



福瑞鲤 2 号不同生长速率个体肌肉组织转录组分析

王兰梅^{1,2}, 朱文彬¹, 傅建军¹, 董在杰^{1,2*}

(1. 中国水产科学研究院淡水渔业研究中心,
农业农村部淡水渔业和种质资源利用重点实验室, 江苏 无锡 214081;
2. 南京农业大学无锡渔业学院, 江苏 无锡 214128)

摘要: 鱼类通过选育获得优良生长性状的遗传分子机制目前尚不清楚。实验分析比较了 2 种不同生长速率福瑞鲤 2 号肌肉组织的转录组文库。组装后共获得 392 238 条测序序列 (contigs)。其中可对比上斑马鱼、弓斑东方鲀、青鲮、三刺鱼、尼罗罗非鱼和松浦镜鲤的蛋白序列的比例为 56.01%~77.71%。2 种不同生长速率福瑞鲤 2 号肌肉转录组比较, 获得 749 个差异表达基因, 包括 348 个上调基因和 401 个下调基因。鉴定出了与肌肉生长相关关键基因, 如 *mb*、*myl2b*、*tnni1*、*fh1*、*lamb3* 和 *pdk2* 等。研究结果为后续基因功能的研究以及鱼类新品种优良生长性状的分子机制解析奠定了基础。

关键词: 鲤; 转录组; 肌肉生长; 差异表达基因

中图分类号: Q 785; S 917.4

文献标志码: A

鲤 (*Cyprinus carpio*) 种类繁多, 是世界上最重要的食用鱼类之一。本实验所用的鲤为福瑞鲤 2 号 (freshwater fisheries research center, FFRC 2 strain), 是以建鲤 (*Cyprinus carpio* var. *jian*)、野生黄河鲤和黑龙江野鲤为原始亲本, 进行 3×3 完全双列杂交, 建立基础群体, 然后运用混合模型估算育种值, 经过连续 5 代选育, 获得的具有生长速率快、体型好、越冬成活率高等稳定遗传性状的鲤新品种。近年来, 关于鲤的遗传方面的研究主要集中在遗传标记^[1-2]、遗传图谱^[3-4]、表达序列标签 (expressed sequence tags, EST)^[5-6] 和互补 DNA (complementary DNA, cDNA) 微阵列^[7] 等。最近报道了鲤的基因组信息和遗传多样性^[8]。但至今通过选育获得的优良经济性状 (特别是生长性能), 鲤的分子机制尚不清楚。

鱼类的骨骼肌占体质量的 40%~50%, 而肌

肉的生长对幼鱼到成鱼的生长发育至关重要^[9]。具有肌源性母细胞的鱼骨骼肌纤维呈层状生长, 镶嵌样增生和肌管生成, 从而使得肌肉增加^[9]。而肌纤维的产生是通过末节分化激活和肌节形成完成的^[10]。在鱼类中, 肌肉发育的早期阶段是通过叠加新的纤维层来实现, 因此在不同时期产生的纤维仍然集中在不同的区域^[11]。对鲤后期肌肉增生的定量分析表明, 鲤肌肉中确实存在典型的卫星细胞, 但仍有 2 个肌源性细胞群存在^[11-12]。此外, 鲤幼鱼对肌肉生长的贡献减少会导致其生长缓慢^[13]。目前虽然已经报道了一些肌肉生长相关的基因^[14-15], 但还没有对鲤肌肉组织进行全面的组学研究和注释。

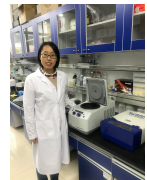
转录组分析是基因组研究和功能基因鉴定的一种有效方法^[16-17]。目前, 本实验室在鲤选育和胰岛素样生长因子 3 (*igf3*) 基因功能的研究上

收稿日期: 2020-04-09 修回日期: 2020-05-29

资助项目: 江苏省博士后科研资助 (2018K208C); 中国水产科学研究院基本科研业务费专项 (2020TD37); 现代农业产业技术体系建设专项 (CARS-45-05)

第一作者: 王兰梅 (照片), 从事渔业资源研究, E-mail: wanglm@ffrc.cn

通信作者: 董在杰, E-mail: dongzj@ffrc.cn



已经做了大量工作^[18-20]。在此基础上,本实验对快速和慢速生长的福瑞鲤2号群体肌肉组织进行了比较转录组分析,以期获得可能参与肌肉生长过程的候选基因和调控通路,从而为解析鱼类选育获得快速生长性状的分子机制提供基础。

1 材料与amp;方法

1.1 实验鱼与样品采集

福瑞鲤2号的选育在中国水产科学研究院淡水渔业研究中心宜兴岷亭试验基地进行。具体的育种过程详参见Dong等^[18]的方法。选择同一世代6月龄的鱼进行取样,首先从不同家系2955尾鱼中分别挑选生长速率快(high-growth, HG)和生长速率慢(slow-growth, SG)的鱼各44条(3%),生长速率快和生长速率慢的鲤平均末体质量分别为(522.3±11.0)g和(314.9±7.2)g。然后再从2组44条鱼中分别随机挑选8条(4雄4雌)鱼,获得其肌肉组织,液氮速冻后置于-80℃冰箱中保存,用于后续RNA提取建库。

1.2 RNA提取, cDNA文库构建和测序

采用TRIzol (Invitrogen, 卡尔斯巴德, 美国)方法提取福瑞鲤2号肌肉组织的总RNA。提取的RNA样品用DNase I酶(New England Biolabs, Frankfurt/M., Germany)消化后,使用Nanodrop2000分光光度计(Thermo Fisher Scientific, BRIMS, Cambridge, MA, 美国)检测RNA纯度。HG和SG组8条鱼的肌肉RNA样品混合后分别建库。使用TruSeq™ RNA样品制备试剂盒(Illumina)的操作步骤构建HG和SG两个文库。采用Illumina HiSeq 2500(Illumina, 圣地亚哥, 美国)测序仪,对每个文库进行2×125 bp双末端测序。

1.3 鲤转录组组装

采用Fastqc软件(<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>)对所有的原始测序片段(raw reads)进行质量控制。去除带接头的、无法确定碱基信息的、比例大于10%的以及低质量(质量值≤5的碱基数占整个测序片段50%以上)的测序片段。过滤后的测序片段(clean reads)采用Tophat软件与鲤的参考基因组(松浦镜鲤)进行比对。然后使用Trinity软件的默认参数将clean reads从头组装成更大的转录本序列(contigs)。

<https://www.china-fishery.cn>

最后,使用cuffmerge软件进行聚类合并,得到去冗余后的转录本信息用于后续分析。

1.4 功能注释, 组装评估和开放阅读框(ORF)鉴定

首先将获得的转录本序列与4个公共数据库进行比对,包括NCBI Nr (non-redundant protein)、Universal Protein Resource database、COG (cluster of orthologous groups of protein)和KEGG (Kyoto encyclopedia of genes and genomes)数据库。后续详细的功能注释和组装评估过程见Zhu等^[21]的方法。采用内部Perl脚本来鉴定比对到Nr数据库的测序序列的ORF区域。对于未比对到数据库的测序序列,使用Transdecoder (<http://transdecoder.sourceforge.net/>)来预测假定的ORFs区域。

1.5 差异表达基因的鉴定

采用Bowtie软件,比对测序片段(clean reads)到鲤的转录本序列,基于比对结果,采用RSEM软件对每个转录本进行定量,最后采用EdgeR软件包,进行两两差异表达分析。两组间差异表达基因的判断标准是 $|\log_2FC$ (fold change 倍数变化) $|\geq 1$ 且FDR (false discovery rate 错误率) ≤ 0.05 。GO(gene ontology)富集性和KEGG代谢通路分析按文献[21-22]方法进行。

1.6 实时荧光定量PCR验证

为了验证测序结果的准确度,实验随机挑选了20个肌肉生长相关的基因进行了实时荧光定量PCR验证,其中包括8个上调基因和12个下调基因。实时荧光定量PCR在ABI PRISM 7500 Real-time PCR系统上进行。反应体系为10 μL: 5 μL 2×SYBR Green MasterMix, 1 μL cDNA模板和上下游引物各0.2 μL(10 μmol/L), 3.6 μL H₂O₂。反应条件: 95℃ 5 min 变性,接着是95℃ 15 s,共40个循环,然后60℃ 45 s 退火。具体参见Zhu等^[21]的方法。

2 结果

2.1 测序和组装结果

测序获得的福瑞鲤2号HG和SG文库的原始测序片段(raw reads)分别为48 868 398和49 494 796条,过滤后得到的clean reads分别为48 267 178和48 924 886条。所有测序结果已经

中国水产学会主办 sponsored by China Society of Fisheries

提交到 NCBI SRA 数据库 (SRP078896)。

Trinity 组装后获得 392 238 条测序序列 (contigs), 对应 348 144 个基因。序列长度为 200~32 499 bp, 平均长度为 603.27 bp。从总体序列的长度分布看, 大部分长度为 200~1 000 bp, 只有 13.37% (52 436) 的序列长度为 1 000~32 499 bp。

2.2 功能注释

利用 BLASTx 将所有组装的序列分别比对到 Nr、UniProt-SwissProt、KEGG 和 COG 4 个数据库, 总共比对上 113 238 条序列, 对应 82 821 个基因。其中比对到 Nr 数据库的序列最多, 为 112 001 条。e 值 (10 为底的指数幂 e 值) 分布显示有 82.66% 的 contigs 的注释 e 值 $<1 \times 10^{-10}$, 表明注释结果的可信度高。根据比对结果所做的柱状图显示, 福瑞鲤 2 号的序列注释结果与斑马鱼 (*Danio rerio*) 的百分比最高 (图 1-a)。

对 102 845 条序列进行 KEGG 分析, 获得 322 条代谢通路。其中序列数量最多的前 20 条 KEGG 通路如图 1-b 所示。

利用 COG 数据库对所有组装的测序序列进行功能和系统分类。分析表明, 只有少部分的序列 (14 803 条) 能被归入 COG 类别中。第一大类为一般功能预测类 (general function prediction) (5 095 条, 23.85%), 其次为 2 707 条序列参与到复制、重组和修复类 (replication, recombination and repair) (12.67%) 和 1 814 个序列与信号转导机制类 (signal transduction mechanisms) 相关联 (8.49%)。而关联最少的类别为细胞运动类 (cell motility) 和核结构 (nuclear structure) 类, 均只有 0.06%。

组装结果与 Nr 数据库蛋白序列比对, 获得序列的功能信息, 进行 GO 注释。结果显示, 有 31 406 条测序序列 (8.01%) 匹配到 GO 类别, 对应 20 826 个基因 (图 1-c)。总共匹配到 63 个 GO 类, 归为 3 大类, 分别是生物进程 (biological process) (23, 36.51%)、细胞组分 (cellular component) (19, 30.16%) 和分子功能 (molecular function) (21, 33.33%)。

2.3 转录组组装评估

利用 BLASTx (以 $e < 1 \times 10^{-5}$ 为比对参数), 将福瑞鲤 2 号肌肉组织的转录组组装序列与其他物种的蛋白序列进行比对, 有 146 148 条 (37.26%) 序列可与斑马鱼、弓斑东方鲀 (*Fugu ocellatus*)、

青鳉 (*Oryzias latipes*)、三刺鱼 (*Gasterosteus aculeatus*)、尼罗罗非鱼 (*Oreochromis niloticus*) 和松浦镜鲤的蛋白序列比对上。基于相似度结果, 有 71 272 条 (18.17%) 序列可比对到所有 6 种硬骨鱼类中, 但与每种鱼的匹配度不同, 为 21.13%~34.82%。福瑞鲤 2 号的肌肉转录组序列可比对到以上 6 种鱼类, 蛋白的百分比为 56.01%~77.71% (表 1)。

2.4 ORF 鉴定和预测

对组装的福瑞鲤 2 号肌肉转录组序列进行 Nr 数据库比对, 发现 112 001 条注释序列包含 ORF。ORF 的长度为 27~32 490 bp, 平均长度为 554.68 bp。在这些序列中, 大部分 (89 478 条, 79.89%) 的 ORF 完整性低于 70%, 只有 5 011 (4.47%) 和 8 096 (7.22%) 条序列的 ORF 完整性为 95% 和 100%。对于未被注释的序列, 利用 Transdecoder 软件来检测潜在的 ORF 区域。结果显示, 有 59 606 条序列包含 ORF 框, 平均长度为 212.55 bp, 最小和最大长度分别为 147 和 2 070 bp。总共有 220 631 条序列不包含 ORF 框, 可能是因为它们来自非编码基因或未完成组装的非编码区。

2.5 福瑞鲤 2 号肌肉转录组差异基因表达

根据 $|\log FC| \geq 1$ 和 $FDR \leq 0.05$ 的标准, 鉴定出 HG 相对于 SG 组鱼的差异表达基因 749 个, 其中 348 个上调基因, 401 个下调基因 (图 2)。在这些差异表达基因中, 观察到了一些肌肉生长相关的基因, 如肌红蛋白 (myoglobin, mb)、肌球蛋白轻链 2b (myosin, light chain 2b, myl2b)、肌球蛋白轻链 4 (myl4)、肌钙蛋白 I 型 (troponin I type, tnni1)、tnni2、四和半 LIM 结构域蛋白 1 (four and a half LIM domains protein 1, fh11)、丙酮酸脱氢酶激酶同工酶 2 (pyruvate dehydrogenase kinase isozyme 2, pdk2) 和层黏连蛋白 $\beta 3$ (laminin subunit beta-3, lamb3) 等。

为了验证测序结果的准确度, 实验随机挑选了 20 个肌肉生长相关的基因进行了实时荧光定量 PCR 验证, 其中包括 8 个上调基因和 12 个下调基因。所有上调基因的表达模式与测序结果一致, 12 个下调基因中的 11 个表达模式与测序结果一致 (图 3), 表明测序结果的可信度高。

2.6 GO 和 KEGG 通路富集分析

对所有差异表达基因进行 GO 富集和 KEGG

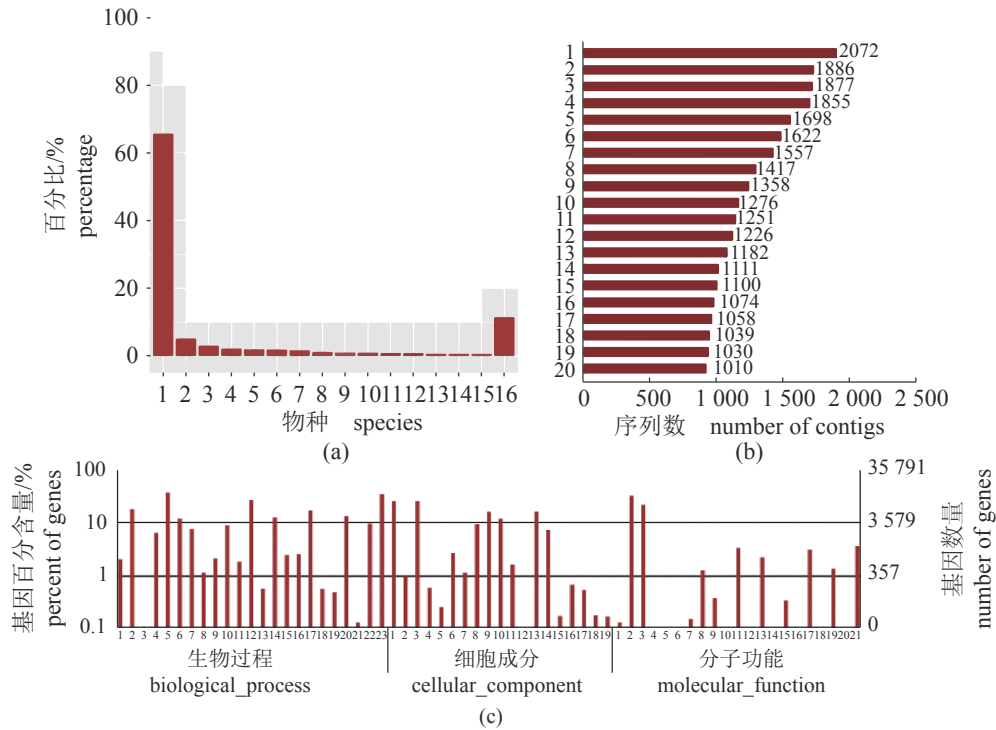


图 1 鲤肌肉转录组功能注释

(a) 基于功能注释结果的同源种分布, 1. 斑马鱼, 2. 墨西哥脂鲤, 3. 虹鳟, 4. 鲤, 5. 大黄鱼, 6. 大西洋鲑, 7. 尼罗罗非鱼, 8. 白斑狗鱼, 9. 美丽硬仆骨舌鱼, 10. 青鳉, 11. 南极鱼, 12. 绿河鲀, 13. 雀鲷, 14. 鲫, 15. 雀鳝, 16. 其他。(b) 前 20 的 KEGG 通路, 1. PI3K-Akt 信号通路, 2. 癌症蛋白聚糖通路, 3. 粘着斑通路, 4. 癌症相关通路, 5. MAPK 信号通路, 6. 紧密连接通路, 7. 肌动蛋白细胞骨架的调节, 8. 甲型流感 A, 9. 钙信号通路, 10. 人类嗜 T 淋巴细胞病毒 I 型感染, 11. 内吞作用, 12. 神经组织配体受体交互作用, 13. 胰岛素信号通路, 14. 阿尔茨海默病, 15. 吞噬体, 16. 亨廷顿氏舞蹈症, 17. 病毒致癌作用, 18. 核糖体, 19. 扩张型心肌病, 20. 病毒性心肌炎。(c) 基因功能注释分类。生物过程: 1. 生物附着, 2. 生物调节, 3. 细胞凋亡, 4. 细胞组分或生物起源, 5. 细胞进程, 6. 发育过程, 7. 定位构建, 8. 生长, 9. 免疫系统过程, 10. 定位, 11. 移动, 12. 代谢过程, 13. 多生物过程, 14. 多细胞生物过程, 15. 生物过程的负调控, 16. 生物过程的正调控, 17. 生物过程调控, 18. 繁殖, 19. 生殖过程, 20. 刺激反应, 21. 节奏过程, 22. 信号, 23. 个体生物过程; 细胞成分: 1. 细胞, 2. 细胞连接, 3. 细胞组分, 4. 细胞外基质, 5. 细胞外基质组分, 6. 胞外区, 7. 胞外区组分, 8. 高分子配合物, 9. 膜, 10. 膜组分, 11. 膜封闭腔, 12. 类核, 13. 细胞器, 14. 细胞器组分, 15. 膜边缘, 16. 突触, 17. 突触组分, 18. 病毒体, 19. 病毒体组分; 分子功能: 1. 抗氧化活性, 2. 结合, 3. 催化活性, 4. 通道调节活性, 5. 化学诱导物活性, 6. 化学排斥物活性, 7. 电子载体活性, 8. 酶调节活性, 9. 鸟苷-核苷酸交换因子活性, 10. 金属伴侣活性, 11. 分子传感器活性, 12. 形态发生素活性, 13. 核酸结合转录因子活性, 14. 营养储备活性, 15. 蛋白质结合转录因子活性, 16. 蛋白标记, 17. 受体活性, 18. 受体调控活性, 19. 生结构分子活性, 20. 翻译调节活性, 21. 转运活性

Fig. 1 Function annotation of *C. carpio* muscle transcriptome assembly

(a) homologous species distribution based on the function annotation results, 1. *Danio rerio*, 2. *Astyanax mexicanus*, 3. *Oncorhynchus mykiss*, 4. *Cyprinus carpio*, 5. *Larimichthys crocea*, 6. *Clupea harengus*, 7. *Oreochromis niloticus*, 8. *Esox lucius*, 9. *Sclerophages formosus*, 10. *Oryzias latipes*, 11. *Notothenia coriiceps*, 12. *Tetraodon nigroviridis*, 13. *Stegastes partitus*, 14. *Carassius auratus*, 15. *Lepisosteus oculatus*, 16. other. (b) top 20 pathways classification of contigs by KEGG pathway, 1. PI3K-Akt signaling pathway, 2. proteoglycans in cancer, 3. focal adhesion, 4. pathways in cancer, 5. MAPK signaling pathway, 6. tight junction, 7. regulation of actin cytoskeleton, 8. influenza A, 9. calcium signaling pathway, 10. HTLV-I infection, 11. endocytosis, 12. neuroactive ligand-receptor interaction, 13. insulin signaling pathway, 14. alzheimer's disease, 15. phagosome, 16. Huntington's disease, 17. viral carcinogenesis, 18. ribosome, 19. dilated cardiomyopathy, 20. viral myocarditis. (c) classification of assembled contigs annotated with Gene ontology; biological process: 1. biological adhesion, 2. biological regulation, 3. cell killing, 4. cellular component organization or biogenesis, 5. cellular process, 6. developmental process, 7. establishment of localization, 8. growth, 9. immune system process, 10. localization, 11. locomotion, 12. metabolic process, 13. multi-organism process, 14. multicellular organismal process, 15. negative regulation of biological process, 16. positive regulation of biological process, 17. regulation of biological process, 18. reproduction, 19. reproductive process, 20. response to stimulus, 21. rhythmic process, 22. signaling, 23. single-organism process; cellular component: 1. cell, 2. cell junction, 3. cell part, 4. extracellular matrix, 5. extracellular matrix part, 6. extracellular region, 7. extracellular region part, 8. macromolecular complex, 9. membrane, 10. membrane part, 11. membrane-enclosed lumen, 12. nucleoid, 13. organelle, 14. organelle part, 15. side of membrane, 16. synapse, 17. synapse part, 18. virion, 19. virion part; molecular function: 1. antioxidant activity, 2. binding, 3. catalytic activity, 4. channel regulator activity, 5. chemoattractant activity, 6. chemorepellent activity, 7. electron carrier activity, 8. enzyme regulator activity, 9. guanyl-nucleotide exchange factor activity, 10. metallochaperone activity, 11. molecular transducer activity, 12. morphogen activity, 13. nucleic acid binding transcription factor activity, 14. nutrient reservoir activity, 15. protein binding transcription factor activity, 16. protein tag, 17. receptor activity, 18. receptor regulator activity, 19. structural molecule activity, 20. translation regulator activity, 21. transporter activity

表 1 福瑞鲤 2 号肌肉转录组的组装评估

Tab. 1 Statistic assessment of FFRC 2 strain *C. carpio* transcriptome

数据库 database	转录组匹配序列数 FFRC 2 strain <i>C. carpio</i> hits	转录组匹配序列百分比/% percentage of total FFRC 2 strain <i>C. carpio</i> contigs	蛋白质/条 unique protein	转录组匹配蛋白百分比/% percentage of total unique proteins
斑马鱼 <i>D. rerio</i>	107 124	27.31	29 247	65.74
弓斑东方鲀 <i>F. ocellatus</i>	82 884	21.13	26 794	56.01
青鳞 <i>O. latipes</i>	86 742	22.11	18 750	75.99
三刺鱼 <i>G. aculeatus</i>	85 657	21.84	20 326	73.71
松浦镜鲤 <i>C. carpio</i>	136 571	34.82	35 403	75.96
尼罗罗非鱼 <i>O. niloticus</i>	90 706	23.13	20 797	77.71

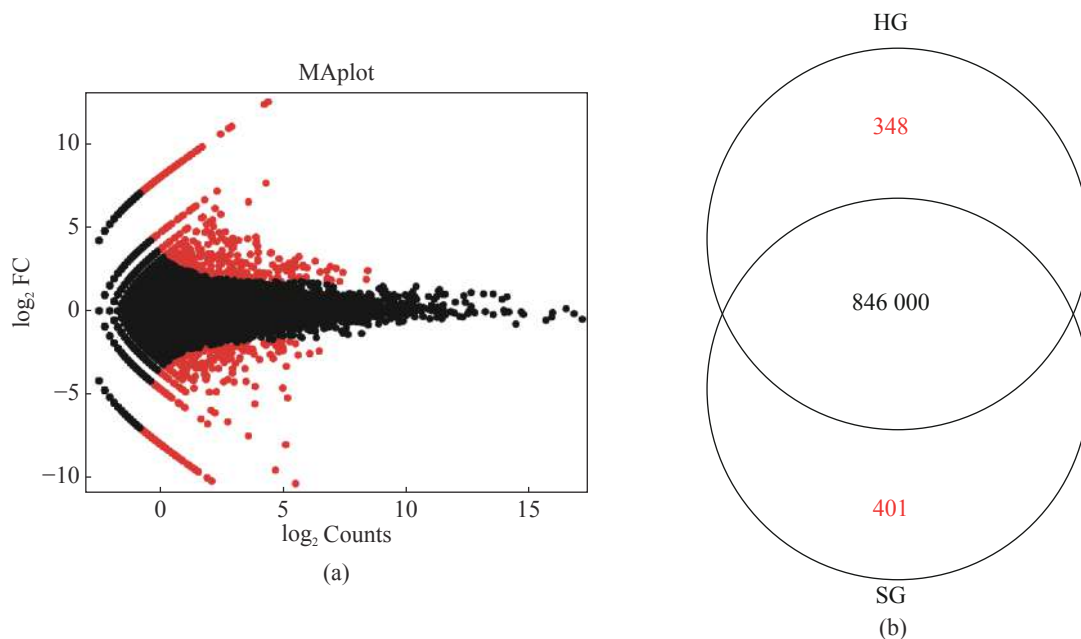


图 2 不同生长速率福瑞鲤 2 号肌肉组织基因表达

(a) M-A 图, y 轴表示折叠变化的对数, x 轴表示计数的对数。(b) 不同生长速率福瑞鲤 2 号肌肉差异表达基因数量的 Venn 图。HG. 快速生长组, SG. 慢速生长组。其中红色的点和数字表示差异表达的上调和下调的基因数, 黑色的点和数字表示无差异表达的基因数

Fig. 2 Gene expression in the muscle of different growth rate FFRC No. 2 strain *C. carpio*

(a) M-A plot, the y-axis represents the logarithm of fold change and the x-axis represents the logarithm of read counts. (b) a Venn diagram with the number of differentially expressed genes at the different growth speed FFRC. HG. high growth, SG. slow growth. The red dots and numbers represent the up-regulated and down-regulated differentially expressed genes, while the black dots and numbers represent the number of genes without differential expression

通路分析, 以获得其潜在的功能注释。本研究发现, 一些差异表达基因匹配到肌肉相关的富集中, 如肌动蛋白细胞骨架 (GO:0015629)、肌丝 (GO:0036379)、肌球蛋白复合体 (GO:0016459)、细胞骨架部分 (GO:0044430)、肌钙蛋白复合体 (GO:0005861)、肌原纤维 (GO:0030016)、收缩纤维 (GO:0043292)、肌原纤维组装 (GO:0030239)、肌细胞发展 (GO:0055001)、骨骼肌细肌丝组装 (GO:0030240) 和肌球蛋白纤维装配 (GO:0031034)

等。这些与肌肉生长相关的差异表达基因信息量丰富, 值得进一步研究。此外, KEGG 通路分析结果显示, 心肌收缩 (cardiac muscle contraction) 和肥厚性心肌症 (hypertrophic cardiomyopathy HCM) 通路显著富集差异表达基因, 其中 *tnni1*、*tnni2*、肌球蛋白结合蛋白 C (myosin-binding protein C, *mybpc1*)、辅肌动蛋白 2 (alpha-actinin-2, *actn2*) 和 β -原肌球蛋白 (tropomyosin beta chain, *tpm2*) 为关键调控蛋白。

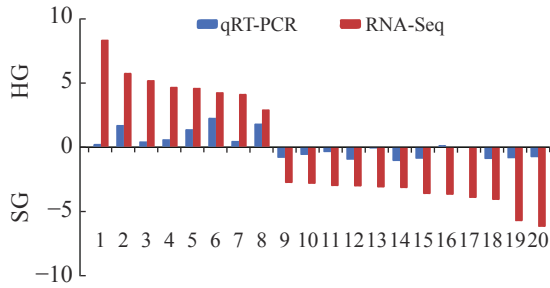


图3 测序和实时荧光定量PCR的基因表达模式

β -actin 为内参基因。1. 丙酮酸激酶 PKM 样亚型 X1 基因, 2. 纤溶酶原激活物表面受体, 3. β 白细胞介素, 4. Krt5 蛋白, 5. 角蛋白 8, 6. 二磷酸果糖激酶, 7. 丙酮酸脱氢酶同工酶 2, 8. 丝裂原活化蛋白激酶 4, 9. 肌红蛋白, 10. 肌钙蛋白 2e, 11. 原肌球蛋白 2, 12. 肌钙蛋白 1, 13. 4 个半 LIM 结构域蛋白 1, 14. 肌钙蛋白 I 型, 15. α 辅肌动蛋白 2, 16. 肌球蛋白结合蛋白 C1, 17. 重链肌球蛋白 1, 18. 轻链肌球蛋白 4, 19. 轻链肌球蛋白 2, 20. 重链肌球蛋白

Fig. 3 Gene expression patterns of RNA-Seq and qRT-PCR

β -actin was used as an internal control and used for the normalization of the expression level of each gene. 1. pyruvate kinase PKM-like isoform X1, *pkm2*; 2. urokinase plasminogen activator surface receptor, *plaur*; 3. interleukin-1 beta, *il1b*; 4. Krt5 protein, *krt5*; 5. keratin 8, *krt8*; 6. 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase 1, *pfkfb1*; 7. pyruvate dehydrogenase kinase isozyme 2, *pdh2*; 8. mitogen-activated protein kinase 4, *map2k4*; 9. myoglobin, *mb*; 10. troponin T2e, *tnnt2e*; 11. tropomyosin, *tpm2*; 12. troponin I type 1b, *tnni1*; 13. four and a half LIM domains protein 1, *fhl1*; 14. troponin I type 2a-1, *tnni2a-1*; 15. alpha-actinin-2, *actn2*; 16. myosin-binding protein C1, *mybpc1*; 17. slow myosin heavy chain 1, *sm1*; 18. myosin light chain 4, *myl4*; 19. myosin light chain 2, *myl2*; 20. myosin heavy chain, *mhc*

3 讨论

本研究中, 所有组装序列分别比对到 Nr、UniProt-SwissProt、KEGG 和 COG 数据库。结果显示, 组装序列与斑马鱼的同源性最高 (65.68%), 这可能是由于模式生物斑马鱼具有完整的基因注释, 且斑马鱼和鲤均为鲤科 (Cyprinidae) 鱼类。

肌生成是一个复杂且精确调控的过程, 可导致许多神经、心脏和肌肉功能障碍性疾病^[23]。在胚胎时期, 起源于体细胞的肌原性前体细胞首先生成肌细胞, 经过增殖、迁移、融合成多核的肌小管, 最终分化成成熟的肌纤维^[24]。出生后骨骼肌的生长则是通过增加纤维体积来实现的^[25]。研究表明, 骨骼肌纤维的大小和肌肉质量 (如肌肉生长或肥大) 会随着生理和病理条件的不同而变化^[26]。骨骼肌和肌纤维的类型对同一刺激的反应不同。一般认为, 肌纤维可分为两

大类: 慢收缩 (I 型) 和快收缩型 (II 型)。本研究中, HG 和 SG 组鱼肌肉组织的差异表达基因中, 与肌肉生长相关的基因包括编码肌红蛋白、肌球蛋白、肌钙蛋白 I 型和角蛋白 5 的基因。据报道, *tnni1*、*mb* 和 *myl2b* 为慢型肌肉蛋白编码基因^[23], 本研究中, 此 3 个基因在 SG 组鱼中的表达量上调, 提示这些基因可能导致鱼生长缓慢。*tnni1* 只表达于成体慢骨骼肌纤维, 其编码的蛋白可以阻止肌肉收缩, 这与本研究中, SG 组鱼肌肉中 *tnni1* 表达上调的结果相一致。在 HG 和 SG 组差异表达的基因中, 还包括编码表皮生长因子、含黏液样受体 2 (module-containing mucin-like receptor 2, *emr2*) 和肌球蛋白结合蛋白 C (myosin-binding protein C, *mybpc*) 基因。*mybpc1* 是肌球蛋白结合蛋白 C 的骨骼肌慢型, 通过向肌球蛋白激活肌肉型肌酸激酶而在肌肉收缩中发挥重要作用^[27]。本研究中, *mybpc1* 在慢生长鱼类肌肉中的表达水平也较高。

本研究还鉴定了几种在 HG 和 SG 组鱼肌肉转录组中差异表达的 LIM 域蛋白编码基因。一些 LIM 蛋白已被证实与肌动蛋白细胞骨架相关, 在各种细胞过程中的肌动蛋白丝的信号转导和组织中发挥作用^[28]。*fhl1* 作为 LIM 结构域蛋白基因, 在小鼠发育过程中低表达^[29]。本研究中, *fhl1* 在 HG 组福瑞鲤 2 号肌肉组织中的低表达, 表明 *fhl1* 可能是肌肉生长和肥大的调节因子。*lamb3* 在层黏连蛋白 5 的形成过程中起重要作用, 层黏连蛋白是一组调节细胞生长、细胞运动和细胞间连接的蛋白质。层黏连蛋白 5 也是一种被称为锚定细丝 (anchoring filaments) 纤维的主要组成部分, 锚定细丝连接两层基底膜并将皮肤连在一起^[30-31]。另一个上调的基因是 *pdh2*, 它在调节葡萄糖和脂肪酸代谢方面起重要作用^[32]。研究认为, 代谢率的差异可能是导致肌肉量不同的一个因素^[23, 33]。本研究中, 相较于 SG 组, HG 组中 *lamb3* 和 *pdh2* 的表达水平更高, 说明这些基因在控制肌肉生长方面起着重要作用, 且 HG 组鱼体内葡萄糖和脂肪酸的代谢可能发生了变化。

GO 和 KEGG 富集分析表明, 不同生长速率福瑞鲤 2 号肌肉转录组的差异表达基因富集在不同的基因簇和代谢通路中。如 *tpm2* 编码的 β -原肌球蛋白是有条纹的肌肉特异性卷曲螺旋二聚体, 其功能是稳定肌动蛋白丝和调节肌肉收缩。据报道, α -原肌球蛋白 (*tpm1*) 更多的存在于快骨

骨骼肌和心肌中, 而 *tpm2* 则更多的存在于慢骨骼肌中^[34]。本研究中, *tpm2* 在 SG 组鱼肌肉中的表达量也高于 HG 鱼。

此外, 在 KEGG 代谢通路中还发现了几个差异不显著但与肌肉发育相关的通路, 包括脂肪消化吸收通路、脂肪酸降解通路和 mTOR 信号通路。本实验还发现了维生素的消化和吸收通路, 表明维生素相关的营养代谢可能参与了肌生成过程。此外, 还发现了胰岛素信号通路、PI3K-Akt 信号通路和肌动蛋白骨架通路的调控。可见, 肌生成是一个多通路调控的复杂过程, 值得进一步研究。

参考文献 (References):

- [1] Ji P F, Zhang Y, Li C, *et al.* High throughput mining and characterization of microsatellites from common carp genome[J]. *International Journal of Molecular Sciences*, 2012, 13(8): 9798-9807.
- [2] Xu J, Ji P F, Zhao Z X, *et al.* Genome-wide SNP discovery from transcriptome of four common carp strains[J]. *PLoS One*, 2012, 7(10): e48140.
- [3] Zhang X F, Zhang Y, Zheng X H, *et al.* A consensus linkage map provides insights on genome character and evolution in common carp (*Cyprinus carpio* L.)[J]. *Marine Biotechnology*, 2013, 15(3): 275-312.
- [4] Sun X W, Liang L Q. A genetic linkage map of common carp (*Cyprinus carpio* L.) and mapping of a locus associated with cold tolerance[J]. *Aquaculture*, 2004, 238(1-4): 165-172.
- [5] Zhang Y J, Stupka E, Henkel C V, *et al.* Identification of common carp innate immune genes with whole-genome sequencing and RNA-Seq data[J]. *Journal of Integrative Bioinformatics*, 2011, 8(2): 169.
- [6] Ji P F, Liu G M, Xu J, *et al.* Characterization of common carp transcriptome: sequencing, *de novo* assembly, annotation and comparative genomics[J]. *PLoS One*, 2012, 7(4): e35152.
- [7] Williams D R, Li V, Hughes M A, *et al.* Genomic resources and microarrays for the common carp *Cyprinus carpio* L.[J]. *Journal of Fish Biology*, 2008, 72(9): 2095-2117.
- [8] Xu P, Zhang X F, Wang X M, *et al.* Genome sequence and genetic diversity of the common carp, *Cyprinus carpio*[J]. *Nature Genetics*, 2014, 46(11): 1212-1219.
- [9] Eldem V, Zararsiz G, Erkan M, *et al.* *De novo* assembly and comprehensive characterization of the skeletal muscle transcriptomes of the European anchovy (*Engraulis encrasicolus*)[J]. *Marine Genomics*, 2015, 20: 7-9.
- [10] Garcia de la Serrana D, Devlin R H, Johnston I A. RNAseq analysis of fast skeletal muscle in restriction-fed transgenic coho salmon (*Oncorhynchus kisutch*): an experimental model uncoupling the growth hormone and nutritional signals regulating growth[J]. *BMC Genomics*, 2015, 16(1): 564.
- [11] Koumans J T M, Akster H A. Myogenic cells in development and growth of fish[J]. *Comparative Biochemistry and Physiology-Part A: Physiology*, 1995, 110(1): 3-20.
- [12] Koumans J T M, Akster H A, Booms G H R, *et al.* Growth of carp (*Cyprinus carpio*) white axial muscle; hyperplasia and hypertrophy in relation to the myonucleus/sarcoplasm ratio and the occurrence of different subclasses of myogenic cells[J]. *Journal of Fish Biology*, 1993, 43(1): 69-80.
- [13] Alami-Durante H, Fauconneau B, Rouel M, *et al.* Growth and multiplication of white skeletal muscle fibres in carp larvae in relation to somatic growth rate[J]. *Journal of Fish Biology*, 1997, 50(6): 1285-1302.
- [14] Yan X C, Ding L, Li Y C, *et al.* Identification and profiling of microRNAs from skeletal muscle of the common carp[J]. *PLoS One*, 2012, 7(1): e30925.
- [15] 雷莹, 苏胜彦, 丁敬波, 等. 背部肌肉注射含 *IGF2b* 基因慢病毒载体对鲤鱼生长的影响[J]. *西北农林科技大学学报(自然科学版)*, 2014, 42(7): 8-14, 28.
- Lei Y, Su S Y, Ding J B, *et al.* Effect of injecting vector containing *IGF2b* to dorsal muscle tissue on growth of common carp[J]. *Journal of Northwest A & F University (Natural Science Edition)*, 2014, 42(7): 8-14, 28(in Chinese).
- [16] 赵文, 高峰英, 石振广. 达氏鳃肌肉组织转录组测序和功能分析[J]. *水产学报*, 2014, 38(9): 1255-1262.
- Zhao W, Gao F Y, Shi Z G. The transcriptome sequencing and functional analysis of muscle tissue of Kaluga (*Huso dauricus*)[J]. *Journal of Fisheries of China*, 2014, 38(9): 1255-1262(in Chinese).

- [17] 罗辉, 叶华, 肖世俊 等. 转录组学技术在水产动物研究中的运用[J]. 水产学报, 2015, 39(4): 598-607.
Luo H, Ye H, Xiao S J, *et al.* Application of transcriptomics technology to aquatic animal research[J]. Journal of Fisheries of China, 2015, 39(4): 598-607(in Chinese).
- [18] Dong Z J, Nguyen N H, Zhu W B. Genetic evaluation of a selective breeding program for common carp *Cyprinus carpio* conducted from 2004 to 2014[J]. *BMC Genetics*, 2015, 16(1): 94.
- [19] Song F B, Wang L M, Zhu W B, *et al.* A novel *igf3* gene in common carp (*Cyprinus carpio*): evidence for its role in regulating gonadal development[J]. *PLoS One*, 2016, 11(12): e0168874.
- [20] Song F B, Wang L M, Zhu W B, *et al.* Long noncoding RNA and mRNA expression profiles following *igf3* knockdown in common carp, *Cyprinus carpio*[J]. *Scientific Data*, 2019, 6(1): 190024.
- [21] Zhu W B, Wang L M, Dong Z J, *et al.* Comparative transcriptome analysis identifies candidate genes related to skin color differentiation in red tilapia[J]. *Scientific Reports*, 2016, 6(1): 31347.
- [22] Untergasser A, Cutcutache I, Koressaar T, *et al.* Primer3-new capabilities and interfaces[J]. *Nucleic Acids Research*, 2012, 40(15): e115.
- [23] Zheng Q, Zhang Y, Chen Y, *et al.* Systematic identification of genes involved in divergent skeletal muscle growth rates of broiler and layer chickens[J]. *BMC Genomics*, 2009, 10(1): 87.
- [24] Buckingham M, Bajard L, Chang T, *et al.* The formation of skeletal muscle: from somite to limb[J]. *Journal of Anatomy*, 2003, 202(1): 59-68.
- [25] Moss F P, Leblond C P. Satellite cells as the source of nuclei in muscles of growing rats[J]. *The Anatomical Record: Advances in Integrative Anatomy & Evolutionary Biology*, 1971, 170(4): 421-435.
- [26] Schiaffino S, Dyar K A, Ciciliot S, *et al.* Mechanisms regulating skeletal muscle growth and atrophy[J]. *The FEBS Journal*, 2013, 280(17): 4294-4314.
- [27] Gurnett C A, Desruisseau D M, McCall K, *et al.* Myosin binding protein C1: a novel gene for autosomal dominant distal arthrogyrosis type 1[J]. *Human Molecular Genetics*, 2010, 19(7): 1165-1173.
- [28] Khurana T, Khurana B, Noegel A A. LIM proteins: association with the actin cytoskeleton[J]. *Protoplasma*, 2002, 219(1-2): 1-12.
- [29] Chu P H, Ruiz-Lozano P, Zhou Q, *et al.* Expression patterns of FHL/SLIM family members suggest important functional roles in skeletal muscle and cardiovascular system[J]. *Mechanisms of Development*, 2000, 95(1-2): 259-265.
- [30] Barzegar M, Mozafari N, Kariminejad A, *et al.* A new homozygous nonsense mutation in *LAMA 3A* underlying laryngo-onycho-cutaneous syndrome[J]. *British Journal of Dermatology*, 2013, 169(6): 1353-1356.
- [31] Hamill K J, Paller A S, Jones J C R. Adhesion and migration, the diverse functions of the laminin $\alpha 3$ subunit[J]. *Dermatologic Clinics*, 2010, 28(1): 79-87.
- [32] Huang B L, Wu P F, Bowker-Kinley M M, *et al.* Regulation of pyruvate dehydrogenase kinase expression by peroxisome proliferator-activated receptor- α ligands, glucocorticoids, and insulin[J]. *Diabetes*, 2002, 51(2): 276-283.
- [33] Xu J X, Han J Y, Epstein P N, *et al.* Regulation of PDK mRNA by high fatty acid and glucose in pancreatic islets[J]. *Biochemical and Biophysical Research Communications*, 2006, 344(3): 827-833.
- [34] Muthuchamy M, Grupp I L, Grupp G, *et al.* Molecular and physiological effects of overexpressing striated muscle β -tropomyosin in the adult murine heart[J]. *Journal of Biological Chemistry*, 1995, 270(51): 30593-30603.

***De novo* transcriptome analysis and comparison of the FFRC No.2 strain common carp (*Cyprinus carpio*) associated with its muscle growth**

WANG Lanmei^{1,2}, ZHU Wenbin¹, FU Jianjun¹, DONG Zaijie^{1,2*}

(1. Freshwater Fisheries Research Centre of Chinese Academy of Fishery Sciences, Key Laboratory of Freshwater Fisheries and Germplasm Resources Utilization, Ministry of Agriculture and Rural Affairs, Wuxi 214081, China;

2. Wuxi Fisheries College, Nanjing Agricultural University, Wuxi 214128, China)

Abstract: In order to explore the molecular basis of the higher growth rate in FFRC No.2 strain common carp (*Cyprinus carpio*), we utilized comparative transcriptomics analysis using next-generation sequencing and examined transcript profiles of the muscles from high-growth (HG) (average final body weight FBW= 522.3 g, SE=11.03) and slow-growth (SG) (average FBW= 314.9 g, SE=7.16) individuals. Two poly (A)⁺ RNA-seq libraries were constructed and sequenced using paired-end (PE 2×125) Illumina technique. All sequence data were submitted to NCBI SRA database (SRP078896). A total of 286 118 576 reads were generated, resulting in 392 238 assembled contigs, corresponding to 348 144 unique genes and their length ranged from 200 to 32 499 bp. We firstly performed a functional annotation on the assembly using first aligning by BLASTx with cut-off e-value < 1×10⁻¹⁰, then all assembled contigs were blasted against the Nr database, UniProt-SwissProt database, KEGG database and COG database, respectively. The contigs hit 56.01%-77.71% of the unique proteins of *Danio rerio*, *Fugu ocellatus*, *Oryzias latipes*, *Gasterosteus aculeatus*, *C. carpio* and *Oreochromis niloticus*. And 93 462 contigs containing 129 047 simple sequence repeats (SSRs) were identified. For the assembled contigs which had hits against with Nr database, 112 001 annotated contigs were found to contain the Open Reading Frame (ORF). The ORF length ranged from 27 to 32 490 bp with an average of 554.68 bp.

To reveal the underline mechanism of the muscle growth rate in FFRC No. 2 strain *C. carpio*, we performed a differential expression analysis between the two muscle transcriptomes. Based on the criteria of $|\log_2FC| \geq 1$ and $FDR \leq 0.05$, we identified 749 differentially expressed genes in HG fishes compared with SG fishes, of which 348 genes were up-regulated and 401 genes were down-regulated genes in HG fishes. Several key genes related to muscle growth, i.e. *mb*, *myl2b*, *tnni1*, *fhl1*, *lamb3*, *pdk2* and *igfbp7* were identified. All these DEGs were subjected to GO term enrichment and KEGG pathway analysis to obtain the potential function annotation. These DEGs were enriched in different gene ontologies and pathways. The KEGG pathway analysis results showed that the cardiac muscle contraction pathway and hypertrophic cardiomyopathy (HCM) pathway were found significantly enriched of DEGs, in which *tnni1*, *tnni2*, *mybpc1*, *actn2* and *tpm2* serve as key regulatory proteins. These findings will facilitate gene function study and help to understand the molecular mechanisms of muscle growth in *C. carpio* by breeding and improve the molecular selection method of the specific strain with fast growth.

Key words: *Cyprinus carpio*; transcriptome; muscle growth; differentially expressed genes

Corresponding author: DONG Zaijie. E-mail: dongzj@ffrc.cn

Funding projects: Jiangsu Provincial Postdoctoral Research Program in 2018 (2018K208C); Central Public-interest Scientific Institution Basal Research Fund, Chinese Academy of Fishery Sciences (2020TD37); China Agriculture Research System (CARS-45-05)