Comparative transcriptomics identifies genes differentially expressed in the intestine of a new fast-growing strain of common carp with higher unsaturated fatty acid content in muscle

Chengfeng Zhang[1,2☯], **Shengyan Su**[2,3☯], **Xinyuan Li**[3☯], **Bing Li**[2], **Baojuan Yang**[2], **Jian Zhu**[2]*, **Weimin Wang**[1]*

1 College of Fisheries, Huazhong Agricultural University, Wuhan, PR China, **2** Key Laboratory of Genetic Breeding and Aquaculture Biology of Freshwater Fishes, Ministry of Agriculture; Freshwater Fisheries Research Center, Chinese Academy of Fishery Sciences, Wuxi, PR China, **3** Wuxi Fisheries College, Nanjing Agricultural University, Wuxi, PR China

☯ These authors contributed equally to this work.
* zhuj@ffrc.cn (JZ); wangwm@mail.hzau.edu.cn (WW)

Abstract

We have created a new, fast-growing strain of common carp with higher unsaturated fatty acid content in muscle. To better understand the impacts of gene regulation in intestinal tissue on growth and unsaturated fatty acid content, we conducted a comparative RNA-Seq transcriptome analysis between intestine samples of Selected and Control groups (and corroborated selected results by PCR). After eight weeks of cage culture, weight gain of the Selected group was 20.84% higher. In muscles of the control group, monounsaturated fatty acids (FAs) were more abundant, whereas polyunsaturated FAs were more abundant in muscles of the Selected group. In total, we found 106 differentially expressed genes (DEGs) between the two groups. Only the endocytosis pathway was significantly enriched in DEGs, with two upregulated genes: *il2rb* and *ehd1*. The latter is involved in the growth hormone/insulin-like growth factor (Gh/Igf) axis, which plays a key role in the regulation of growth in animals. *tll2*, which is known to be associated with intestinal regeneration, was extremely highly upregulated in both transcriptomic (infinite) and qPCR (610.70) analyses. Two of the upregulated genes are associated with the fatty acid metabolism, several genes are likely to be indicators of heightened transcription levels, several are associated with metabolic and developmental roles, several with neuronal functions (including two with vision), several with the immune system, and two downregulated genes with the development of vasculature. The higher growth rate of the Selected group is likely to be at least partially attributed to increased endocytosis efficiency and genetically-driven behavioural differences (higher aggression levels). There are some indications that this new strain might have slightly impaired immune responses, and a higher propensity for inherited diseases leading to sight impairment, as well for neurodegenerative diseases in general, but these indications still need to be confirmed.

Introduction

Growth rate, regulated by environmental factors and genetics, is a primary trait of interest in selection programs of most cultured fish species due to its intrinsic link with productivity and profitability of aquaculture enterprises [1]. The common carp (*Cyprinus carpio* Linnaeus 1758) is probably the oldest and most common cultured fish species, with third largest total aquaculture production output globally. In 2014, the total global output was over four million tonnes (almost 10% of the global annual freshwater aquaculture production), producing the economic value of almost six billion dollars, with China producing around 70% of the total output [2].

Growth of skeletal muscles of fish is the primary topic of interest for the aquaculture industry [3]. Several growth-related features set the fishes apart from mammals: continuous accretion of muscle tissue results in the increase of body length and mass (albeit at a slowing rate) until mortality or senescence occur, both hyperplasia and hypertrophy contribute to muscle growth, and fishes predominantly accumulate functional protein (as opposed to storing the excessive energy as adipose tissue) [3–5]. Although some molecular aspects of growth in fish are relatively well-studied, such as the growth hormone—insulin-like growth factor I (GH-IGFI) axis of the neuroendocrine system [3,6–8], our understanding of the molecular regulation of growth in fish remains fragmentary [3,9].

Fatty acids are essential cellular components, whose composition is important for the nutritive value and taste of meat [10]. In humans, saturated FAs are risk factors for cardiovascular diseases, whereas unsaturated and especially polyunsaturated FAs, such as omega-3 and -6 (ω-3 and ω-6), have antiatherosclerotic effects and a number of other beneficial health effects [11,12]. Fish meat is generally the main source of polyunsaturated fatty acids (FAs) in human diet, and thus of remarkable significance in human nutrition [13,14].

The gastrointestinal tract acts as a selectively permeable barrier for dietary nutrients, electrolytes and water, while maintaining an effective defense against pathogens [15]. Regardless of anatomical and physiological differences (common carp is an agastric, omnivorous fish), teleost fishes are characterised by the presence of at least two intestinal segments: the first is responsible for the absorption of lipids and the second for pinocytotic uptake of macromolecules, including proteins [3,4]. Uptake of nutrients from food via the intestinal barrier is a prerequisite for growth, and impacts of different diets on intestinal transcriptome have been studied in fish [16,17]. However, individual differences in the efficiency of this process in fish, and particularly the genetic background thereof, remain almost completely unstudied.

The molecular control of growth in fish muscles is relatively well-understood [5,9], but molecular control of intestinal nutrient uptake, although likely to play a part in the growth rate [9], remains poorly understood. High-throughput RNA sequencing technology (RNA-Seq) has been used to study the molecular mechanisms underlying growth in bacteria [18], plants [19], mammals [20], and fish [21–26]. Furthermore, detailed RNA-Seq transcriptome analyses have been applied to study gene expression profiles in 19 different tissues of the common carp, but the analysis did not include intestinal samples [27]. Intestinal transcriptomes have been used to search for drug candidates [28], study regeneration mechanism [29], and genome polyadenylation [30], but (to our knowledge) none of previous transcriptome-based studies of fish growth have used intestine among the sampled tissues [21,22].

A new, fast-growing strain of common carp with higher unsaturated fatty acid content in the muscle was recently created by combining the best linear unbiased prediction and molecular markers-based breeding methods [31]. To better understand the overall gene expression levels in the intestine of common carp, as well as the impacts of gene regulation in intestinal tissue on the growth and unsaturated fatty acid content in the muscles of this important cultured fish species, we conducted a comparative RNA-Seq transcriptome analysis of intestinal

tissues of the new strain and a control group. This study also aims to contribute to the understanding of the genetic basis of growth and unsaturated fatty acid content in fish and vertebrates in general.

Results and discussion

Growth performance and effects of selection on fatty acid composition of muscles

Growth performance test of 48 families showed that after eight weeks of cage culture experiment, weight gain of the Selected group was 20.84% higher than that of the Control group (Table 1). Fatty acid composition analysis showed that saturated FAs (SFA) were more abundant (albeit non-significantly) in the muscles of Control group (Table 1). Among eight SFAs, six were more abundant in the Control group, but only one significantly. Regarding the two that were higher in the Selected group (C:18 and C:20), only C:18 was significantly higher. Unsaturated FAs (UFA), however, were more abundant in the Selected group: among 15 UFAs, only four were more abundant in the Control group, and only C18:1 significantly. Among the remaining eleven UFAs, three (C20:4, C22:3 and C22:4) were significantly higher in the Selected group. Intriguingly, this difference was driven exclusively by polyunsaturated FAs (PUFA), as monounsaturated FAs (MUFA) were significantly more abundant in the Control group: three out of four were more abundant. Among the PUFAs, only one (C18:3 *n-3*) was more abundant in the Control group (Table 1).

A large number of factors play a part in determining the fatty acid composition of muscles of common carp [32,33]. However, as environmental factors were controlled for in this experiment, the differences observed should be largely attributable to genetics [33]. The observed reduction in the level of oleic acid (MUFA: C18:1) is not a desirable result, as this FA is beneficial both for the total cholesterol and low-density lipoprotein cholesterol levels in plasma in humans and for the fish meat flavour [34]. However, the lower proportion of this MUFA in the Selected group is compensated by the increase in essential PUFAs, the consumption of which is associated with numerous health benefits [11,13]. Therefore, we can conclude that selection for faster growth did not negatively affect the meat quality (at least not in terms of FA composition) of the Selected line.

Transcriptome analyses

The sequencing of twelve intestine samples, which includes eight specimens from the Selected group and four specimens from the Control group, yielded 93.78 Gb of clean data, reaching 6.17 Gb for a single sample, with Q30 over 87% (S1 Table). Clean reads of each sample were compared to the designated common carp reference genome, with mapping successfulness rates ranging from 62.97% to 64.91%. On the basis of the results of alternative splicing prediction, gene structure optimization, and queries of seven different databases, 5 346 new genes were discovered. Differentially expressed genes (DEGs) were identified by statistically comparing the results between the two groups. Functional annotation and enrichment analysis of 106 discovered DEGs was carried out subsequently. Data are deposited in the NCBI's BioProject (PRJNA414702) and SRA (SRR6202422, SRR6202423, SRR6202418, SRR6202419, SRR6202424, SRR6202425, SRR6202421, SRR6202426, SRR6202420, SRR6202429, SRR6202427, SRR6202428) databases.

Overall distribution of gene ontology terms following the GO database [35] annotation results are comparable to those previously reported for liver samples of bighead carp [23], with the exception of a strong increase in the percentage of DEGs associated with synapse (cellular

Table 1. Effect of selection on growth and long-chain fatty acid (FA) composition (% of total FA) of muscles.

Growth/FA	Control group	Selected group
IBW	11.88±2.63	16.09±1.90
FBW	38.08±7.84[a]	63.83±5.66[b]
BWG	26.20±6.87[a]	47.74±4.96[b]
C12:0	0.0175±0.0016[1]	0.0145±0.0007
C14:0	0.6256±0.0305	0.5794±0.034
C15:0	0.163±0.0111[a]	0.1365±0.0043[b]
C16:0	20.443±0.4523	19.321±0.4607
C17:0	0.2525±0.0162	0.2478±0.0112
C18:0	6.6548±0.0458[b]	7.4656±0.1645[a]
C20:0	0.2143±0.0118	0.2151±0.009
C22:0	0.0815±0.008	0.0714±0.0035
C16:1	1.4435±0.1591	1.0863±0.103
C18:1	20.511±0.5896[a]	17.236±0.7601[b]
C20:1	1.0023±0.0495	1.003±0.0234
C22:1	0.241±0.0314	0.2033±0.0074
C18:2	25.674±1.141	25.998±0.6794
C18:3 n-6	0.3213±0.0268	0.3878±0.0229
C18:3 n-3	1.5308±0.0725	1.4283±0.0453
C20:2	1.1575±0.0349	1.2028±0.0497
C20:3	2.398±0.0805	2.4934±0.0572
C20:4	3.9213±0.2575[b]	5.3394±0.2921[a]
C20:5	1.315±0.0422	1.3968±0.093
C22:3	0.3418±0.0152[b]	0.4541±0.0267[a]
C22:4	1.6068±0.0451[b]	2.3871±0.0909[a]
C22:5	0.8765±0.0422	0.9205±0.0391
C22:6	9.1105±0.2154	10.331±0.4868
SFA[2]	28.549±0.4591	28.132±0.3424
MUFA[3]	23.198±0.7656[a]	19.529±0.8618[b]
PUFA[4]	48.254±0.9015[b]	52.339±1.0159[a]
UFA[5]	71.452±0.4591	71.868±0.3424

IBW = initial body weight (g), FBW = final body weight (g), BWG = body weight gain (%). Within a row, different letters (superscript) indicate that the means are significantly different between two groups ($P<0.05$).

[1] All values are presented as mean % ± SE

[2] SFA = sum of saturated FAs (C12:0 to C22:0)

[3] MUFA = sum of monounsaturated FAs (C16:1 to C22:1)

[4] PUFA = sum of polyunsaturated FAs (C18:2 to C22:6)

[5] UFA = sum of unsaturated FAs (MUFA + PUFA).

component) in the intestinal samples of common carp. Comparison of functional category distribution between DEGs and the background transcriptome revealed notable differences in the percentages of genes associated with the following biological processes: biological adhesion, reproductive process and growth, all of which represented between 4 and 8% of all annotated genes, but were completely absent from the DEGs. In the cellular components category, notable differences were observed in the percentage of genes associated with macromolecular complexes, membrane-enclosed lumen, extracellular region part, extracellular matrix and synapse part, all of which appear to be notably less abundant, or even completely absent, from the
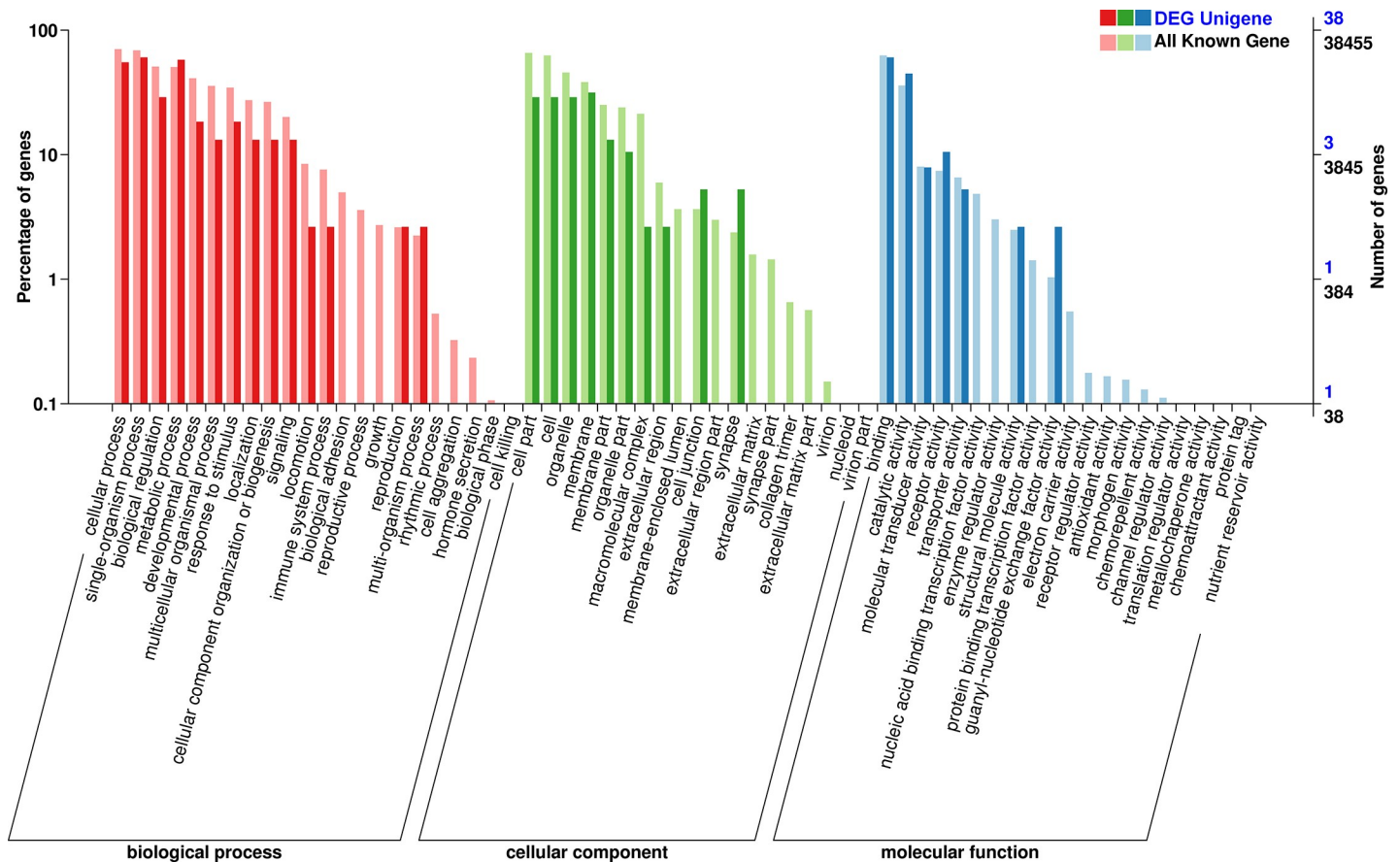
Fig 1. GO functional classification of annotated genes in transcriptome profiles of selection and control groups. 'DEG Unigene' refers to differentially expressed genes and 'All Known Gene' to all annotated genes in both groups.

DEGs. In the molecular function category, notable differences were observed in the percentages of genes associated with nucleic acid binding transcription factor activity and enzyme regulator activity, both of which represented between 3 and 7% of all annotated genes, but were completely absent from the DEGs. Among the categories with higher proportion of DEGs than all annotated genes were (in approximately descending order): guanyl-nucleotyde exchange factor activity (1% among all genes and $\approx$ 5% among DEGs), synapse, cell junction, receptor activity, metabolic process, multi-organism process, catalytic activity, and structural molecule activity (Fig 1).

Functions of DEGs were also predicted and classified by searching COG [36] and KOG [37] databases (Fig 2). Results of these two functional classifications were only partially congruent: by far the highest proportion of DEGs could be merely assigned to the 'general function prediction only' functional class in both classifications. In KOG classification, this was followed by 'signal transduction mechanisms', which was completely absent from COG, but 'replication, recombination and repair' and 'amino acid transport and metabolism' were found in both classifications. A small proportion of DEGs were associated with categories likely to be implicated in growth, such as carbohydrate and lipid (absent from COG) transport and metabolism. Intriguingly, although it may be expected that faster growth would be reflected in a number of DEGs related to the energy production, cell growth and division, DEGs classified as 'energy production and conversion', 'cell cycle control, cell division, chromosome

partitioning', 'cell wall biogenesis', and 'extracellular structures' were absent from both classifications. This scarcity of functional categories of DEGs becomes even more obvious when the results are compared with liver profiles of fast and slow-growing bighead carp, where all functional categories were represented among the discovered DEGs [23]. This is almost certainly at least partially a reflection of the central regulatory role the liver has in the growth of fish [7], as opposed to very specific functions of intestines. However, importantly for the objectives of our study, pathway enrichment analysis showed that (only) endocytosis pathway was significantly enriched in DEGs, compared to the whole genome background (S1 Fig). This finding indicates that endocytosis efficiency might indeed contribute to the observed differences in the growth rates of two groups.

Specific differentially regulated genes

Among the 106 DEGs (S2 Table), we have selected a subset of 25 genes which might be particularly relevant for the objectives of this study (Table 2). Overall, among the DEGs, there were more downregulated than upregulated genes. To corroborate the results of the transcriptome analysis, the expression of a further subset of ten of these DEGs was studied using qPCR (Fig 3). Correlation between transcriptomic and qPCR results was relatively low (S2 Fig), but this was mostly caused by a number of outliers, whereas some genes exhibited almost perfectly congruent results between the two methods (S3 Table). As discussed before [38], this might be a consequence of different data normalization methods used for qPCR (reference gene) and RNA-seq (RPKM) data analyses [39,40]. Gene expression patterns vary widely between tissues in common carp, which poses a problem for RNA-Seq analyses, which assume that samples are comparable [27,39]. These high individual gene expression differences may also explain the relatively low number of DEGs observed, as a large SD would influence the statistical significance calculations. Furthermore, gene expression analyses in teleost fishes are additionally encumbered by a large number of co-expressed, highly similar paralogs [38]: as a result of a relatively recent (≈8 MYA) genome duplication specific to this taxonomic group, teleost fishes possess a large number of duplicated genes [41]. Although many of the superfluous gene copies have been lost in the course of evolution, in many cases two (or more) paralogs have retained their functions and continue to be expressed [42,43]. In common carp, which is an allotetraploid, possessing a large number of functional paralogs [27,44,45], this can be up to four functional paralogs. As observed in our transcriptomic results as well, sometimes paralogs forming teleost gene families are differentially expressed, which indicates that this functional redundancy may have resulted in the evolution of slightly diversified functions and very complex regulation [5,46,47]. Apart from presenting a problem for gene annotation, these paralogs can also produce apparently inconsistent expression results. For example, in cases where qPCR analysis indicates upregulation, whereas the transcriptomics indicates downregulation, it is possible that the primers designed for qPCR may have non-specifically amplified both (or more) differentially expressed paralogs present in the genome.

A number of regulated genes could be associated with metabolism, growth, development and proliferation. Most importantly for this study, two upregulated genes were associated with endocytosis pathway: *interleukin-2 receptor subunit beta* (*il2rb*; infinite upregulation) and *EH domain-containing protein 1* (*ehd1* or *intersectin-1*; +1.6). The *il2rb* gene is involved in receptor-mediated endocytosis and transduces the mitogenic signals of *il2* [48]. This gene has not been associated with animal growth before (to our knowledge), but a very recent study indicated that it might be implicated in jaw diversity in pupfishes [49]. The Ehd1 protein interacts with Insulin-like growth factor 1 (Igf-1) receptor. The level of Igf-1 in serum is known to be positively correlated with growth rate in fishes [3]. Generally, in most vertebrates, growth
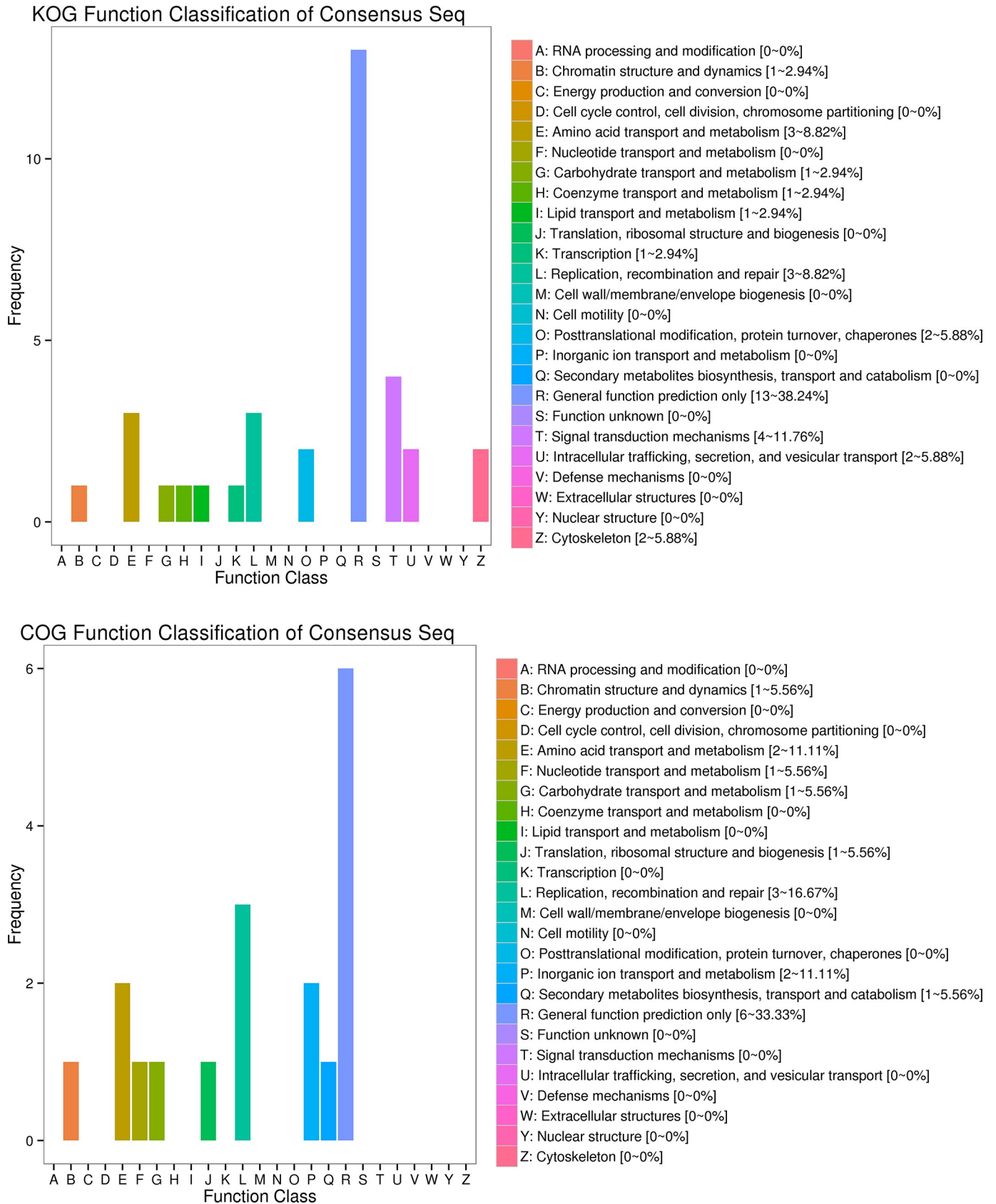
KOG Function Classification of Consensus Seq

A: RNA processing and modification [0~0%]
B: Chromatin structure and dynamics [1~2.94%]
C: Energy production and conversion [0~0%]
D: Cell cycle control, cell division, chromosome partitioning [0~0%]
E: Amino acid transport and metabolism [3~8.82%]
F: Nucleotide transport and metabolism [0~0%]
G: Carbohydrate transport and metabolism [1~2.94%]
H: Coenzyme transport and metabolism [1~2.94%]
I: Lipid transport and metabolism [1~2.94%]
J: Translation, ribosomal structure and biogenesis [0~0%]
K: Transcription [1~2.94%]
L: Replication, recombination and repair [3~8.82%]
M: Cell wall/membrane/envelope biogenesis [0~0%]
N: Cell motility [0~0%]
O: Posttranslational modification, protein turnover, chaperones [2~5.88%]
P: Inorganic ion transport and metabolism [0~0%]
Q: Secondary metabolites biosynthesis, transport and catabolism [0~0%]
R: General function prediction only [13~38.24%]
S: Function unknown [0~0%]
T: Signal transduction mechanisms [4~11.76%]
U: Intracellular trafficking, secretion, and vesicular transport [2~5.88%]
V: Defense mechanisms [0~0%]
W: Extracellular structures [0~0%]
Y: Nuclear structure [0~0%]
Z: Cytoskeleton [2~5.88%]

COG Function Classification of Consensus Seq

A: RNA processing and modification [0~0%]
B: Chromatin structure and dynamics [1~5.56%]
C: Energy production and conversion [0~0%]
D: Cell cycle control, cell division, chromosome partitioning [0~0%]
E: Amino acid transport and metabolism [2~11.11%]
F: Nucleotide transport and metabolism [1~5.56%]
G: Carbohydrate transport and metabolism [1~5.56%]
H: Coenzyme transport and metabolism [0~0%]
I: Lipid transport and metabolism [0~0%]
J: Translation, ribosomal structure and biogenesis [1~5.56%]
K: Transcription [0~0%]
L: Replication, recombination and repair [3~16.67%]
M: Cell wall/membrane/envelope biogenesis [0~0%]
N: Cell motility [0~0%]
O: Posttranslational modification, protein turnover, chaperones [0~0%]
P: Inorganic ion transport and metabolism [2~11.11%]
Q: Secondary metabolites biosynthesis, transport and catabolism [1~5.56%]
R: General function prediction only [6~33.33%]
S: Function unknown [0~0%]
T: Signal transduction mechanisms [0~0%]
U: Intracellular trafficking, secretion, and vesicular transport [0~0%]
V: Defense mechanisms [0~0%]
W: Extracellular structures [0~0%]
Y: Nuclear structure [0~0%]
Z: Cytoskeleton [0~0%]

Fig 2. KOG and COG functional classification of differentially expressed genes between the selection and control groups.

Table 2. A subset of 25 putatively relevant genes among those differentially expressed in Selected and Control groups.

Gene Name	Nr or Swiss Prot[2] annotation	eggNOG class annotation	FDR	log2FC	Reg
phyh	Phytanoyl-CoA dioxygenase, peroxisomal precursor [Osmerus mordax]	Lipid transport and metabolism	0.003956	2.31	up
il2rb	Interleukin-2 receptor subunit beta precursor [Danio rerio]	Extracellular structures	0.000423	Inf	up
H3.3	Histone H3.3 [Salmo salar]	Chromatin structure and dynamics	7.54E-07	Inf	up
itsn1	Intersectin or EH-domain containing 1a [D. rerio]	Signal transduction[1]	0.003146	1.63	up
novel β-ketoacyl	Novel protein containing a beta-ketoacyl synthase, N-terminal domain [D. rerio]	Lipid transport and metabolism	0.002079	11.54	up
flad1	FAD synthase region [D. rerio][2]	Coenzyme transport and metabolism	2.05E-17	5.77	up
tll2	PRED: cubilin [D. rerio]; Tolloid-like protein 2 [Xenopus laevis][2]	Coenzyme transport and metabolism	0.00392	Inf	up
cd276	CD276 antigen homolog (Precursor) [X. laevis][2]	General function prediction only	3.27E-07	Inf	up
calm	Calmodulin [Oreochromis mossambicus] [2]	Signal transduction mechanisms	0.004969	7.96	up
nlrp3-like	PRED: NACHT, LRR and PYD domains-containing protein 3-like [D. rerio][3]	General function prediction only	1.16E-05	2.43	up
nlrp3	PRED: NACHT, LRR and PYD domains-containing protein 3[5] [D. rerio]	General function prediction only	5	5	down
nlrp12	PRED: NACHT, LRR and PYD domains-containing protein 12[5] [D. rerio]	General function prediction only	5	5	down
setd8-a	PRED: N-lysine methyltransferase SETD8-like [D. rerio]	General function prediction only	0.000502	-5.01	down
rpgrip1l	RPGRIP1-like (K16550)[6]	Embryonic pattern specification[7]	0.000491	-inf	down
atp10-b	PRED: probable phospholipid-transporting ATPase VB [D. rerio]	Inorganic ion transport and metabolism	2.63E-06	-inf	down
epd-2	PRED: ependymin-2-like [Astyanax mexicanus]	General function prediction only	0.003711	-1.93	down
ldlrad3	PRED: low-density lipoprotein receptor class A domain-containing protein 4-like [D. rerio]	Signal transduction mechanisms	0.001095	-3.99	down
glomulin	PRED: glomulin, FKBP associated protein b isoform X1 [D. rerio]	Function unknown	5.38E-14	-6.16	down
aggf1	Angiogenic factor with G patch and FHA domains 1 [D. rerio]	General function prediction only	0.00088	-1.73	down
cxcl14	Cxcl14 protein [D. rerio]	General function prediction only	0.000458	-1.49	down
tln2	PRED: talin-2 isoform X2 [D. rerio]	Cytoskeleton	0.00066	-1.50	down
svop	PRED: synaptic vesicle glycoprotein 2C-like [A. mexicanus]	General function prediction only	1.43E-05	-5.91	down
st8sia1	alpha 2,8-sialyltransferase ST8Sia I/V/VI-r2 [D. rerio]	Carbohydrate transport and metabolism	0.000165	-6.39	down
proteoglycan 4-like	PRED: proteoglycan 4-like, partial [Poecilia reticulata]	-	8.88E-05	-inf	down
trim25	PRED: E3 ubiquitin/ISG15 ligase TRIM25 isoform X3 [D. rerio][4]	Posttranslational modification, protein turnover, chaperones	1.60E-07	-inf	down

PRED = predicted

[1] KOG class annotation

[2] Swiss Prot annotation

[3] SwissProt annotation is Neoverrucotoxin subunit alpha [*Synanceia verrucosa*, Reef stonefish]

[4] Annotation of this protein was ambiguous, depending on the database: E3 ubiquitin-protein ligase ORTHRUS 2-like in KEGG annotation, Stonustoxin subunit alpha [*Synanceia horrida*, Estuarine stonefish] in SwissProt

[5] Eight isoforms of this gene were downregulated, five of which were annotated as protein 3(-like) and three were protein 12(-like). FDR values ranged from 1.17E-07 to 5.16E-03 (2.08E-03 on average) for protein 12-like, and from 8.92E-07 to 0.000787 (0.000251 on average) for protein 3-like. Regulation ranged from -2.02 to -2.31-fold (-2.17 on average) for protein 12-like, and from -2.01 to -5.74-fold (-3.61 on average) for protein 3-like. Two of the isoforms (one protein-3 and one protein-12) were annotated as Protein NLRC5 (gene name = nlrc5) [*Ictalurus punctatus*, Channel catfish] in SwissProt database

[6] KEGG annotation

[7] GO annotation.

hormone/insulin-like growth factor (Gh/Igf) axis (a major component of which is Igf-1) plays a key role in the regulation of growth, proliferation and differentiation [6,7,9,21,50]. Particularly interesting for this study is that Gh supports muscle protein synthesis indirectly by
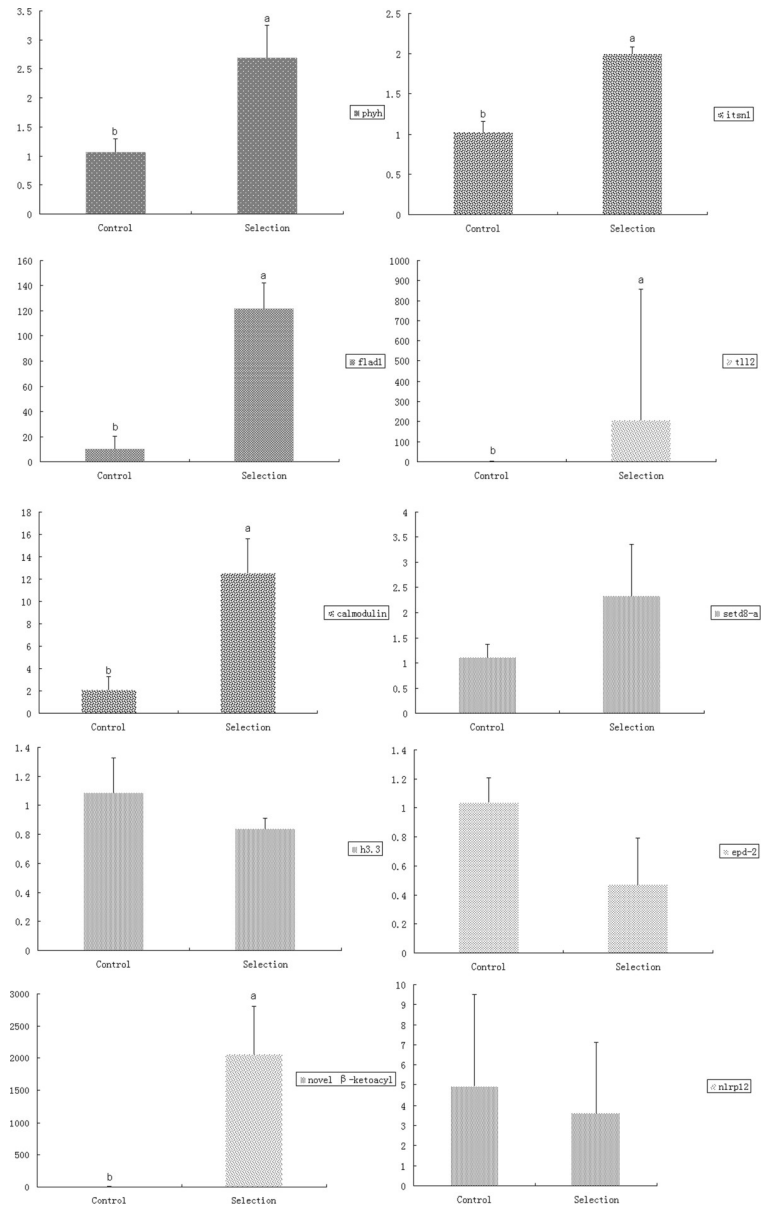
Fig 3. qPCR analysis of a subset of ten differentially expressed genes. Data were normalized to *gapdh* as the reference gene and presented as a fold change between the Selected and Control groups (mean±SD). Results were analyzed by the t-test, where different letters indicate significant (*P* < 0.05) differences.

enhancing rates of amino acid uptake from the intestine and by stimulating intestinal growth [3,51]. This indicates that the higher growth rate observed in the Selected group is likely to be at least partially attributed to increased endocytosis efficiency.

Two of the upregulated genes are known to be involved in the fatty acid metabolism: the enzyme encoded by *phytanoyl-CoA dioxygenase, peroxisomal* (*phyh*) gene (transcriptome = 2.3, qPCr = 2.5) is critical for the normal function of peroxisomes, which play a key role in the breakdown of some fatty acids [48,52]. A novel gene was annotated on the basis of its similarity with zebrafish genes: protein containing a beta-ketoacyl synthase, N-terminal domain. This gene, very highly upregulated in both transcriptome (11.5) and qPCR (710.79) analyses, is

likely to be responsible for the chain-elongation step of dissociated (type II) fatty-acid biosynthesis, i.e. the addition of two C atoms to the fatty-acid chain [53].

Several other genes upregulated in the Selected group are likely to be indicators of heightened transcription levels: *histone h3.3* gene was extremely highly upregulated in the transcriptome analysis (infinite), but only mildly upregulated (0.77) in the qPCR analysis. This gene is believed to represent an epigenetic imprint of transcriptionally active chromatin [48]. Three more upregulated genes are associated with transcription and RNA processing [54]: *rna-binding protein 5* (2.01), *n-lysine methyltransferase setd8-a* (4.15) and *fad synthase* region (*flad1*), which was very highly upregulated in both transcriptome (5.77) and qPCR (11.94) analyses. Intriguingly, two *setd8-a* paralogs were downregulated in the transcriptomic analysis (-5 and -infinite), which indicates that *setd8* paralogs might have slightly different functions and complex regulation.

Several more DEGs are associated with metabolic and developmental roles: *tolloid-like protein 2* (*tll2*) gene encodes a protease which processes procollagen C-propeptides, such as chordin. It is required for the embryonic development, where it influences dorso-ventral patterning and skeletogenesis [48,55,56]. It appears to have a broad range of functions, as it has been linked with behaviour in mice and bipolar disorder in humans [57]. Intriguingly, a knockout of this gene resulted in increased muscle mass in mice [58], which is not in agreement with our observations (positive correlation). This gene was extremely highly upregulated in both analyses: infinitely in transcriptome and 610.70 in qPCR. As tolloids were also associated with gut regeneration in sea cucumber; the authors proposed that Bmp1/Tll axis might be involved in folding of the luminal epithelium and gut looping [59]. This raises some interesting hypotheses about the correlation between *tll2* expression, proliferation of intestinal epithelium, and growth in common carp, but they need to be further experimentally tested. *Calmodulin* was very highly upregulated in both analyses: 7.96 in transcriptome and 5.92 in qPCR. It mediates the control of a large number of enzymes, including protein kinases and phosphatases [48]. Calmodulin plays an important role in the activation of phosphorylase kinase, which ultimately leads to glucose being cleaved from glycogen by glycogen phosphorylase. It also plays an important role in lipid metabolism by activating calcitonin, which is a hormone that lowers blood Ca^{2+} levels and activates G Protein cascades that lead to the generation of cAMP [48]. In fish, calmodulin has been proposed as a molecular stress indicator [60]. *Phospholipid-transporting ATPase IC* gene (*atp10b*), a component of the P4-ATPase flippase complex, which catalyzes the hydrolysis of ATP coupled to the transport of aminophospholipids from the outer to the inner leaflet of various membranes and ensures the maintenance of asymmetric distribution of phospholipids [48], was strongly downregulated (- infinite). Phospholipid translocation also seems to be implicated in vesicle formation and in uptake of lipid signaling molecules [48]. The protein is found in brain and in low levels in testis, but its expression is known to be enhanced in intestines in humans [61,62]. However, the reason for its strong downregulation in the Selected group remains unclear.

Intriguingly, two downregulated genes are associated with vision in animals: *rpgrip1-like* gene (- infinite) encodes a photoreceptor protein that is a key component of cone and rod photoreceptor cells. Mutations in this gene lead to autosomal recessive congenital blindness [54,63,64]. *Ependymin-2-like* (*epd*) was downloaded in the transcriptome analysis (-1.9) and slightly upregulated (0.46) in the qPCR analysis. It belongs to the family of proteins predominant in the cerebrospinal fluid of teleost fishes, associated with neuroplasticity and (optic nerve) regeneration [65]. This particular gene (*epd-2*), however, is expressed in non-brain tissues in fishes. In intestines, it is overexpressed in colon cancer, and its overexpression might be associated with intestinal regeneration [65]. It has also been associated with aggression levels in teleosts: inactivation of *ependymin* in subdominant fish resulted in a substantial increase

in aggression in parallel with an enhanced competitive ability [66]. Therefore, lower expression of this gene in the Selected group may result in more aggressive behaviour during feeding [67] and ultimately faster growth of the fish in this group. Two more downregulated genes have been associated with neuronal function: *ST8 Alpha-N-Acetyl-Neuraminide Alpha-2,8-Sialyl-transferase 1* (*st8sia1*: -6.4) has been associated with multiple sclerosis, as well as with metabolism of membrane-bound sphingolipids, important for cell adhesion and growth of cultured malignant cells [68]. *Synaptic vesicle 2-related* (*svop*; -5.9) is involved in neuron formation, maturation, or neuronal function [69]. Another gene associated with multiple functions in the synaptic vesicle cycle [70], *itsn1*, was upregulated in both transcriptome (1.63) and qPCR (1.94) analyses. *Low-density lipoprotein receptor class A domain-containing protein 3* (*ldlrad3*), an important component of a pathway associated with neurodegenerative diseases, including Alzheimer's [54,71], was strongly downregulated (-4.0). Similarly, *rhoGEF and pleckstrin domain-containing protein 1* (*farp1*) gene, also relatively strongly downregulated (-2.7), plays a role in the assembly and disassembly of dendritic filopodia, the formation of dendritic spines, regulation of dendrite length and ultimately the formation of synapses [48]. It is not clear whether mRNAs of these genes originate from the enteric nervous system, or whether some of them are expressed in cells of the gastrointestinal wall. Regardless, these results indicate that genetically-driven behavioural differences (higher aggression levels) may also have contributed to the higher growth of the Selected group. However, there are also indications that this population might have a higher propensity for inherited diseases leading to sight impairment, as well higher propensity towards neurodegenerative diseases in general.

A number of regulated genes were associated with the immune system: *cd276* antigen homolog (upregulated: infinite) modulates immune responses, *trim25* gene (downregulated: -infinite) is involved in innate immune defense against viruses, and *cxcl14* (downregulated: -1.5) belongs to the cytokine gene family, which encode secreted proteins involved in immunoregulatory and inflammatory processes [48]. *cxcl14* might also be involved in the homeostasis of monocyte-derived macrophages, rather than in inflammation [54]. Nine gene isoforms belonging to the NACHT, LRR and PYD domains-containing protein (*nlrp*) family were also regulated: three *nlrp12* gene isoforms and six *nlrp3* gene isoforms. Apart from one *nlrp3* isoform (upregulated: 2.43), all other isoforms were downregulated: *nlrp3* from -2.01 to -5.74, and *nlrp12* from -2.02 to -2.31. Whereas in mammals this family contains only several members, in teleost fish, NLR proteins have expanded into a huge family containing hundreds of genes, which mostly act as innate immune sensors for pathogen-associated stress signals [72]. Therefore, if we presume that studied carps were exposed to similar pathogens, downregulation of these paralogs indicates a possibility of slightly impaired immune responses of the Selected group.

Two downregulated genes are associated with the development of vasculature: mutations in *glomulin* (-6.2) have been associated with glomuvenous malformations [73], whereas *angiogenic factor with G patch and FHA domains 1 g*ene (*aggf1*; -1.73) promotes angiogenesis and proliferation of endothelial cells. Another two downregulated genes might be an indication of a higher likelihood of health-related problems in the Selected group: *talin-1* (*tln1*; -1.5) is involved in connections of major cytoskeletal structures to the plasma membrane and cell-cell contacts [48], whereas *proteoglycan 4-like* (-infinite) prevents protein deposition onto cartilage from synovial fluid by controlling adhesion-dependent synovial growth and inhibiting the adhesion of synovial cells to the cartilage surface [48].

To sum up the results: two upregulated genes are known to be involved in the fatty acid metabolism, several genes are likely to be indicators of heightened transcription levels and proliferation, several are associated with metabolic and developmental roles, several with neuronal

functions (including two with vision), several with the immune system, and two downregulated genes with the development of vasculature.

Conclusions

Gene expression analyses indicate that the higher growth rate of the Selected group is likely to be at least partially attributed to increased endocytosis efficiency and genetically-driven behavioural differences (higher aggression levels). Among the detected DEGs, *il2rb* is the most likely candidate gene to explain the increased growth rate, while *phyh* is the most likely candidate to explain the increased unsaturated fatty acid content in the muscles of the new Huanghe carp strain. Our results contribute to the understanding of the genetic basis for growth and unsaturated fatty acid content in fish, as well as vertebrates in general.

Materials and methods

Strain creation

Following the national aquaculture development strategy, a new strain of common carp was created at the Nanquan farm of the Freshwater Fishery Research Center (FFRC) of Chinese Fisheries Academy in the period between 2010 and 2016 with the aim of achieving higher growth rate and unsaturated fatty acid muscle content, and lower feeding costs [31]. The basis for the strain creation was a population (n = 1600) of the 'Yellow River' (Yuxuan Huanghe) strain specimens introduced from the Fisheries Science Research Institute of the Henan Academy of Fishery Sciences and a local Jian carp population (n = 20). A subpopulation of about 600 Huanghe and 20 Jian specimens exhibiting a comparatively high growth rate was selected in 2010 for the creation of the new strain (age ≈ 17-month-old, average weight = 197.40±75.08 g) via artificial breeding (S3 Fig). The best model of parental pair selection was inferred via BLUP analysis [74] on the basis of their breeding values and inbreeding coefficients (<0.02). 'Selected' line was created in the following way: families and individuals in each family were ranked according to their breeding values, then male and female specimens with highest breeding values were selected from the top-ranking families, and their inbreeding coefficient values were calculated. If the inbreeding coefficient value of a given pair was below 0.02 it was chosen for the Selected line family. A total of approximately 60 Selected line families (the number varied among generations) were created in this way, with each family represented by 50 (offspring) specimens produced by a selected parental pair. The control line (average number of families per generation = 20) was generated by random mating of specimens with inbreeding coefficient < 0.02. Their breeding values and the best model for parental pair selection were inferred via a BLUP analysis. When they reached the size of about 10 g (≈ 3 months of age), F1 juveniles were PIT (Passive Integrated Transponder) tagged for subsequent identification, and growth performance data (length and weight) were simultaneously collected. At 8 months of age, growth performance of the F1 offspring was measured again, which was followed by a comprehensive screening for molecular markers, BLUP analysis, and unsaturated fatty acid muscle content analysis, to estimate the breeding value of the broodstock. For the molecular markers we selected five markers (SNPs and SSRs) which we previously found to be associated with body weight in carp: *IGF2a*[4#], *IGF2R* intron 1, *D-LOOP*253, *COI*626, and Koi42 [75]. DNA samples of parents and selected (high-performing) 500 offspring were also collected. F2 parents were selected on the basis of juvenile growth performance and their molecular markers. The best model for parental pair selection was obtained via data integration with the help of SAS/STAT 8.0 (SAS Institute Inc., Cary, NC, USA) and R3.1.14 [76] programs. The breeding value was estimated using ASReml2 [77], DMU [78], Python [79], and two unpublished (but copyrighted) in-house developed programs, DNA compare1.0 and

COECAL2.1. This process was repeated for three generations (2010–2016, two years per generation; S3 Fig): one-to-one artificial propagation of the core population was conducted in the early May of 2012, 2014 and 2016 following the standard procedure [31,80]. In 2016, after three generations of selection and breeding, we had a stock of 4000 (80 × 50) selected F3 specimens, which were released into the same pond for the breeding program (not further discussed in this study). In addition to these, we randomly selected five specimens from 48 families (5 × 48 = 240 specimens) for the growth performance test (cage culture experiment) and unsaturated fatty acid muscle content analyses. Among these 48 families, 38 were from the Selected line and 10 were from the control group. When they reached the size of about 10 g ($\approx$ 3 months of age), all 240 juveniles were PIT (Passive Integrated Transponder) tagged for subsequent identification.

Growth performance and tissue sampling

To test the growth performance of the new strain, the aforementioned 240 healthy one-year-old F3 generation specimens were reared in eight 1×1×1.2 m cages (approx. 30 fish per cage) placed in earthen ponds from early September to mid-November 2016 (8 weeks). Selected and Control lines were reared together, and there was no significant difference in the initial weight (at the onset of the cage-rearing experiment) between the eight groups (cages). Light/dark schedule was natural. Fish were fed pelleted feed (S4 Table) three times daily (8:00, 12:00, 16:00) to satiation (until fish would stop taking floating pellets from the surface). The amount of feed was, thus, variable. Hand clapping was used as the feeding signal. The last feeding was on the 56th day, 8:00am, at which point the fish was eight months-old. After starving the fish for 24h, at 8:30am of the 57th day all 240 individuals (mortality rate = 0.0) were tranquilised in 20–30 mg/L MS-222 (Sigma-Aldrich, USA) as described [81]. Selected morphometric traits were recorded: weight (g, 2 kg scale), standard length, body depth, body width (all to mm, using a ruler) (S5 Table). One family from both lines (Control and Selected) was randomly chosen for tissue sampling, followed by random selection of eight specimens from the Selected group and four specimens from the Control group (S1 and S5 Tables). These twelve specimens (5 male + 7 female, S1 Table) were euthanised in buffered MS-222 at 350 mg/L concentration [81]. The fish were then dissected and a caudal portion of the epaxial muscle (> 10 g) sampled using sterile scissors and forceps, and placed in sterile plastic bags. After weighing the entire viscera, the intestine was rinsed three times with sterile phosphate buffered saline (PBS, pH = 7.0) to remove the ingesta, and foregut, midgut and hindgut separated as described [82]. Samples were immediately (<3min) placed in labelled sterile plastic bags, flash-frozen in liquid nitrogen, and stored at -80˚C. The handling of animals was conducted in accordance with the guidelines for the care and use of animals for scientific purposes set by the Institutional Animal Care and Use Committee of the Freshwater Fisheries Research Center, Chinese Academy of Fishery Sciences, Wuxi, China and the EU Directive 2010/63/EU for animal experiments. The permit to conduct this study was obtained from the institutional Animal Care and Use Committee of the Freshwater Fisheries Research Center, Chinese Academy of Fishery Sciences.

Analysis of fatty acid composition

To determine the effects of selection for fast growth on fatty acid (FA) composition of carp meat, muscle samples of eight specimens from the Selected group and four specimens from the Control group were collected (described above) and analysed. As fish muscles are mostly composed of long-chain fatty acids [33,83], we focused on the composition of long-chain (C12 to C22) FAs in this experiment. A portion ($\approx$4 g) of each muscle sample was manually crushed in liquid nitrogen using mortar and pestle and sent (on ice) to the State Key Laboratory of

Food Science (Jiangnan University, Beijing), where the total lipid extraction and analysis was performed as described before [33,84]. Briefly, lipids were extracted using chloroform/methanol (v/v = 2:1), methylated with 10% (v/v) methanolic HCl at 60˚C for 3 h, extracted with n-hexane, and analysed by gas chromatography using Agilent DB-WAXETR column (Agilent Technologies, Santa Clara, CA, US), according to the manufacturer's protocol. Pentadecanoic acid (15:0) was used as the internal standard. Statistical analysis was conducted in Excel, with the significance threshold set at 0.05.

Transcriptome analysis

RNA isolation, cDNA library construction and Illumina sequencing. Total RNA for each sample was extracted from the intestine samples, cDNA libraries constructed and sequenced, and transcriptome assembled and annotated by the Beijing Biotech Co. Ltd. (Beijing, China) roughly as described before [85,86]. In total, eight specimens from the Selected group and four specimens from the Control group were used for the transcriptome analyses. To ensure that foregut, midgut and hindgut segments are represented in the transcriptome, RNA was extracted (Trizol, Invitrogen, US) separately from the three segments, and equal amounts (of the three segments belonging to the same specimen) were then pooled together. RNA integrity and concentration were assessed using an Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA). cDNA libraries were generated using NEBNext UltraTM RNA Library Prep Kit for Illumina (NEB, USA) following the manufacturer's recommendations. One μg of RNA per sample was used as input material. Index codes were added to attribute sequences to each sample. Briefly: poly-T oligo-attached magnetic beads (NEB, E7490) were used to purify the mRNA from total RNA. Fragmentation was carried out using divalent cations under elevated temperature in NEBNext First Strand Synthesis Reaction Buffer (5X). First strand cDNA was synthesized using random hexamer primer and M-MuLV Reverse Transcriptase (RNase H). Second strand cDNA synthesis was subsequently performed using DNA Polymerase I and RNase H. Remaining overhangs were converted into blunt ends via exonuclease/polymerase activities. After adenylation of 3' ends of DNA fragments, NEBNext adaptors with hairpin loop structure were ligated to prepare for hybridisation. In order to select cDNA fragments of preferentially 200–250 bp in length, library fragments were purified with AMPure XP system (Beckman Coulter, Beverly, MA, USA). Following this, 3 μl of USER Enzyme (NEB, USA) was incubated with size-selected, adaptor-ligated cDNA at 37˚C for 15 min, followed by 5 min at 95˚C. PCR was then performed with Phusion High-Fidelity DNA polymerase, universal PCR primers and Index (X) Primer. Finally, PCR products were purified (AMPure XP system) and library quality assessed on the Agilent Bioanalyzer 2100 system. Clustering of the index-coded samples was performed on a cBot Cluster Generation System using TruSeq PE Cluster Kit v4-cBot-HS (Illumina) according to the manufacturer's instructions. After cluster generation, the library preparations were sequenced on an Illumina Hiseq 2500 platform and paired-end reads were generated.

Transcriptome assembly, annotation, ontology, and differential gene expression. Raw data (raw reads) were first processed through in-house perl scripts WipeAadpter.pl and Fastq_filter (Biomarker Technologies, Beijing): adaptor-only reads, reads containing poly-N stretches (>5% total N), and low-quality reads (Q20-value≤20) were filtered out, thereby producing 'clean reads'. These clean reads of each sample were compared to the designated reference common carp genome downloaded from the CarpBase [87] using TopHat2 [88], based on Bowtie2 algorithm [89]. The numbers of reads produced by RNA-Seq analyses were normalized to FPKM (fragments per kilobase of transcripts per million fragments mapped) to compute gene expression levels [90], and differentially expressed genes (DEGs) between the

Table 3. Primers used for qPCR.

Gene	Primer sequence (5'-3')	T_m (°C)	Fragment length
phyh	F: 5`-GCTCGTTGACTCTGTGCTGG-3`	60	400
	R: 5`- TCGTAATGCTTCGGCTTGTG-3`		
itsn1	F: 5`-GGCAGATGTTGACAAAGACG-3`	59	118
	R: 5`- CACAAGGTGAGCAGGCAGT-3`		
flad1	F: 5`-GCACCAAATCCTCCGACAG-3`	59	142
	R: 5`- TACCAGCAGAACAGCAACAA-3`C		
tll2	F: 5`-TCCAAAGGAGGATTTACTG-3`	52	179
	R: 5`- GATTTGATGGTGCCTGTAT-3`		
calmodulin	F: 5`-AGACAACTACCTGAGCACCTC-3`	57	130
	R: 5`- GACAAATAGCAGCCATCCA-3`		
setd8-a	F: 5`-ACAAGAGGCAGAAGAAGTA-3`	52	145
	R: 5`- TGCTAAAGTAGGAGGAAAC-3`		
h3.3	F: 5`-GCGAGTGCTGTTTGTTCTG-3`	57	95
	R: 5`- CTCGGTGGACTTCTGGTAG-3`		
epd-2	F: 5`-GTCTCGTCGCAGGGTGTAAT-3`	59	135
	R: 5`- CCAGACCAGGAACGGACAT-3`		
novel *β-ketoacyl*	F: 5`-TTCCCTACCCACGATACCA-3`	58	123
	R: 5`- ACATTGAACTGCCTGTGCC-3`		
nlrp12	F: 5`-TGACCAACAGAAGGAGCAG-3`	57	226
	R: 5`- CTGTATCCGCAGGAAGTGA-3`		
18S[1]	F: 5`-TTCCCTACCCACGATACCA-3`	60	
	R: 5`- ACATTGAACTGCCTGTGCC-3`		
beta actin[1]	F: 5`-TGACCAACAGAAGGAGCAG-3`	60	
	R: 5`- CTGTATCCGCAGGAAGTGA-3`		
gapdh[1]	F: 5`-CCGTTCATGCTATCACAGCTACACA-3`	62	310
	R: 5`- GTGGATACCACCTGGTCCTCTG-3`		

[1] reference genes.

https://doi.org/10.1371/journal.pone.0206615.t003

two groups (S and C) detected using EBseq software [91]. Benjamini-Hochberg procedure [92] was used to control the false discovery rate (FDR). Genes were defined as differentially expressed when they exhibited the following parameters: |fold change|$\geq$ 2 and FDR < 0.01. Genes were queried against a number of databases using BLASTx [93]: RefSeq [94], UniProt [95], GO [35], COG [36], KOG [37], Pfam [96], and KEGG [97]. After using KOBAS 2.0 [98] to obtain the KEGG Orthology results, HMMER [99] web server was employed to obtain the annotation for new genes from the predicted amino acid sequences using Pfam database. Pathway enrichment analysis was conducted using KEGG database as described before [100].

qPCR

This method was used to further investigate the expression of a subset of ten genes significantly differentially expressed between the two groups (eight specimens for the Selected group and four specimens for the Control group). Genes were selected based on their relevance for the objectives of this study and regulation magnitude. Total RNA was extracted prepared as described above from the same samples that were used for the transcriptome sequencing. cDNA libraries were prepared using PrimeScriptTM RT reagent Kit with gDNA Eraser (Takara) following the manufacturer's protocol. cDNA libraries were diluted five-fold and

used as templates for qPCR with the primers listed in Table 3. Three reference genes, *gapdh*, *beta actin* and *18S*, were tested for the stability of expression among all samples; as reflected in no significant differences between Selection and Control groups (S6 Table), all three genes exhibited a stable expression. On the basis of our previous experiences and published studies [101,102], we selected *gapdh* as the reference gene. Primers were designed using Primer Premier 5 software (Premier Biosoft, USA) on the basis of sequences of these ten genes obtained from the transcriptome data, and synthesized by the WCcgene company (Shanghai, China). qPCR was performed using ABI VIIA@7 instrument (ABI, USA) and GoTaq qPCR Master Mix (Promega, USA). In brief: the total qPCR mixture reaction volume of 10 μL contained 5 μL GoTaq qPCR Master Mix 2X, 3 μL ddH$_2$O, 0.75 μL of each primer and 0.5 μL of cDNA template. qPCR procedure: pre-incubation at 95˚C for 10 min, followed by 40 cycles of 15 s at 95˚C, 30 s at 60˚C, and 30 s at 72˚C. Melt curve analysis was performed at the end. Reactions were performed in triplicate for each sample. *gapdh* was chosen as the reference gene according to earlier recommendations [103,104]. Expression levels are presented as fold changes relative to the target gene expression in the control group, calculated using the $2^{-\Delta\Delta Ct}$ method [105]. qPCR data were analyzed statistically using Microsoft Excel and SPSS. Student-T test, implemented in SPSS, was applied to test the statistical significance of differences between the two groups, where significance thresholds were set at $P < 0.05$ (significant) and $P < 0.01$ (highly significant).

Supporting information

S1 Table. Transcriptome sequencing statistics for the 12 samples.
(DOCX)

S2 Table. Details of all DEGs.
(XLSX)

S3 Table. Comparison of gene expression results between transcriptomic and qPCR data.
Log2FC is the result of transcriptome analysis, Reg is regulation, and inf is infinite.
(DOCX)

S4 Table. Diet composition.
(DOCX)

S5 Table. Identification details and growth parameters for the 12 specimens selected for RNA-seq and fatty acid composition analyses. BWe = body weight (g), TL = total length (mm), CF = condition factor, BWi = body width (mm), BH = body height (mm). The last row in each group contains the average values of morphometric characteristics.
(DOCX)

S6 Table. Ct values of three tested reference genes in all samples.
(DOCX)

S1 Fig. Endocytosis pathway enrichment analysis in the KEGG database. Significantly enriched genes are highlighted in red.
(PNG)

S2 Fig. Correlation between transcriptome and qPCR data. Transcriptome results are presented on the x-axis (log$_2$FC) and qPCR on the y-axis. The following outliers were removed from the analysis: *h3.3* and *tll2* (inf) and novel *β-ketoacyl* (>700) (see S3 Table).
(PDF)

S3 Fig. New Huanghe carp strain selection workflow.
(PDF)

Author Contributions

Conceptualization: Baojuan Yang.

Data curation: Chengfeng Zhang, Shengyan Su.

Formal analysis: Shengyan Su, Xinyuan Li, Baojuan Yang.

Funding acquisition: Chengfeng Zhang, Shengyan Su.

Investigation: Shengyan Su, Xinyuan Li, Bing Li.

Methodology: Shengyan Su.

Project administration: Shengyan Su, Baojuan Yang.

Resources: Shengyan Su, Xinyuan Li.

Software: Shengyan Su.

Supervision: Shengyan Su, Jian Zhu, Weimin Wang.

Validation: Shengyan Su, Bing Li.

Writing – original draft: Chengfeng Zhang, Shengyan Su.

Writing – review & editing: Xinyuan Li, Jian Zhu, Weimin Wang.

References

1. Gui J, Zhu Z. Molecular basis and genetic improvement of economically important traits in aquaculture animals. Chinese Sci Bull. SP Science China Press; 2012; 57: 1751–1760. https://doi.org/10.1007/s11434-012-5213-0

2. FAO. The State of World Fisheries and Aquaculture, 2014. FAO Fisheries Proceedings. 2014. 92-5-105177-1

3. Mommsen TP. Paradigms of growth in fish. Comp Biochem Physiol Part B Biochem Mol Biol. 2001; 129: 207–219. https://doi.org/10.1016/S1096-4959(01)00312-8

4. Sire MF, Vernier J-M. Intestinal absorption of protein in teleost fish. Comp Biochem Physiol Part A Physiol. 1992; 103: 771–781. https://doi.org/10.1016/0300-9629(92)90180-X

5. Johnston IA, Bower NI, Macqueen DJ. Growth and the regulation of myotomal muscle mass in teleost fish. J Exp Biol. 2011; 214: 1617–28. https://doi.org/10.1242/jeb.038620 PMID: 21525308

6. Duan C. The Insulin-like Growth Factor System and Its Biological Actions in Fish. Am Zool. 1997; 37: 491–503. https://doi.org/10.1093/icb/37.6.491

7. Reinecke M. Influences of the environment on the endocrine and paracrine fish growth hormone-insulin-like growth factor-I system. J Fish Biol. Blackwell Publishing Ltd; 2010; 76: 1233–1254. https://doi.org/10.1111/j.1095-8649.2010.02605.x PMID: 20537012

8. Boeuf G, Boujard D, Ruyet JP- L. Control of the somatic growth in turbot. J Fish Biol. Blackwell Publishing Ltd; 1999; 55: 128–147. https://doi.org/10.1111/j.1095-8649.1999.tb01051.x

9. Vélez EJ, Lutfi E, Azizi S, Perelló M, Salmerón C, Riera-Codina M, et al. Understanding fish muscle growth regulation to optimize aquaculture production. Aquaculture. 2017; 467: 28–40. https://doi.org/10.1016/j.aquaculture.2016.07.004

10. Wood JD, Enser M, Fisher AV, Nute GR, Sheard PR, Richardson RI, et al. Fat deposition, fatty acid composition and meat quality: A review. Meat Sci. 2008; 78: 343–358. https://doi.org/10.1016/j.meatsci.2007.07.019 PMID: 22062452

11. Lunn J, Theobald HE. The health effects of dietary unsaturated fatty acids. Nutr Bull. 2006; 31: 178–224. https://doi.org/10.1111/j.1467-3010.2006.00571.x

12. Williams CM. Dietary fatty acids and human health. Ann Zootech. 2000; 49: 165–180. https://doi.org/10.1051/animres:2000116

13. Steffens W. Aquaculture produces wholesome food: cultured fish as a valuable source of n-3 fatty acids. Aquac Int. Springer International Publishing; 2016; 24: 787–802. https://doi.org/10.1007/s10499-015-9885-8

14. Stancheva M, Merdzhanova A. Fatty acid composition of common carp, rainbow trout and grey mullet fish species. Agric Sci Technol. Trakia University; 2011; 3: 285–289.

15. Martin SAM, Dehler CE, Król E. Transcriptomic responses in the fish intestine. Dev Comp Immunol. 2016; 64: 103–117. https://doi.org/10.1016/j.dci.2016.03.014 PMID: 26995769

16. Betancor MB, Li K, Sprague M, Bardal T, Sayanova O, Usher S, et al. An oil containing EPA and DHA from transgenic Camelina sativa to replace marine fish oil in feeds for Atlantic salmon (Salmo salar L.): Effects on intestinal transcriptome, histology, tissue fatty acid profiles and plasma biochemistry. Soengas JL, editor. PLoS One. Taylor & Francis, CRC Press; 2017; 12: e0175415. https://doi.org/10.1371/journal.pone.0175415 PMID: 28403232

17. De Santis C, Taylor JF, Martinez-Rubio L, Boltana S, Tocher DR, Bron J. Influence of Development and Dietary Phospholipid Content and Composition on Intestinal Transcriptome of Atlantic Salmon (Salmo salar). Fernandes JMO, editor. PLoS One. Elsevier; 2015; 10: e0140964. https://doi.org/10.1371/journal.pone.0140964 PMID: 26488165

18. Matsumoto Y, Murakami Y, Tsuru S, Ying B- W, Yomo T. Growth rate-coordinated transcriptome reorganization in bacteria. BMC Genomics. 2013; 14: 808. https://doi.org/10.1186/1471-2164-14-808 PMID: 24252326

19. Huang G, Han M, Yao W, Wang Y. Transcriptome analysis reveals the regulation of brassinosteroids on petal growth in *Gerbera hybrida*. PeerJ. PeerJ Inc.; 2017; 5: e3382. https://doi.org/10.7717/peerj.3382 PMID: 28584713

20. Puig-Oliveras A, Ramayo-Caldas Y, Corominas J, Estell? J, P?rez-Montarelo D, Hudson NJ, et al. Differences in Muscle Transcriptome among Pigs Phenotypically Extreme for Fatty Acid Composition. Guillou H, editor. PLoS One. 2014; 9: e99720. https://doi.org/10.1371/journal.pone.0099720 PMID: 24926690

21. Sun Y, Guo C-Y, Wang D-D, Li XF, Xiao L, Zhang X, et al. Transcriptome analysis reveals the molecular mechanisms underlying growth superiority in a novel grouper hybrid (Epinephelus fuscogutatus♀ × E. lanceolatus♂). BMC Genet. 2016; 17: 24. https://doi.org/10.1186/s12863-016-0328-y PMID: 26785614

22. Gao Y, Zhang H, Gao Q, Wang L, Zhang F, Siva VS, et al. Transcriptome Analysis of Artificial Hybrid Pufferfish Jiyan-1 and Its Parental Species: Implications for Pufferfish Heterosis. Yue GH, editor. PLoS One. Public Library of Science; 2013; 8: e58453. https://doi.org/10.1371/journal.pone.0058453 PMID: 23520511

23. Fu B, Wang X, Feng X, Yu X, Tong J. Comparative transcriptomic analyses of two bighead carp (Hypophthalmichthys nobilis) groups with different growth rates. Comp Biochem Physiol Part D Genomics Proteomics. 2016; 20: 111–117. https://doi.org/10.1016/j.cbd.2016.08.006 PMID: 27639030

24. Ma D, Ma A, Huang Z, Wang G, Wang T, Xia D, et al. Transcriptome Analysis for Identification of Genes Related to Gonad Differentiation, Growth, Immune Response and Marker Discovery in The Turbot (Scophthalmus maximus). Song L, editor. PLoS One. Public Library of Science; 2016; 11: e0149414. https://doi.org/10.1371/journal.pone.0149414 PMID: 26925843

25. He L, Pei Y, Jiang Y, Li Y, Liao L, Zhu Z, et al. Global gene expression patterns of grass carp following compensatory growth. BMC Genomics. 2015; 16: 184. https://doi.org/10.1186/s12864-015-1427-2 PMID: 25887225

26. Tymchuk W, Sakhrani D, Devlin R. Domestication causes large-scale effects on gene expression in rainbow trout: Analysis of muscle, liver and brain transcriptomes. Gen Comp Endocrinol. 2009; 164: 175–183. https://doi.org/10.1016/j.ygcen.2009.05.015 PMID: 19481085

27. Kolder ICRM, van der Plas-Duivesteijn SJ, Tan G, Wiegertjes GF, Forlenza M, Guler AT, et al. A full-body transcriptome and proteome resource for the European common carp. BMC Genomics. 2016; 17: 701. https://doi.org/10.1186/s12864-016-3038-y PMID: 27590662

28. W J., D A., G X., L Z., M R., M M., et al. The hookworm Ancylostoma ceylanicum intestinal transcriptome provides a platform for selecting drug and vaccine candidates. Parasites and Vectors. 2016. p. 518. https://doi.org/10.1186/s13071-016-1795-8 PMID: 27677574

29. Sun L, Yang H, Chen M, Ma D, Lin C. RNA-Seq Reveals Dynamic Changes of Gene Expression in Key Stages of Intestine Regeneration in the Sea Cucumber Apostichopus japonicas. PLoS One. 2013; 8. https://doi.org/10.1371/journal.pone.0069441 PMID: 23936330

30. Blazie SM, Babb C, Wilky H, Rawls A, Park JG, Mangone M. Comparative RNA-Seq analysis reveals pervasive tissue-specific alternative polyadenylation in Caenorhabditis elegans intestine and muscles. BMC Biol. 2015; 13: 4. https://doi.org/10.1186/s12915-015-0116-6 PMID: 25601023

31. Su S-Y, Zhang C-F, Dong Z-J, P X, Y X-H. The breeding gain of Huanghe carp (Cyprinus carpio hacmalopterus Temminck et Schlegel) new strain G3- The effect of higher carbohydrate diet on growth and fatty acid profile of huanghe carp new strain. J Yangzhou Univ (Agricultural Life Sci Ed. 2018; 39: 63–66.

32. Mráz J, Pickova J. Factors influencing fatty acid composition of common carp (Cyprinus carpio) muscle. Neuro Endocrinol Lett. 2011; 32 Suppl 2: 3–8.

33. Buchtová H, Svobodová Z, Kocour M, Velíšek J. Chemical Composition of Fillets of Mirror Crossbreds Common Carp (Cyprinus carpio L.). Acta Vet Brno. University of Veterinary and Pharmaceutical Sciences; 2010; 79: 551–557. https://doi.org/10.2754/avb201079040551

34. Mir Z, Paterson LJ, Mir PS. Fatty acid composition and conjugated linoleic acid content of intramuscular fat in crossbred cattle with and without Wagyu genetics fed a barley-based diet. Can J Anim Sci. 2000; 80: 195–197. https://doi.org/10.4141/A98-113

35. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. Nat Genet. 2000; 25: 25–29. https://doi.org/10.1038/75556 PMID: 10802651

36. Tatusov RL, Koonin E V., Lipman DJ. A Genomic Perspective on Protein Families. Science (80-). 1997; 278: 631–637. https://doi.org/10.1126/science.278.5338.631

37. Koonin E V, Fedorova ND, Jackson JD, Jacobs AR, Krylov DM, Makarova KS, et al. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. Genome Biol. 2004; 5: R7. https://doi.org/10.1186/gb-2004-5-2-r7 PMID: 14759257

38. Prisingkorn W, Prathomya P, Jakovlić I, Liu H, Zhao Y-HH, Wang W-MM. Transcriptomics, metabolomics and histology indicate that high-carbohydrate diet negatively affects the liver health of blunt snout bream (Megalobrama amblycephala). BMC Genomics. 2017; 18: 856. https://doi.org/10.1186/s12864-017-4246-9 PMID: 29121861

39. Fonseca NA, Marioni J, Brazma A. RNA-Seq gene profiling—A systematic empirical comparison. PLoS One. 2014;9. https://doi.org/10.1371/journal.pone.0107026 PMID: 25268973

40. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001; 25: 402–8. https://doi.org/10.1006/meth.2001.1262 PMID: 11846609

41. Taylor JS, Van de Peer Y, Braasch I, Meyer A. Comparative genomics provides evidence for an ancient genome duplication event in fish. Philos Trans R Soc B Biol Sci. 2001; 356: 1661–1679. https://doi.org/10.1098/rstb.2001.0975 PMID: 11604130

42. Pasquier J, Cabau C, Nguyen T, Jouanno E, Severac D, Braasch I, et al. Gene evolution and gene expression after whole genome duplication in fish: the PhyloFish database. BMC Genomics. 2016; 17: 368. https://doi.org/10.1186/s12864-016-2709-z PMID: 27189481

43. Crollius HR, Weissenbach J. Fish genomics and biology. Genome Research. 2005. pp. 1675–1682. https://doi.org/10.1101/gr.3735805 PMID: 16339365

44. David L, Blum S, Feldman MW, Lavi U, Hillel J. Recent duplication of the common carp (Cyprinus carpio L.) genome as revealed by analyses of microsatellite loci. Mol Biol Evol. 2003; 20: 1425–34. https://doi.org/10.1093/molbev/msg173 PMID: 12832638

45. Xu P, Zhang X, Wang X, Li J, Liu G, Kuang Y, et al. Genome sequence and genetic diversity of the common carp, Cyprinus carpio. Nat Genet. 2014;advance on. https://doi.org/10.1038/ng.3098 http://www.nature.com/ng/journal/vaop/ncurrent/abs/ng.3098.html#supplementary-information PMID: 25240282

46. Jakovlić I, Liu H, Wang W- M. Identification, origin and evidence for retained functionality of two IκBα paralogs in Megalobrama amblycephala. Dev Comp Immunol. 2016; 62: 89–96. https://doi.org/10.1016/j.dci.2016.05.002 PMID: 27155355

47. Jakovlić I, Wang W- M. Expression of Hox paralog group 13 genes in adult and developing Megalobrama amblycephala. Gene Expr Patterns. 2016; 21: 63–68. https://doi.org/10.1016/j.gep.2016.07.001 PMID: 27476414

48. Consortium U. UniProt: the universal protein knowledgebase. Nucleic Acids Res. Oxford University Press; 2017; 45: D158–D169. https://doi.org/10.1093/nar/gkw1099 PMID: 27899622

49. Lencer ES, Warren WC, Harrison R, McCune AR. The Cyprinodon variegatus genome reveals gene expression changes underlying differences in skull morphology among closely related species. BMC Genomics. 2017; 18: 424. https://doi.org/10.1186/s12864-017-3810-7 PMID: 28558659

50. Mullis PE. Genetic control of growth. Eur J Endocrinol. 2005; 152: 11–31. https://doi.org/10.1530/eje.1.01797 PMID: 15762183

51. Farmanfarmaian A, Sun LZ. Growth hormone effects on essential amino acid absorption, muscle amino acid profile, N-retention and nutritional requirements of striped bass hybrids. Genet Anal Biomol Eng [Genet Anal Biomol Eng]. 1999; 15: 107–113.

52. Zomer AWM, Jansen GA, van der Burg B, Verhoeven NM, Jakobs C, van der Saag PT, et al. Phyta-noyl-Coa hydroxylase activity is induced by phytanic acid. Eur J Biochem. Blackwell Science Ltd; 2000; 267: 4063–4067. https://doi.org/10.1046/j.1432-1327.2000.01451.x PMID: 10866807

53. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, De Castro E, et al. ExPASy: SIB bioinformatics resource portal. Nucleic Acids Res. 2012; 40. https://doi.org/10.1093/nar/gks400 PMID: 22661580

54. Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D. GeneCards: A novel functional genomics compendium with automated data mining and query reformulation support. Bioinformatics. 1998; 14: 656–664. https://doi.org/10.1093/bioinformatics/14.8.656 PMID: 9789091

55. Connors SA, Trout J, Ekker M, Mullins MC. The role of tolloid/mini fin in dorsoventral pattern formation of the zebrafish embryo. Development. 1999; 126: 3119–30. PMID: 10375503

56. Branam AM, Hoffman GG, Pelegri F, Greenspan DS. Zebrafish chordin-like and chordin are functionally redundant in regulating patterning of the dorsoventral axis. Dev Biol. 2010; 341: 444–458. https://doi.org/10.1016/j.ydbio.2010.03.001 PMID: 20226780

57. de Mooij-van Malsen JG, van Lith HA, Laarakker MC, Brandys MK, Oppelaar H, Collier DA, et al. Cross-species genetics converge to TLL2 for mouse avoidance behavior and human bipolar disorder. Genes, Brain Behav. 2013; 12: 653–657. https://doi.org/10.1111/gbb.12055 PMID: 23777486

58. Lee SJ. Genetic analysis of the role of proteolysis in the activation of latent myostatin. PLoS One. 2008; 3. https://doi.org/10.1371/journal.pone.0001628 PMID: 18286185

59. Mashanov VS, Zueva OR, Garcia-Arraras JE. Expression of Wnt9, TCTP, and Bmp1/Tll in sea cucumber visceral regeneration. Gene Expr Patterns. NIH Public Access; 2012; 12: 24–35. https://doi.org/10.1016/j.gep.2011.10.003 PMID: 22079950

60. Alves RN, Cordeiro O, Silva TS, Richard N, de Vareilles M, Marino G, et al. Metabolic molecular indicators of chronic stress in gilthead seabream (Sparus aurata) using comparative proteomics. Aquaculture. 2010; 299: 57–66. https://doi.org/10.1016/j.aquaculture.2009.11.014

61. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based map of the human proteome. Science (80-). 2015; 347: 1260419–1260419. https://doi.org/10.1126/science.1260419 PMID: 25613900

62. The Human Protein Atlas—ATP10B [Internet]. [cited 26 Jun 2017]. Available: http://www.proteinatlas.org/ENSG00000118322-ATP10B/tissue

63. Kostic C, Arsenijevic Y. Animal modelling for inherited central vision loss. J Pathol. John Wiley & Sons, Ltd; 2016; 238: 300–310. https://doi.org/10.1002/path.4641 PMID: 26387748

64. Miyadera K, Brierley I, Aguirre-Hernández J, Mellersh CS, Sargan DR. Multiple Mechanisms Contribute to Leakiness of a Frameshift Mutation in Canine Cone-Rod Dystrophy. PLoS One. 2012; 7. https://doi.org/10.1371/journal.pone.0051598 PMID: 23251588

65. Suárez-Castillo EC, García-Arrarás JE. Molecular evolution of the ependymin protein family: a necessary update. BMC Evol Biol. 2007; 7: 1–20.

66. Sneddon LU, Schmidt R, Fang Y, Cossins AR. Molecular correlates of social dominance: A novel role for ependymin in aggression. PLoS One. 2011; 6. https://doi.org/10.1371/journal.pone.0018181 PMID: 21483679

67. Sneddon LU. The bold and the shy: Individual differences in rainbow trout. J Fish Biol. 2003; 62: 971–975. https://doi.org/10.1046/j.1095-8649.2003.00084.x

68. Husain S, Yildirim-Toruner C, Rubio JP, Field J, Schwalb M, Cook S, et al. Variants of ST8SIA1 are associated with risk of developing multiple sclerosis. PLoS One. 2008; 3. https://doi.org/10.1371/journal.pone.0002653 PMID: 18612409

69. Logan MA, Steele MR, Vetter ML. Expression of synaptic vesicle two-related protein SVOP in the developing nervous system of Xenopus laevis. Dev Dyn. 2005; 234: 802–807. https://doi.org/10.1002/dvdy.20618 PMID: 16217738

70. Pechstein A, Shupliakov O, Haucke V, Heuser JE, Reese TS, Murthy VN, et al. Intersectin 1: a versatile actor in the synaptic vesicle cycle. Biochem Soc Trans. 2010; 38: 181–6. https://doi.org/10.1042/BST0380181 PMID: 20074056

71. Ranganathan S, Noyes NC, Migliorini M, Winkles JA, Battey FD, Hyman BT, et al. LRAD3, A Novel Low-Density Lipoprotein Receptor Family Member That Modulates Amyloid Precursor Protein Trafficking. J Neurosci. 2011; 31: 10836–10846. https://doi.org/10.1523/JNEUROSCI.5065-10.2011 PMID: 21795536

72. Howe K, Schiffer PH, Zielinski J, Wiehe T, Laird GK, Marioni JC, et al. Structure and evolutionary history of a large family of NLR proteins in the zebrafish. Open Biol. 2016; 6: 160009. https://doi.org/10.1098/rsob.160009 PMID: 27248802

73. Brouillard P, Boon LM, Mulliken JB, Enjolras O, Ghassibé M, Warman ML, et al. Mutations in a Novel Factor, Glomulin, Are Responsible for Glomuvenous\rMalformations ("Glomangiomas"). Am J Hum Genet. 2002; 70: 866–874. https://doi.org/10.1086/339492 PMID: 11845407

74. Henderson CR. Applications of Linear Models in Animal Breeding Models. Guelph: University of Guelph; 1984. https://doi.org/10.1002/9780470316856.ch7

75. Chengfeng Z, Su S, Zhu J, Shi L. Candidate parents selection containing multiple genotypes with higher performance and correlated molecular markers' genetic effects analysis. Acta Hydrobiol Sin. 2017; 41: 79–85.

76. Harrell FE. Regression modeling strategies With Applications to Linear Models, Logistic Regression, and Survival Analysis. Springer Series in Statistics. Springer; 2014. https://doi.org/10.1007/978-1-4757-3462-1

77. Gilmour AR, Butler D, Cullis BR, Thompson R, others. ASReml 2 and SAMM 2-mixed models for many occasions. Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Minas Gerais, Brazil, 13–18 August, 2006. 2006. pp. 12–27.

78. Madsen P, Sørensen P, Su G, Damgaard LH, Thomsen H, Labouriau R, et al. DMU-a package for analyzing multivariate mixed models. 8th World Congress on Genetics Applied to Livestock Production. 2006.

79. Van Rossum G, others. Python 2.7. 10 Language Reference. Samurai Media Limited; 2015.

80. Szabó T, Ditrói B, Szabó K, Bokor Z, Urbányi B. Comparison of the Efficiency of Common Carp and Silver Carp Pituitary in the Breeding of Common Carp (Cyprinus carpio) and Northern Pike (Esox lucius). Turkish J Fish Aquat Sci. 2014; 14: 841–844.

81. Topic Popovic N, Strunjak-Perovic I, Coz-Rakovac R, Barisic J, Jadan M, Persin Berakovic A, et al. Tricaine methane-sulfonate (MS-222) application in fish anaesthesia. J Appl Ichthyol. 2012; 28: 553–564. https://doi.org/10.1111/j.1439-0426.2012.01950.x

82. Harder W, Sokoloff S. Anatomy of fishes, part I: text. Stuttgart: Schweizerbartsche Verlagsbuchhandlung; 1975.

83. Yeganeh S, Shabanpour B, Hosseini H, Reza Imanpour M, Shabani A. Comparison of Farmed and Wild Common Carp (Cyprinus carpio): Seasonal Variations in Chemical Composition and Fatty Acid Profile. Czech J Food Sci. 2012; 30: 503–511.

84. Tang X, Chen H, Chen YQ, Chen W, Garre V, Song Y, et al. Comparison of Biochemical Activities between High and Low Lipid-Producing Strains of Mucor circinelloides: An Explanation for the High Oleaginicity of Strain WJ11. Bahn Y-S, editor. PLoS One. CRC press; 2015; 10: e0128396.

85. Yue H, Li C, Du H, Zhang S, Wei Q. Sequencing and De Novo Assembly of the Gonadal Transcriptome of the Endangered Chinese Sturgeon (Acipenser sinensis). PLoS One. Public Library of Science; 2015; 10: e0127332. https://doi.org/10.1371/journal.pone.0127332 PMID: 26030930

86. Liu X, Xu X, Li B, Wang X, Wang G, Li M. RNA-seq transcriptome analysis of maize inbred carrying nicosulfuron-tolerant and nicosulfuron-susceptible alleles. Int J Mol Sci. 2015; 16: 5975–5989. https://doi.org/10.3390/ijms16035975 PMID: 25782159

87. Williams DR, Li W, Hughes MA, Gonzalez SF, Vernon C, Vidal MC, et al. Genomic resources and microarrays for the common carp Cyprinus carpio L. J Fish Biol. Blackwell Publishing Ltd; 2008; 72: 2095–2117. https://doi.org/10.1111/j.1095-8649.2008.01875.x

88. 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 Biol. 2013; 14: R36. https://doi.org/10.1186/gb-2013-14-4-r36 PMID: 23618408

89. Langmead B, Trapnell C, Pop M, Salzberg S. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009; 10: R25. https://doi.org/10.1186/gb-2009-10-3-r25 PMID: 19261174

90. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 2008; 5: 621–8. https://doi.org/10.1038/nmeth.1226 PMID: 18516045

91. Leng N, Dawson JA, Thomson JA, Ruotti V, Rissman AI, Smits BMG, et al. EBSeq: An empirical Bayes hierarchical model for inference in RNA-seq experiments. Bioinformatics. 2013; 29: 1035–1043. https://doi.org/10.1093/bioinformatics/btt087 PMID: 23428641

92. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B. 1995; 57: 289–300. https://doi.org/10.2307/2346101

93. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. 1997; 25: 3389–3402. https://doi.org/10.1093/nar/25.17.3389 PMID: 9254694

94. Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. 2007; 35. https://doi.org/10.1093/nar/gkl842 PMID: 17130148

95. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, et al. UniProt: the Universal Protein knowledgebase. Nucleic Acids Res. 2004; 32: 115D–119. https://doi.org/10.1093/nar/gkh131 PMID: 14681372

96. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: Towards a more sustainable future. Nucleic Acids Res. 2016; 44: D279–D285. https://doi.org/10.1093/nar/gkv1344 PMID: 26673716

97. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Research. 1999. pp. 29–34. https://doi.org/10.1093/nar/27.1.29 PMID: 9847135

98. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, et al. KOBAS 2.0: A web server for annotation and identification of enriched pathways and diseases. Nucleic Acids Res. 2011;39. https://doi.org/10.1093/nar/gkr483 PMID: 21715386

99. Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F, et al. HMMER web server: 2015 Update. Nucleic Acids Res. 2015; 43: W30–W38. https://doi.org/10.1093/nar/gkv397 PMID: 25943547

100. Chen L, Chu C, Lu J, Kong X, Huang T, Cai YD. Gene ontology and KEGG pathway enrichment analysis of a drug target-based classification system. Wang J, editor. PLoS One. Wiley-Interscience; 2015; 10: e0126492. https://doi.org/10.1371/journal.pone.0126492 PMID: 25951454

101. Wang Y, Li Z, Lu Y, Hu G, Lin L, Zeng L, et al. Molecular characterization, tissue distribution and expression, and potential antiviral effects of TRIM32 in the common carp (Cyprinus carpio). Int J Mol Sci. 2016; 17: 1693. https://doi.org/10.3390/ijms17101693 PMID: 27735853

102. Özdemir S, Altun S, Arslan H. Imidacloprid exposure cause the histopathological changes, activation of TNF-α iNOS, 8-OHdG biomarkers, and alteration of caspase 3, iNOS, CYP1A, MT1 gene expression levels in common carp (Cyprinus carpio L.). Toxicol Reports. 2018; 5: 125–133. https://doi.org/10.1016/j.toxrep.2017.12.019 PMID: 29321977

103. MacKenzie S, Ribas L, Pilarczyk M, Capdevila DM, Kadri S, Huntingford FA. Screening for coping style increases the power of gene expression studies. PLoS One. 2009; 4. https://doi.org/10.1371/journal.pone.0005314 PMID: 19390591

104. Zhang W, Jia Y, Ji X, Zhang R, Liang T, Du Q, et al. Optimal Reference Genes in Different Tissues, Gender, and Gonad of Yellow River Carp (Cyprinus carpio var) at Various Developmental Periods. Pakistan J Zool. 2016; 48: 1615–1622.

105. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the2-DDCT method. Methods. 2001; 25: 402–408. https://doi.org/10.1006/meth.2001.1262 PMID: 11846609