene–trait association was left out in turn for testing, and the other gene–trait associations were used as training samples for model learning. An important part of implementing this kind of validation is to obtain nearly all the samples for training, which results in a distribution of the training sample subjects similar to the distribution of the population samples. Also, there were no random factors affecting data during the experiment. Thus, the process of our validation is totally replicable. However, our research also had shortcomings. The cost of calculation was high because almost all the samples were used to train model for each round. When the sample population was large, LOOCV was difficult to implement in practice unless the training procedure for models was fast or was done in parallel in order to save time.

If the rank of a given testing sample was higher than the given threshold, the prediction was considered successful. We set different thresholds in order to obtain corresponding true positive rates (TPR, sensitivity) and false positive rates (FPR, 1-specificity) [13]. Thus, the receiver–operating characteristics (ROC) curve could be plotted by drawing TPR versus FPR at different thresholds. Areas under the ROC curve were calculated in order to evaluate the prediction performance of the KATZ-YC. A perfect prediction was an AUC value of 1. An AUC value of 0.5 indicated random performance [14].

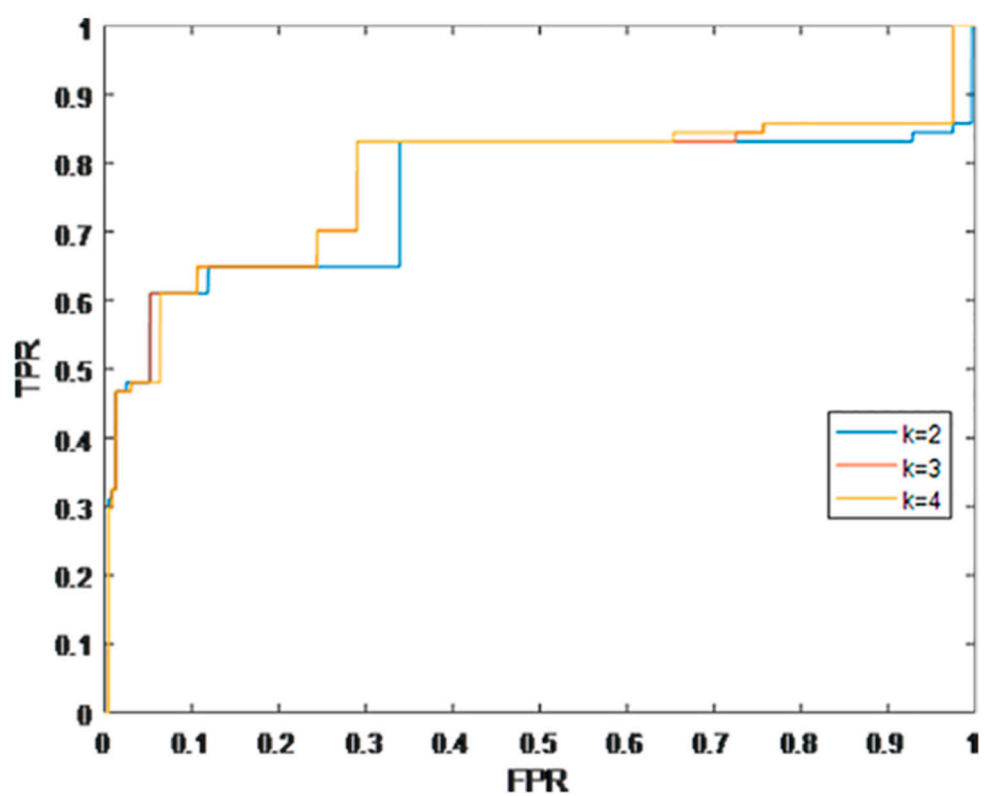
For the KATZ-YC, k is an important parameter, which represent the number of walks. The k value affected the prediction performance for gene–trait associations. We performed a series of comparative experiments in order to estimate the influence of k . As a result, the KATZ-YC achieved a better prediction performance when k was set as 2 (Table 1 and Figure 1a). Specifically, when k was set as 3, the KATZ-YC achieved AUC of 0.7766 in the framework of global LOOCV.

2.2. K-Fold Cross Validation

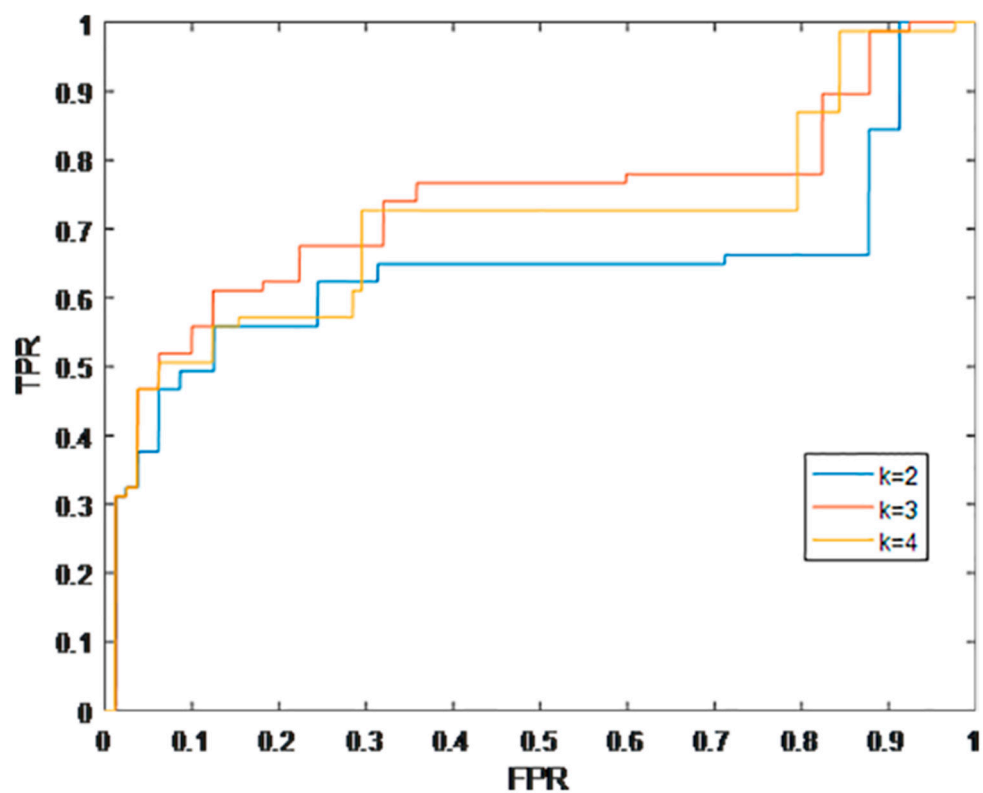
K-fold cross validation consists of dividing all samples into K copies. Generally, the number of each copy is equal or similar. One copy is taken as the test sample and the remaining $K-1$ as the training sample. This process repeats K times, and the final average test results can estimate the performance of the model. The final result was obtained by calculating the average result of these K results, or by using another way to obtain a single result. An advantage of implementing K-fold validation is that it randomly selects subsamples for training and validation repeatedly, and takes the results of each test into consideration. For further evaluation of the KATZ-YC prediction performance, we used 2-fold and 5-fold cross validation. Random sample division for performance evaluation can cause potential bias. As much as possible, in order to eliminate the potential bias, we divided the gene–trait associations many times and repeated experimental verification of the corresponding ROC curves. AUCs were obtained in a similar way as for LOOCV. The KATZ-YC achieved a prediction performance with average AUCs of 0.7183 and 0.7476 (with a standard deviation of 0.03) when using 2-fold (Table 2 and Figure 1b) and 5-fold cross validations (Table 3 and Figure 1c), respectively. We used the LOOCV and K-fold cross validation to test whether the method of the KATZ-YC prediction performance was trustworthy and valid. The gene–trait associations with higher ranks would be verified by experimental observations in the future.

Table 1. Performance comparison among different parameter settings ($k = 2, 3, 4$) in the framework of LOOCV.

K = 2	K = 3	K = 4
0.7564	0.7766	0.7760



(a)



(b)

Figure 1. Cont.

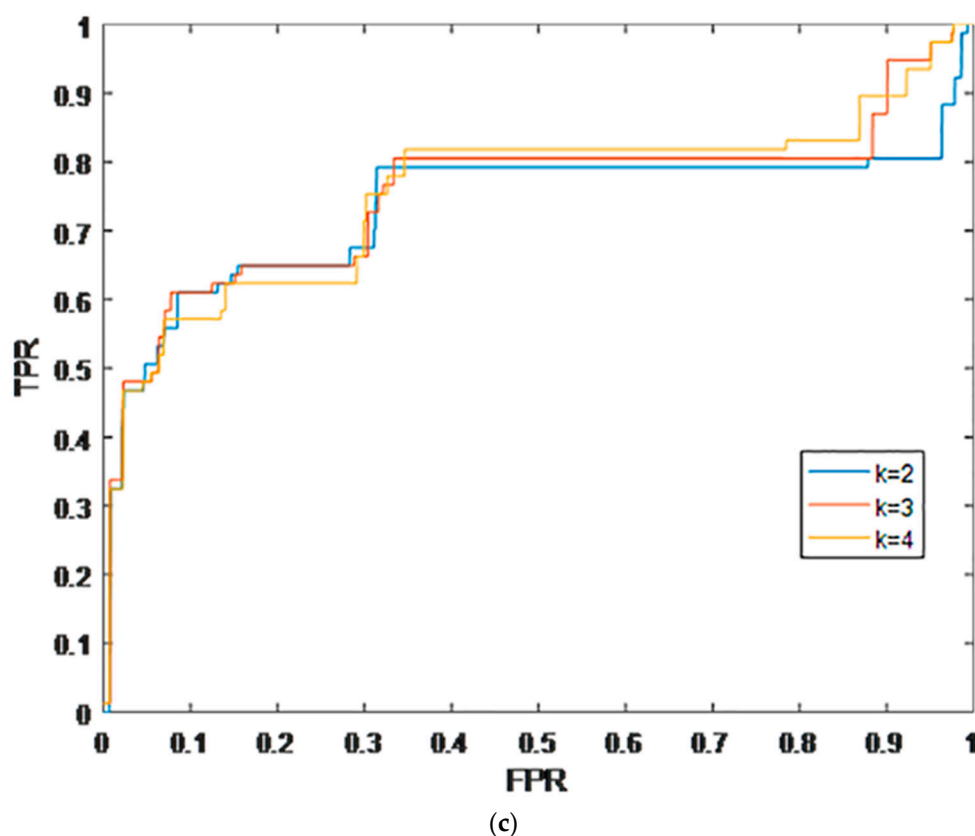


Figure 1. Prediction performance of KATZ-YC with different parameter settings ($k = 2, 3$ and 4) in terms of the receiver–operating characteristics (ROC) curve and area under the curve (AUC), based on K-fold cross validation: (a) based on Leave One Out Cross Validation (LOOCV); (b) based on 2-fold cross validation; and (c) based on 5-fold cross validation.

Table 2. Performance comparison among different parameter settings ($k = 2, 3, 4$) in the framework of 2-fold cross validation.

K = 2	K = 3	K = 4
0.7036 ± 0.03	0.7183 ± 0.03	0.7132 ± 0.05

Table 3. Performance comparison among different parameter settings ($k = 2, 3, 4$) in the framework of 5-fold cross validation.

K = 2	K = 3	K = 4
0.7338 ± 0.03	0.7476 ± 0.03	0.7357 ± 0.04

3. Materials and Methods

3.1. Materials

The gene–trait associations were mainly collected from published papers related to the large yellow croaker. The database contained information about the relevant genes, the corresponding traits that have been generated, and the causes for the trait (e.g., genes of *PPAR α* [13], *COX2* [15], *IL-1 β* [16] and traits of adipose [17], defense [18], and growth [19], respectively). As a result, 77 gene–trait associations (including 24 large yellow croaker traits and 102 genes) were collected from more than 1000 papers (see Supplementary Materials). We constructed a gene–trait matrix of size of 102×24 as the information source, among which 1 represented the gene and trait which have an association.

3.2. Methods

3.2.1. Problem to Formalize [8]

We needed to determine how to predict a new link based on a known gene–trait association network. We assumed that set $X_g = \{g_1, g_2, g_3 \dots g_i\}$ and $X_t = \{t_1, t_2, t_3 \dots t_i\}$ represented genes and traits, respectively. After collecting existing data, some associations between genes and traits were determined, which helped us build a heterogeneous network to express connections between genes and traits. In this network, we treated genes and traits as nodes and connections as edges. We used binary adjacency matrix to express this network: if A_{ij} was 1, then gene i and trait j were connected, or they were not (see Figure 2).

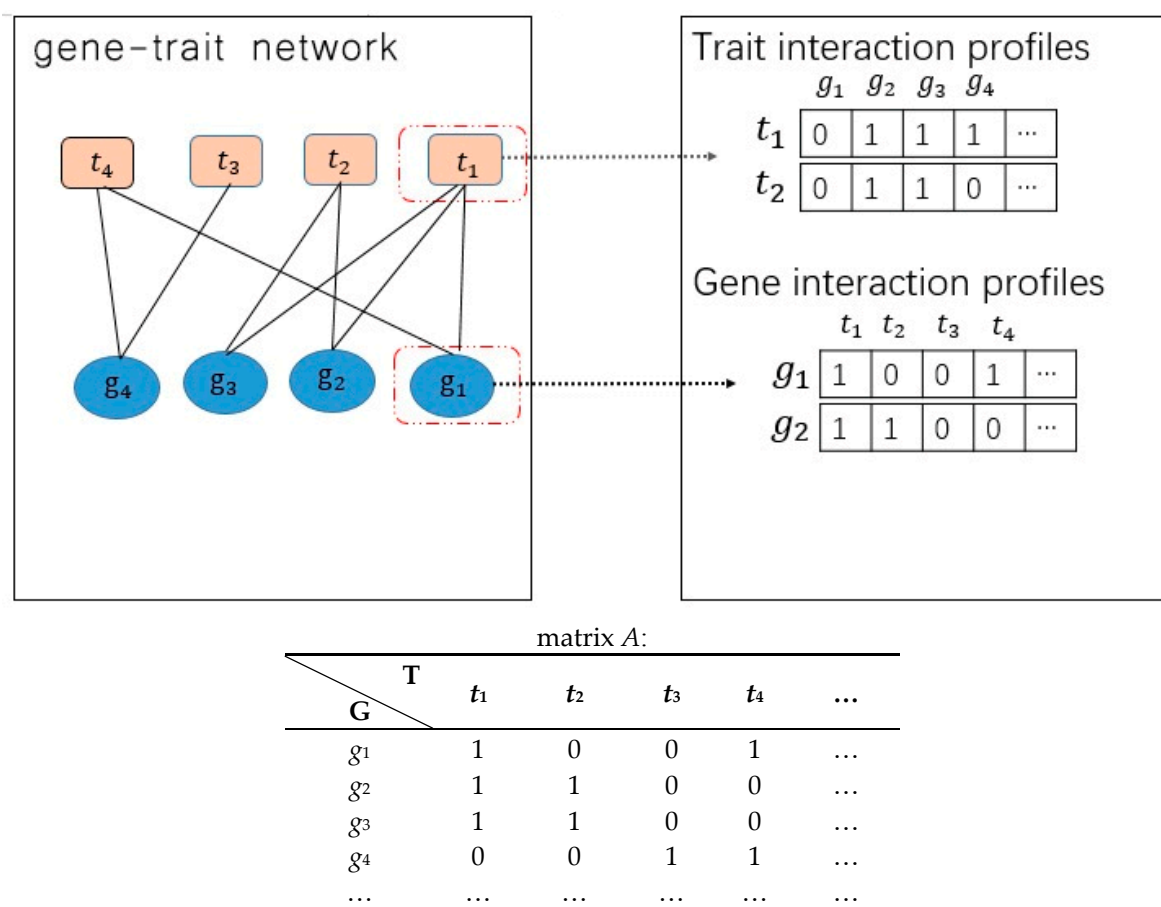


Figure 2. Description the use of a known gene–trait network to obtain trait interaction profiles, gene interaction profiles and the corresponding binary adjacency matrix A . The orange node represents the trait, and the blue node represents the gene in the network.

3.2.2. Gaussian Interaction Profile Kernel Similarity for Genes

We made the following assumption before our experiments [20]: that functionally similar genes share the same pattern. If a gene g and a trait t are related, then genes functionally similar to gene g are more likely to be related, with traits similar to the specific trait t . The Gaussian interaction profile kernel similarity for genes was constructed from a known gene–trait association network. We used a similar assumption with relation to genes. By using genes, we proposed that the interaction profile $IP(g_i)$ was the binary vector encoding the existence or lack of interaction with every target in the considered gene–trait network [8]. Given a gene i , the interaction profile $IP(g_i)$ would be defined as the i th column of the matrix A . The interaction profiles extracted from a gene–trait interaction network

can be used as feature vectors for a classifier. Figure 2 illustrates the construction of interaction profiles [2]. Then, based on the distribution of the relationship between the gene and the trait, we used the following formula to calculate the Gaussian interaction profile kernel similarity, KG , between each gene pair (i.e., g_i and g_j):

$$KG(g_i, g_j) = \exp\left(-\gamma_g \|IP(g_i) - IP(g_j)\|^2\right) \quad (1)$$

$$\gamma_g = \gamma'_g / \left(\frac{1}{n_g} \sum_{k=1}^{n_g} \|IP(g_k)\|^2 \right) \quad (2)$$

where γ_g is the normalized kernel bandwidth based on the new bandwidth parameter γ'_g ; n_g is the number of genes; and $KG(g_i, g_j)$ is the Gaussian interaction profile kernel similarity between gene g_i and g_j .

3.2.3. Gaussian Interaction Profile Kernel Similarity for Trait

By making an analogy between the calculating methods of gene associations and trait associations, we can use the same method to obtain connections of traits. The Gaussian interaction profile kernel similarity for traits (KT) was computed in a similar way as for genes, as follows:

$$KT(t_i, t_j) = \exp\left(-\gamma_t \|IP(t_i) - IP(t_j)\|^2\right) \quad (3)$$

$$\gamma_t = \gamma'_t / \left(\frac{1}{n_t} \sum_{k=1}^{n_t} \|IP(t_k)\|^2 \right) \quad (4)$$

where γ_t is the normalized kernel bandwidth based on the new bandwidth parameter γ'_t ; n_t is the number of traits; and $KT(t_i, t_j)$ is the Gaussian interaction profile kernel similarity between trait t_i and t_j . A relationship network of all traits of similarity is presented in Figure 3. According to the experimental results, and in order to better display the similarity of the traits in the graph, we kept similarity correlations of greater than 0.35.

3.2.4. KATZ-YC

KATZ measure has been successfully applied in many areas, such as link prediction of some microbial-diseases [20] associations, social network prediction [6], and the prediction of lncRNA-disease [21] associations. KATZ measure is a web-based approach that predicts the relevance of related nodes in a network by calculating the similarity between nodes in a heterogeneous network [20]. In this paper, we proposed a new model based on the KATZ algorithm in order to explore the association of genetic traits in the large yellow croaker (*Pseudosciaena crocea*) [22]. The KATZ method calculates the similarity between nodes based on the number and length of paths. Thus, we calculated the relationships between the genes and traits of the large yellow croaker in such a way as to represent them as the number of paths between genes and traits in heterogeneous networks. Similarities between nodes would be affected by number of steps and step length, which means that shorter distances indicate a high possibility of correlations and longer ones indicate a low possibility of correlations. Therefore, we incorporated steps and step lengths as effective information for the process of predicting possible associations between genes and traits. This method considers both the quantity and length of paths between nodes in a graph as valid similarity metrics. Thus, we transformed the problem of measuring gene–trait associations by counting the number of walks of connections between each gene and each trait in the heterogeneous network. Heterogeneous networks consist of genetic similarity networks and feature similarity networks. The genetic similarity network was based on the gene Gaussian interaction profile, and the trait similarity network was based on the trait Gaussian interaction

profile. The known gene–trait association network was constructed based on the collected databases and literature.

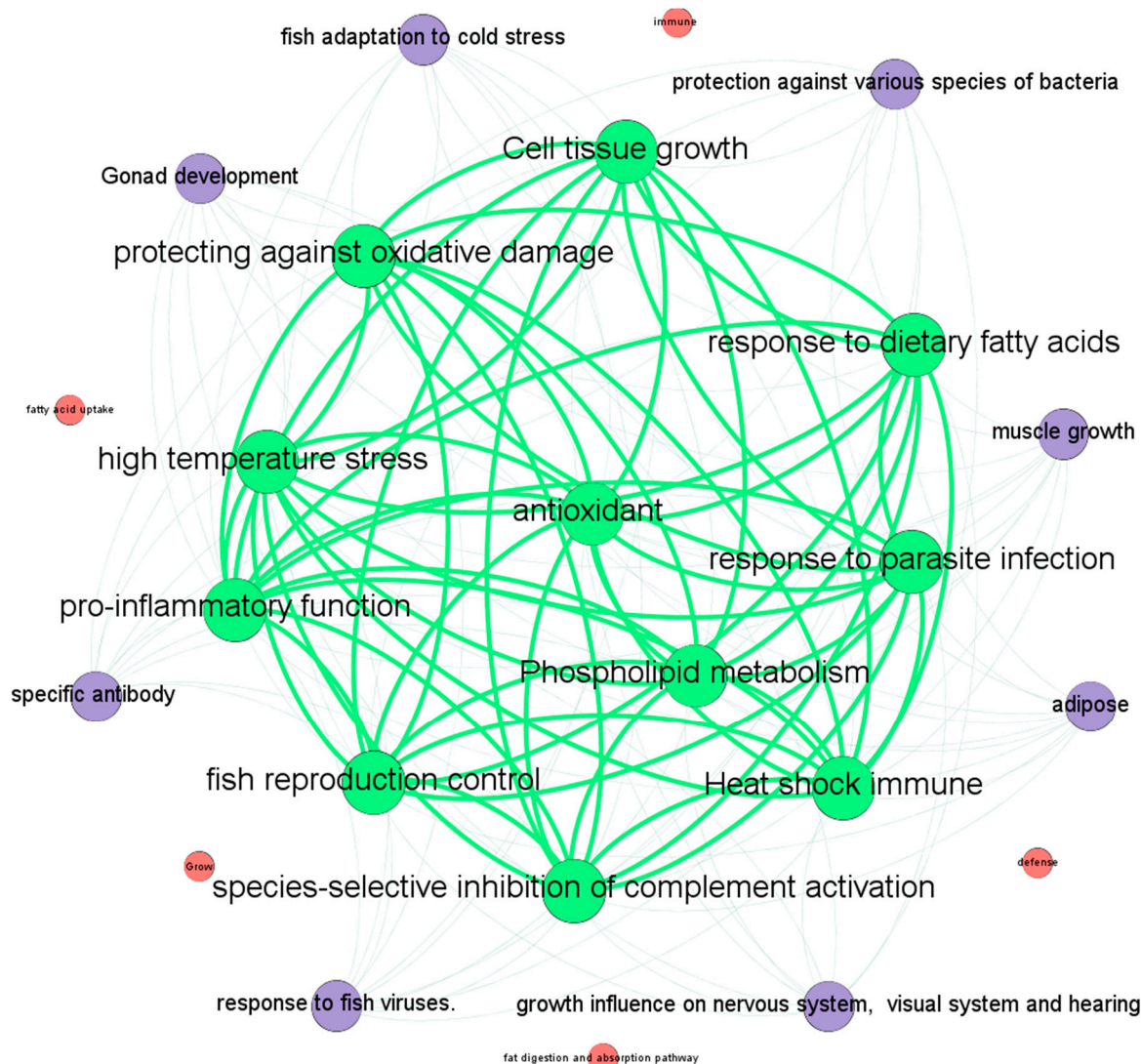


Figure 3. The similarity network between traits, where the nodes represent traits, the weighted edges represent the similarity associations between nodes, the size of a node corresponds to the number of neighbors of the node, and nodes in the network with identical colors have the same similarity value.

The KATZ-YC first calculated the number of walks between gene nodes (i.e., g_i) and trait nodes (i.e., t_i) in the known gene–trait association network. As mentioned, we stored the gene–trait association heterogeneous network in an adjacent matrix A . We calculated A_{ij}^l to obtain the number of l -length walks between g_i (gene) and t_i (trait). By using the Gaussian interaction profile kernel, we obtained homogeneous networks of similarities in genes and traits. In order to implement the similarities of genes and traits into prediction of links related to genes and traits, we further integrated the gene similarity matrix KG , the trait similarity matrix KT , and the adjacent matrix A as follows:

$$A^* = \begin{bmatrix} KT & A \\ A^T & KG \end{bmatrix} \quad (5)$$

The new synthesis matrix A^* here can be used to further predict gene–trait associations, which allows the application of the KATZ-YC to new genes and traits with unknown associations if we

could acquire further similarity information without relying on known gene–trait network topology information. All walks of different lengths were integrated in order to obtain a single measurement of each gene–trait pair. However, taking into account the different contributions of different steps, greater similarity was obtained between the nodes that are closer, whereas the correlation between nodes with longer steps may be smaller. Thus, we introduced a nonnegative coefficient Sequence ∂_l to control the effect of longer steps. The corresponding coefficients of shorter walks in ∂_l would be larger than those corresponding to longer walks. We supposed that the step between the pair of nodes was l_1 and that l_2 ($l_1 < l_2$) corresponded to $\partial_{l_1} > \partial_{l_2}$.

Therefore, we proposed the entity $S(g_i, t_i)$ of matrix S to calculate the potential association probability between gene g_i and trait t_i , as follows:

$$S(g_i, t_j) = \sum_{l=1}^k \partial_l A^{*l}(i, j) \quad (6)$$

Here, we replaced ∂_l by ∂^l and transformed this equation into the matrix form:

$$S = \sum_{l \geq 1} \partial^l A^{*l} = (I - \partial A^*)^{-1} - I \quad (7)$$

Here, matrix S with dimensions of $(24 + 102) \times (24 + 102)$ depicted the association possibilities of all the gene–trait pairs. Furthermore, we represented the partitioned matrix A^* in a form similar to matrix S' as follows:

$$S' = \begin{bmatrix} S11 & S12 \\ S21 & S22 \end{bmatrix} \quad (8)$$

We finally predicted the result for matrix S within the matrix $S12$, which could provide the correlative probability between each gene and trait. However, we collected a only small amount of data because the gene–trait association network is still sparse. Walks with long lengths in a sparse network may be unimportant and could perturb the link prediction. Thus, we set parameter k to be 2, 3 and 4 and evaluated the parameter k 's influence on the prediction performance. We calculated the respective unknown association probability when k was set as 2, 3 and 4 by the following formula. Here, the final prediction result matrix could be represented by matrices A , KT , and KG .

$$S_{K=2} = \partial \cdot A + \partial^2 \cdot (KT \cdot A + A \cdot KG) \quad (9)$$

$$S_{K=3} = S_{K=2} + \partial^3 (A \cdot A^T \cdot A + KM^2 \cdot A + KG \cdot A \cdot KT + A \cdot KT^2) \quad (10)$$

$$S_{K=4} = S_{K=3} + \partial^4 \cdot (KG^3 \cdot A + A \cdot A^T \cdot KG \cdot A + KG \cdot A \cdot A^T \cdot A + A \cdot KT \cdot A^T \cdot A) + \partial^4 \cdot (A \cdot A^T \cdot A \cdot KT + KG^2 \cdot A \cdot KT + KG \cdot A \cdot KT^2 + A \cdot KD^3) \quad (11)$$

4. Discussion and Conclusions

Considering research about gene–trait associations of large yellow croakers was scarce, and that databases of this species were lacking, we achieved the following two tasks in this paper. First, we collected genes related to economic traits of large yellow croakers by searching various literatures and collated expressions, protein sequences, and other information associated with these genes. Second, we used a heterogeneous network to describe gene and trait similarity based on gathered known information about gene–trait associations. We further used that information to implement the KATZ-YC algorithm, improved from KATZ, in order to perform the prediction work on unknown gene–trait associations.

With the development of molecular technology, more correlation between genes and traits will be found. We can build a database by collecting more data based on the abovementioned ones, and we

can execute predictive models. Through the association between genes and traits, we have obtained a possible existing link between some economic traits. This work contributes to filling the void in associations between genes and economic traits of large yellow croakers. In the aquaculture of the large yellow croaker, the probability of occurrence of other traits can be predicted according to the appearance of some traits. These traits could effectively promote growth and reproduction or the prevention of possible illnesses. This understanding can further accelerate the development of the large yellow croaker breeding industry.

Acknowledgments: This work was supported by the National Natural Science Foundation of China (No. 61472333, 61772441 and 61272152), Natural Science Foundation of the Higher Education Institutions of Fujian Province (No. JZ160400), Natural Science Foundation of Fujian Province (No. 2015J06019), President Fund of Xiamen University (No. 20720170054).

Author Contributions: X.Z. and S.J. conceived and designed the experiments; S.J. collected the dataset; Z.X. performed the experiments; S.J. wrote the paper; X.Z. and J.J. analyzed the data; K.H. contributed reagents/materials/analysis tools; X.M. and X.L. discussed the results and improved the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Brown, S.M. Molecular Biology Technology. In *Essentials of Medical Genomics*, 2nd ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2008; pp. 31–51.
2. Shen, B.; Hu, Y.; Zhang, S.; Zheng, J.; Zeng, L.; Zhang, J.; Zhu, A.; Wu, C. Molecular characterization and expression analyses of three RIG-I-like receptor signaling pathway genes (MDA5, LGP2 and MAVS) in *Larimichthys crocea*. *Fish Shellfish Immunol.* **2016**, *55*, 535. [[CrossRef](#)] [[PubMed](#)]
3. Zou, P.F.; Huang, X.N.; Yao, C.L.; Sun, Q.X.; Li, Y.; Zhu, Q.; Yu, Z.X.; Fan, Z.J. Cloning and functional characterization of IRAK4 in large yellow croaker (*Larimichthys crocea*) that associates with MyD88 but impairs NF- κ B activation. *Fish Shellfish Immunol.* **2017**, *63*, 452–464. [[CrossRef](#)] [[PubMed](#)]
4. Liu, X.; Kang, L.; Liu, W.; Lou, B.; Wu, C.; Jiang, L. Molecular characterization and expression analysis of the large yellow croaker (*Larimichthys crocea*) chemokine receptors CXCR2, CXCR3, and CXCR4 after bacterial and poly I:C challenge. *Fish Shellfish Immunol.* **2017**, *70*, 228–239. [[CrossRef](#)] [[PubMed](#)]
5. Kankainen, M.; Setälä, J.; Kause, A.; Quinton, C.; Airaksinen, S.; Koskela, J. Economic values of supply chain productivity and quality traits calculated for a farmed European whitefish breeding program. *Aquac. Econ. Manag.* **2016**, *20*, 131–164. [[CrossRef](#)]
6. Katz, L. A new status index derived from sociometric analysis. *Psychometrika* **1953**, *18*, 39–43. [[CrossRef](#)]
7. Chen, X.; Yan, G.Y. Novel human lncRNA-disease association inference based on lncRNA expression profiles. *Bioinformatics* **2013**, *29*, 2617. [[CrossRef](#)] [[PubMed](#)]
8. Van Laarhoven, T.; Nabuurs, S.B.; Marchiori, E. Gaussian interaction profile kernels for predicting drug-target interaction. *Bioinformatics* **2011**, *27*, 3036–3043. [[CrossRef](#)] [[PubMed](#)]
9. Li, J.Q.; Rong, Z.H.; Chen, X.; Yan, G.Y.; You, Z.H. MCMDA: Matrix completion for MiRNA-disease association prediction. *Oncotarget* **2017**, *8*, 21187–21199. [[CrossRef](#)] [[PubMed](#)]
10. Dong, L.; Li, Y.; Yin, H.; Le, H.; Rui, M. The Algorithm of Link Prediction on Social Network. *Math. Probl. Eng.* **2013**, *2013*, 125123. [[CrossRef](#)]
11. Laarhoven, T.V.; Nabuurs, S.B.; Marchiori, E. *Gaussian Interaction Profile Kernels for Predicting Drug-Target Interaction*; Oxford University Press: Oxford, UK, 2011.
12. Shao, Z.; Meng, J.E.; Wang, N. An Efficient Leave-One-Out Cross-Validation-Based Extreme Learning Machine (ELOO-ELM) With Minimal User Intervention. *IEEE Trans. Cybern.* **2016**, *46*, 1939–1951. [[CrossRef](#)] [[PubMed](#)]
13. Dong, X.; Tan, P.; Cai, Z.; Xu, H.; Li, J.; Ren, W.; Xu, H.; Zuo, R.; Zhou, J.; Mai, K.; et al. Regulation of FADS2 transcription by SREBP-1 and PPAR- α influences LC-PUFA biosynthesis in fish. *Sci. Rep.* **2017**, *7*, 40024. [[CrossRef](#)] [[PubMed](#)]
14. Chen, X.; You, Z.H.; Yan, G.Y.; Gong, D.W. IRWRLDA: Improved random walk with restart for lncRNA-disease association prediction. *Oncotarget* **2016**, *7*, 57919. [[CrossRef](#)] [[PubMed](#)]

15. Liao, K.; Yan, J.; Mai, K.; Ai, Q. Dietary lipid concentration affects liver mitochondrial DNA copy number, gene expression and DNA methylation in large yellow croaker (*Larimichthys crocea*). *Comp. Biochem. Physiol. Part B* **2015**, *193*, 25–32. [[CrossRef](#)] [[PubMed](#)]
16. Zhu, Q.; Li, C.; Yu, Z.X.; Zou, P.F.; Meng, Q.X.; Yao, C.L. Molecular and immune response characterizations of IL-6 in large yellow croaker (*Larimichthys crocea*). *Fish Shellfish Immunol.* **2016**, *53*, 263–273. [[CrossRef](#)] [[PubMed](#)]
17. Qiu, H.; Jin, M.; Li, Y.; Lu, Y.; Hou, Y.; Zhou, Q. Dietary Lipid Sources Influence Fatty Acid Composition in Tissue of Large Yellow Croaker (*Larmichthys crocea*) by Regulating Triacylglycerol Synthesis and Catabolism at the Transcriptional Level. *PLoS ONE* **2017**, *12*, e0169985. [[CrossRef](#)] [[PubMed](#)]
18. Sun, P.; Bao, P.; Tang, B. Transcriptome analysis and discovery of genes involved in immune pathways in large yellow croaker (*Larimichthys crocea*) under high stocking density stress. *Fish Shellfish Immunol.* **2017**. [[CrossRef](#)] [[PubMed](#)]
19. Chen, Y.; Wang, Y.; He, S.; Zhu, Z. Cloning and sequencing of the growth hormone gene of large yellow croaker and its phylogenetic significance. *Biochem. Genet.* **2004**, *42*, 365. [[CrossRef](#)] [[PubMed](#)]
20. Chen, X.; Huang, Y.A.; You, Z.H.; Yan, G.Y.; Wang, X.S. A novel approach based on KATZ measure to predict associations of human microbiota with non-infectious diseases. *Bioinformatics* **2016**, *33*, 733–739. [[CrossRef](#)] [[PubMed](#)]
21. Chen, X. KATZLDA: KATZ measure for the lncRNA-disease association prediction. *Sci. Rep.* **2015**, *5*, 16840. [[CrossRef](#)] [[PubMed](#)]
22. Xiao, S.; Wang, P.; Zhang, Y.; Fang, L.; Liu, Y.; Li, J.T.; Wang, Z.Y. Gene map of large yellow croaker (*Larimichthys crocea*) provides insights into teleost genome evolution and conserved regions associated with growth. *Sci. Rep.* **2015**, *5*, 18661. [[CrossRef](#)] [[PubMed](#)]

Sample Availability: Not available.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).