Multi-omics data reveal genes that are relevant for muscular amino acids in the common carp

Youxiu Zhu

Shanghai Ocean University

Hanyuan Zhang

Chinese Academy of Fishery Sciences

Peng Xu

Xiamen University

Zixia Zhao

Chinese Academy of Fishery Sciences

Jianxin Feng

Henan Academy of Fishery Science

Biyin Wu

Shanghai Ocean University

Yanliang Jiang

Chinese Academy of Fishery Sciences

Chuanju Dong

Henan Normal University

Jian Xu (✉ xuj@cafs.ac.cn)

Chinese Academy of Fishery Sciences <https://orcid.org/0000-0003-0274-4268>

Research article

Keywords: common carp, muscular amino acids, GWAS, transcriptome, methylation

Posted Date: June 8th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-27730/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

Fish muscular amino acids are a series of essential nutrients that embrace essential amino acids, branched-chain amino acids and flavorful amino acids. Previous studies have found that amino acids have important physiological effects on fish growth and development, as they are involved in maintaining nitrogen balance and in the formation of enzymes and hormones. Amino acids, such as aspartic acid, glutamic acid, glycine and alanine, that can have a significant effect on fish umami taste are called flavorful amino acids. Nevertheless, the studies on the genetic mechanisms of amino acid metabolism in the common carp (*Cyprinus carpio*) are still limited.

Results

The purpose of this study was to examine the divergent patterns at the genomic, transcriptomic and epigenomic levels in fish with different amino acid contents. Genome-wide association analysis using 195 individuals of common carp was conducted, and 62 genes were identified to be associated with glycine, proline, and tyrosine content. RNA-Seq of samples with extreme contents of essential amino acids, branched-chain amino acids and flavorful amino acids was applied using brain, liver and muscle tissues, resulting in 1,643 differentially expressed genes. Whole-genome bisulfite sequencing identified 3,108 genes with differentially methylated promoters. Through the enrichment analysis of transcriptome and DNA methylation results, we screened out a series of enriched pathways associated with amino acid metabolism, including various categories of pathways spanning growth regulation, lipid metabolism, the citrate cycle and other signaling pathways. Integrated studies demonstrated prominent correlations between DGE and DMP for amino acid contents trait in brain and muscle tissues.

Conclusion

In summary, the multi-omics data revealed candidate genes and pathways correlated with amino acid metabolism. These results will promote the process of the genomic selection and breeding strategy in muscular amino acid contents of fish.

Background

Common carp (*Cyprinus carpio*) is one of the most widely cultured freshwater fishes all over the world. Most *C. carpio* are cultured in China, with an annual production of around three million tons. There are abundant populations of *C. carpio* in China, such as Yellow River carp, Songpu mirror carp, Hebao red carp, Xingguo red carp, Oujiang color carp and many hybrid strains [1]. Among these, the Yellow River carp occupies an important position due to its high nutritional value. The composition and content of amino acids are the key factors affecting the nutritional value and taste of the Yellow River carp.

Amino acids have been traditionally divided into essential amino acids (EAAs) and nonessential amino acids (NEAAs) according to whether the organism can synthesize the amino acids *in vivo* from the food resources. Presently, the definition of functional amino acids (FAAs) has received increasing attention. Functional amino acids are those that influence health, survival, growth, development, cell signaling, immune response and reproduction of the organisms by modulating key metabolic pathways [2]. Arginine, cysteine, leucine, methionine, tryptophan, tyrosine, aspartate, glutamic acid, glycine, proline and taurine have been defined as FAAs in human nutrition [3]. Amino acids are also divided into branched-chain amino acids (BCAAs, including leucine, valine and isoleucine) and straight-chain amino acids according to their chain shape. BCAAs regulate many key signaling pathways—the activation of the mTOR signaling pathway, for example [4]. In recent years, novel functions of BCAAs have been demonstrated, such as lipolysis and glucose consumption [4]. BCAAs, especially leucine, improve the ability of muscle protein synthesis through insulin-dependent and insulin-independent pathways [5]. Aspartic acid, glutamic acid, glycine and alanine, which have significant effects on fish umami taste, are called flavorful amino acids (FLAs).

In addition to the effects on meat quality and taste, the composition and content of amino acids are of great biological significance [6]. Fish growth is largely regulated by hormones and nutritional levels in a complex regulatory network. However, in this complex regulatory network, growth hormone (GH) and insulin-like growth factors (IGFs) play major roles. Amino acids can directly or indirectly act on the GH/IGF growth axis, thereby affecting fish growth [7]. The α -ketoacid, the catabolic product of amino acid, is regulated by the metabolic pathways of sugar or lipids with different characteristics. The α -ketoacid can be resynthesized into new amino acids, converted into sugars or fats, or oxidized and decomposed into CO_2 and H_2O in the tricarboxylic acid cycle with the release of energy [8].

The protein content of most fish is 15%–20% of the total mass, and EAAs account for 20%–50% of the total amino acid content [6]. Researchers studying the fatty acid composition, mineral composition and EAA content of 12 fish species found that the protein content of these fish was 18.7%–25.5%, and the EAA content was 1.93 to 8.73 mg/g. The content of crude protein and fat in teleosts varied widely, and the content of EAAs ($35.8\% \pm 2.6\%$) only accounted for half the amount of NEAAs ($64.2\% \pm 2.6\%$) [9]. The quality of food protein is determined by two parameters of amino acids: composition and content. EAAs are obtained from the diet, and each amino acid plays different roles in human metabolism. A recent study demonstrated that differences in fish muscle composition are under genetic control that encodes production of different types of proteins or amino acids [10]. However, due to the complexity of amino acid traits, it is difficult to significantly enhance the nutritional composition of muscle by traditional breeding strategies.

The early research was often limited to specific genes related to amino acid synthesis, but such an approach is insufficient to fully understand the mechanisms and pathways that underlie the traits. With the prompt progress of high-throughput sequencing technology, more research has adopted the methods of resequencing, transcriptome, methylation, proteome and metabolome analyses to explore the molecular network mechanisms of traits from different perspectives. The genome-wide association study

(GWAS) on the citrulline metabolism of watermelon identified two key candidate genes (*ferrochelatase* and *acetolactate synthase*) that were significantly related to citrulline content [11]. A multi-locus GWAS model was used to dissect biosynthetic processes of the free amino acids (FAAs) in bread wheat [12]. A total of 15 candidate genes affecting the anabolism of FAAs were identified. One candidate gene encoding tryptophan decarboxylase was associated with the content of tryptamine *in vitro*. A GWAS of amino acid content in oysters obtained 417 significant SNPs, and the genetic network analysis found that three node SNP regions were associated with glycogen, protein and aspartate content [13].

Solexa/Illumina RNA-Seq was employed to generate 52.5 million reads of mRNA sequences from Chinese perch fast muscle samples, and 21 amino acid transporter genes were annotated using protein and gene ontology databases [14]. Despite these results from previous studies, the gene regulatory network is complicated and remains obscure due to insufficient multi-omics evidence. Multi-omics studies usually combine two or more types of omics data, such as transcriptomics, genomics, epigenetics, proteomics and metabolomics. In teleosts, some biological phenomena have been uncovered through multi-omics methods. To examine molecular mechanisms of cold tolerance in *T. fasciatus*, Wen *et al.* integrated mRNA-Seq and metabolomics of *T. fasciatus* liver tissue to establish an mRNA-protein-metabolite interaction network associated with cold tolerance; their study identified differentially expressed metabolites, genes and proteins involved in membrane transport, fatty acid metabolism and signal transduction [15]. These results indicated that nutrient substances could be added to the fodder to help the overwintering of *T. fasciatus*. Wei *et al.* integrated transcriptomic and metabolomic analyses of flesh quality in the large yellow croaker; their results showed that feeding dietary hydroxyproline could improve flesh quality through changes in comprehensive metabolic pathways, including lipid metabolism, collagen synthesis metabolism and glycolysis [16]. A multi-omics analysis of the dynamics of muscular metabolites in leopard coral grouper that combined transcriptomic and metabolomic methods indicated that branched-chain amino acids played a role in energy production in the fish muscle tissues during the fasting period [17]. Zhang *et al.* conducted DNA methylation sequencing and comparative transcriptome analyses that uncovered the molecular mechanisms of body color variation in the crucian carp [18]. Integrating the methods of GWAS, transcriptome and methylation analyses, Zhang *et al.* identified candidate genes and pathways associated with fatty acid metabolism in the common carp [19]. Here, we conducted multi-omics analysis of muscular amino acid content in the Yellow River carp using genomic, transcriptomic and epigenomic data. This study demonstrates the potential metabolic mechanisms of amino acids in the Yellow River carp and provides an experimental foundation for improving the meat quality of Yellow River carp by genome-assisted breeding.

Results

Genotyping and phenotyping of *C. carpio* samples

Based on previous work [1, 19, 20], 199 samples were randomly collected from a cultivated population of *C. carpio*. The raw genotyping data with 184,978 SNPs for the 199 samples were obtained from SNP genotyping. A total of 95,400 polymorphic SNPs from 195 samples passed the quality control threshold. As shown in **Figure 1A** and **Table S1**, the contents of glutamic (GLU), aspartic (ASP), leucine (LEU) and

lysine (LYS) accounted for a relatively large proportion of the total amino acid content, and the distribution of the other 16 amino acids was relatively balanced. The gender information of samples was collected and it showed no significant differences between males and females across all traits. To discover the potential relationship among multiple traits, the contents of 19 amino acids were analyzed using the correlation heatmap shown in **Figure 1B**. GLU, ASP, THR, LYS and PHE contents were dramatically associated with each other, possibly indicating similar functions within the amino acid content regulating mechanism.

Genome-wide association analysis

In total, 36, 6, and 1 SNPs were identified for glycine (GLY), proline (PRO) and tyrosine (TYR), respectively, with a threshold of $P < 5.241 \times 10^{-7}$ (**Table S2**). The Manhattan plot showed the most promising results for GLY content, while the Q-Q plot indicated the reliability of the data analysis (**Figure 2A and B**). Similar results for PRO and TYR content are shown in **Figure S1**. The distributions of SNPs on each chromosome were calculated as shown in **Figure 2C**, which reflects the relatively even distribution among all of the chromosomes. Genes were annotated through the new common carp genome [21], and 54, 10 and one genes were identified for GLY, PRO and TYR traits, respectively.

Transcriptomic analysis of divergent amino acid content

We examined the amino acid contents of 20 newly collected fish samples and selected six individuals with relatively extreme amino acid content. The EAA, BCAA and FLA contents of the 20 samples are presented in **Table S3**. Taken from our previously published data, the quality of RNA sequencing data is shown in **Tables S4 and S5**. As EAA and BCAA content were highly correlated, these two categories were integrated as one group in the following analyses.

Plenty of differentially expressed genes were found for the trait EAA (or BCAA), with 236, 81 and 960 genes in brain, liver and muscle tissues, respectively (**Figure 3A-C and Tables S6, S8 and S10**). The Venn diagram for EAA (or BCAA) trait showed that two, thirteen and eight genes were shared in pairwise comparisons among three tissues (**Figure 3D**). The cluster analyses of the differentially expressed genes for EAA (or BCAA) in three tissues are shown in **Figures 3E and S2**.

For the analysis of the FLA trait, 818, 74 and 3 differentially expressed genes were found in three tissues, respectively (**Figure 4; Tables S6, S8 and S10**). Very limited numbers of genes were shared among three the tissues, which could be due to the diversified gene functions in different tissues (**Figure 4D**). The cluster studies of the differentially expressed genes for FLA in three tissues are shown in **Figures 4E and S2**.

The network analyses of differentially expressed genes for the three classified amino acids in multiple tissues are shown in **Figures S3-7 and Tables S7, S9 and S11**. The network of GO enrichment pathways are illustrated in **Figures 5, S8, S9 and Tables S12 and S13**, while the KEGG enrichment analysis is shown in **Table S14**.

Differential methylation analysis

The epigenetic difference analyses between the high and low amino acid content groups were conducted and the DNA was isolated from muscle tissues of six samples for WGBS. Taken from our previously reported data, the quality of WGBS data was shown in **Table S15**. A volcano plot of the EAA (or BCAA) shows the distribution of methylation differences in all the DMRs (**Figure 6A**). DMRs were classified into several genomic regions including exons, introns, promoters and intergenic repeat regions (**Figure 6B**). **Figure 6C** shows the length distribution of DMRs and most regions were shorter than 1000 bp. Similar results were observed in the DMRs identification for the FLA trait (**Figure 6D–F**). The DMRs within the promoter regions (differentially methylated promoters, DMPs) were chosen for further functional enrichment analysis due to the significance of DMPs in the regulation of transcription [22]. The GO enrichment analysis and KEGG enrichment analysis are shown in **Tables S17 and S18**.

Multi-omics analysis of DGE and DMP results

Through DGE and DMP studies, some related genes and pathways have been found, but the interactions between methylated sites and altered expression are still unclear. Obviously it is considerable to implement a multi-omics analysis through the DGE and DMP aspects. Because very few intersections were observed between the liver DGE and DMP results, we concentrated on the relatedness between DGE and DMP results in brain and muscle tissues (**Table S19**). After Pearson correlation analysis, significant linear correlations (**Figure 7**) were found in the EAA (or BCAA) and FLA traits of muscle and brain tissues. These genes were divided into two types, positively related and negatively related. For the EAA (or BCAA) trait in muscle, 57 genes were included in the correlation analysis. In addition, 36 genes of the FLA trait in brain tissues were obtained in the correlation analysis. For the EAA (or BCAA) trait in muscle tissues, 24 positively related genes and 33 negatively related genes were presented in **Figure 7A**. Totally 12 positively related genes and 24 negatively related genes for FLA trait in brain tissues were found in **Figure 7B**.

Discussion

Genes associated with muscular amino acid content by GWAS

Among the identified 36 SNPs associated with GLY content in *C. carpio*, 32 were located on chromosome 25. Two SNPs (carp102954 and carp102953) were located within the intron and exon regions of the *glud1* (glutamate dehydrogenase 1) gene. Glutamate dehydrogenase is a zinc protein that plays a major role in amino acid metabolism in multiple tissues such as liver, heart, muscle and kidney. Glutamate dehydrogenase, as a crucial enzyme in crayfish, was reported to be involved in bridging amino acid metabolism and carbohydrate metabolism, and the enzyme was regulated by reversible phosphorylation in severe hypoxia-tolerant conditions [23]. A recent study in rice revealed that glutamate dehydrogenase mediated amino acid metabolism after ammonium uptake and also enhanced rice growth under aeration conditions [24]. A number of amino acid or peptide transporters were observed in the associated genes list, including *abcb10* (ATP-binding cassette sub-family B member 10) and *slc1a4* (neutral amino acid transporter A). The gene *abcb10* encoded a protein that was involved in heme biosynthesis [25], while

slc1a4 encoded a protein that was reported as an important transporter for multiple neutral amino acids [26]. The gene *myoz1* (myozenin-1) was identified near the downstream area of the SNP carp022265, which was previously reported to function in myofibrillogenesis, and may affect myofibrillar protein content in muscle [27, 28]. Only a few SNPs and genes were identified to be associated with PRO and TYR content, and no significant markers were found to be associated with other amino acid contents, probably due to the limited number of samples. Nevertheless, dozens of SNPs exceeded the suggestive threshold in other traits, indicating that more associated SNPs might be identified with larger samples in the future. As the genes discovered in GWAS analysis may not be enough to illustrate the underlying mechanism of muscular amino acid metabolism, and no associated SNPs were found in the classified amino acids (EAA, BCAA, and FLA), we sought to uncover more evidence for the three categories of amino acids by performing gene expression and epigenetic changes analyses.

Network analysis based on RNA-Seq data

Both BCAAs and EAAs play crucial roles in the physiological functions of organisms, while the FLAs are the main source of the fish umami and flavor. We focused on these three types of classified amino acids using RNA-Seq data analysis and gene network construction. A large number of genes were found differentially expressed in muscle for EAA (or BCAA), corresponding to the fact that most EAAs or BCAAs are metabolized in muscle tissue [29]. As thousands of genes were used for the network analyses (**Figures S3-7**), it was not easy to identify the most significant node genes from the network map. Functional enrichment of genes generated GO and KEGG pathways that played central roles in the regulation of amino acid contents. As shown in **Figure 5A**, in muscle tissue for EAA or BCAA trait, various pathways were enriched, such as “skeletal muscle cell differentiation,” “skeletal muscle tissue development,” “striated muscle tissue development,” “myotube cell development” and “regulation of protein ubiquitination.” KEGG enrichment for EAA or BCAA in muscle resulted in 36 pathways, including “valine, leucine and isoleucine degradation,” “mTOR signaling pathway,” “fatty acid metabolism” and “alanine, aspartate and glutamate metabolism.” Enrichment results for the FLA trait in brain tissue showed a distinct pattern of pathways related to stem cell or hematopoietic stem cell differentiation (**Figure 5B**), including “embryonic skeletal system development,” “regulation of hemopoiesis,” and “stem cell differentiation.” A number of neuro-related pathways (“nerve development,” “neural crest cell migration,” and “positive regulation of neurogenesis”) were also observed in the GO enrichment network of FLA, partially because some amino acids (e.g., GLY) within the FLA category served as neurotransmitters [30]. KEGG enrichment for the FLA trait in brain tissue contained 20 pathways comprising “wnt signaling pathway,” “neuroactive ligand-receptor interaction,” “glycine, serine and threonine metabolism” and several signaling pathways. Only one pathway, “glutathione metabolism,” was enriched in liver tissue for the FLA trait, while glutathione is a tripeptide that plays a pivotal role in critical physiological processes affecting amino acid contents [31]. These network results will facilitate further research on the hub genes and pathways that are involved in the regulation of gene expression associated with the relevant traits.

Gene identification with DMP and integration incorporating DEGs

It has been widely accepted that genes regulating amino acid metabolism are closely related to other pathways, as many products are shared in the pathways of fatty acid metabolism, growth regulation and other physiological functions. As shown in **Table S18**, 69 pathways were enriched for the EAA or BCAA trait. Interestingly, several amino acid metabolism related pathways were enriched, such as “lysine degradation,” “valine, leucine and isoleucine degradation,” “beta-alanine metabolism,” and “arginine and proline metabolism.” Multiple important biological processes were also screened out by KEGG enrichment, indicating the wide interactions among these physiological functions, such as “pyrimidine and purine metabolism,” “insulin signaling pathway,” “adipocytokine signaling pathway” and “citrate cycle (TCA cycle).” Similarly, 66 pathways were enriched for the FLA trait (**Table S18**), which showed comparable patterns with the above results from EAA or BCAA. The widespread methylation differences in promoter regions of selected genes indicated that epigenomic features were common in amino acid metabolism regulation.

Among those interactive results shared by DMP and DGE, genes with more highly methylated promoters and decreased expression levels (or vice versa) were paid more attention due to their multi-omics evidence. Taking the gene *agtpbp1* (cytosolic carboxypeptidase 1) for example, the promoter showed a higher methylation ratio in the H group of EAA (or BCAA), corresponding to decreased expression (**Figure 7A**). Likewise, for the integrated analysis of the FLA trait, the gene *gpt2l* (alanine aminotransferase 2-like) was screened out with decreased methylation ratio and elevated expression in the H group. The protein alanine aminotransferase 2-like catalyzes the reversible transamination between alanine and 2-oxoglutarate to form pyruvate and glutamate, engaging in metabolism of multiple amino acids.

Conclusions

In this study, we conducted the GWAS, transcriptome and whole genome methylation analysis on muscular amino acid content in *C. carpio*. The GWAS investigation of amino acid content revealed some genes that are related with amino acid metabolism. Through studies of the RNA-Seq results and gene profiling with DMPs, we uncovered key genes which were enriched in EAA, BCAA and FLA metabolism pathways. Integrative results of RNA-Seq and DNA methylation indicated significant correlations at two multi-omics levels. The comprehensive analysis also identified genes related to amino acid synthesis, transportation and utilization. It is desirable to validate these results by extensive sequencing with a larger population. The results will accelerate the application of molecular selection breeding for common carp fleshy with high nutritional value.

Methods

Sample collection

A total of 199 individuals of Yellow River carp were randomly collected from Henan Academy of Fishery Sciences, Henan, China. Based on statistical power calculation by GCTA [32], these samples could be used for reliable GWAS analysis. All fishes were euthanized in MS222 solution before the samples were

collected, and processing was finished within several minutes for each fish. For each individual, the blood sample was taken for molecular analysis. For amino acid quantification, dorsal muscle tissue was obtained for amino acid content determination. Brain, liver and muscle tissues from 20 samples were collected and used for RNA extraction.

SNP genotyping

The genomic DNA of 199 samples was extracted from the blood for SNP analysis. The quality and concentration of DNA were examined by agarose gel electrophoresis and colorimetry. The genotyping and quality control were conducted following the previously reported procedures [19].

Genome-wide association analysis

The TASSEL version 5.0 software [33] was applied for genome-wide association analysis of amino acid content using the genotyping data. The significant P -value threshold was set to 5.241×10^{-7} , and the SNPs associated with amino acid content were selected. The Manhattan plots, Q-Q plots and SNP density were produced by the CMplot package.

Transcriptome sequencing and differential expression analysis

Brain, liver and muscle tissues of 20 fish were used to perform the RNA sequencing based on the three respective traits EAA, BCAA and FLA contents under RNase-free conditions. According to the EAA, BCAA and FLA contents, three fish with the overall lowest amino acids content were selected as the control group, and correspondingly three fish with the highest contents were selected as the experimental group. The total RNA was isolated, and the RNA quantity was verified using 1.5% agarose gel. A cDNA library was constructed and sequenced through an Illumina HiSeq2500 sequencing system (Illumina, San Diego, CA, USA) and 150 bp paired-end reads were produced.

Low-quality reads and adaptors were removed through Trimmomatic v.0.32 [34]. A Bowtie index was built by Bowtie2 v.2.3.4.2 [35] and Tophat2 was used to aligned transcriptome data to the genome sequence [36]. Samtools [37] was applied to index the bam files, and Cufflinks [38] was applied to assemble individual transcripts from RNA-Seq data. Individual transcripts were then merged using Cuffmerge to generate an integrated assembly. Based on FPKM, differentially expressed transcripts were identified using Cuffdiff. The differentially expressed genes were inferred from a volcano plot that showed the overall distribution using the R software. The genes with $\log_2$ fold-change >1 and the q value < 0.05 were identified as significant expression abundance. The heatmaps and Venn diagram were constructed by the pheatmap package and VennDiagram package, respectively. The relationships between Gene Ontology (GO) pathways enriched by differentially expressed genes were demonstrated by the Cytoscape software [39].

DNA methylation sequencing and differential methylation analysis

The genomic DNA was used for whole genome bisulfite sequencing (WGBS). The genomic DNA was sonicated with sodium sulfite, which first converted unmethylated cytosine to uracil and then to thymine [40]. WGBS was conducted according to previously reported method [19]. Bismark [41] was applied to align the methylation data to the reference genome, and swDMR [42] was used to analyze the differentially methylated regions (DMR) through a sliding window method. The process of DMR detection and annotation followed the protocols that were previously reported [15, 19]. The correlation analysis of DEG and DMP was performed using the R package ggplot2.

Abbreviations

EAA: essential amino acid; NEAA: nonessential amino acid; FAA: functional amino acid; FLA: flavorful amino acid; BCAA: branched-chain amino acids; GWAS: genome-wide association study; WGBS: whole genome bisulfite sequencing; GO: gene ontology; KEGG: Kyoto encyclopedia of genes and genomes

Declarations

Acknowledgments

The authors would like to thank all staffs at the breeding center of Henan Academy of Fishery Sciences (Zhengzhou, China) for the help of sample collection.

Authors' contributions

JX initiated and coordinated the research project. YZ, HZ and JX conceived and conducted the analysis, and drafted the manuscript. PX engaged in sample collection and genotyping analysis. ZZ, JF, BW, CD and YJ took part in trait measurement, tissue manipulation and enrichment analysis. All authors read and approved the final manuscript.

Funding

This work was supported by the National Key Research and Development Program (2018YFD0900102) for genotyping and phenotyping; the Special Scientific Research Funds for Central Non-profit Institutes, Chinese Academy of Fishery Sciences (2020TD24, 2015C005) for RNA-Seq and WGBS; and the National Infrastructure of Fishery Germplasm Resources of China (No. 2019DKA30470) for sample collection.

Availability of data and materials

The sequencing datasets of all samples have been deposited at NCBI (PRJNA493161).

Ethics approval and consent to participate

This study was approved by the Animal Care and Use Committee (ACUC) of the Centre for Applied Aquatic Genomics at the Chinese Academy of Fishery Sciences. All experiments complied with the

ARRIVE guidelines and were carried out in accordance with the National Institutes of Health guide for the care and use of laboratory animals (NIH Publications No. 8023, revised 1978).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Xu J, Zhao Z, Zhang X, Zheng X, Li J, Jiang Y, Kuang Y, Zhang Y, Feng J, Li C, et al. Development and evaluation of the first high-throughput SNP array for common carp (*Cyprinus carpio*). *BMC Genomics*. 2014;15:307.
2. Wu G. Functional amino acids in growth, reproduction, and health. *Advances in nutrition* (Bethesda Md). 2010;1(1):31–7.
3. Wu G. Functional amino acids in nutrition and health. *Amino Acids*. 2013;45(3):407–11.
4. Zhang S, Zeng X, Ren M, Mao X, Qiao S. Novel metabolic and physiological functions of branched chain amino acids: a review. *J Anim Sci Biotechnol*. 2017;8:10.
5. Andersen SM, Waagbo R, Espe M. Functional amino acids in fish nutrition, health and welfare. *Front Biosci*. 2016;8:143–69.
6. Gao Y, Tian L, Liu Y. Research progress of amino acids in fish. *Guangdong Feed*. 2012;21(a01):53–7.
7. Wheelhouse NM, Stubbs AK, Lomax MA, MacRae JC, Hazlerigg DG. Growth hormone and amino acid supply interact synergistically to control insulin-like growth factor-I production and gene expression in cultured ovine hepatocytes. *J Endocrinol*. 1999;163(2):353–61.
8. Mihalik SJ, Michaliszyn SF, de las Heras J, Bacha F, Lee S, Chace DH, DeJesus VR, Vockley J, Arslanian SA. Metabolomic profiling of fatty acid and amino acid metabolism in youth with obesity and type 2 diabetes: evidence for enhanced mitochondrial oxidation. *Diabetes Care*. 2012;35(3):605–11.
9. Chalamaiah M, Dinesh Kumar B, Hemalatha R, Jyothirmayi T. Fish protein hydrolysates: proximate composition, amino acid composition, antioxidant activities and applications: a review. *Food Chem*. 2012;135(4):3020–38.
10. Ali A, Al-Tobasei R, Lourenco D, Leeds T, Kenney B, Salem M. Genome-Wide Association Study Identifies Genomic Loci Affecting Filet Firmness and Protein Content in Rainbow Trout. *Frontiers in genetics*. 2019;10:386.
11. Joshi V, Shinde S, Nimmakayala P, Abburi VL, Alaparthi SB, Lopez-Ortiz C, Levi A, Panicker G, Reddy UK. **Haplotype Networking of GWAS Hits for Citrulline Variation Associated with the Domestication of Watermelon.** *International journal of molecular sciences* 2019, 20(21).

12. Peng Y, Liu H, Chen J, Shi T, Zhang C, Sun D, He Z, Hao Y, Chen W. Genome-Wide Association Studies of Free Amino Acid Levels by Six Multi-Locus Models in Bread Wheat. *Frontiers in plant science*. 2018;9:1196.
13. Meng J, Song K, Li C, Liu S, Shi R, Li B, Wang T, Li A, Que H, Li L, et al. Genome-wide association analysis of nutrient traits in the oyster *Crassostrea gigas*: genetic effect and interaction network. *BMC Genomics*. 2019;20(1):625.
14. Wu P, Li Y, Cheng J, Chen L, Zeng M, Wu Y, Wang J, Zhang J, Chu W: **Transcriptome Analysis and Postprandial Expression of Amino Acid Transporter Genes in the Fast Muscles and Gut of Chinese Perch (*Siniperca chuatsi*)**. *PloS one* 2016, **11**(7):e0159533.
15. Wen X, Hu Y, Zhang X, Wei X, Wang T, Yin S. Integrated application of multi-omics provides insights into cold stress responses in pufferfish *Takifugu fasciatus*. *BMC Genomics*. 2019;20(1):563.
16. Wei Z, Zhou H, Zhang Y, Zhang Q, Zhang W, Mai K. Integrative analysis of transcriptomics and metabolomics profiling on flesh quality of large yellow croaker *Larimichthys crocea* fed a diet with hydroxyproline supplementation. *Br J Nutr*. 2018;119(4):359–67.
17. Mekuchi M, Sakata K, Yamaguchi T, Koiso M, Kikuchi J. Trans-omics approaches used to characterise fish nutritional biorhythms in leopard coral grouper (*Plectropomus leopardus*). *Scientific reports*. 2017;7(1):9372.
18. Zhang Y, Liu J, Fu W, Xu W, Zhang H, Chen S, Liu W, Peng L, Xiao Y. Comparative Transcriptome and DNA methylation analyses of the molecular mechanisms underlying skin color variations in Crucian carp (*Carassius carassius* L.). *BMC Genet*. 2017;18(1):95.
19. Zhang H, Xu P, Jiang Y, Zhao Z, Feng J, Tai R, Dong C, Xu J. Genomic, Transcriptomic, and Epigenomic Features Differentiate Genes That Are Relevant for Muscular Polyunsaturated Fatty Acids in the Common Carp. *Frontiers in genetics*. 2019;10:217.
20. Xu P, Zhang X, Wang X, Li J, Liu G, Kuang Y, Xu J, Zheng X, Ren L, Wang G, et al. Genome sequence and genetic diversity of the common carp, *Cyprinus carpio*. *Nat Genet*. 2014;46(11):1212–9.
21. Xu P, Xu J, Liu G, Chen L, Zhou Z, Peng W, Jiang Y, Zhao Z, Jia Z, Sun Y, et al. The allotetraploid origin and asymmetrical genome evolution of the common carp *Cyprinus carpio*. *Nature communications*. 2019;10(1):4625.
22. Haque MM, Nilsson EE, Holder LB, Skinner MK. Genomic Clustering of differential DNA methylated regions (epimutations) associated with the epigenetic transgenerational inheritance of disease and phenotypic variation. *BMC Genomics*. 2016;17:418.
23. NJ D, KB S. An enzymatic bridge between carbohydrate and amino acid metabolism: regulation of glutamate dehydrogenase by reversible phosphorylation in a severe hypoxia-tolerant crayfish. *J Comp Physiol B*. 2012;182(3):331–40.
24. Xiaochuang C, Meiyang W, Chunquan Z, Chu Z, Junhua Z, Lianfeng Z, Lianghuan W, Qianyu J. Glutamate dehydrogenase mediated amino acid metabolism after ammonium uptake enhances rice growth under aeration condition. *Plant cell reports*. 2020;39(3):363–79.

25. Seguin A, Takahashi-Makise N, Yien YY, Huston NC, Whitman JC, Musso G, Wallace JA, Bradley T, Bergonia HA, Kafina MD, et al. Reductions in the mitochondrial ABC transporter Abcb10 affect the transcriptional profile of heme biosynthesis genes. *J Biol Chem*. 2017;292(39):16284–99.
26. Gibson C, de Ruijter-Villani M, Rietveld J, Stout TAE. Amino acid transporter expression in the endometrium and conceptus membranes during early equine pregnancy. *Reprod Fertil Dev*. 2018;30(12):1675–88.
27. Roberts MD, Romero MA, Mobley CB, Mumford PW, Roberson PA, Haun CT, Vann CG, Osburn SC, Holmes HH, Greer RA, et al. Skeletal muscle mitochondrial volume and myozenin-1 protein differences exist between high versus low anabolic responders to resistance training. *PeerJ*. 2018;6:e5338.
28. Ren RM, Liu H, Zhao SH, Cao JH. **Targeting of miR-432 to myozenin1 to regulate myoblast proliferation and differentiation.** *Genetics and molecular research: GMR* 2016, 15(4).
29. Adeva-Andany MM, Lopez-Maside L, Donapetry-Garcia C, Fernandez-Fernandez C, Sixto-Leal C. Enzymes involved in branched-chain amino acid metabolism in humans. *Amino Acids*. 2017;49(6):1005–28.
30. Qin X, Akter F, Qin L, Xie Q, Liao X, Liu R, Wu X, Cheng N, Shao L, Xiong X, et al. MicroRNA-26b/PTEN Signaling Pathway Mediates Glycine-Induced Neuroprotection in SAH Injury. *Neurochem Res*. 2019;44(11):2658–69.
31. Minich DM, Brown BI. **A Review of Dietary (Phyto)Nutrients for Glutathione Support.** *Nutrients* 2019, 11(9).
32. Visscher PM, Hemani G, Vinkhuyzen AA, Chen GB, Lee SH, Wray NR, Goddard ME, Yang J. Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. *PLoS Genet*. 2014;10(4):e1004269.
33. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*. 2007;23(19):2633–5.
34. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–20.
35. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature methods*. 2012;9(4):357–9.
36. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology*. 2013;14(4):R36.
37. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9.
38. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols*. 2012;7(3):562–78.

39. Malinsky M, Svardal H, Tyers AM, Miska EA, Genner MJ, Turner GF, Durbin R. Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature ecology evolution*. 2018;2(12):1940–55.
40. Wang Y, Zheng W, Luo J, Zhang D, Zuhong L. In situ bisulfite modification of membrane-immobilized DNA for multiple methylation analysis. *Anal Biochem*. 2006;359(2):183–8.
41. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*. 2011;27(11):1571–2.
42. Wang Z, Li X, Jiang Y, Shao Q, Liu Q, Chen B, Huang D. swDMR: A Sliding Window Approach to Identify Differentially Methylated Regions Based on Whole Genome Bisulfite Sequencing. *PloS one*. 2015;10(7):e0132866.

Additional Files

Additional file 1: Table S1-19. Phenotypes, genes, and enriched pathways by GWAS, DGE, and DMP analyses.

Additional file 2: Figure S1. GWAS of 195 individuals for PRO and TYR content. (A) Manhattan plot of the PRO content association. (B) Q-Q plot for the association results of PRO content. (C) Manhattan plot of the TYR content association. (D) Q-Q plot for the association results of TYR content.

Additional file 3: Figure S2. Cluster analysis of DEGs in two tissues. (A) Clustering of DEGs in brain tissue for EAA (or BCAA) trait. (B) Clustering of DEGs in liver tissue for EAA (or BCAA) trait. (C) Clustering of DEGs in liver tissue for FLA trait. The color bars represent the relative expression levels of transcripts. Highly expressed transcripts are marked in red, while low expressed transcripts are assigned in blue.

Additional file 4: Figure S3. The network of the differentially expressed genes in brain tissue for EAA (or BCAA) trait. The combined score between the DEGs is indicated by the thickness of the edge between the genes. The color of each node represents the q -value of the genes. The genes with q -value < 0.05 are displayed.

Additional file 5: Figure S4. The network of the differentially expressed genes in liver tissue for EAA (or BCAA) trait. The combined score between the DEGs is indicated by the thickness of the edge between the genes. The color of each node represents the q -value of the genes. The genes with q -value < 0.05 are displayed.

Additional file 6: Figure S5. The network of the differentially expressed genes in muscle tissue for EAA (or BCAA) trait. The combined score between the DEGs is indicated by the thickness of the edge between the genes. The color of each node represents the q -value of the genes. The genes with q -value < 0.05 are displayed.

Additional file 7: Figure S6. The network of the differentially expressed genes in brain tissue for FLA trait. The combined score between the DEGs is indicated by the thickness of the edge between the genes. The

color of each node represents the q -value of the genes. The genes with q -value < 0.05 are displayed.

Additional file 8: Figure S7. The network of the differentially expressed genes in liver tissue for FLA trait. The combined score between the DEGs is indicated by the thickness of the edge between the genes. The color of each node represents the q -value of the genes. The genes with q -value < 0.05 are displayed.

Additional file 9: Figure S8. The network of enriched GO terms in muscle tissue for EAA (or BCAA) trait. The thickness of the edge between the GO terms represents the Kappa Score between the terms. The q -value of the GO terms is manifested through the color of each node. The size of each node shows the number of enriched genes for every GO term. The GO terms with q -value < 0.05 are displayed.

Additional file 10: Figure S9. The network of enriched GO terms in brain tissue for FLA trait. The thickness of the edge between the GO terms represents the Kappa Score between the terms. The q -value of the GO terms is manifested through the color of each node. The size of each node shows the number of enriched genes for every GO term. The GO terms with q -value < 0.05 are displayed.

Figures

Figure 1

Phenotypic distribution and correlation clustering analysis. (A) Boxplots for phenotypic distributions. The lines inside the boxes denote the median. The mean is shown as the rhombus in the boxes. (B) The correlation analysis of different amino acid traits. The color shows the relative coefficient R^2 values. The red color indicates a positive correlation, whereas negative correlations are shown in blue.

Figure 2

GWAS of 195 individuals for GLY content. (A) Manhattan plot displaying the GWAS result of the GLY content. (B) Q-Q plot for the GWAS result of the GLY content. (C) Density distribution of SNPs associated with GLY content on each chromosome.

Figure 3

Transcriptomics analysis of EAA (or BCAA) trait. (A) Volcano plots of DEGs in brain tissues for EAA (or BCAA) trait. Red dots represent upregulated genes, while blue dots indicate downregulated genes. The dashed horizontal line represents the threshold value of $q = 0.05$. The vertical dashed line indicates the absolute log_2 fold-change (FC) value = 1. (B) Volcano plots of DEGs in liver tissues for EAA (or BCAA) trait. (C) Volcano plots of DEGs in muscle tissues for EAA (or BCAA) trait. (D) Venn diagrams of DEGs describing the numbers of shared and unique genes in three tissues for EAA (or BCAA) trait. (E) Clustering of DEGs in muscle tissues for EAA (or BCAA) trait. The color bars represent the relative expression levels of transcripts. Highly expressed transcripts are marked in red, while low expressed transcripts are assigned in blue.

Figure 4

The transcriptomics analysis of FLA trait. (A) Volcano plots of DEGs in brain tissues for FLA trait. Red dots represent upregulated genes, while blue dots indicate downregulated genes. The dashed horizontal line represents the threshold value of $q = 0.05$. The vertical dashed line indicates the absolute log_2 fold-change (FC) value = 1. (B) Volcano plots of DEGs in liver tissues for FLA trait. (C) Volcano plots of DEGs in muscle tissues for FLA trait. (D) Venn diagrams of DEGs describing the numbers of shared and unique genes in three tissues for FLA trait. (E) Clustering of DEGs in muscle tissues for FLA trait. The color bars represent the relative expression levels of transcripts. Highly expressed transcripts are marked in red, while low expressed transcripts are assigned in blue.

Figure 5

The network of significantly enriched GO terms in muscle and brain tissue for EAA (or BCAA) content and FLA trait. The thickness of the edge between the GO terms represents the Kappa Score between the terms. The q -value of the GO terms is manifested through the color of each node. The size of each node shows the number of enriched genes for every GO term. The GO terms with q -value < 0.05 are displayed.

Figure 6

Volcano plots of DMRs, DMR distribution in genomic regions and the length distribution of DMRs. (A) Volcano plots of DMRs for EAA (or BCAA) trait. The X-axis indicates the difference in methylation ratio. The Y-axis represents the $-\log_{10}(\text{P-value})$. Red dots represent upregulated genes, while blue dots indicate downregulated genes. (B) Numbers for each DMR in different genomic regions for EAA (or BCAA) trait. (C) Counts for each DMR in different lengths for EAA (or BCAA) trait. (D) Volcano plots of DMRs for FLA trait. (E) Numbers for each DMR in different genomic regions for FLA traits. (F) Counts for each DMR of different lengths for FLA trait.

Figure 7

Pearson's correlation of gene expression fold changes and differential methylation rates for DEGs. Scatter plots (A) for EAA (or BCAA) trait in muscle tissues; (B) for FLA trait in brain tissues. The X-axis means difference in methylation ratio between high and low EAA (or BCAA) content groups. The Y-axis indicates the logarithm of fold changes (log2FC). The red dots indicate the negatively correlated genes, while the blue dots represent positively correlated genes.

Supplementary Files

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

- [supplement1.pdf](#)
- [supplement2.xlsx](#)
- [supplement3.tif](#)
- [supplement4.tif](#)
- [supplement5.tif](#)
- [supplement6.tif](#)
- [supplement7.tif](#)
- [supplement8.tif](#)
- [supplement9.tif](#)
- [supplement10.tif](#)

- [supplement11.tif](#)