

达氏鳇肌肉组织转录组测序和功能分析

赵文^{1*}, 高峰英¹, 石振广²

(1. 大连海洋大学辽宁省水生生物学重点实验室, 辽宁 大连 116023;

2. 云南阿穆尔鲟鱼集团有限公司, 云南 昆明 050228)

摘要:为发掘达氏鳇的功能基因,采用新一代高通量测序技术对达氏鳇肌肉组织转录组进行测序。结果获得原始数据 14 447 211 200 bp,拼接获得了 55 531 条单基因序列(unigene),长度范围 300~32 613 bp,平均长度 941 bp。利用生物信息学方法对 unigene 进行了非冗余蛋白质数据库(Nr)相似搜索,此外还进行了 GO 的功能注释和 KEGG 代谢通路分析。结果一共有 20 735 条 unigene(37.34%)与 Nr 数据库中的已知基因同源;根据 GO 功能可分为生物过程、细胞组分和分子功能 3 大类 56 分支;依据 KEGG 代谢通路分析可以分成 290 类。

关键词:达氏鳇;肌肉组织;基因;转录组

中图分类号:Q 785; S 965.1

文献标志码:A

达氏鳇(*Huso dauricus*)隶属硬骨鱼纲(Osteichthyes),鲟形目(Acipenseriformes),鲟科(Acipenseridae),鳇属^[1],与欧洲鳇(*Beluga sturgeon*)是世界上仅存的两种鳇鱼,其自然种群主要分布在黑龙江中下游^[2],用其卵所生产的鱼子酱有“黑黄金”的美誉,是一种营养价值、经济价值和科研价值都很高的大型淡水鱼类^[3]。由于环境污染和利益驱动,渔民对鳇鱼捕捞强度的逐年加大等原因,自然水域达氏鳇种群已经成为濒危物种,种群的数量越来越少,捕捞个体规格越来越小^[4],严重威胁其产业的健康持续发展,因此,应用分子生物学手段,加快开展达氏鳇生长、免疫、抗逆和性别等重要经济性状相关的功能基因的筛选研究,进而克隆全长并研究基因水平的表达和调控,对于揭示达氏鳇的生长、抗病和抗逆能力,培育优势养殖品种和保护这一濒危物种资源至关重要。

转录组技术利用高通量测序对特定组织或细胞中所有 RNA 反转录而成的 cDNA 文库进行测序,并可在没有研究物种基因信息的情况下,直接对任何物种的转录组进行分析^[5],是目前进行基因发现、查找控制特异性状的潜在主效基因和研究基

因表达调控的有效方法,而且还可用于单核苷酸多态性(single nucleotide polymorphism, SNP)鉴定^[6]、简单重复序列(simple sequence repeat, SSR)鉴定^[7]、差异表达基因鉴定^[8]、可变拼接位点的识别^[9]等。应用转录组技术对鱼类的研究工作也已经陆续展开,包括湖鲟(*Acipenser fulvescens*)^[10]、虹鳟(*Oncorhynchus mykiss*)^[11]、大西洋鳕(*Gadus morhua*)^[12]及鲤(*Cyprinus carpio*)^[13]等,并已经成功应用于基因的大规模筛选和鉴定。

目前,达氏鳇基因组研究较少,本研究利用转录组技术构建达氏鳇养殖群体肌肉组织转录组文库,旨在为进一步揭示达氏鳇相关功能基因特点,永久保存其基因资源奠定基础,对解决达氏鳇基因进化、遗传育种及其资源恢复等诸多方面问题提供重要信息。

1 材料与方法

1.1 实验材料

实验用达氏鳇取自河北省涉县泉溪鲟鱼生态园有限公司,2013 年 5 月在养殖场取 20 尾刚刚性成熟、体质量 100 kg 左右、生长状态良好的健

收稿日期:2014-03-07 修回日期:2014-06-25

资助项目:横向联合项目“达氏鳇基础生物学研究”(2012-2014)

通信作者:赵文, E-mail: zhaowen@dlou.edu.cn

康雌性达氏鳊,取背鳍下方肌肉组织,取样规格 100 mg,液氮保存样品。这批达氏鳊苗种于 2004 年取自黑龙江的野生亲鱼。

1.2 实验方法

总 RNA 的提取与处理 取 100 mg 达氏鳊肌肉组织,不间断加入液氮迅速研磨成粉末状,加入 1 mL RNAiso Plus 裂解细胞,经氯仿抽提、异丙醇沉淀、75% 乙醇洗涤后,用适量 RNase-free 水溶解 RNA。使用 DNase I 经 37 °C 水浴 30 min 消化总 RNA 中残留的 DNA,琼脂糖凝胶电泳检测其完整性,紫外分光光度计检测其纯度及浓度。

mRNA 的富集与片段化 RNA 提取后,使用 Promega 公司的 PolyATtract mRNA 分离系统,按照试剂盒说明书操作,应用磁珠吸附原理,利用 Oligo(dT) 磁珠和磁分离器分离出 mRNA,用 Tris-HCl 缓冲液洗脱得到纯化的 mRNA,将 mRNA 与打断试剂 RNA Fragmentation Reagents 均匀混合,70 °C 水浴 15 min,立即在冰上用乙醇回收打断产物,并溶于 RNase-free 水中。

转录组文库的制备 采用 TaKaRa 公司的 PrimeScript™ Double Strand cDNA Synthesis Kit 试剂盒,按照说明书体系操作:使用 PrimeScript RTase、随机引物,以片段化的 mRNA 为模版经 42 °C 水浴 1 h 合成一链 cDNA,总体系 20 µL;使用 RNaseH 和 DNA Polymerase I 经 16 °C 反应 2 h,70 °C 加热 10 min 合成双链 cDNA,总体系 146 µL;加 4 µL T₄ DNA Polymerase 经 37 °C 水浴反应 10 min 补平末端,总体系 150 µL,并经苯酚/氯仿/异戊醇进行纯化回收。使用 TaKaRa 的 DNA A-Tailing Kit 经 72 °C 反应 20 min 进行 cDNA 末端加 A 尾,总体系 50 µL,并按照 TaKaRa MiniBEST DNA Fragment Purification Kit Ver. 4.0 试剂盒操作进行纯化回收;利用 T₄ DNA Ligase 将 Illumina PE adapter 接头连接到加 A 尾的 cDNA 片段的 3'末端;连接产物经琼脂糖凝胶电泳回收(200 ± 25) bp 的 cDNA 片段,胶回收按照 TaKaRa MiniBEST Agarose Gel DNA Extraction Kit Ver. 4.0 试剂盒操作说明;利用 Illumina 引物 PCR Primer PE 1.0 和 PCR Primer PE 2.0 进行 15 个循环的 PCR 扩增,总体系 20 µL,扩增条件为 98 °C 30 s;98 °C 10 s、60 °C 30 s、72 °C 30 s,15 个循环,72 °C 延伸 5 min;最后

用试剂盒进行 PCR 产物的纯化。

测序 转录组文库的测序,由北京基因组研究所提供测序服务,测序平台为 Illumina/Hiseq2000,样品为刚刚性成熟的生长状态良好的雌性达氏鳊肌肉组织混合转录组文库,测序模式为 2 × 100 bp。

数据处理与分析 对测序所得原始数据进行统计及质量评估,去除 Reads 中含有的测序接头序列,使用 Trinity 软件对序列进行拼接,将拼接得到的基因参考序列进行进一步比对分析:(1) 将拼接得到的基因参考序列(所有的 unigene)与 NCBI 的 Nr 蛋白数据库比对获取基因注释信息(Blastx, E-value ≤ 10⁻⁵)。(2) 把拼接得到的基因参照序列与 Rfam 数据库进行比对,应用 RSEM 软件 v1.2.6 进行基因表达定量分析,基因表达量用 FPKM 值(fragments per kilobase of transcript per million fragments mapped)表示。(3) 根据参照基因序列与 NCBI 的 Nr 数据库 Blast 比对结果,用 Blast2Go 提取 GO 注释信息。(4) 把拼接得到的参照基因序列与 KEGG 数据库中的蛋白序列进行 Blastx 比对(E-value ≤ 10⁻⁵)从中提取 KEGG 注释信息。

2 结果

2.1 总 RNA 的提取

提取的总 RNA 经琼脂糖凝胶电泳检测,显示 28S、18S、5S 条带清晰,且 28S:18S 亮度为 2:1,表明总 RNA 的完整性良好。超微量分光光度计测得 A260/A280 为 1.9 ~ 2.0,表明提取到的 RNA 纯度很高,可以进行后续试验。

2.2 测序数据统计

达氏鳊肌肉组织转录组测序得原始数据 14 447 211 200 bp,对得到的测序数据,分别计算 Reads 上每个位置碱基测序质量的平均质量,得到碱基质量分布(图 1,图 2);分别计算每条 Reads 的平均质量值,并统计相应质量值的 Reads 数,通过累积分布曲线反映数据的平均质量分布,得到质量分布(图 3,图 4)。测序质量统计结果显示,测序质量值 20 以上的碱基 Reads1 占 96.60%,Reads2 占 92.97%;测序质量值 30 以上的碱基 Reads1 占 90.73%,Reads2 占 85.92%,从数据的各项指标可以看出测序质量较高,达到了后续数据组装分析的要求。

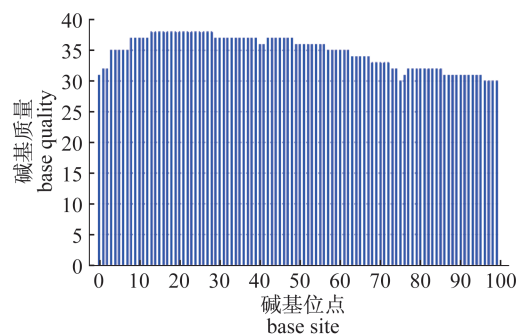


图 1 Reads1 的碱基质量分布

Fig.1 Base quality distribution of Reads1

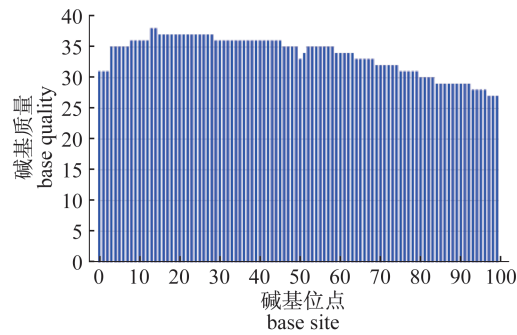


图 2 Reads2 的碱基质量分布

Fig.2 Base quality distribution of Reads2

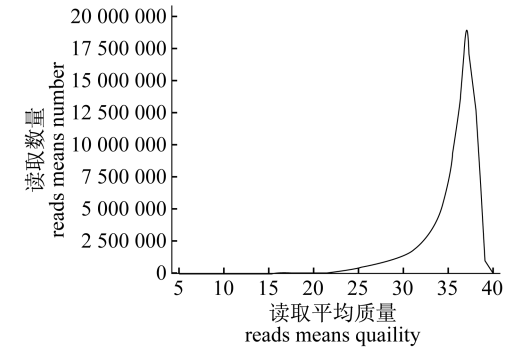


图 3 Reads1 的质量分布

Fig.3 Quality distribution of Reads1

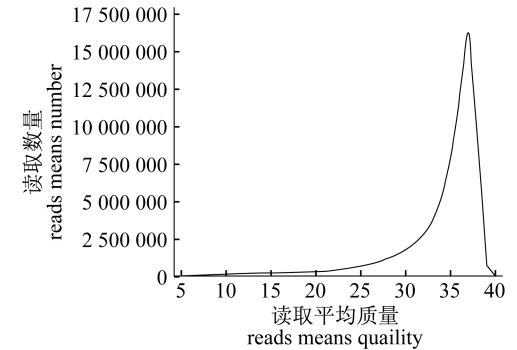


图 4 Reads2 的质量分布

Fig.4 Quality distribution of Reads2

2.3 序列拼接与比对

去除 Reads 中含有的测序接头序列,将各样品的数据合并,用 Trinity 软件按默认参数进行拼接。在拼接结果序列中,选最长的可变剪切序列作为该基因的代表转录本 (unigene)。结果拼接获得了 55 531 条单基因序列 (unigene),长度范围从 300 ~ 32 613 bp,平均长度 941 bp(图 5)。应用 bowtie1.0.0 软件将测序 Reads 序列与拼接出的基因参照序列 (unigene) 进行比对,总体 Mapping Rate 为 85.61%,可以进行后续分析。

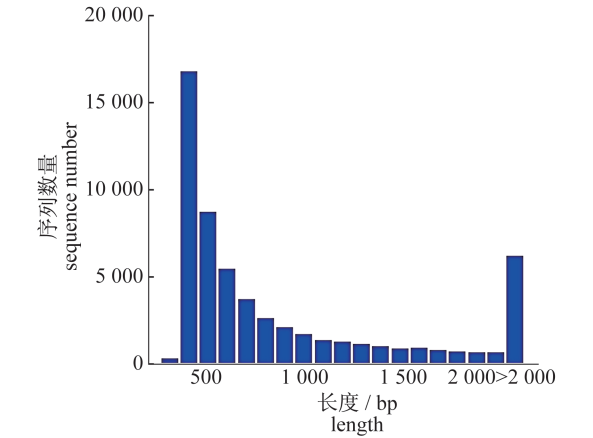


图 5 达氏鳊 unigene 的长度分布

Fig.5 Length distribution of patchwork sequence in *H. dauricus*

将拼接得到的所有的 unigene 与 NCBI 的 Nr 蛋白数据库比对,结果一共有 20 735 个 unigene (37.34%) 与数据库中已知基因同源,而其余 34 796 个 unigene (62.66%) 与数据库中的已知基因同源性较低,可能属于新基因。本研究重点关注与达氏鳊重要经济性状相关的功能基因,对比结果显示,有 175 条与生长功能相关的同源序列和 54 条与免疫相关的同源序列,此外发现 HSP(热休克蛋白)基因家族成员 5 种,包括 *HSP40*、*HSP60*、*HSP70*、*HSP90* 及 *HSPB1*,共涉及基因数量 10 条, *SOX* 基因家族成员 11 个,包括 *SOX30*、*SOX8*、*SOX7*、*SOX9*、*SOX5*、*SOX6*、*SOX12*、*SOX13*、*SOX18A*、*SOX4b*,其中关注度最高的 *SOX9* 基因 2 个。

2.5 RNA 分类及基因表达定量分析

把拼接得到的基因参照序列与 Rfam 数据库进行比对,将 RNA 分类为 mRNA、tRNA、rRNA、ncRNA 等。结果表明 mRNA 含量最多,占 74.88%; tRNA 占 22.19%; rRNA 占 2.44%; miRNA 占 0.04%; snoRNA 占 0.02%。应用

RSEM 软件 v1.2.6 进行基因表达定量分析,基因表达量用 FPKM 值表示,将基因表达量按照极低表达基因(0~0.5)、低表达基因(0.5~5)、中等表达基因(5~50)、高表达基因(50~500)、极高表达基因(>500)分为 5 个等级。结果显示,141 个基因未检测到表达量,转录本表达丰度集中在极低表达(50.8%),其次是低表达(37.42%),10.14% 为中等表达,高表达和极高表达占 1.38%。分析测序饱和度和可以看出,随测序量的增加检测到的表达基因数量的增加趋向平缓,所以样品的测序量接近饱和,完全可以用于后续基因分析(图 6)。

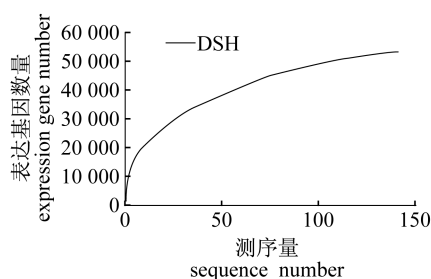


图 6 达氏鳇测序饱和度曲线

Fig. 6 Saturation curve of sequencing in *H. dauricus*

2.6 GO 功能注释及分类

根据参照基因序列与 NCBI 的 Nr 数据库 Blast 比对结果,用 Blast2Go 提取 GO 注释信息。结果表明,一共有 10 017 条 unigene(18.04%)与数据库中的基因具有相似性。GO 功能大致可分为生物过程(biological process)、细胞组分(cellular component)和分子功能(molecular function)3 大类 56 亚类(图 7),例如生化过程、代谢过程、生长、发育和免疫防御等过程。数量统计结果显示,达氏鳇肌肉组织转录组中细胞杀伤、拟核、病毒粒子、通道调节活性、化学引诱物活性、化学排斥物活性、金属伴侣活性、形态发生素活性、蛋白标记、受体调节器活性和翻译调节活性 11 个功能亚类的 unigene 较少,其余功能亚类的 unigene 分布比较均衡。其中,分子功能中具有代表性的功能类别为与整合、催化活性、酶活性调节、分子转导活性和转运活性等相关;生物学过程中占优势的为生物调节、细胞过程、代谢过程和单生物体过程等;细胞组分的代表类型为细胞部分、细胞膜和细胞器等。

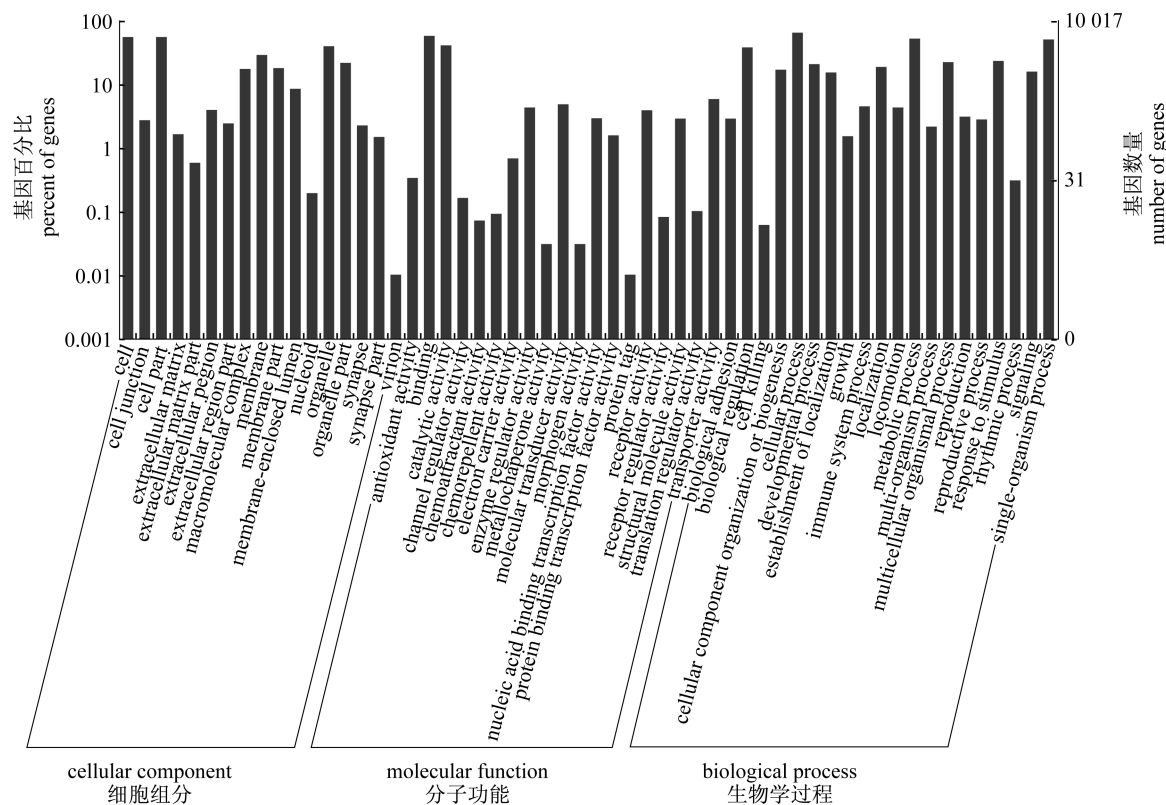


图 7 达氏鳇基因的 GO 功能分类

Fig. 7 Gene ontology classification of putative functions of the *H. dauricus*

2.7 KO 代谢途径注释及富集分析 (KEGG)

KEGG 代谢通路分析有助于进一步了解基因的生物学功能以及基因产物的相互作用。把拼接得到的参照基因序列与 KEGG 数据库中的蛋白序列进行 Blastx 比对 ($E\text{-value} \leq 10^{-5}$) 从中提取 KEGG 注释信息。分析结果,一共有 15 312 条 unigene (27.57%) 被注释,并被归类到 290 个 KEGG 通路。基因数量排名靠前的代谢途径包括 MAPK 信号通路、肌动蛋白细胞骨架调节、胞吞作用、嘌呤代谢和钙信号途径等。

3 讨论

3.1 转录组文库测序分析

转录组技术是利用高通量测序法对组织或细胞中所有 RNA 反转录而成的 cDNA 文库进行测序。由于和传统的基因芯片技术相比,其可在没有研究物种基因信息的情况下,直接对任何物种的转录组进行分析,而且提供更高的检测通量^[14]、更精准的数字信号、更广泛的检测范围、花费更低成本和时间^[15],使得转录组测序成为一种极其有效的新基因发现途径^[16]。近年来,已广泛应用于各种植物和动物的转录组研究中,对于鱼类转录组的研究工作也已经陆续展开,包括湖鲟鱼、虹鳟、欧洲鳗鲡 (*Anguilla anguilla*)^[17]、大西洋鳕、鲤及大菱鲆 (*Scophthalmus maximus*)^[18] 等,但未见关于达氏鳢转录组的报道。本研究通过对 20 尾刚刚性成熟的健康雌性达氏鳢肌肉组织进行转录组文库构建,并采用 Illumina/Hiseq2000 高通量测序平台,在没有参照基因组的情况下对其转录组进行测序,获得原始数据 14 447 211 200 bp,测序质量值 20 以上的碱基 Reads1 占 96.60%,Reads2 占 92.97%;测序质量值 30 以上的碱基 Reads1 占 90.73%,Reads2 占 85.92%,从测序数据的各项指标可以看出测序质量较高,达到了后续数据组装分析的要求。经过对测序数据的组装和聚类拼接,获得了 55 531 条非冗余 unigene 序列,其中长度大于 1 000 bp 的有 15 000 条,大于 2 000 bp 有 6 000 条,平均长度 941 bp,而且测序 Reads 序列与拼接出的基因参照序列 (unigene) 进行比对,总体 Mapping Rate 为 85.61%,这说明本研究的数据组装质量较高,可以进行后续生物信息学分析,这些数据极大地丰富了达氏鳢肌肉组织基因序列数据资源,为达氏鳢乃至鳢鱼类后续功能基因和基因组学的研究

奠定了基础。

3.2 转录组文库功能分析

将拼接的所有 unigene 序列与 NCBI 的 Nr 蛋白数据库比对获取基因注释信息,结果一共有 20 735 条 unigene 序列与数据库中已知基因同源,占 37.34%,而其余 62.66% (34 796 条) unigene 序列与数据库中的已知基因相似性较低,未被注释,同样的问题也出现在其他生物转录组中,在赵晓霞^[19]构建的马氏珠母贝 (*Pinctada martensii*) 珍珠囊转录组文库中 102 762 条 unigene 序列,只有 36 989 条 unigene 序列 (约 35.99%) 被注释;Eli 等^[20]构建的珊瑚幼虫转录组文库的 62 657 条 unigene 序列只有 5 330 (约 8.5%) 条注释,其余的 unigene 序列比对相似性较低,未被注释,造成这一问题的原因是目前国际公共基因数据库中收录的相近物种基因注释信息较少。由于鳢鱼类分子基础研究薄弱,遗传背景了解不足,因此本研究得到的达氏鳢转录组基因注释信息较少,其中未被注释的基因序列中,那些比较短的 unigene 序列,有可能是 3' 或者 5' 的非翻译区,非编码 RNA,或者是不含有功能结构域的序列,而那些超过 500 bp 长度的 unigene 序列,有很大的可能是未被记录的新基因,对这些未知基因的研究有可能揭示鱼类某些重要的生物学过程,也将是今后需要重点关注的基因。

本实验还对达氏鳢肌肉组织转录组文库组装的 55 531 条 unigene 序列进行了 GO 功能分类和 KEGG 代谢通路等分析,结果一共有 10 017 条 unigene (约 18.04%) 与 GO 数据库中的基因具有相似性,包括生物学过程、细胞组分和分子功能 3 大类 56 亚类,主要的 GO 类型包括催化活性、酶活性调节、生物调节、细胞过程等;代谢通路分析得到 15 312 条 unigene 序列 (约 27.57%) 注释,并被归类到 290 个 KEGG 通路,包括转录和翻译调控、信号转导、物质代谢和次生代谢产物合成等,尽管本实验得到的注释信息相对较少,但是,通过这些注释,获得了大量的达氏鳢肌肉组织转录组信息,初步阐明了达氏鳢肌肉组织基因的功能、参与的生物过程、所处的代谢途径或信号通路等,这对于今后深入了解基因的功能很重要,为发掘达氏鳢相关功能基因、研究相关生理功能提供了基础数据。

3.3 部分候选功能基因的分析

达氏鳢是一种营养价值、经济价值和科研价

值都很高的大型淡水鱼类,近年来达氏鳇自然种群资源遭到严重破坏,已经濒临灭绝,严重威胁其产业健康可持续发展,因此加快开展达氏鳇生长、免疫、抗逆和性别等重要经济性状相关的功能基因的筛选尤为重要。本实验应用 BLAST 相似性搜索注释,在构建的达氏鳇肌肉组织转录组中筛选出 175 条与生长功能相关的同源序列和 54 条与免疫相关的同源序列,这些基因可用于后续生长和免疫基因在达氏鳇不同发育时期的差异表达分析,对生长相关基因的研究为揭示达氏鳇的生长机理提供理论依据;对免疫基因的研究为提高达氏鳇防病、抗病能力提供重要的理论依据,这些基因序列将来可以制成基因表达谱芯片,用来检测达氏鳇的生长和免疫水平,作为达氏鳇优良品种选育的指标。

此外,本实验还筛选出 HSP 基因家族成员 5 种,包括 HSP40、HSP60、HSP70、HSP90 及 HSPB1,共涉及基因数量 10 条,其中 HSP70 是国内外学者的研究热点,它的主要作用是分子伴侣作用、协同免疫作用、抗细胞凋亡作用和抗氧化作用^[21]。目前关于鲟鳇鱼的 HSP70 的研究较少,只是克隆了杂交鲟、匙吻鲟 (*Polyodon spathula*)^[22] 和西伯利亚鲟 (*Acipenser baerii*) HSP70^[23] 基因全长,但关于 HSP70 在抗逆环境胁迫中的作用一直未有深入了解。由于达氏鳇栖息在水中,环境比较复杂,更容易受到环境因子的影响,因此,关于达氏鳇转录组中热休克蛋白的研究,将有利于增强对达氏鳇在不良环境下抵抗力水平的认识。

鱼类由于进化上的原始性,性别决定机制比较复杂,具有原始性、多样性和易变性^[24]。近年来研究比较多的是与性别决定相关的基因,如 SOX 基因家族、DMRT1 基因^[25]等。本研究中发现 SOX 基因家族成员 11 个,包括 SOX30、SOX8、SOX7、SOX9、SOX5、SOX6、SOX12、SOX13、SOX18A、SOX4b,其中关注度最高的可能与性别决定相关的 SOX9 基因 2 个。在虹鳟中发现存在两种 SOX9 基因并且只在精巢中表达^[26],在斑马鱼 (*Danio rerio*) 中存在两种 SOX9 基因,其中 SOX9a 在脑、肾、肌肉、精巢和胸鳍中呈现泛表达模式,而 SOX9b 仅仅在卵巢中被检测到^[27];SOX9 基因在 1~3 龄史氏鲟的大脑、心脏、肝脏、眼睛、胰脏、肾脏、精巢和卵巢等 8 种组织中均有表

达,只是表达量随发育阶段和组织不同而稍有差异^[28]。SOX9 在西伯利亚鲟中的表达具明显的组织差异性和时间差异性^[29]。基于上述研究,达氏鳇肌肉组织中发现的 SOX9 基因与其性别决定和分化是否相关还有待证实,目前只作为一个重点关注的候选基因。

Blast 相似性搜索注释是一种推测查询基因序列生物学功能的快速、有效的方法,但其也只是为确定其编码蛋白质的功能提供暂定的线索,要彻底认识和了解一个基因的功能,还需要做基因表达定位的研究和基因功能验证。另外,本研究对达氏鳇肌肉组织转录组中筛选出的这些基因的功能不能明确定位,还需进行更深入的研究。

参考文献:

- [1] Wang J Q, Jiang Z Q, Hu H X. Biological characteristics of the main breeding sturgeon [J]. Fishery Sciences, 1998, 17(6): 34-38. [王吉桥,姜志强,胡红霞. 主要养殖鲟鱼的生物学特性. 水产科学, 1998, 17(6): 34-38.]
- [2] Shi Z G, Dong S L, Lu H S, et al. A preliminary study on juvenile growth performance of kaluga (*Huso dauricus*) hybrid under artificial farming conditions [J]. Periodical of Ocean University of China, 2008, 38(1): 33-38. [石振广,董双林,鲁宏申,等. 人工养殖条件下达氏鳇杂交种生长特性的初步研究. 中国海洋大学学报, 2008, 38(1): 33-38.]
- [3] Niu C J, Hu H X, Luo J, et al. Genetic diversity in the reserved parent fish broodstocks of *Asipenser schrenckii* and *Huso dauricus* [J]. Journal of Fisheries of China, 2010, 34(12): 1795-1798. [牛翠娟,胡红霞,罗静,等. 史氏鲟和达氏鳇养殖亲鱼群体遗传多样性分析. 水产学报, 2010, 34(12): 1795-1798.]
- [4] Shi Z G. A study on the key technique of artificial reproduction and farming of kaluga *Huso dauricus* [D]. Qingdao: Ocean University of China, 2008. [石振广. 达氏鳇人工繁殖和养殖关键技术研究. 青岛: 中国海洋大学, 2008.]
- [5] Liu H L, Zhen L M, Liu Q Q, et al. Studies on the transcriptomes of non-model organisms [J]. Hereditas, 2013, 35(8): 955-970. [刘红亮,郑丽明,刘青青,等. 非模式生物转录组研究. 遗传, 2013, 35(8): 955-970.]
- [6] Gao Z X, Luo W, Liu H, et al. Transcriptome

- analysis and SSR/SNP markers information of the blunt snout bream (*Megalobrama amblycephala*) [J]. Public Library of Science One, 2012, 7 (8):e42637.
- [7] Garg R, Patel R K, Tyagi A K, Jain M. De novo assembly of chickpea transcriptome using short reads for gene discovery and marker identification [J]. DNA Research, 2011, 18(1):53–63.
- [8] Jeukens J, Renaut S, St-Cyr J, et al. The transcriptomics of sympatric dwarf and normal lake whitefish (*Coregonus clupeaformis* spp., Salmonidae) divergence as revealed by next-generation sequencing [J]. Molecular Ecology, 2010, 19(24):5389–5403.
- [9] Huh J W, Kim Y H, Park S J, et al. Large-scale transcriptome sequencing and gene analyses in the crab-eating macaque (*Macaca fascicularis*) for biomedical research [J]. Biomed Central Genomics, 2012, 13(1):163.
- [10] Hale M C, McCormick C R, Jackson J R, et al. Next-generation pyrosequencing of gonad transcriptomes in the polyploid lake sturgeon *Cipenser fulvescens*: The relative merits of normalization and rarefaction in gene discovery [J]. Biomed Central Genomics, 2009, 10(1):203.
- [11] Salem M, Rexroad C E, Wang J N, et al. Characterization of the rainbow trout transcriptome using Sanger and 454-pyrosequencing approaches [J]. Biomed Central Genomics, 2010, 11(1):564.
- [12] Johansen S D, Karlsen B O, Furmanek T, et al. RNA deep sequencing of the Atlantic cod transcriptome [J]. Comparative Biochemistry and Physiology-Part D: Genomics Proteomics, 2011, 6(1):18–22.
- [13] Ji P F, Liu G M, Xu J, et al. Characterization of common carp transcriptome: Sequencing, de novo assembly, annotation and comparative genomics [J]. Public Library of Science One, 2012, 7(4):e35152.
- [14] Yan S P, Yang R H, Leng S J, et al. High-fluxed DNA sequencing technology and its application in agricultural science research [J]. Chinese Agricultural Science Bulletin, 2012, 28(30):171–176. [闰绍鹏, 杨瑞华, 冷淑娇, 等. 高通量测序技术及其在农业科学研究中的应用. 中国农学通报, 2012, 28(30):171–176.]
- [15] Zeng D G, Chen X L, Xie D X, et al. Deep sequencing-based transcriptome analysis of *Litopenaeus vannamei* [J]. Genomics and Applied Biology, 2013, 32(3):308–313. [曾地刚, 陈秀荔, 谢达祥, 等. 基于高通量测序的凡纳滨对虾的转录组分析. 基因组学与应用生物学, 2013, 32(3):308–313.]
- [16] Kaur S, Cogan N, Pembleton L, et al. Transcriptome sequencing of lentil based on second-generation technology permits large-scale unigene assembly and SSR marker discovery [J]. Biomed Central Genomics, 2011, 12(1):265.
- [17] Coppe A, Pujolar J M, Maes G E, et al. Sequencing, de novo annotation and analysis of the first *Anguilla anguilla* transcriptome: EelBase opens new perspectives for the study of the critically endangered European eel [J]. Biomed Central Genomics, 2010, 11(1):635.
- [18] Pereiro P, Balseiro P, Romero A, Dios S, et al. High-throughput sequence analysis of turbot (*Scophthalmus maximus*) transcriptome using 454-pyrosequencing for the discovery of antiviral immune genes [J]. Public Library of Science One, 2012, 7(5):e35369.
- [19] Zhao X X. Studies on Transcriptome and digital gene expression of pearl sac in *Pinctada martensii* by RNA sequencing [D]. Zhanjiang: Guangdong Ocean University, 2011. [赵晓霞. 基于 RNA 测序技术的马氏珠母贝珍珠囊转录组及数字基因表达谱分析. 湛江: 广东海洋大学, 2011.]
- [20] Eli M, Galina V, Shi W, et al. Sequencing and denovo analysis of a coral larval transcriptome using 454 GSFlx [J]. Biomed Central Genomics, 2009, 10(19):1–18.
- [21] Neal F, Gordon N F, Clark B. Heat shock proteins and immune response, The challenges of bringing autologous HSP-based vaccines to commercial reality [J]. Methods, 2004, 2(1):63–69.
- [22] Gao Y, Yuan G L, Li D P, et al. Cloning and sequence analysis of HSP70 cDNA from hybrid sturgeon and paddle fish (*Polyodon spathula*) [J]. Journal of Huazhong Agricultural University, 2010, 29(1):85–89. [高宇, 袁改玲, 李大鹏, 等. 杂交鲟和匙吻鲟 HSP70 cDNA 克隆与序列分析. 华中农业大学学报, 2010, 29(1):85–89.]
- [23] Tian Z H, Xu S G, Wang W. Molecular cloning, characterization and tissue profiles of a HSP70 cDNA in Siberian sturgeon *Acipenser baerii* [J]. Journal of Dalian Ocean University, 2012, 27(2):150–157. [田照辉, 徐绍刚, 王巍. 西伯利亚鲟热休克蛋白 HSP70 cDNA 的克隆、序列分析和组织分布. 大连海洋大学学报, 2012, 27(2):150–157.]
- [24] Wen A Y, You F, Xu Y L, et al. Research progress

- on the study of sex determination and differentiation related to gene of fish [J]. Marine science, 2008, 32 (1): 74 – 80. [文爱韵, 尤 锋, 徐永立, 等. 鱼类性别决定与分化相关基因研究进展. 海洋科学, 2008, 32 (1): 74 – 80.]
- [25] Tong J G, Zhu J H, Guan H S. A mini-review of studies on genetic basis of sex determination in fish [J]. Journal of Fisheries of China, 2003, 27 (2): 169 – 176. [童金苟, 朱嘉濠, 关海山. 鱼类性别决定的遗传基础研究概况. 水产学报, 2003, 27 (2): 169 – 176.]
- [26] Takamatsu N, Kanda H, Ito M, *et al.* Rainbow trout SOX9: cDNA cloning, gene structure and expression [J]. Gene, 1997, 202 (2): 167 – 70.
- [27] Chiang E, Pai C, Wyatt M, *et al.* Two sox9 genes on duplicated zebrafish chromosomes: Expression of similar transcription activators in distinct sites [J]. Developmental Biology, 2001, 231 (1): 149 – 163.
- [28] Chen J P, Yuan H M, Wang B, *et al.* cDNA cloning, mRNA transcription of SOX9 gene at early developmental stages in amur sturgeons (*Acipenser schrenckii*) [J]. Zoological Research, 2004, 25 (6): 527 – 533. [陈金平, 袁红梅, 王斌, 等. 史氏鲟 SOX9 基因 cDNA 的克隆及在早期发育过程不同组织中的表达. 动物学研究, 2004, 25 (6): 527 – 533.]
- [29] Shi Z Y, Cheng Q Q, Song J K, *et al.* Cloning, analysis and expression detection of the full-length cDNA of Sox9 from the Siberian sturgeon (*Acipenser baerii* Brandt) [J]. Journal of Fisheries of China, 2010, 34 (5): 664 – 672. [施志仪, 程千千, 宋佳坤, 等. 西伯利亚鲟 Sox9 基因 cDNA 全长克隆、序列分析及其表达检测. 水产学报, 2010, 34 (5): 664 – 672.]

The transcriptome sequencing and functional analysis of muscle tissue of Kaluga (*Huso dauricus*)

ZHAO Wen^{1*}, GAO Fengying¹, SHI Zhengguang²

(1. Key Laboratory of Hydrobiology in Liaoning Province, College of Fisheries and Life Science, Dalian Ocean University, Dalian 116023, China;

2. Yunnan Amur Sturgeon Group Limited Company, Kunming 050228, China)

Abstract: The transcriptome of the *Huso dauricus*' muscle tissue was sequenced by a new generation of high throughput sequencing technology in order to fill a gap which has hitherto existed in the research field of gene sequences of *H. dauricus* and to preserve the gene resources, at the same time to lay the ground for discovering functional genes and cloning new genes, etc. Obtaining 14 447 211 200 bp of the original data, it generated 55 531 Unigenes which ranged from 300 bp to 32 613 bp in length, the average length was 941 bp. Similar searches were made by Bioinformatics methods which was against the Nr database of these Unigenes, in addition, it made gene functional description of the GO and analysis of metabolic pathways of KEGG. The result indicated that 20 735 (37.34%) of these Unigenes have significant matches. The unigenes GO functions in the transcriptome library were divided into 3 categories: biological process, cellular component and molecular function with 56 branches; it could be divided into 290 classes taking the KEGG database as a reference.

Key words: *Huso dauricus*; muscle tissue; gene; transcriptome

Corresponding author: ZHAO Wen. E-mail: zhaowen@dlou.edu.cn