Open Journal of Fisheries Research 水产研究, 2021, 8(4), 175-184
Published Online December 2021 in Hans. http://www.hanspub.org/journal/ojfr
https://doi.org/10.12677/ojfr.2021.84021

Hans汉斯

大洋河刀鲚鳃转录组测序与分析

张伯序，胡宗云，张　健，王建军，杨培民*

辽宁省淡水水产科学研究院，辽宁省水生动物病害防控重点实验室，辽宁　辽阳

收稿日期：2021年11月14日；录用日期：2021年11月30日；发布日期：2021年12月14日

摘　要

为探究盐度影响与刀鲚(*Coilia nasus*)腮组织转录组的关系，本实验对取自大洋河流域不同盐度水体的刀鲚的腮组织进行了转录组测序和分析。结果显示：一共有70,964条unigenes被注释；刀鲚的unigene序列与大西洋鲱(*Clupea harengus*)最为接近；有关细胞进程、生物调节、代谢过程的unigenes占比很高；腮组织部有关代谢功能的转录组表达十分旺盛；生物体系统(Organismal Systems)与环境信息处理的信号通路(Environmental Information Processing)的unigene数最多。

关键词

大洋河，刀鲚，转录组测序

Transcriptome Sequencing and Analysis for Gills of *Coilia nasus* from Dayang River

Boxu Zhang, Zongyun Hu, Jian Zhang, Jianjun Wang, Peimin Yang*

Liaoning Key Laboratory for Prevention and Treatment of Aquatic Animal Diseases, Freshwater Fisheries Research Institute of Liaoning Province, Liaoyang Liaoning

Received: Nov. 14th, 2021; accepted: Nov. 30th, 2021; published: Dec. 14th, 2021

Abstract

In order to explore the relationship between salinity effects and transcriptome of (*Coilia nasus*) gill tissues, transcriptome sequencing and analysis were performed on the gill tissues of *Coilia nasus* from waters with different salinity in the Dayang River basin. The results showed that 70964 unigenes were annotated. The unigene sequence of *Coilia nasus* was closest to that of Atlantic Her-

*通讯作者。

文章引用: 张伯序, 胡宗云, 张健, 王建军, 杨培民. 大洋河刀鲚鳃转录组测序与分析[J]. 水产研究, 2021, 8(4): 175-184. DOI: 10.12677/ojfr.2021.84021

ring (*Clupea harengus*). The proportion of unigenes related to cell process, biological regulation and metabolic process was high. Transcriptome expression related to metabolic function was very vigorous in the gill tissue. Organismal Systems and Environmental Information Processing have the largest number of unigene.

Keywords

Dayang River, *Coilia nasus*, RNA-Seq Sequencing

1. 引言

刀鲚(*Coilia nasus*)属鲱形目(Clupeiformes)鳀科(Engraulidae)鲚属(Coilia)，俗称刀鱼、河刀[1]。刀鲚主要分布在我国黄渤海和东海一带[2]，辽宁境内的辽河、鸭绿江及大洋河亦有分布[3]。刀鲚味道鲜美，长久以来都有“长江三鲜”的美誉，是重要的经济型鱼类，但近年来，由于水利工程兴建以及环境污染等不利因素，导致刀鲚的生存繁殖环境遭到严重的破坏。特别是近 10 年间，由于不科学的过度捕捞导致刀鲚种群量逐年递减，现已被世界自然保护联盟(IUCN)列为濒危物种[4]。辽宁境内水域是我国刀鲚分布的最北线，有关其研究相对较少，仅见有繁殖生物学、遗传多样性及活体运输等方面的报道[5] [6] [7]：利用 DNA 分子技术探究长江刀鲚、湖鲚的遗传多样性[8]；利用转录组测序技术对刀鲚嗅觉上皮组织进行比较[9]。本文以大洋河流域的洄游型刀鲚为研究对象，对来自不同盐度水域刀鲚鳃的转录组进行了比较分析，探究在转录组水平上刀鲚鳃部组织对盐度变化的响应机制，为今后研究大洋河刀鲚的生活习性和环境适应性机制提供研究数据和理论参考。

2. 材料与方法

2.1. 样品采集

2020 年 8 月用流刺网在大洋河下游不同河段采捕刀鲚，本试验取河口咸水区(盐度为 12‰，样品标记为 DE)和石山桥河段淡水区(盐度为 0.2‰，样本标记为 SG)大小相近的鲜活刀鲚各 5 尾，刀鲚全长和体重分别为(27.52 ± 0.81) cm、(68.49 ± 6.96) g。剪去鳃部组织置于 RNAfixer (Bioteke 北京)保存液中带回实验室备用。

2.2. 总 RNA 提取与检测

取保存的鳃部组织 100 mg 左右，采用 TRizol 法提取 total RNA。利用 Thermo Nanodrop2000 对所提 RNA 的浓度和纯度进行检测，并用 1.5%琼脂糖凝胶电泳检测 RNA 完整性。采用 Agilent2100 测定 RIN 值，单次建库要求 RNA 量 1 ug，浓度 ≥ 50 ng/μL，OD260/280 介于 1.8~2.2 之间。

2.3. 文库构建

利用带有 Oligo (dT)的磁珠与 ployA 进行 A-T 碱基配对，总 RNA 中分离出 mRNA，加入 fragmentation buffer，可以将 mRNA 随机断裂成 300 bp 左右的小片段。通过逆转录酶加入六碱基随机引物(random

hexamers)，以 mRNA 为模板反转合成一链 cDNA，随后进行二链合成，形成稳定的双链结构。双链的 cDNA 结构为粘性末端，加入 End Repair Mix 将其补成平末端，随后在 3’末端加上一“A”碱基，用于连接 Y 字形的接头。

2.4. 转录组测序、质控与组装

本实验采用 Illumina HiSeq2500 测序平台完成转录组测序，Illumina 测序基于循环可逆终止技术工作，流程如下：

1) 技术代表目前新兴的基因组 DNA 提纯后被随机打断。这一步可以通过物理方法完成，如声波法、剪切法，或者雾化法，通常进一步通过长度分选随机打断的 DNA 片段。在两端都接上接头。

2) 单链 DNA 片段共价连接到流动细胞通道的表面。

3) 加入 DNA 聚合酶和未标记的脱氧核苷酸产生固相“桥扩增”，其中模板 DNA 使两端连接到通道表面形成 U 形环。

4) 双链桥生成。双链分子变性，然后继续扩增以形成高度簇集的模板 DNA。

5) 加入四个标记的可逆终端(包含引物和 DNA 聚合酶)。在给定的循环中，一个可逆终端只能被加入一个模板。在特殊的不能延长的碱基处会产生链终止。

6) 在激光的激发下，第一个碱基的身份被记录。

7) 在第二个循环中，可逆终端被去除(保护)。所有四个标记的可逆终端和聚合酶再次被加入流细胞中。这个循环被重复。

之后对获得的测序数据进行质量控制(QC)，之后利用生物信息学手段对转录组数据进行分析。其中使用 fastx_toolkit_0.0.14 软件对对每一个样本的碱基质量、碱基错误率以及碱基分布进行分析；然后使用软件 SeqPrep (https://github.com/jstjohn/SeqPrep)和 Sickle (https://github.com/najoshi/sickle)去除低质量 reads 得到高质量的质控数据(clean data)；最后由 Trinity (https://github.com/trinityrnaseq/trinityrnaseq/wiki)软件将 clean data 进行从头组装。

2.5. 数据处理

将测序过程中的图像信号经 CASAVA 碱基识别(Base Calling)转换成文字信号，并将其以 fastq 格式储存起来作为原始数据。根据 index 序列区分各个样本的数据，以便进行后续分析。单次运行能产生数十亿级的 reads。将测序获得的 reads 与 unigenes 通过 Bowtie 2.3.5 进行比对，然后结合比对结果，经由 RSEM 1.2.2 软件的分析得到表达量水平的结果估计。最终输入不同样本的拷贝数(read counts)信息构成的矩阵，从而得到表达量信息差异分析结果。

2.6. 功能注释

通过 BLAST 软件将该次转录组测序获得的所有转录本与六大数据库(NR, Swiss-Prot, Pfam, COG, GO 和 KEGG 数据库)进行比对，获得在各数据库的注释信息，并使用 HMMER 与 Pfam 软件对各数据库注释情况进行统计和分析；并对 SNP 功能区域和 SSR 进行统计。

3. 结果与分析

3.1. 刀鲚鳃组织 RNA 检测结果

提取得到的 RNA 样品具有完整清晰的 28S、18S 和 5S 带型，A260/280 值为 2.02，28S:18S 值为 1.80，RIN 值为 8.0，说明本次提取得到的刀鲚组织样品 RNA 质量较好，可以用于制备测序文库。

3.2. 测序数据与组装结果

利用 Trinity 将拼接过滤后得到的 reads 片段进行聚类及进行拼接组装，共得到 70,964 条 unigenes 和 100,022 条转录本(transcript)，平均长度分别为 1061.73 bp 和 1140.11 bp。transcript 与 unigenes 的 N50 (重叠群序列累加后长度超过转录组总长度一半时的重叠群序列长度)分别为 2227 和 2209 (表 1)。其中长度在 0~500 bp 之间的 unigenes 数量为 36,872 条，占总数量的 52%。长度在 501~1000 bp 为 12,779 条，总 unigenes 数量的 18%；1000 bp 长度以上的 unigene 数量均不超过 10% (表 2)。

Table 1. Statistical table of assembly results
表 1. 组装结果统计表

类别	Unigene 数	Transcript 数
Total number	70,964	100,022
Total base	75,344,738	114,036,134
Largest length (bp)	19,645	19,645
Smallest length (bp)	201	201
Average length (bp)	1061.73	1140.11
N50 length (bp)	2227	2209
GC percent (%)	47.71	47.75

Table 2. Sequence length distribution of unigenes
表 2. Unigenes 序列长度分布情况

序列长度	Unigene 数量	Unigene 百分比
0~500	36,872	52%
501~1000	12,779	18%
1001~1500	5613	8%
1501~2000	4204	6%
2001~2500	3069	4%
2501~3000	2348	3%
3001~3500	1761	2%
3501~4000	1186	2%
4001~4500	907	1%
>4500	2225	3%

3.3. Unigene 功能注释统计

将测序获得的所有转录本在各数据库注释的整体情况进行统计，结果表明，有 22,342 个 unigenes 获得了 GO 注释，占全部 unigenes 的 31.48% (表 3)。GO 注释的功能分析由 3 大部分组成，可以对基因和基因产物按照其参与的 BP (Biological Process，生物过程)、MF(Molecular Function，分子功能)及 CC (Cellular Component，细胞组分)方面进行分类注释(图 1)。在这三个大分支下面又分很多小层级，功能上的细分更

Table 3. Annotation result statistics of unigenes
表 3. Unigenes 注释概况统计表

数据库	Unigene 数量	Unigene 百分比
GO	22,342	31.48%
NR	29,750	41.92%
KEGG	18,997	26.77%
COG	27,529	38.79%
Swiss-Pro	24,766	34.9%
Pfam	24,156	34.04%
Total_anno	31,069	43.78%
Total	70,964	100%

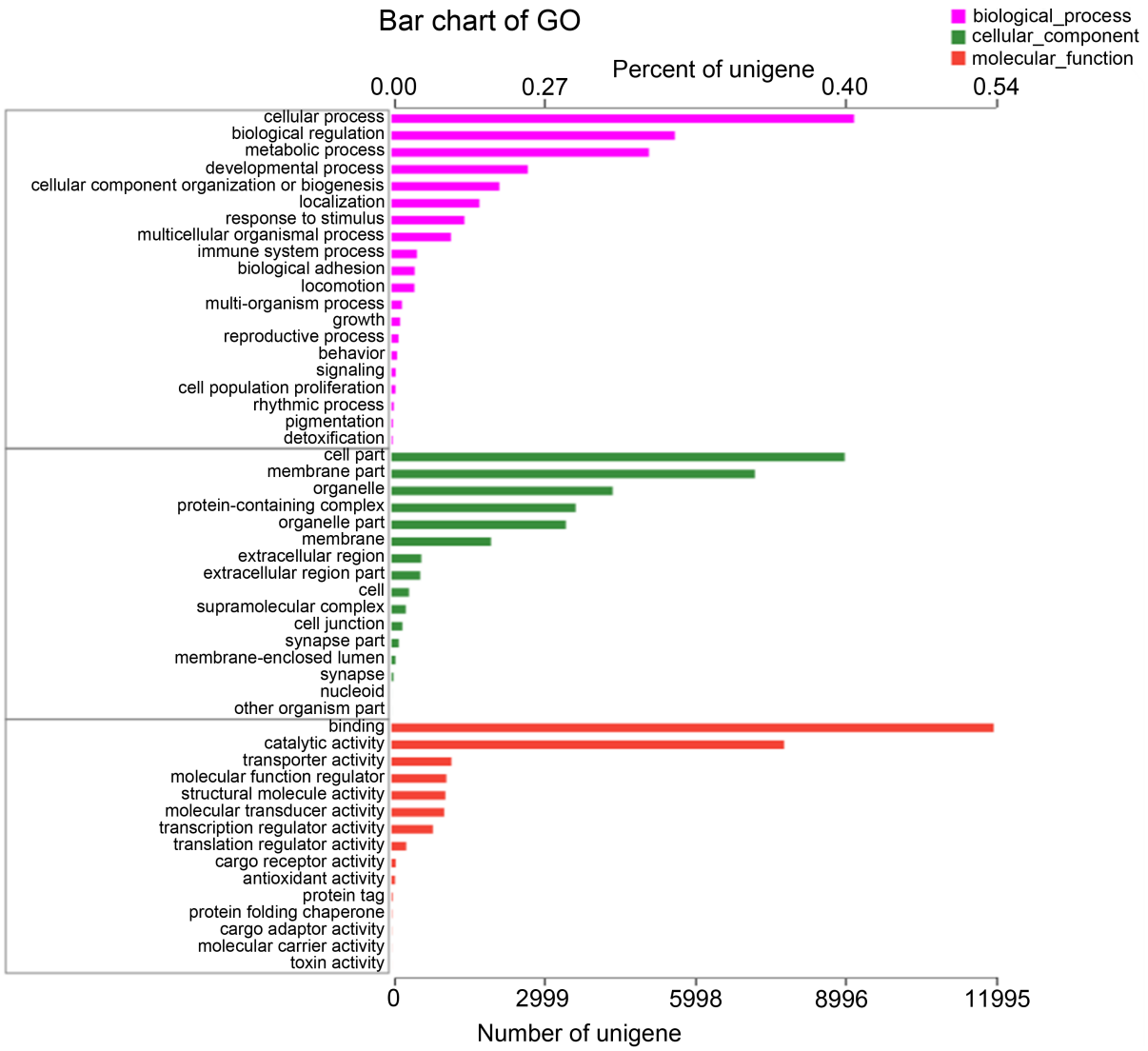


Figure 1. Annotated diagram of GO
图 1. GO 注释分析图

有助于从整体上了解全部基因产物的功能分类。本次实验重点关注了生物过程这一分支，在有关生物过程下属的层级的注释中(图 2)：细胞进程(celluar process)共 9214 条占 29.22%；生物调节(biological regulation)共 5646 条，占 17.90%；代谢过程(metabolic process)共 5127 条，占 16.26%，NCBI_NR(NCBI 非冗余蛋白库)为综合数据库，可通过比对查看本物种转录本序列与相近物种的相似情况，以及同源序列的功能信息。如图 3 所示，刀鲚(*Coilia nasus*)的 unigene 序列大西洋鲱(*Clupea harengus*)相似度最高，为 58.68%，另外还有 19.60%的序列被注释到其他物种之中。通过与 KEGG 数据库比对，可获得某基因或转录本可能参与的具体生物学通路情况，这些信息有助于从系统水平解读基因的生物学功能。如图 4 所示，6 大代谢途径中，参与生物体系统(Organismal Systems)的信号通路的 unigene 所占比例最高，遗传信息处理(Genetic Information Processing)所占比例最低。通过对 unigene 的注释和蛋白差异表达分析，为后续了解基因功能和解释表型差异提供数据基础。

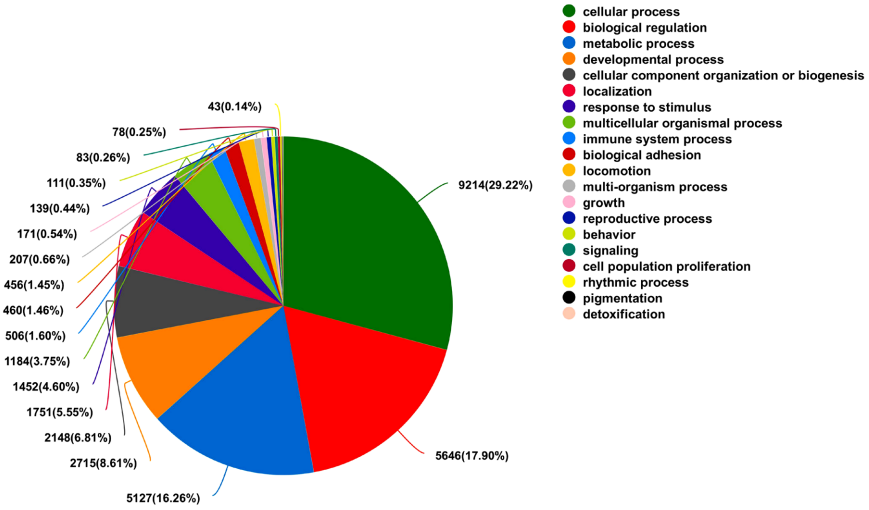


Figure 2. GO (BP) analysis pie charts
图 2. GO (BP)分析饼图

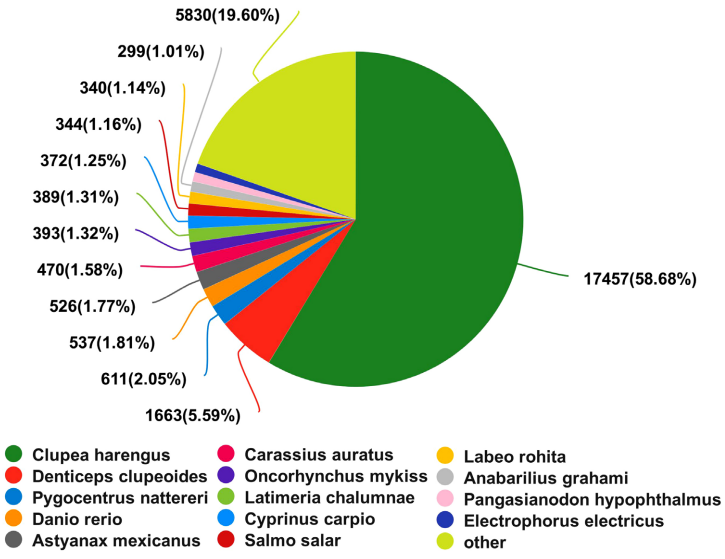


Figure 3. Annotated pie chart of NR
图 3. NR 注释物种分布饼图

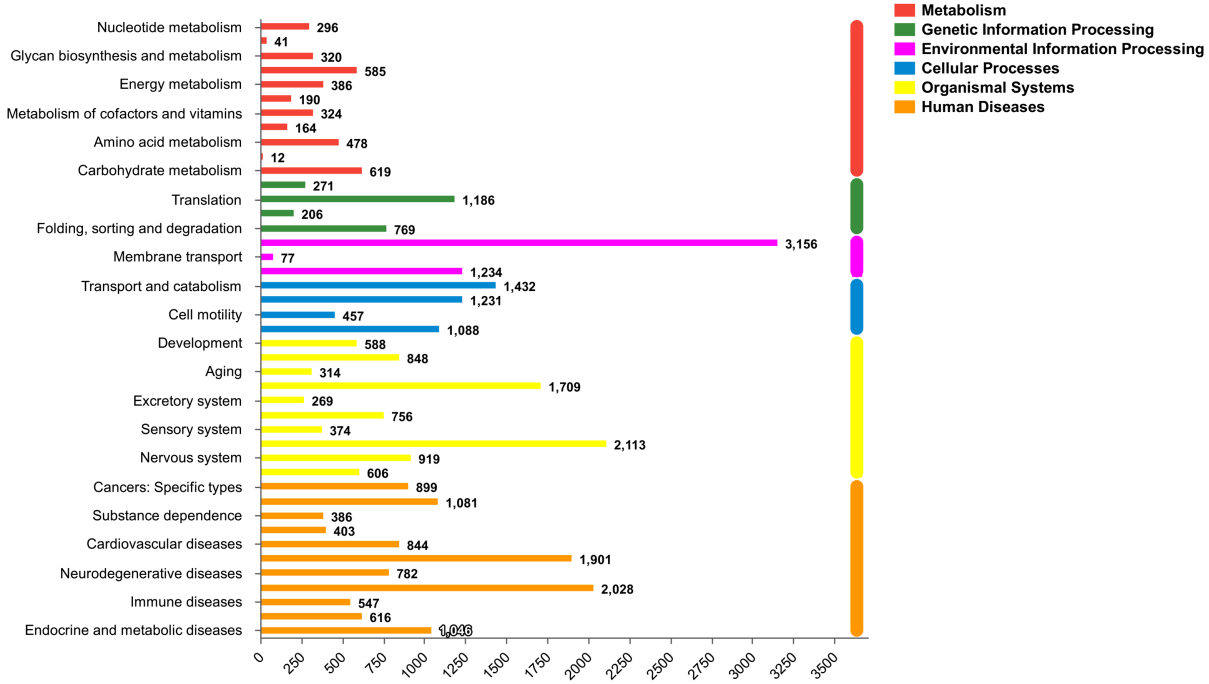
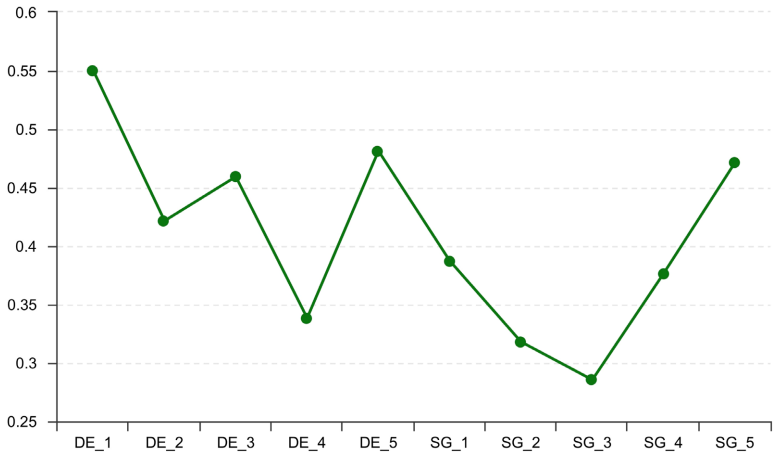


Figure 4. KEGG analysis bar chart
图 4. KEGG 分析柱状图

3.4. 表达量统计

RSEM，Kallisto，Salmon 是三款常用的转录组定量分析软件，本次实验采用 RSEM 进行转录组的定量，并完成转录组和组装结果的比对。三款软件的输出结果包含表达量 TPM 或 FPKM (仅 RSEM)信息和 read counts (比对到基因上的 reads 个数)信息。不同样本的 read counts 信息构成的矩阵可以用于差异分析的输入，表达量信息用于后续的样本聚类等分析。如图 5 所示，样本 DE_1 的表达量最高；样本 SG_3 表达量最低，同盐度样本中，样本之间的表达量变化较大，总体来看不同盐度的样本之间，随着盐度的变化的总体表达量无明显的差异。



注：横坐标为样本名称，纵坐标为样本表达量(log10 TPM+1)

Figure 5. Statistical line chart of expression quantity
图 5. 表达量统计折线图

3.5. 样本间 Venn 分析

Venn 分析展示样本间或组别间共有和特有表达的基因/转录本，可简单呈现样本间相关性，同一组别样本表达基因/转录本的数目不应差别很大。从图 6 可以看出，DE 组的表达 unigene 数目为 34,768；SG 组的表达 unigene 数目为 23,546；两组共有的组的表达 unigene 数目为 20,626，总体上开看 DE 组的 unigene 的总表达量数量要高于 SG 组。

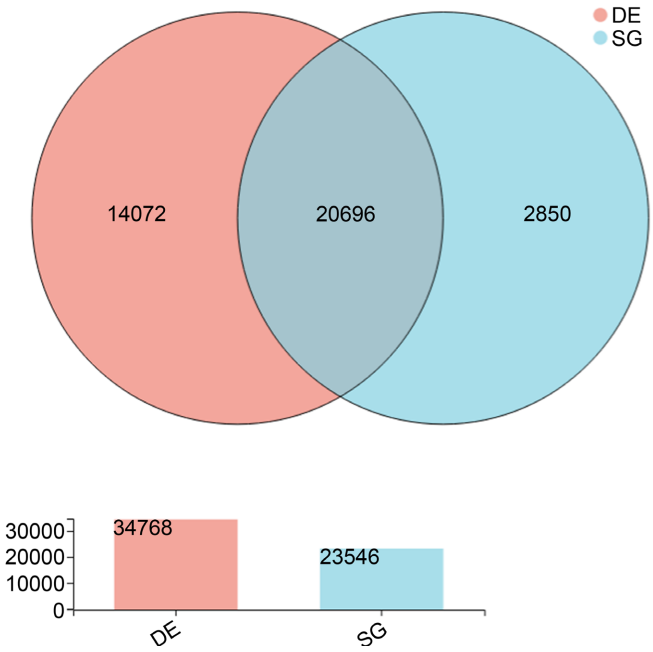


Figure 6. Expression quantity Venn diagram
图 6. 表达量 Venn 图

4. 讨论

全基因组测序技术的发展和测序数据免费公开极大地改变了生物学的研究方式，测序和其他基因组学数据具有推动农业、环境科学和生态学的巨大潜力。近些年得益于高通量测序技术的快速发展和普及，基因组测序的水平有了显著的提高[10] [11]，对于刀鲚这类非模式生物而言，NGS (高通量测序)是一个方便且可以快速得到其基因序列的方法。本文通过对大洋河不同水域刀鲚样本的鳃组织进行转录组测序，得到可信的 unigenes

1) 其中长度在 0~500 bp 之间的 unigenes 数量为 36872 条，占总数量的 52%。长度在 501~1000 bp 为 12779 条，占总 unigenes 数量的 18%。测序样本的 Mapped read 数分别为 SG-1:42641542、SG-2:38980922、SG-3:44444358、SG-4:41879374、SG-5:41038376、DE-1:44473856、DE-2:43234696、DE-3:43123052、DE-4:47355458、DE-5:40274730，序列在 NR、Swiss-Prot、Pfam、COG、GO 和 KEGG 公共数据库中进行注释。

2) NR 结果表明刀鲚的 uingene 序列与大西洋鲱(*Clupea harengus*)的最为接近，达到了 58.68%，符合二者同属鲱形目的动物学分类。其次有 5.59%的 unigene 在齿鲱(*Depticeps clupeoides*)中得到注释，说明虽同属鲱形目，但是二者之间差异巨大。另外还有 19.60%的 unigene 被注释到其他物种中。

3) 在 GO 注释的生物过程 unigenes 表明，细胞进程、生物调节、代谢过程这三类 unigene 占据很大比重，这些注释的 unigenes 均是参与调控和维持细胞正常生理活动的基因，这三类的高占比，在一定程

度上体现了刀鲚的鳃细胞代谢旺盛，增殖快，新陈代谢速率快等特点。这些基因既保证了细胞的正常运转，也可能与调控刀鲚鳃细胞的增殖、分化密切相关。另外 GO 的注释结果中有相当部分的未知功能的 unigenes 没有得到注释，说明对于刀鲚 unigenes 的研究还有还大的空间。

4) KEGG 通路注释分类结果中，生物体系统(Organismal Systems)的信号通路 unigene 数量最多，表明在生长发育及生命活动过程中的代谢活动非常旺盛，这些信号通路起着至关重要的生理调控作用；另外参与环境信息处理(Environmental Information Processing)的信号通路的 unigene 数也很多，说明刀鲚鳃组织细胞对环境因子的敏感程度很高。某些硬骨鱼类鳃组织的酶活性与盐度的变化有一定关系[12]，鳃丝表皮细胞对维持渗透压等生理功能有重要的调节作用[13]，可能与刀鲚适应海水盐度或者水温的频繁变化与维持自身生理系统稳定和提高环境自适应能力的 unigene 信号通路数量占比高，具有一定的联系。

在生物体内，不同基因产物之间通过有序的相互协调来行使其具体的生物学功能，基因表达量的变化一定程度上体现了细胞代谢的强弱，某些酶的表达对鳃的生理功能有着重要的作用，例如在某些硬骨鱼类中鳃丝 Na^+/K^+-ATP 酶活力会随着盐度的变化而变化，而某些硬骨鱼类从半海水过渡到淡水或者海水的过程中，其鳃丝 Na^+/K^+-ATP 酶活力没有明显的变化；HEIJDEN [14]等认为，盐度对鱼类鳃丝 Na^+/K^+-ATP 酶活力影响并不显著。通过对转录组定量的分析可以看出 DE_1 的表达量最高，SG_3 的表达量最低，从总体来看 DE 组的表达量大于 SG 组的表达量(图 6)，但不同盐度之间的表达量差异并不与盐度呈现相关性，从结果来看盐度对刀鲚鳃部基因的表达量的影响并不显著。大洋河刀鲚属于洄游性鱼类，不同于其他广盐性鱼类，大洋河刀鲚活动范围仅限于近海，区别于其他鱼类，刀鲚在生存环境上的差异不如其他鱼类那样显著。目前针对环境因子对刀鲚鳃组织基因表达和生理功能的影响的研究较少，对于刀鲚鳃组织基因表达蛋白通路的作用机制尚不明确，今后对于盐度对刀鲚鳃组织转录组影响的研究还有很大的空间。

5. 展望

目前关于转录组的报道和研究主体还是以四大家鱼、鲆鲽类、大西洋鲑等市场经济型鱼类为主[15] [16] [17]关于刀鲚及其种属的研究目前尚不多见。刀鲚作为一种名贵鱼类，其经济价值潜力巨大。日后旨在探究在洄游过程中，不同盐度对刀鲚基因表达及生理功能的影响，还应在各个层次水平开展更为深入的研究，为刀鲚的资源养护提供更多的理论基础，提升完善整个刀鲚的理论体系，为日后的发展创造更好的理论条件。

基金项目

辽宁省农业农村厅项目(JH20-210000-39754)、辽宁省农科院项目(2021HQ1918)及辽宁省科技厅项目(2021JH2/10200031)。

参考文献

[1] 解玉浩. 东北地区淡水鱼类[M]. 沈阳: 辽宁科学技术出版社, 2007, 148.

[2] Liu, D., Li, Y.Y., Tang, W.Q., *et al*. (2014) Population Structure of *Coilia nasus* in the Yangtze River Revealed by Insertion of Short Interspersed Elements. *Biochemical Systematics and Ecology*, **54**, 103-112. https://doi.org/10.1016/j.bse.2013.12.022

[3] 张健, 杨培民, 胡宗云, 等. 大洋河刀鲚繁殖生物学特性[J]. 淡水渔业, 2021, 51(6): 3-11.

[4] Hata, H. (2018) The IUCN Red List of Threatened Species. Page Bros, Norwich.

[5] 程方圆, 陶紫玉, 李晨虹. 应用单核苷酸多态性(SNP)标记鉴定短颌鲚、湖鲚和刀鲚[J]. 上海海洋大学学报, 2019, 28(1): 10-19.

[6] 王生, 方春林, 周辉明, 等. 鄱阳湖刀鲚的渔汛特征及渔获物分析[J]. 水生态学杂志, 2017, 38(6): 82-87.

[7] 施永海, 张根玉, 张海明, 等. 刀鲚的全人工繁殖及胚胎发育[J]. 上海海洋大学学报, 2015, 24(1): 36-43.

[8] 杨巧莉. 中国鱿属鱼类进化关系及刀鳞、凤鱿的分子系统地理学研究[D]: [博士学位论文]. 青岛: 中国海洋大学, 2012.

[9] Zhu, G., Wang, L., Tang, W., *et al*. (2014) *De Novo* Transcriptomes of Olfactory Epithelium Reveal the Genes and Pathways for Spawning Migrationin Japanese Grenadier Anchovy (*Coilia nasus*). *PLoS ONE*, **9**, e103832. https://doi.org/10.1371/journal.pone.0103832

[10] Li, S., Shen, L., Sun, L., *et al*. (2017) Small RNA-Seq Analysis Reveals microRNA-Regulation of the Imd Pathway during *Escherichia coli* Infection in *Drosophila*. *Developmental & Comparative Immunology*, **70**, 80-87. https://doi.org/10.1016/j.dci.2017.01.008

[11] Jia, Z., Wang, Q., Wu, K., *et al*. (2017) *De Novo* Transcriptome Sequencing and Comparative Analysis to Discover Genes Involved in Ovarian Maturity in *Strongylocentrotus nudus*. *Comparative Biochemistry and Physiology Part D*: *Genomics and Proteomics*, **23**, 27-38. https://doi.org/10.1016/j.cbd.2017.05.002

[12] Pelis, R.M. and McCormick, S.D. (2001) Effects of Growth Hormone and Cortisol on Na^+-K^+-2Cl^- Cotransporter Localization and Abundance in the Gills of Atlantic Salmon. *General and Comparative Endocrinology*, **124**, 134-143. https://doi.org/10.1006/gcen.2001.7703

[13] 王艳, 胡先成. 不同盐度下鲈鱼稚鱼鳃的显微结构观察[J]. 海洋科学, 2009, 33(12): 138-142.

[14] Heijden, A., Verbost, P., Eygenstcyn, J., *et al*. (1997) Mitochondria-Rich Cells in Gills of Tilapia (*Oreochro mismossambicus*) Adapted to Fresh Water or Sea Water: Quantification by Confocal Laser Scanning Microscopy. *Journal of Experimental Biology*, **200**, 55-64. https://doi.org/10.1242/jeb.200.1.55

[15] Dang, Y., Xu, X., Shen, Y., *et al*. (2016) Transcriptome Analysis of the Innate Immunity-Related Complement System in Spleen Tissue of *Ctenopharyngodon idella* Infected with *Aeromonas hydrophila*. *PLoS ONE*, **11**, e0157413. https://doi.org/10.1371/journal.pone.0157413

[16] Pereiro, P., Balseiro, P., Romero, A., *et al*. (2012) High-Throughput Sequence Analysis of Turbot (*Scophthalmus maximus*) Transcriptome Using 454-Pyrosequencing for the Discovery of Antiviral Immune Genes. *PLoS ONE*, **7**, e35369. https://doi.org/10.1371/journal.pone.0035369

[17] Sutherland, B.J., Koczka, K.W., Yasuike, M., *et al*. (2014) Comparative Transcriptomics of Atlantic *Salmo salar*, Chum *Oncorhynchus keta* and Pink Salmon *O*. *gorbuscha* during Infections with Salmon Lice *Lepeophtheirus salmonis*. *BMC Genomics*, **15**, Article No. 200. https://doi.org/10.1186/1471-2164-15-200