

## 刀鲚基因家族鉴定及扩张与收缩

张金鹏<sup>1,2</sup>, 高淑芳<sup>1,2</sup>, 施永海<sup>3</sup>, 严银龙<sup>3</sup>, 齐明<sup>1,2</sup>, 刘其根<sup>1,2\*</sup>

(1. 上海海洋大学, 水产科学国家级实验教学示范中心, 上海 201306;

2. 上海海洋大学, 农业农村部鱼类营养与环境生态研究中心, 上海 201306;

3. 上海市水产研究所, 上海市水产技术推广站, 上海 200433)

**摘要:** 为增加对刀鲚在进化过程中因核心基因家族改变而产生的特殊机制的了解。实验利用比较基因组学, 通过比较刀鲚及其近缘物种斑马鱼、三刺鱼、青鳉、红鳍东方鲀和绿河鲀基因组, 鉴定刀鲚基因家族; 利用 CAFÉ v 4.2 软件进行基因家族扩张和收缩分析; 最后将鉴定的扩张基因家族基因与 GO、KEGG 等数据库进行比对、注释和通路分析。结果显示, 刀鲚具有 11 872 个基因家族, 包含 16 470 个基因, 有 2 963 个基因未形成基因家族。同其他 5 个物种相比, 刀鲚有 150 个特有基因家族, 包含 419 个基因。刀鲚扩张和收缩基因家族分析结果表明, 刀鲚有 1 200 个基因家族扩张了, 7 543 个基因家族收缩了。刀鲚显著扩张的基因家族有 39 个, 包含 508 个基因; 显著收缩的基因家族有 36 个, 包含 21 个基因。刀鲚显著扩张基因家族 GO 富集分析条目中鉴定较多的是, 代谢过程、催化活性、细胞、细胞组分等。KEGG 富集分析条目中鉴定较多的是紧密连接、心肌细胞的肾上腺素能信号、心肌收缩、细胞黏附分子等。与渗透压调节相关的通路有紧密连接、心肌收缩等; 与生殖洄游行为相关的通路有 GnRH 信号通路、嗅觉转导通路等。对显著扩张基因家族基因进行基因组定位, 结果显示 23 号染色体上基因数目显著多于其他染色体。定位在 23 号染色体的基因多为组蛋白基因, 其功能与染色质结构调控有关。研究表明, 刀鲚渗透调节、信号转导、代谢、性腺发育和嗅觉等方面均可能产生了适应性进化, 且其染色质结构调节也可能发生改变。

**关键词:** 刀鲚; 基因家族; 收缩; 扩张; 通路分析

**中图分类号:** Q 785; S 917.4

**文献标志码:** A

刀鲚 (*Coilia nasus*) 隶属于鲚科 (Engraulidae) 鲚属 (*Coilia*), 是我国重要的经济鱼类。其营养丰富、肉嫩味鲜, 与河鲀、鲥鱼并称为“长江三鲜”, 并一度成为世界上最贵的鱼类<sup>[1-2]</sup>。除了极高的经济和营养价值, 刀鲚还具有特殊的生态学研究价值。中国鲚属鱼类生态型多样, 既有生活在海洋河口的七丝鲚 (*C. grayii*)、凤鲚 (*C. mystus*), 也有纯淡水生活的短颌鲚 (*C. brachygnathus*) 及若干陆

封群体 (如湖鲚) 等, 还有河海洄游的刀鲚<sup>[3]</sup>。这使得中国鲚属鱼类成为很好的鱼类适应性进化研究模型, 而刀鲚也成为极好的河海洄游鱼类研究模型。

在溯河洄游过程中, 鱼类不仅要经历剧烈的渗透压变化, 还要进行定位、调控性腺发育, 这些行为或生理上的变化依赖于生物体内因和外因的双重调控<sup>[4]</sup>, 并会产生适应性的进化。关于洄

收稿日期: 2020-11-04 修回日期: 2021-04-13

资助项目: 上海海洋大学农业农村部鱼类营养与环境生态研究中心专项 (A1-3201-19-300204); 上海市科委项目 (17391900300)

第一作者: 张金鹏 (照片), 从事鱼类生殖洄游和适应性进化研究, E-mail: kim1103@foxmail.com

通信作者: 刘其根, 从事渔业资源与保护生物学研究, E-mail: qgliu@shou.edu.cn



游鱼类适应性进化的研究较多<sup>[5-7]</sup>,然而刀鲚因洄游产生的适应性进化研究较少,仅见 Xu 等<sup>[8]</sup>利用群体基因组测序确定了 150 个迁移适应性候选基因。唐文乔等<sup>[9-10]</sup>通过研究嗅觉受体 *MOR-51I2* 和 *MOR-2AK2* 来探究刀鲚洄游机制。吴利红等<sup>[11]</sup>利用刀鲚体内脂肪迁移探讨其洄游距离等。分析扩张和收缩基因家族可以从整个基因组层面检测物种在进化过程中核心基因家族的变化,从而帮助了解物种特异的适应性机制。例如 Ren 等<sup>[12]</sup>分析扩张基因家族,对褐尾麻蝇 (*Sarcophaga peregrina*) 卵胎生繁殖和腐肉进食的适应性进行了讨论。本研究对刀鲚扩张和收缩基因家族进行鉴定和分析,以增加对刀鲚适应性进化的了解。

## 1 材料与方法

### 1.1 数据下载

从 NCBI 数据库 (<https://www.ncbi.nlm.nih.gov/>) 中下载刀鲚基因组、蛋白质序列文件及基因注释文件。根据刀鲚基因组基因注释表,选取斑马鱼 (*Danio rerio*)、三刺鱼 (*Gasterosteus aculeatus*)、青鳉 (*Oryzias latipes*)、红鳍东方鲀 (*Takifugu rubripes*) 和绿河鲀 (*Tetraodon nigroviridis*) 作为刀鲚的近缘物种。从 Ensembl 数据库下载 (<http://asia.ensembl.org/index.html>) 斑马鱼、三刺鱼、青鳉和绿河鲀的基因组、蛋白质序列文件及基因注释文件,从 NCBI 数据库中下载红鳍东方鲀的基因组、蛋白质序列文件及基因注释文件。

### 1.2 基因家族鉴定及物种系统发育树的构建

判断基因注释文件中 mRNA 和 gene 数目是否一致,若 mRNA 数目多于 gene 数目,则存在可变剪切。存在可变剪切的情况下,利用 Gffread v 0.9.9<sup>[13]</sup> 软件提取最长转录本对应的蛋白质序列文件用于后续分析。利用 Blast+ v2.9.0 对蛋白文件进行比对,利用 OrthoFinder v 2.4.0 对比软件<sup>[14]</sup> 对结果进行基因家族聚类分析。

将 OrthoFinder 软件运行结果中的单拷贝蛋白文件提取出来,使用 MAFFT v 7.429 软件<sup>[15]</sup> 对单拷贝蛋白文件进行比对,利用 Gblocks v0.9.1b 对比软件<sup>[16]</sup> 对结果进行优化。将优化后的蛋白序列连接成一条蛋白序列用于物种树的构建。选择最大似然法构建物种系统发育树,利用 IQ-Tree v1.6.11 软件<sup>[17]</sup> 进行建树,bootstrap 值设为 1 000。使用 PAML v4.9i 软件中的 MCMCTREE 模块进行

分歧时间估计,利用 TimeTree (<http://www.timetree.org/>) 网站中提供的斑马鱼 [179.3~222.4 百万年前 (Mya)] 和红鳍东方鲀 (206~252 Mya) 化石时间进行校正。此步独立运行 2 次以检验有分化时间的进化树的可靠性。

### 1.3 扩张与收缩家族鉴定

利用脚本过滤掉超过 200 个基因拷贝的基因家族。根据基因家族聚类结果和具有分歧时间的物种系统发育树,利用 CAFÉ v 4.2 软件<sup>[18]</sup> 进行扩张和收缩基因家族的鉴定,*P* 参数设置为 0.05。

### 1.4 扩张与收缩家族基因、刀鲚独有家族基因注释和通路分析

将鉴定到的扩张与收缩家族基因、刀鲚独有家族基因与 NR、GO、KEGG、Swissprot、TrEMBL、COG、Pfam、KOG 等相关数据库进行比对,获得序列的注释信息。并对扩张和收缩家族基因进行 GO (<http://geneontology.org/>) 和 KEGG 通路分析 (<https://www.genome.jp/kegg/pathway.html>)。利用 TBtools<sup>[19]</sup> 对显著扩张基因家族的基因进行基因组定位。

## 2 结果

### 2.1 基因家族鉴定及物种系统发育树的构建

对 6 个物种的基因家族进行统计,并鉴定出刀鲚具有 11 872 个基因家族,包含 16 470 个基因,有 2 963 个基因未形成基因家族;其余 5 个物种的基因家族数均大于 14 279,高于刀鲚基因家族数 (图 1)。此外,鉴定出刀鲚有 150 个特有基因家族,包含 419 个基因。利用 3 919 个单拷贝直系同源序列构建物种系统发育树,最大似然法构建的系统发育树 bootstrap 百分比为 100%,建树结果可信。独立运行时间树 2 次,计算两棵树时间节点的相关性,得到  $R^2$  为 0.849 8,相关性较高,时间树建树结果可信,系统发育树表明青鳉、三刺鱼、红鳍东方鲀和绿河鲀聚为一支;刀鲚与斑马鱼分歧时间发生在 226.787 9 Mya (95% HSD: 222.397, 231.496) (图 2)。

### 2.2 扩张基因家族鉴定结果及通路分析

刀鲚有 1 200 个基因家族扩张,7 543 个基因家族收缩。刀鲚显著扩张 ( $P < 0.05$ ) 的基因家族有 39 个,包含 508 个基因;显著收缩 ( $P < 0.05$ ) 的基因家族有 36 个,包含 21 个基因 (图 3)。刀鲚显著

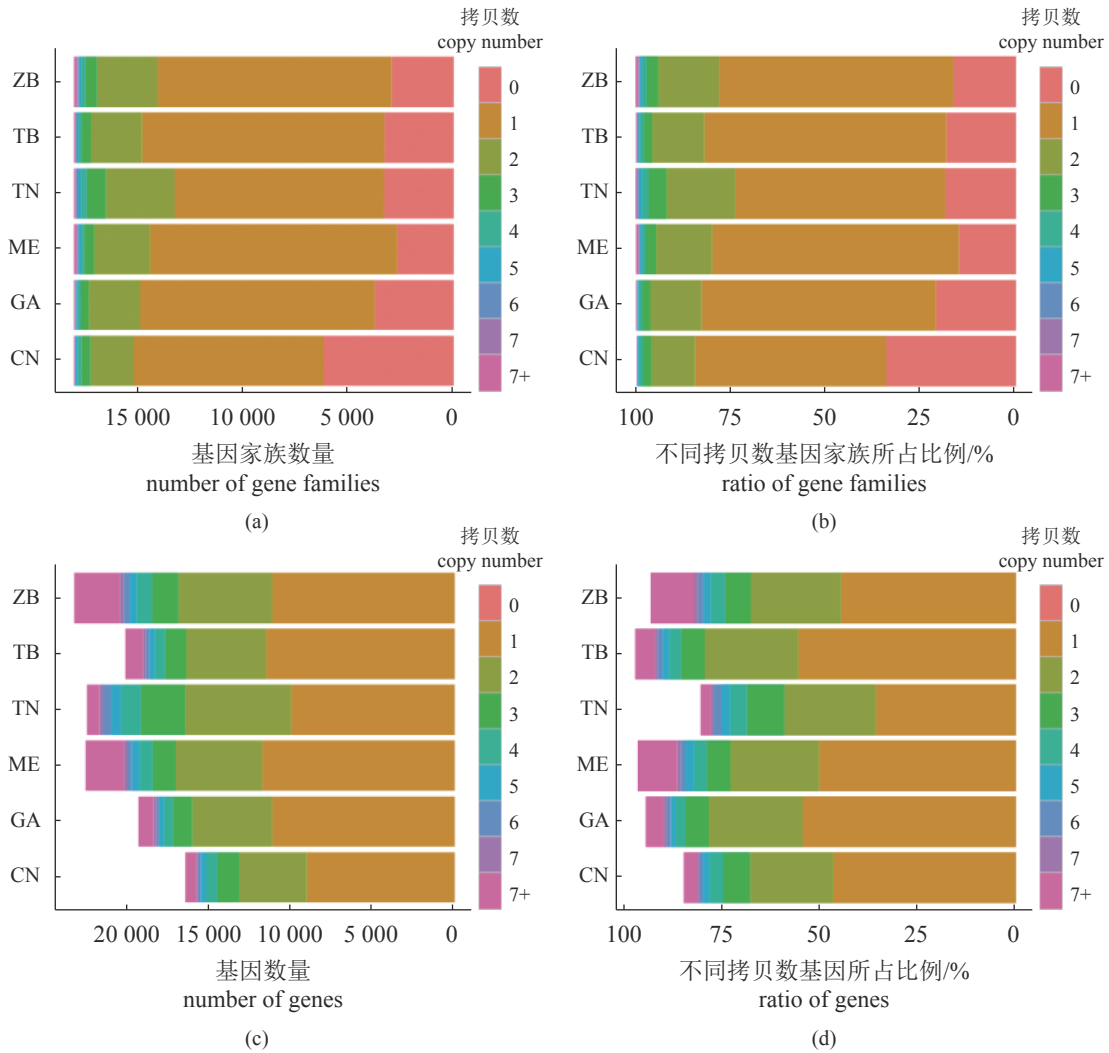


图 1 6 个物种的基因家族统计图

图中不同颜色代表基因拷贝数, CN.刀鲚, ZB.斑马鱼, ME.青鳞, GA.三刺鱼, TR.红鳍东方鲀, TN.绿河鲀, 下同

Fig. 1 Statistics of gene families of six species

Color in the figure represents gene copy number; CN. *C. nasus*, ZB. *D. rerio*, ME. *O. latipes*, GA. *G. aculeatus*, TR. *T. rubripes*, TN. *T. nigroviridis*, the same below

扩张基因家族与显著收缩基因家族数目相近, 其他 5 个物种的显著扩张基因家族数均远高于显著收缩家族。

对显著扩张基因家族的基因进行基因组定位, 23 号染色体扩张基因家族基因数为 70, 显著多于其他染色体具有的扩张基因数 (图 4)。

对扩张基因家族进行注释。COG 分类显示, 转位因子、一般功能预测基因和转录为前三分类 (图 5)。GO 分类显示, 代谢过程、催化活性和结合为前三分类 (图 6)。紧密连接、心肌细胞的肾上腺素能信号和心肌收缩分列 KEGG 分类前三 (表 1)。

2.3 刀鲚独有 150 个基因家族通路分析

对刀鲚独有 150 个基因家族的 419 个基因进

行注释和通路分析。COG 前三分类条目分别为信号转导机制、一般功能预测基因和碳水化合物的运输和代谢。GO 分类显示, 代谢进程、催化活性和细胞进程为前三条目。KEGG 结果中, 富集较多的基因通路分别为神经活性配体-受体相互作用、心肌细胞的肾上腺素能信号、钙信号通路和 MAPK 信号通路。

3 讨论

本研究对刀鲚基因家族进行了初步鉴定, 并对刀鲚显著扩张和收缩基因家族与独有基因家族进行了注释和通路分析。

刀鲚显著扩张基因家族 COG 分类显示转位

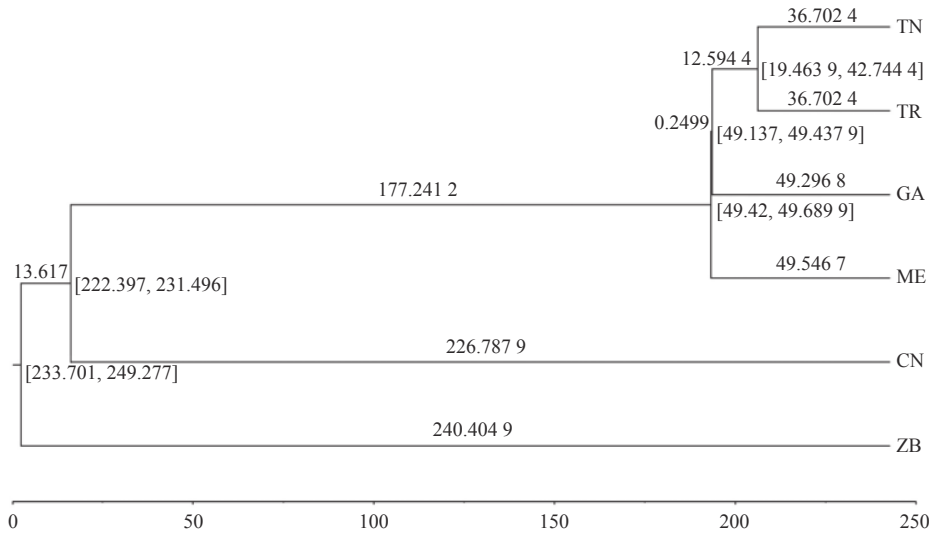


图 2 6 个物种的系统发育树

分歧时间 (Mya) 和 95% HPD 分别标注在分支和分支节点上

Fig. 2 Phylogenetic trees of six species

Divergence time (Mya) and 95% HPD are marked on the branch and branch nodes, respectively

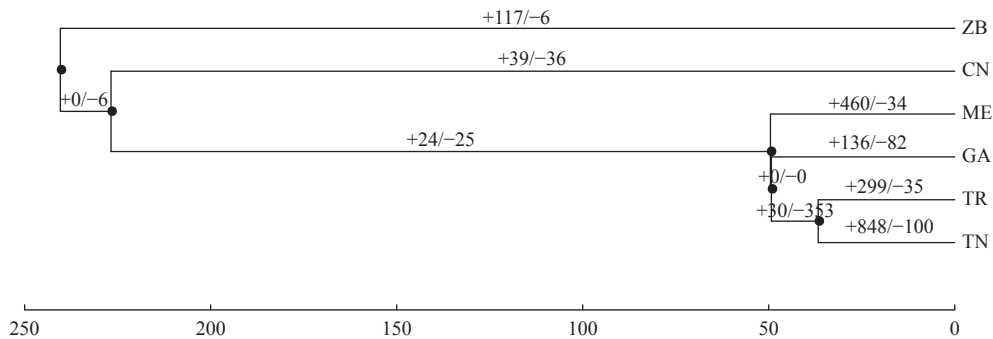


图 3 6 个物种的显著扩张和收缩基因家族

Fig. 3 Significantly expanded and contracted gene families of six species

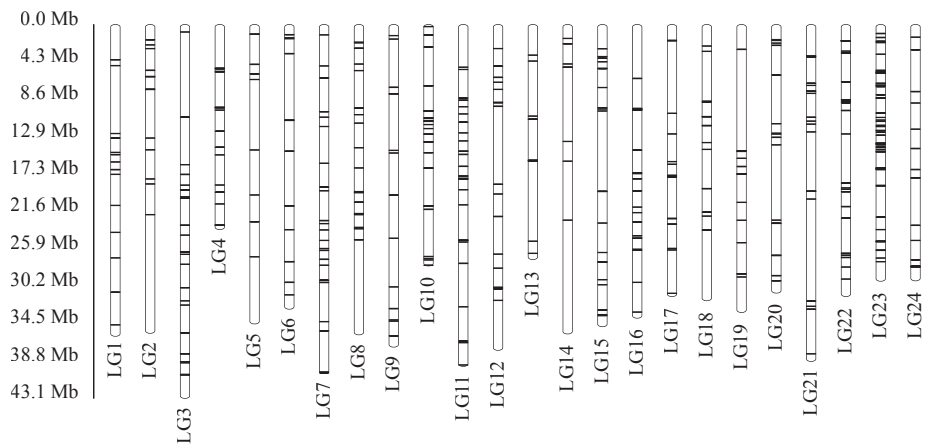


图 4 刀鲚显著扩张基因家族基因的基因组定位

Fig. 4 Locations of significantly expanded gene families on the *C. nasus*'s genome

因子、一般功能预测基因和转录这三个分类条目远高于其他分类。在进化过程中，生物需要遗传

物质变化来为适应环境提供材料。而遗传物质的变化方式主要包括突变和基因重排，这两种方式

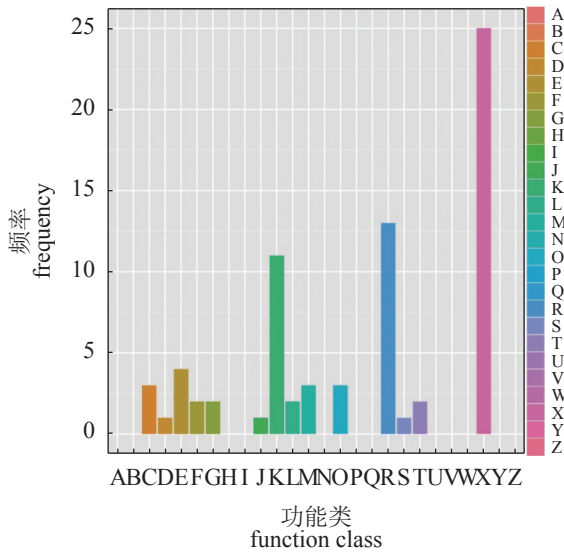


图5 刀鲚显著扩张基因家族基因 COG 分类图

A. RNA 的加工和修饰, B. 染色质结构和动力学, C. 能量生产和转化, D. 细胞周期控制, 细胞分裂, 染色体分割, E. 氨基酸运输与代谢, F. 核苷酸的运输和代谢, G. 碳水化合物运输与代谢, H. 辅酶运输与代谢, I. 脂质运输与代谢, J. 翻译, 核糖体结构和生物发生, K. 转录, L. 复制、重组和修复, M. 细胞壁/膜/包膜生物发生, N. 细胞运动, O. 翻译后修饰、蛋白质转换、分子伴侣, P. 无机离子转运与代谢, Q. 次生代谢物的生物合成、运输和分解代谢, R. 一般功能预测, S. 功能未知, T. 信号转导机制, U. 细胞内运输、分泌和囊泡运输, V. 防御机制, W. 细胞外结构, X. 动员组: 原噬菌体, 转座子, Y. 核结构, Z. 细胞骨架

Fig. 5 COG function class of the significantly expanded gene families' genes of *C. nasus*

A. RNA processing and modification, B. chromatin structure and dynamics, C. energy production and conversion, D. cell cycle control, cell division, chromosome partitioning, E. amino acid transport and metabolism, F. nucleotide transport and metabolism, G. carbohydrate transport and metabolism, H. coenzyme transport and metabolism, I. lipid transport and metabolism, J. translation, ribosomal structure and biogenesis, K. transcription, L. replication, recombination and repair, M. cell wall/membrane/envelope biogenesis, N. cell motility, O. posttranslational modification, protein turnover, chaperones, P. inorganic ion transport and metabolism, Q. secondary metabolites biosynthesis, transport and catabolism, R. general function prediction only, S. function unknown, T. signal transduction mechanisms, U. intracellular trafficking, secretion, and vesicular transport, V. defense mechanisms, W. extracellular structures, X. mobilome: prophage, transposons, Y. nuclear structure, Z. cytoskeleton

均可以引起生物变异或产生新性状。转位因子可以引起基因重排, 与物种的生物特征、种群结构和栖息环境密切相关, 并且其丰富度随着基因组大小和机体复杂度增加而增加<sup>[20]</sup>。Kido 等<sup>[21]</sup>发现原始鲑鳟鱼类经历 *Hpa* I 家族、*Sma* I 家族和 *Fok* I 家族的扩增及散布而形成今天的鳟和鲑。刀鲚具有丰富的转位因子, 表明其保持着高度的遗传多样性, 具有较强的环境适应能力。同时转位

因子也会引起居群产生分化、形态生理变化和生殖隔离。一些陆封型刀鲚群体丢失或不拥有广泛存在于其他群体的转位因子, 或转位因子诱发刀鲚基因组的重新装配等, 这些情况有可能会造成刀鲚群体发生分化, 甚至生殖隔离, 进而产生新的物种。注释到转位因子条目的基因主要为逆转座子基因和转座子整合酶基因。Chen 等<sup>[22]</sup>发现南极鱼类逆转座子在寒冷条件下活化, 使其适应南极寒冷环境。刀鲚的逆转座子与其渗透调节等生物特性的关系值得进一步研究。GO 富集结果表明, 刀鲚显著扩张基因家族主要在代谢过程、催化活性、结合、细胞、高分子配合物、细胞器、膜等条目富集。陈强等<sup>[23]</sup>比较陆封型和洄游型香鱼 (*Plecoglossus altivelis*) 肾脏蛋白质组, 发现陆封型香鱼糖酵解相关酶和应激蛋白表达量较洄游型高, 蛋白代谢相关酶较洄游型低, 陆封型和洄游型香鱼可能在糖代谢和氨基酸代谢方面发生了改变。涂志英等<sup>[24]</sup>研究表明, 洄游过程会对巨须裂腹鱼 (*Schizothorax macropogon*) 代谢产生影响。GO 富集分析条目中代谢过程、催化活性等可能与其洄游适应性相关。KEGG 注释的通路则涉及到紧密连接、心肌细胞的肾上腺素能信号、心肌收缩、细胞黏附分子、MAPK 信号通路、钙信号通路、黑素合成、卵母细胞减数分裂、GnRH 信号通路、三羧酸循环、碳代谢、嗅觉转导等通路。紧密连接、心肌收缩等通路与渗透压调节有关<sup>[25-26]</sup>, 且其通路条目下基因显著多余其他通路条目。其他通路则涉及到信号转导、代谢、性腺发育及嗅觉等。鱼类的嗅觉几乎与所有的行为相关<sup>[27]</sup>, 其在鱼类洄游的研究中也较多<sup>[28-29]</sup>, 刀鲚显著扩张基因家族与嗅觉转导通路相关基因也值得进一步研究。这些结果说明刀鲚为适应洄游产生了基因层面的适应性进化。

对刀鲚显著扩张基因家族基因进行染色体定位, 结果显示, 显著扩张基因家族基因在 23 号染色体的数目最多, 为 70 个; 其次为 22 号染色体, 为 39 个; 其余染色体分布的基因数目为 8~31 个。定位在 23 号染色体的基因数远高于其他染色体, 这 70 个基因主要为组蛋白家族基因。组蛋白通过与染色质结合或被移除, 参与基因转录调控、DNA 复制与修复、细胞周期调控等诸多生命活动<sup>[30]</sup>。而刀鲚 23 号染色体组蛋白家族基因的显著扩张, 表明刀鲚染色质空间结构的调节也可能发生适应性的改变。

对刀鲚特有的 150 个基因家族进行通路分析,

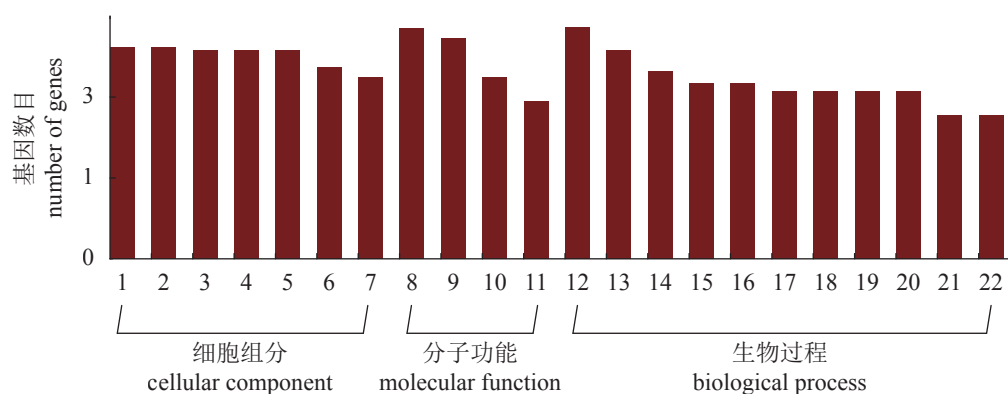


图6 刀鲚显著扩张基因家族基因GO分类图

1.细胞, 2.细胞组分, 3.分子复合物, 4.细胞器, 5.细胞器组分, 6.膜, 7.膜组分, 8.催化活性, 9.结合, 10.受体活性, 11.分子传感器活性, 12.代谢过程, 13.细胞过程, 14.单组织过程, 15.信号转导, 16.细胞成分组织或生物发生, 17.多细胞生物过程, 18.发育过程, 19.刺激反应, 20.生物调节, 21.生物附着, 22.运动

Fig. 6 GO function class of the significantly expanded gene families' genes of *C. nasus*

1. cell, 2. cell part, 3. macromolecular complex, 4. organelle, 5. organelle part, 6. membrane, 7. membrane part, 8. catalytic activity, 9. binding, 10. receptor activity, 11. molecular transducer activity, 12. metabolic process, 13. cellular process, 14. single-organism process, 15. signaling, 16. cellular component organization or biogenesis, 17. multicellular organismal process, 18. developmental process, 19. response to stimulus, 20. biological regulation, 21. biological adhesion, 22. locomotion

表1 刀鲚显著扩张基因家族基因KEGG分类表

Tab. 1 KEGG pathway of the significantly expanded gene families' genes of *C. nasus*

通路 pathway	通路ID pathway ID	基因数目/条 gene number
紧密连接 tight junction	ko04530	25
心肌细胞中的肾上腺素能信号 adrenergic signaling in cardiomyocytes	ko04261	10
心肌收缩 cardiac muscle contraction	ko04260	7
细胞黏附分子 cell adhesion molecules (CAMs)	ko04514	6
MAPK信号通路 MAPK signaling pathway	ko04010	4
钙信号通路 calcium signaling pathway	ko04020	4
黑素合成 melanogenesis	ko04916	4
单纯疱疹病毒感染 herpes simplex infection	ko05168	4
嘌呤代谢 purine metabolism	ko00230	3
ErbB信号通路 ErbB signaling pathway	ko04012	3
卵母细胞减数分裂 oocyte meiosis	ko04114	3
吞噬体 phagosome	ko04145	3
Wnt信号通路 Wnt signaling pathway	ko04310	3
GnRH信号通路 GnRH signaling pathway	ko04912	3
柠檬酸循环 citrate cycle (TCA cycle)	ko00020	2
碳代谢 carbon metabolism	ko01200	2
内吞作用 endocytosis	ko04144	2
病毒性心肌炎 viral myocarditis	ko05416	2

GO注释结果与显著扩张基因家族基因注释结果类似。这表明刀鲚特有基因家族与显著扩张基因家族功能类似,可能刀鲚为了适应生殖洄游等特性,相关基因家族进行了扩张,并产生了新的基因家族。

## 4 结论

实验分别对刀鲚基因家族、其扩张和收缩基因家族进行了鉴定,对刀鲚显著扩张基因家族和特有基因家族进行了通路分析。通路分析结果表

明刀鲚渗透调节、信号转导、代谢、性腺发育和嗅觉等方面均可能产生了适应性改变, 且其染色质空间结构调节也可能发生了改变。

(作者声明本文无实际或潜在的利益冲突)

### 参考文献 (References):

- [ 1 ] Jiang T, Yang J, Lu M J, *et al.* Discovery of a spawning area for anadromous *Coilia nasus* Temminck et Schlegel, 1846 in Poyang Lake, China[J]. *Journal of Applied Ichthyology*, 2017, 33(2): 189-192.
- [ 2 ] 魏广莲, 徐钢春, 顾若波, 等. 刀鲚的生物学及人工养殖研究进展[J]. 长江大学学报(自然科学版), 2012, 9(7): 31-35.  
Wei G L, Xu G C, Gu R B, *et al.* Research progress on the biology and artificial breeding of *Coilia nasus*[J]. *Journal of Yangtze University (Natural Science Edition)*, 2012, 9(7): 31-35 (in Chinese).
- [ 3 ] 袁传宓, 林金榜, 秦安龄, 等. 关于我国鲚属鱼类分类的历史和现状—兼谈改造旧鱼类分类学的几点体会 [J]. 南京大学学报(自然科学版), 1976(2).  
Yuan C F, Lin J B, Qin A F, *et al.* The history and current status of the taxonomy of corilia in China—and some experiences on reforming the old fish taxonomy[J]. *Journal of Nanjing University (Natural Science Edition)*, 1976(2) (in Chinese).
- [ 4 ] Lennox R J, Paukert C P, Aarestrup K, *et al.* One hundred pressing questions on the future of global fish migration science, conservation, and policy[J]. *Frontiers in Ecology and the Environment*, 2019, 7: 286.
- [ 5 ] Bloom D D, Burns M D, Schriever T A. Evolution of body size and trophic position in migratory fishes: a phylogenetic comparative analysis of Clupeiformes (anchovies, herring, shad and allies)[J]. *Biological Journal of the Linnean Society*, 2018, 125(2): 302-314.
- [ 6 ] Kess T, Bentzen P, Lehnert S J, *et al.* Modular chromosome rearrangements reveal parallel and nonparallel adaptation in a marine fish[J]. *Ecology and Evolution*, 2020, 10(2): 638-653.
- [ 7 ] O'Malley K G, Vaux F, Black A N. Characterizing neutral and adaptive genomic differentiation in a changing climate: the most northerly freshwater fish as a model[J]. *Ecology and Evolution*, 2019, 9(4): 2004-2017.
- [ 8 ] Xu G C, Bian C, Nie Z J, *et al.* Genome and population sequencing of a chromosome-level genome assembly of the Chinese tapertail anchovy (*Coilia nasus*) provides novel insights into migratory adaptation[J]. *GigaScience*, 2020, 9(1): giz157.
- [ 9 ] 王晓梅, 朱国利, 唐文乔. 刀鲚嗅觉受体基因MOR-51I2克隆、序列分析及组织表达[J]. *水生生物学报*, 2017, 41(1): 33-42.  
Wang X M, Zhu G L, Tang W Q. Cloning, sequence analysis and tissues expression of *Coilia nasus* olfactory receptor gene MOR-51I2[J]. *Acta Hydrobiologica Sinica*, 2017, 41(1): 33-42 (in Chinese).
- [10] 王聪, 朱国利, 唐文乔, 等. 刀鲚MOR-2AK2的克隆、序列分析及组织表达[J]. *水产学报*, 2016, 40(1): 11-23.  
Wang C, Zhu G L, Tang W Q, *et al.* Cloning, sequence analysis and tissue expression of *Coilia nasus* olfactory receptor gene MOR-2AK2[J]. *Journal of Fisheries of China*, 2016, 40(1): 11-23 (in Chinese).
- [11] 吴利红, 唐文乔, 张亚. 从体内脂肪的转移过程探讨凤鲚和刀鲚溯河产卵洄游距离的差异性[J]. *水产学报*, 2017, 41(2): 212-220.  
Wu L H, Tang W Q, Zhang Y. Research on the differences of anadromous migratory distance between *Coilia mystus* and *C. nasus* based on the transfer process of body lipid[J]. *Journal of Fisheries of China*, 2017, 41(2): 212-220 (in Chinese).
- [12] Ren L P, Shang Y J, Yang L, *et al.* Chromosome-level de novo genome assembly of *Sarcophaga peregrina* provides insights into the evolutionary adaptation of flesh flies[J]. *Molecular Ecology Resources*, 2021, 21(1): 251-262.
- [13] Perteza G, Perteza M. GFF Utilities: GffRead and Gff-Compare [version 2; peer review: 3 approved][J]. *F1000Research*, 2020, 9: 304.
- [14] Emms D M, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy[J]. *Genome biology*, 2015, 16(1): 157.
- [15] Kazutaka K, Standley D M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability[J]. *Molecular Biology and Evolution*, 2013, 30(4): 772-780.
- [16] Gerard T, Jose C. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments[J]. *Systematic Bio-*

- logy, 2007, 56(4): 564-577.
- [17] Nguyen LT, Schmidt H A, Von Haeseler A, *et al.* IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies[J]. *Molecular biology and evolution*, 2015, 32(1): 268-74.
- [18] De B, Tijn, *et al.* CAFE: a computational tool for the study of gene family evolution[J]. *Bioinformatics*, 2006, 22(10): 1269-1271.
- [19] Chen C J, Chen H, Zhang Y, *et al.* TBtools: an integrative toolkit developed for interactive analyses of big biological data[J]. *Molecular Plant*, 2020, 13(8): 1194-1202.
- [20] 李树深. 转位因子和物种形成[J]. *动物学研究*, 1999, 20(5): 385-390.
- Li S S. Transposable element and the speciation[J]. *Zoological Research*, 1999, 20(5): 385-390 (in Chinese).
- [21] Kido Y, Aono M, Yamaki T, *et al.* Shaping and reshaping of salmonid genomes by amplification of tRNA-derived retrotransposons during evolution[J]. *Proceedings of the National Academy of Sciences*, 1991, 88(6): 2326-2330.
- [22] 许强华, 吴智超, 陈良标. 南极鱼类多样性和适应性进化研究进展[J]. *生物多样性*, 2014, 22(1): 80-87.
- Xu Q H, Wu Z C, Chen L B. Biodiversity and adaptive evolution of Antarctic notothenioid fishes[J]. *Biodiversity science*, 2014, 22(1): 80-87 (in Chinese).
- [23] 陈强, 陆新江, 黄左安, 等. 陆封型和洄游型香鱼体肾组织的蛋白质组学分析[J]. *生物学杂志*, 2014, 31(2): 13-17.
- Chen Q, Lu X J, Huang Z A, *et al.* The proteome analysis of the trunk kidney between land-locked type and amphidromous type ayu (*Plecoglossus altivelis*)[J]. *Journal of Biology*, 2014, 31(2): 13-17 (in Chinese).
- [24] 涂志英, 袁喜, 王从锋, 等. 亚成体巨须裂腹鱼游泳能力及活动代谢研究[J]. *水生生物学报*, 2012, 36(4): 682-688.
- Tu Z Y, Yuan X, Wang C F, *et al.* Swimming capability and activity metabolism of subadult schizothorax macropogon[J]. *Acta Hydrobiologica Sinica*, 2012, 36(4): 682-688 (in Chinese).
- [25] Jia Y H, Yin S W, Li L, *et al.* iTRAQ proteomic analysis of salinity acclimation proteins in the gill of tropical marbled eel (*Anguilla marmorata*)[J]. *Fish Physiology and Biochemistry*, 2016, 42(3): 935-946.
- [26] Guo B Y, Tang Z R, Wu C G, *et al.* Transcriptomic analysis reveal an efficient osmoregulatory system in Siberian sturgeon *Acipenser baeri* in response to salinity stress[J]. *Scientific reports*, 2018, 8: 14353.
- [27] 魏凯, 陈春山, 张旭光, 等. 鱼类嗅觉系统功能与行为生态研究进展[J]. *海洋渔业*, 2017, 39(6): 723-730.
- Wei K, Chen C S, Zhang X G, *et al.* Review on olfactory organ and behaviour in fish[J]. *Marine Fisheries*, 2017, 39(6): 723-730 (in Chinese).
- [28] Dittman A, Quinn T. Homing in Pacific salmon: mechanisms and ecological basis[J]. *The Journal of Experimental Biology*, 1996, 199(Pt 1): 83-91.
- [29] Hino H, Iwai T, Yamashita M, *et al.* Identification of an olfactory imprinting-related gene in the lacustrine sockeye salmon, *Oncorhynchus nerka*[J]. *Aquaculture*, 2007, 273(2-3): 200-208.
- [30] 张艳君, 李炳杰. 组蛋白H2AX研究进展及其与肿瘤的关系[J]. *医药前沿*, 2013(16): 312.
- Zhang Y J, Li B J. Research progress of histone H2AX and its relationship with tumor[J]. *Medical Frontier*, 2013(16): 312 (in Chinese).



## Identification and analysis of gene families expansion and contraction in *Coilia nasus*

ZHANG Jinpeng<sup>1,2</sup>, GAO Shufang<sup>1,2</sup>, SHI Yonghai<sup>3</sup>, YAN Yinlong<sup>3</sup>, QI Ming<sup>1,2</sup>, LIU Qigen<sup>1,2\*</sup>

(1. National Demonstration Center for Experimental Fisheries Science Education,  
Shanghai Ocean University, Shanghai 201306, China;

2. Centre for Research on Environmental Ecology and Fish Nutrition of the Ministry of Agriculture and  
Rural Affairs, Shanghai Ocean University, Shanghai 201306, China;

3. Shanghai Fisheries Research Institute, Shanghai Fisheries Technology Extension Station, Shanghai 200433, China)

**Abstract:** *Coilia nasus* belongs to genus *Coilia*, family Engraulidae, order Clupeiformes, which is migratory fish with high nutritional and economic value in China. *C. nasus* has both migratory and land-locked populations, which is an excellent model for studying reproductive migration and adaptive evolution. This study identified and analyzed the expansion and contraction of gene families in *C. nasus*, in order to explore the special mechanism of *C. nasus* caused by changes in the core gene families in the evolutionary process. Comparative genomics was used to identify the *C. nasus*'s gene families by using the genomes of *Danio rerio*, *Gasterosteus aculeatus*, *Oryzias latipes*, *Takifugu rubripes* and *Tetraodon nigroviridis*. Then CAFE V4.2 software was used for gene family expansion and contraction analysis. Finally, GO, KEGG and other database were applied for annotations and pathway analysis of the significantly expanded gene families' genes. 11 872 gene families were identified in *C. nasus*, including 16 470 genes. Besides, 2 963 unclustered genes were identified. Compared with the other 5 species, 150 gene families containing 419 genes were found unique to *C. nasus*. In addition, 1 200 gene families of *C. nasus* expanded and 7 543 gene families contracted. *P* values of 39 expanded gene families (including 508 genes) and 36 contracted gene families (including 21 genes) are less than 0.05. 508 genes of significantly expanded gene families were utilized for pathway analysis. Metabolic process, catalytic activity, cell and cell part are the top GO terms. Tight junctions, adrenergic signaling in cardiomyocytes, cardiac muscle contraction, cell adhesion molecules (CAMs), etc. are the top KEGG pathways. Tight junctions and cardiac muscle contraction pathways are related to osmoregulation. GnRH signaling pathway and olfactory transduction pathway are involved in reproductive migration behavior. Furthermore, we mapped 508 genes of significantly expansion gene families on the genome, and found that chromosome 23 had much more genes than other chromosomes. And the genes of significantly expanded gene families located on chromosome 23 were mostly histone genes, which were related to the regulation of chromatin structure. Pathway analysis of the expansion and exclusive gene families of *C. nasus* indicate that the osmotic regulation, signal transduction, metabolism, gonad development and olfactory transduction may have undergone adaptive evolution, and the regulation of its chromatin structure may also have been altered. These results have strengthened our understanding of the adaptive evolution of *C. nasus*.

**Key words:** *Coilia nasus*; gene family; expansion; shrink; pathway analysis

**Corresponding author:** LIU Qigen. E-mail: qgliu@shou.edu.cn

**Funding projects:** Project of Centre for Research on Environmental Ecology and Fish Nutrition (CREEFN) of the Ministry of Agriculture and Rural Affairs, Shanghai Ocean University (A1-3201-19-300204); Project of Science and Technology Commission Shanghai Municipality (17391900300)