

doi: 10.7541/2021.2019.208

基于转录组的池蝶蚌系统发育及其遗传分析

周伟 张万昌 周叶 王小敏 胡蓓娟 王军花 洪一江

(南昌大学生命科学学院, 江西省水产动物资源与利用重点实验室, 南昌 330031)

摘要: 池蝶蚌(*Hyriopsis schlegelii*)和三角帆蚌(*Hyriopsis cumingii*)的亲缘关系一直存在争议, 为了更深入地对淡水珠蚌进行系统发育分析, 研究对池蝶蚌及其3个近缘物种三角帆蚌、褶纹冠蚌(*Cristaria plicata*)和背角无齿蚌(*Anodonta woodiana*)的转录组进行了比较分析。对它们共有单拷贝同源基因进行多序列联配后计算这4种淡水珠蚌之间的遗传距离, 结果表明池蝶蚌和三角帆蚌的遗传距离只有0.00746(小于1%), 而其他淡水珠蚌之间的遗传距离几乎是它的10倍(平均大于6%)。系统发育分析表明, 小方蚌亚科与无齿蚌亚科大约在5144万年前发生分歧, 而池蝶蚌和三角帆蚌发生分化的时间大约在424万年前。进一步的遗传差异分析鉴定了池蝶蚌转录组中的特有基因, 其中KEGG 通路富集的结果表明这些基因主要富集在免疫细胞膜分子和蛋白消化吸收等通路上。同时, GO注释结果表明有很多特有基因与池蝶蚌的优良性状相关, 如与发育相关的生长发育、系统发育和动物器官发育等生物学过程, 及与免疫相关的免疫系统过程、抗原处理和呈递等生物学过程。以上结果表明, 池蝶蚌和三角帆蚌具有更近的亲缘关系, 它们可能是同一物种的不同亚种或不同种群, 这对淡水珠蚌种质资源的保护和遗传育种具有重要参考意义。此外, 相比于三角帆蚌, 池蝶蚌可能具有一些与生长发育及免疫相关的特有基因, 这为探究相关的分子机制提供了线索。

关键词: 池蝶蚌; 三角帆蚌; 转录组; 系统发育; 遗传差异

中图分类号: Q349⁺.13

文献标识码: A

文章编号: 1000-3207(2021)03-0514-09



池蝶蚌(*Hyriopsis schlegelii*)和三角帆蚌(*Hyriopsis cumingii*)是目前在我国养殖的优质淡水珍珠育珠蚌, 它们具有相似的外形和生理功能, 但人们在分类学上对它们有不同的认识。有的认为它们是同属不同种^[1], 有的认为是同种不同亚种^[2]。随着分子生物学和生物化学的发展, 有研究者对池蝶蚌和三角帆蚌的同工酶和*ITS1*基因等进行了分析, 这些研究表明池蝶蚌和三角帆蚌的亲缘关系很近, 达到了种内分化的水平^[2, 3]。此外, 已有的研究表明: 相比于三角帆蚌, 池蝶蚌在生长速度、育珠性能和抗病性等方面更具优势^[4, 5]。然而, 鲜有基于基因组或转录组对二者进行比较分析的报道。

虽然同工酶和*ITS1*基因等分析表明池蝶蚌和三角帆蚌具有更近的亲缘关系, 但是它们的种间关

系还存在争议, 因为这些传统的分子进化研究方法几乎都是选取一个或多个大分子的序列为标准以研究其在各物种同源序列之间的差异, 如16S rRNA和同工酶等。然而, 一个物种的基因编码了数以万计的序列, 以其中一个序列的差异代表整个生物体的差异是不全面的。因此, 有必要从生物体基因组或转录组水平来探讨池蝶蚌和三角帆蚌的亲缘关系。此外, 池蝶蚌生长、育珠和抗病等优势性能相关的分子机制也尚不清楚。然而, 池蝶蚌和三角帆蚌的基因组和转录组信息是未知的。

比较转录组学是对不同物种或近缘物种的转录组进行比较分析。在缺乏基因组信息时, 比较转录组学是分析近缘物种系统发育和同源基因家族的有力手段, 基于比较转录组学的分析主要应用于

收稿日期: 2019-09-24; **修订日期:** 2020-03-10

基金项目: 国家贝类产业技术体系淡水珍珠贝养殖与育珠项目(CARS-49); 国家自然科学基金(31660337); 国家重点研发计划“蓝色粮仓科技创新”重点专项(2018YFD0901400)资助 [Supported by the National Shellfish Industrial Technology System Freshwater Pearl Oyster Culture and Cultivate Pearl Post (CARS-49); the National Natural Science Foundation of China (31660337); the Chinese Ministry of Science and Technology through the National Key Research and Development Program of China (2018YFD0901400)]

作者简介: 周伟(1991—), 男, 硕士研究生; 主要研究方向为水产遗传育种。E-mail: 13657093870@163.com

通信作者: 洪一江, 教授; E-mail: yijianghong@126.com

近缘物种的进化关系和一些动植物适应性进化相关的研究^[6, 7]。在本研究中, 我们对包括池螺蚌和三角帆蚌在内的4种淡水珠蚌转录组进行了二代测序(Next Generation Sequencing, NGS)和*de novo*组装。在此基础上, 通过比较转录组学的方法对池螺蚌及其近缘物种进行了系统发育分析, 这旨在更深入地理解池螺蚌和三角帆蚌的亲缘关系。随后, 我们对池螺蚌中的特有基因进行了富集分析, 这有助于我们进一步研究池螺蚌在生长发育、育珠和抗病性等方面的分子机制。

1 材料与方法

1.1 样品收集、总RNA提取和转录组测序

池螺蚌取自江西省抚州市池螺蚌良种场, 三角帆蚌取自浙江省金华三角帆蚌良种场, 褶纹冠蚌(*Cristaria plicata*)和背角无齿蚌(*Anodonta woodiana*)分别采集于江西鄱阳湖和山东省微山湖。对这4种蚌每种随机选取健康的三龄蚌10只, 置于充分曝气的水族箱内暂养一周, 水温控制在18—25℃。

从水族箱中分别取出4种淡水珠蚌, 每种蚌为一组, 每组各3只。分别取每组每只蚌的多种组织(包括闭壳肌、心脏、肾脏、鳃、斧足、性腺、肝脏、肠胃和外套膜)各100 mg, 并分别将取出的三份相同组织混合。随后采用Trizol试剂法提取4种淡水珠蚌的总RNA。以每种蚌各组织提取到的总RNA浓度为依据, 取等量的各组织总RNA进行混合, 最后分别得到4种淡水珠蚌混合组织的总RNA样品。本研究采用的建库类型为真核链特异性文库(250—300 bp), 之后在Illumina HiSeqXten平台上进行双端测序。

1.2 *De novo*组装和功能注释

用Trimmomatic^[8]移除reads中的Illumina测序接头和引物序列, 去除质量过低的序列(Phred数值小于20), 过滤过短(序列长度小于50 bp)和不合格的序列。采用Trinity软件^[9]分别对4种淡水珠蚌转录组测序数据进行*de novo*组装, 所有参数设置为默认参数。首先用CD-HIT软件去除冗余转录本序列^[10], 再用Corset软件进行分层聚类^[11], 将分层聚类后得到的转录本序列进行后续的分析。对转录本序列进行统计并对其拼接质量进行评估。

首先用TransDecoder软件(Find Coding Region Within Transcripts, <https://transdecoder.github.io/>)对转录本序列进行ORF框预测并翻译成蛋白序列, 然后用EggNOG-mapper将4种淡水珠蚌的蛋白序列分别比对到EggNOG数据库, 依据同源序列的相似性进行基因功能注释^[12—14]。

1.3 直系同源基因鉴定和基因家族聚类

用Orthofinder对4种淡水育珠蚌的蛋白序列进

行同源基因家族分析^[15], 参数设置使用Diamond^[14]以默认的 E value值 $1e-3$ 进行all-vs-all搜索, 寻找潜在的同源基因。使用马尔可夫聚类算法(Markov Cluster Algorithm, MCL)对基因进行聚类, 划分直系同源簇。提取4种淡水珠蚌中各自特有的基因家族, 结合之前得到的注释信息获取4种淡水珠蚌特有基因对应的功能。基因家族聚类的结果可以直接从orthofinder分析结果中获取。

1.4 系统发育分析

对4种蚌的单拷贝同源基因进行多序列联配(Multiple sequence alignment, MSA), 在MEGA软件中采用pairwise uncorrected p-distance模型对多序列联配得到的长片段序列的遗传距离进行计算^[16]。从Uniprot数据库(<https://www.uniprot.org/proteomes/>)下载光滑双脐螺(*Biomphalaria glabrata*)、太平洋牡蛎(*Crassostrea gigas*)、霸王莲花青螺(*Lotia gigantea*)、虾夷扇贝(*Mizuhopecten yessoensis*)和加利福尼亚双斑章鱼(*Octopus bimaculoides*)这5个物种的蛋白质组数据, 结合本研究中的4种淡水珠蚌蛋白序列数据共9个贝类物种以供进一步的分析。随后利用Orthofinder对这9个物种蛋白质组序列筛选出的1317个共有单拷贝同源基因重建了贝类的系统发育。默认使用mafft^[17, 18]进行多序列联配, 用fasttree推断系统发生树^[19]。

通过生命进化时间的公共信息库TimeTree网站(<http://www.timetree.org/>)查找到霸王莲花青螺与光滑双脐螺分歧时间大约在471(423—519)百万年前(Millions of years ago, Mya), 太平洋牡蛎与虾夷扇贝分歧时间大约在455(421—497)百万年前。以上述2个已知的分歧时间作为固定值和校正值, 根据上一步得到的系统发生树用最大似然法(r8s软件)预测它们的分歧时间^[20]。

1.5 池螺蚌特有基因分析

首先用AnnotationForge^[21]依据池螺蚌功能注释信息构建OrgDb数据包, OrgDb包用于存储不同数据库基因ID之间的对应关系及基因与GO、KEGG等注释的相对关系。根据同源基因家族的分析结果提取池螺蚌中特有基因, 随后基于本研究构建的池螺蚌OrgDb数据包用clusterprofiler^[22]对池螺蚌中特有基因进行KEGG和GO富集分析。

2 结果

2.1 转录组组装和注释

通过Illumina测序获得了池螺蚌、三角帆蚌、背角无齿蚌和褶纹冠蚌的转录组数据, 质控后分别得到12.97、13.58、13.41和11.36 Gb转录组测序数据, 各含有clean reads约4323万、4525万、4471万和

3785万条。对这4种蚌的转录组数据进行了 $de\ novo$ 组装, 其中池螺蚌转录组组装后得到138948个转录本和104803个Unigenes, reads回贴率(reads回贴到转录本的比率)达85.62%, Unigenes的平均长度为1373.20 bp, N50长度为1928 bp(表 1)。以后生动物单拷贝同源基因集作为比对数据库, 使用Busco软件对转录组完整性进行分析, 结果表明: 在4种淡水珠蚌的转录本中, 超过98%都是完整的, 只有不到2%的转录本是部分比对或比对不上, 这说明我们组装结果完整度较好, 转录本序列可用于后续分析(表 2)。

将得到的池螺蚌、三角帆蚌、背角无齿蚌和褶纹冠蚌基因翻译成蛋白序列后比对到EggNOG数据库, 基于相似性($E\text{-value}<10^{-3}$), 这4种蚌分别有34143、28180、28609和24778个基因得到了注释, 在蛋白质直系同源簇(Clusters of orthologous groups

of proteins, COG)分类统计表明, 这4种育珠蚌的COG类别分布大致相同, 其中, 富集基因最多的2个分类是“功能未知(S. Function unknown)”和“信号转导机制(T. Signal transduction mechanisms)”(图 1)。

表 1 转录组组装数据统计

Tab. 1 The statistic of transcriptome assembled data				
指标Index	池螺蚌 <i>H. schlegelii</i>	三角帆蚌 <i>H. cumingii</i>	背角无齿蚌 <i>A. woodiana</i>	褶纹冠蚌 <i>C. plicata</i>
转录本数 Number of transcripts	138948	135564	126601	100731
基因数 Number of unigenes	104803	102380	89663	75504
平均长度 Average length (bp)	1373.20	1371.93	1395.53	1526.87
N50 (bp)	1928	1965	1992	2178
回贴率 Mapping rate (%)	85.62	86.45	87.01	88.36

表 2 组装转录组的BUSCO分析结果

物种Species	同源基因Orthologs(完整性Complete)			片段同源数Fragmented	丢失数Missing
	总数Total	单拷贝数Single-copy	多拷贝数Duplicated		
池螺蚌 <i>H. schlegelii</i>	960(98.2%)	705(72.1%)	255(26.1%)	9(0.9%)	9(0.9%)
三角帆蚌 <i>H. cumingii</i>	965(98.7%)	738(75.5%)	227(23.2%)	9(0.9%)	4(0.4%)
背角无齿蚌 <i>A. woodiana</i>	969(99.1%)	701(71.7%)	268(27.4%)	3(0.3%)	6(0.6%)
褶纹冠蚌 <i>C. plicata</i>	968(99.0%)	798(81.6%)	170(17.4%)	4(0.4%)	6(0.6%)

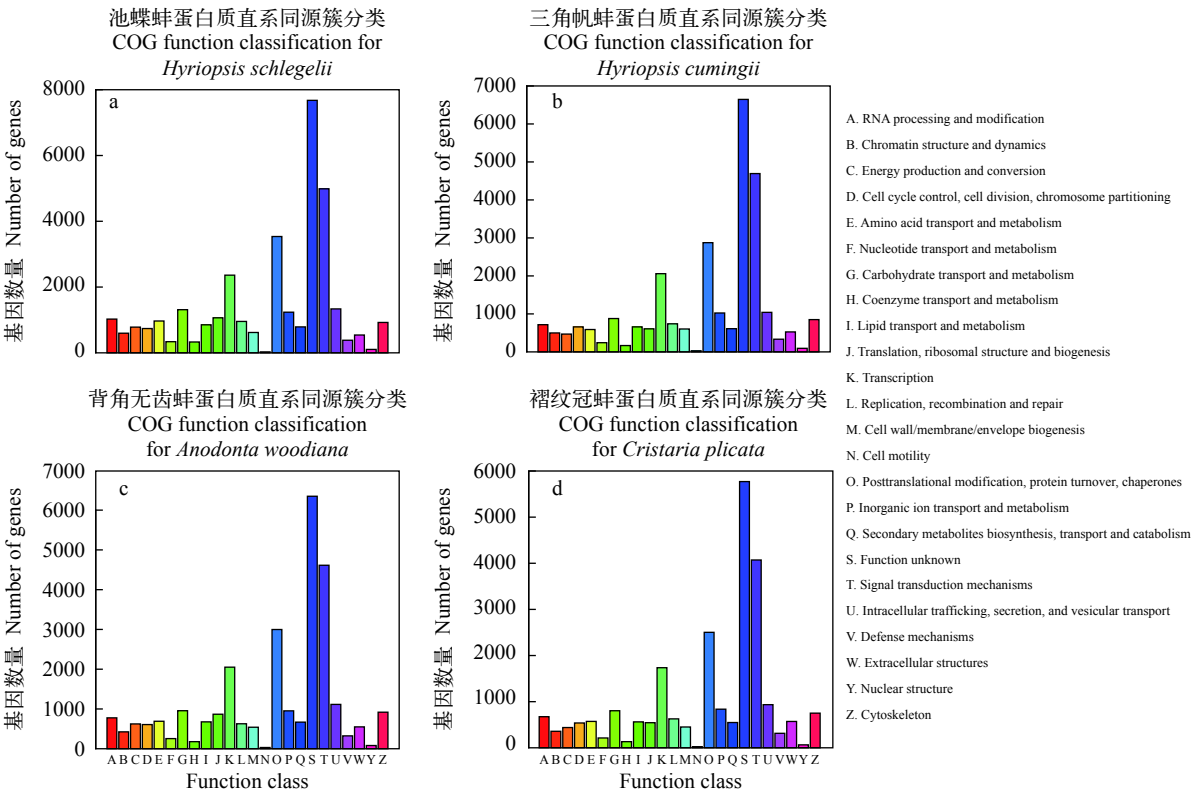


图 1 池螺蚌(a)、三角帆蚌(b)、背角无齿蚌(c)和褶纹冠蚌(d)的蛋白质直系同源簇(COG)分类

Fig. 1 Clusters of Orthologous Group of proteins (COG) classifications for *H. schlegelii* (a), *H. cumingii* (b), *A. woodiana* (c) and *C. plicata* (d)

2.2 直系同源基因的鉴定及基因家族聚类

通过Orthofinder在4种淡水珠蚌中共鉴定到29468个潜在的同源基因家族, 其中4个物种共享的16001个同源基因家族构成了淡水珠蚌核心基因集, 这些核心基因参与广泛的基础活动, 几乎在每一个GO功能中都有分布。只存在于池螺蚌和三角帆蚌中的同源基因家族有5175个(图 2)。另外, 池螺蚌、三角帆蚌、褶纹冠蚌和背角无齿蚌4种育珠蚌间的各自特有的基因家族数量分别为12、11、9和15个。注释结果表明, 池螺蚌特有基因主要与生长发育(*TRIM45*、*TRIM33*和*JAG1*等)和免疫(*ABCB9*)等过程相关, 三角帆蚌的特有基因与抗肿瘤(*CUZD1*)和甾醇代谢(*HDLBP*)等功能有关, 褶纹冠蚌特有基因主要与离子平衡(*SLC12A2*)和免疫(*CEACAM1*)功能相关, 背角无齿蚌特有基因主要与疾病(*TMPRSS12*和*CLCA2*)、细胞内通讯和信号传递(*FLNB*)等功能相关(表 3)。

2.3 系统发育分析

我们对池螺蚌、三角帆蚌、背角无齿蚌和褶纹冠蚌的单拷贝同源基因进行了多序列联配, 采用pairwise uncorrected p-distance模型对多序列联配得到的长片段序列计算遗传距离($E\text{-value}<1e^{-3}$), 结果表明池螺蚌和三角帆蚌的遗传距离小于1%, 而池螺蚌与其他物种的遗传距离几乎是它与三角帆蚌的10倍(表 4)。结合Uniprot数据库中已发布的光滑

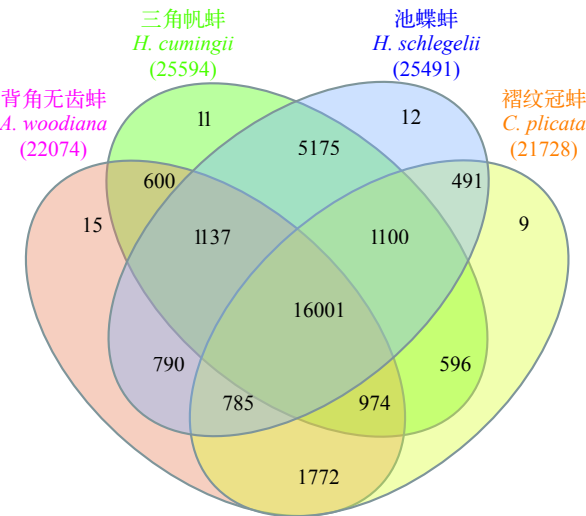


图 2 池螺蚌、三角帆蚌、褶纹冠蚌和背角无齿蚌中的基因家族聚类韦恩图

Fig. 2 Gene families clustering venn diagram in *H. schlegelii*, *H. cumingii*, *C. plicata* and *A. woodiana*

韦恩图中显示的数字是由orthofinder在4种淡水珠蚌鉴定的同源基因家族数量

The venn diagram showing the number of orthologs identified in the four freshwater mussels

双脐螺、太平洋牡蛎和虾夷扇贝等5个物种的蛋白质组数据和本研究获得的4种淡水珠蚌的Uni-genes对应的蛋白质组数据, 我们筛选了包含1317个共有单拷贝直系同源基因家族的蛋白集, 随后利用

表 3 四种淡水珠蚌中物种特有基因列表

Tab. 3 The list of species-specific genes in four freshwater mussels

基因名称 Symbol	物种Species	描述Description
<i>Fndc7</i>	<i>Hyriopsis schlegelii</i>	Fibronectin type III domain containing 7
<i>TRIM45</i>	<i>Hyriopsis schlegelii</i>	Tripartite motif containing 45
<i>ZNF724P</i>	<i>Hyriopsis schlegelii</i>	Zinc finger protein 724
<i>ASTL</i>	<i>Hyriopsis schlegelii</i>	Astacin-like metalloendopeptidase
<i>TRIM33</i>	<i>Hyriopsis schlegelii</i>	Tripartite motif containing 33
<i>JAG1</i>	<i>Hyriopsis schlegelii</i>	Notch ligand involved in the mediation of Notch signaling
<i>ACS</i>	<i>Hyriopsis schlegelii</i>	Acyl-CoA synthetase
<i>ABCB9</i>	<i>Hyriopsis schlegelii</i>	ATP-binding cassette, sub-family B
<i>CUZD1</i>	<i>Hyriopsis cumingii</i>	CUB and zona pellucida-like domains 1
<i>ZFP112H</i>	<i>Hyriopsis cumingii</i>	Zinc finger protein 112 homolog
<i>MUC</i>	<i>Hyriopsis cumingii</i>	Oligomeric mucus gel-forming
<i>HDLBP</i>	<i>Hyriopsis cumingii</i>	High density lipoprotein binding protein
<i>SMPDL3B</i>	<i>Hyriopsis cumingii</i>	Sphingomyelin phosphodiesterase acid-like
<i>AAP1</i>	<i>Cristaria plicata</i>	Amino acid permease
<i>SLC12A2</i>	<i>Cristaria plicata</i>	Solute carrier family 12, member 2
<i>CEACAM1</i>	<i>Cristaria plicata</i>	Carcinoembryonic antigen-related cell adhesion molecule
<i>TMPRSS12</i>	<i>Anodonta woodiana</i>	transmembrane (C-terminal) protease, serine 12
<i>MAMDC4</i>	<i>Anodonta woodiana</i>	MAM domain containing 4
<i>CLCA2</i>	<i>Anodonta woodiana</i>	chloride channel
<i>FLNB</i>	<i>Anodonta woodiana</i>	Filamin-B, bata
<i>TXN</i>	<i>Anodonta woodiana</i>	Thioredoxin

表 4 基于4种淡水珠蚌单拷贝同源基因序列的遗传分化距离

Tab. 4 Genetic differentiation distance based on single copy ortholog sequences of four freshwater mussel

遗传距离 Genetic distance	池螺蚌 <i>H. schlegelii</i>	三角帆蚌 <i>H. cumingii</i>	背角无齿蚌 <i>A. woodiana</i>	褶纹冠蚌 <i>C. plicata</i>
池螺蚌 <i>H. schlegelii</i>		0.00746	0.08912	0.07975
三角帆蚌 <i>H. cumingii</i>	0.00746		0.08964	0.08032
背角无齿蚌 <i>A. woodiana</i>	0.08912	0.08964		0.03449
褶纹冠蚌 <i>C. plicata</i>	0.07975	0.08032	0.03449	

这些单拷贝同源蛋白集重建了贝类的系统发育。结果表明,池螺蚌与三角帆蚌的亲缘关系最近,2个物种聚为一支(帆蚌属)后再与冠蚌属、无齿蚌属聚在一起,瓣鳃纲的真瓣鳃目与异柱目聚为一支(瓣鳃纲),然后它们再与腹足纲、头足纲聚在一起(图3)。随后,我们对这9个物种进行了分歧时间评估,结果表明,池螺蚌与三角帆蚌大约在424万年前发生分歧,帆蚌属与冠蚌属、无齿蚌属发生分歧的时间大约在5144万年前,真瓣鳃目与异柱目大约在5.72亿年前发生分歧,瓣鳃纲、腹足纲和头足纲大约在6.18亿年前发生分歧(图4)。此外,我们还对它们的同源基因家族扩张和收缩情况进行了分析,发现池螺蚌在淡水珠蚌中发生了最多的同源基因家族扩张。

2.4 池螺蚌相对于三角帆蚌的特有基因

通过比较池螺蚌和三角帆蚌的转录组,得到了池螺蚌转录组中相对于三角帆蚌特异性表达的基

因3721条,其中345个特有基因得到了注释。这些特有基因可能与池螺蚌的生长发育、育珠及抗病能力相关。我们对这些特有基因进行了KEGG和GO富集分析。KEGG 通路富集分析结果表明,池螺蚌特有基因主要富集在免疫细胞膜分子(CD molecules)和蛋白消化吸收(Protein digestion and absorption)等通路上(图5)。此外,池螺蚌特有基因GO注释结果表明:有很多特有基因与池螺蚌的优良性状相关,GO类别包括生长发育过程(91个),系统发育(69个),动物器官发育(51个),免疫系统过程(25个),抗原的处理和呈递(4个,图6)。

3 讨论

3.1 池螺蚌与三角帆蚌的亲缘关系

池螺蚌和三角帆蚌的亲缘关系一直存在争议,这是由于我国的淡水双壳类蚌科分类系统主要依

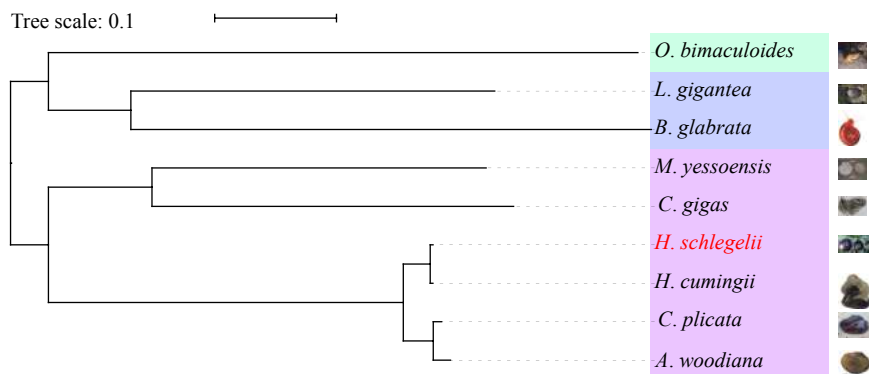


图3 九种贝类动物的系统发育树

Fig. 3 Phylogenetic tree of nine species in mollusca

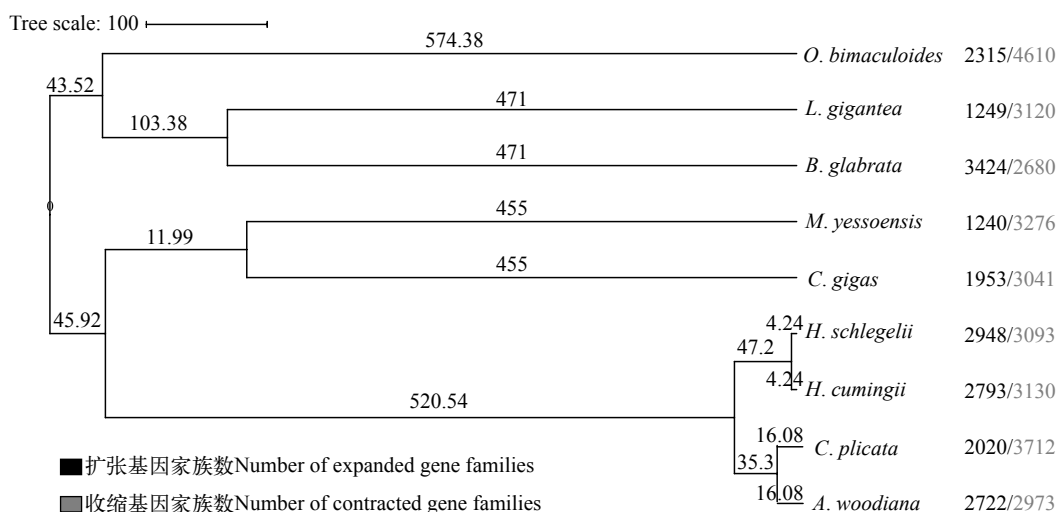


图4 九种贝类动物的分歧时间及其基因家族扩张/收缩情况

Fig. 4 Time of divergence of nine species in mollusca and their expansion/contraction of gene families

分枝上的数字代表分歧时间(95%置信区间), 单位为Mya。物种名称后面的黑色数字代表扩张基因家族, 灰色数字代表收缩基因家族。The number on the branch represents the divergence time (the 95% confidence interval), and the unit is Mya. The black number after the species name represents the expanded gene families, and the grey number represents the contracted gene families

据传统的形态学分类, 而蚌科动物存在趋同现象且贝壳具有高度可塑性, 因此传统的形态学分类并不能完全反映各类群之间的亲缘关系。近年来, 分子

生物学技术常用于解决一些分类学上的疑难问题。在淡水珠蚌中的分子进化研究方法, 大多采用某个大分子的序列为标准, 通过比较相应物种的同

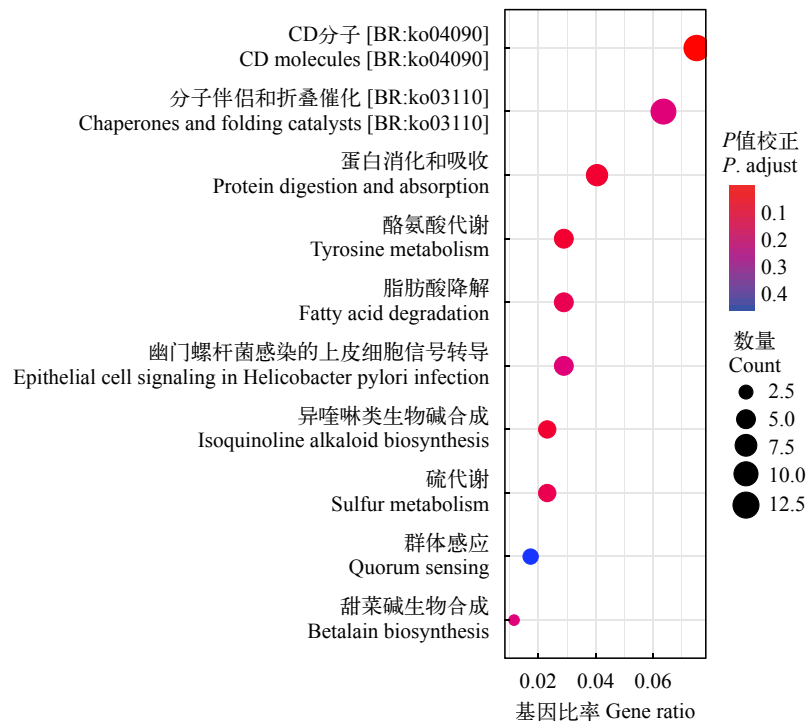


图 5 池螺蚌特有基因的KEGG富集分析

Fig. 5 KEGG enrichment analysis of species-specific genes in *H. schlegelii*

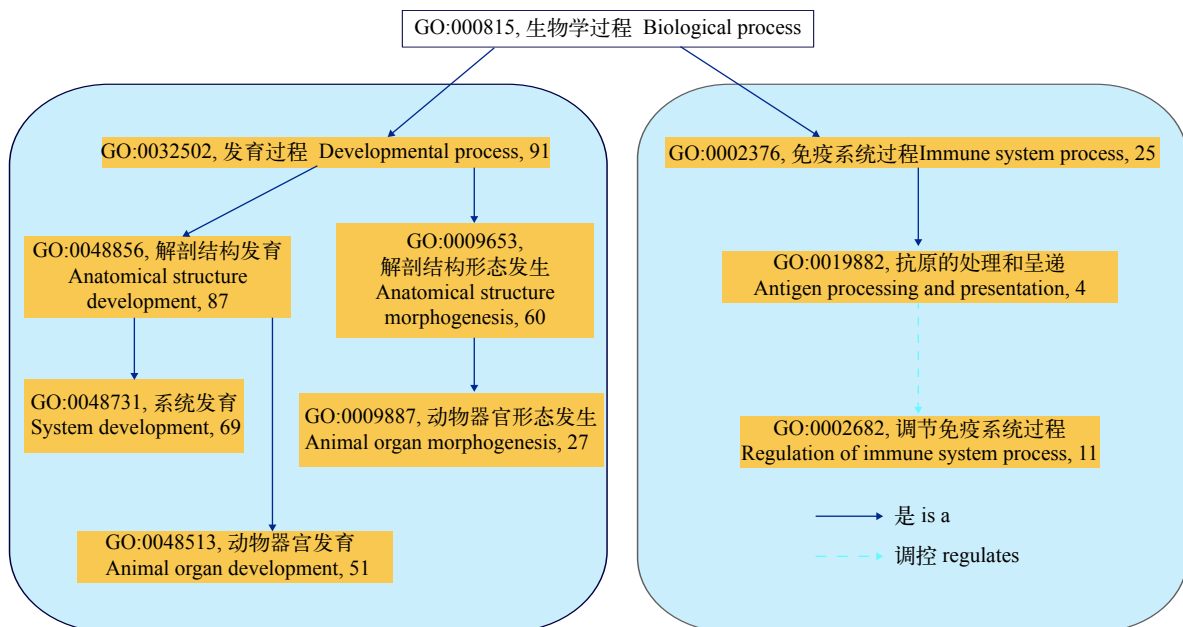


图 6 池螺蚌优势性能相关的特有基因GO富集

Fig. 6 GO enrichment of specific genes related to ecological adaptation in *H. schlegelii*

方框中的数字代表每个GO条目中富集的基因数。图中实线箭头表示“是”的GO关系, 代表低一级层次的条目是高一级层次条目的子类型或亚类型; 虚线箭头表示“调控”的GO关系, 代表一个过程对另一个过程的调控

The Arabic numbers in boxes show the enriched number of genes in each term. The solid line arrow represents the GO relation of “is a”, and the low-hierarchy terms are the subtype of the high-hierarchy terms; the dotted line arrow represents the GO relation of “regulation” and represents the regulation of one process to another

源序列分析物种的进化关系。以往池螺蚌和三角帆蚌的分子进化研究主要是基于同工酶^[2]、*ITS1*^[3]和线粒体DNA^[23]的比较分析,但这样单个基因或少数基因的比较不能代表整体差异,而且可能由于遗传变异存在一定的偶然性,不足以使人信服。比较基因组或比较转录组可以最大限度地对不同物种的遗传差异进行分析,在缺乏基因组信息的非模式生物中,往往可以基于转录组数据解决物种的进化关系^[24, 25]。我们基于转录组水平的系统发育分析表明池螺蚌与三角帆蚌的遗传距离不到1%,大大低于其他淡水珠蚌间的遗传距离,这与传统的形态学分类有所不同,但与Sakai等^[2]的研究结果一致,很大程度上支持了池螺蚌与三角帆蚌可能是不同亚种或不同种群的假设,二者的形态学差异可能是由于生存环境差异造成的。进一步的分析表明,小方蚌亚科与无齿蚌亚科大约在5144万年前发生分歧,而池螺蚌和三角帆蚌发生分化的时间大约在424万年前,这表明可能约在424万年前由于某些环境变化使池螺蚌和三角帆蚌发生地理隔离导致它们产生一些遗传上的差异,但这些差异还处于种内分化的水平。

3.2 池螺蚌中的特有基因

特有基因的分析能帮助我们理解表型性状产生的分子机制,同时可以揭示基因表达模式与表型形成之间的关系。有研究表明,池螺蚌生长速度快,其壳宽、外套膜的厚度、晶杆的长度和直径都大于三角帆蚌^[4]。个体的生长及其形态与生长发育过程息息相关,如外套膜的生长与发育。本研究对池螺蚌和三角帆蚌转录组中的基因进行比较分析,结果表明,根据GO分类,多达51个池螺蚌特有基因被标注为动物器官发育。其中,基因 α -I型胶原质(*COL1A1*)主要负责编码I型胶原的主要成分,即大部分结缔组织中的纤维胶原,包括软骨^[26]。另有基因*JAG1*和Notch信号的调控有关,该基因在胚胎发育中起着重要作用,可能涉及造血过程中的细胞命运决定和抑制成肌细胞的分化^[27]。结缔组织中的纤维胶原可能与池螺蚌外套膜的厚度相关,而胚胎发育对有机体的生长发育至关重要。这表明池螺蚌中可能存在一些特有基因或者调控这些特有基因表达的相关基因,使得池螺蚌在生长发育方面优于三角帆蚌。

此外,有研究表明池螺蚌对蚌瘟病和嗜水气单胞菌的抵抗力明显强于三角帆蚌^[5, 28]。有机体的抗病能力主要依赖免疫系统发挥防御作用,这个过程包括抗原的识别和免疫系统的调控。GO富集结果表明有25个池螺蚌特有基因涉及免疫系统过程,这

个结果与富集在免疫细胞膜分子通路中的13个基因相一致。作为免疫系统过程的重要环节,抗原的处理和呈递发挥着重要的作用。例如,基因*ABCB9*是ATP依赖的低亲和力肽转运体,它将广谱的多肽从胞浆转移到溶酶体腔^[29]。类似地,基因*SEC24A*编码衣壳蛋白复合物II的组成成分,促进内质网转运囊泡的形成。衣壳主要有两种功能,一是内质网膜在囊泡中的物理变形,二是选择转运到高尔基体的货物分子,这与异质蛋白在内质网-高尔基体界面被识别、分拣和捕获有关。抗原的有效识别和处理决定了免疫系统能否有效地发挥作用^[30]。基于以上结果,我们推测池螺蚌优良的抗病能力可能与池螺蚌中特有的免疫基因有关。

4 结论

本研究通过转录组学分析对池螺蚌和三角帆蚌进行了比较,为厘清池螺蚌和三角帆蚌的亲缘关系提供了重要的证据和帮助,并为研究池螺蚌优良性状相关的分子机制提供了新的视角和有利的支持。在本研究中,我们的计算表明池螺蚌和三角帆蚌的遗传距离不到1%,大约只有其他蚌种间遗传距离的十分之一,这与种内分化水平相当。另外,本研究对包括池螺蚌、三角帆蚌、背角无齿蚌和褶纹冠蚌在内的9种贝类重建了贝类系统发育分析,结果表明,池螺蚌和三角帆蚌发生分歧的时间大约在424万年前。池螺蚌产自日本,而三角帆蚌是中国本土物种,可以推断,它们可能在那个时间点由于某些事件发生了地理上的隔离。池螺蚌在一些性能上优于三角帆蚌,这很可能是遗传差异导致的。池螺蚌转录组中特有基因富集分析结果说明池螺蚌具有潜在的优势性能,如生长发育和免疫。这很可能是池螺蚌在生长速度、外套膜厚度和抗病能力等方面具有优势的关键所在。由于转录组学分析难以获得全部和完整的遗传信息,因此需要进一步的研究。尽管如此,本研究依然为池螺蚌和三角帆蚌的亲缘关系提供了重要参考价值,为进一步研究池螺蚌生长发育和免疫等相关的分子机制提供了有益线索。

参考文献:

- [1] Simpson C T. Synopsis of the naiades, or pearly fresh-water mussels [J]. *Science*, 1901, **13**(338): 983-984.
- [2] Sakai H, Ujiie M, Mizutani E, *et al.* Allozyme comparison between Japanese and Chinese limnetic pearl mussels [J]. *Journal of the Shimonoseki University of Fisheries*, 1997, **46**(2): 101-104.
- [3] Wu H S, Ouyang S, Wu X P, *et al.* The taxonomical

- status of *Hyriopsis cumingii* and *Hyriopsis schlegelii* inferred from *ITS-1* sequences [J]. *Journal of Nanchang University* (Engineering & Technology), 2008, **30**(3): 234-237. [吴红松, 欧阳珊, 吴小平, 等. 基于 *ITS-1* 基因序列探讨许氏帆蚌和三角帆蚌的分类地位 [J]. 南昌大学学报(工科版), 2008, **30**(3): 234-237.]
- [4] Zhou C H, Xu M X, Ouyang S, *et al.* On the comparative studies of some biological characteristics between *Hyriopsis schlegelii* and *H. cumingii* [J]. *Jiangxi Science*, 2003, **21**(2): 122-124. [周春花, 徐毛喜, 欧阳珊, 等. 池螺蚌(*Hyriopsis schlegelii*)与三角帆蚌(*H. cumingii*)若干生物学性状比较研究 [J]. 江西科学, 2003, **21**(2): 122-124.]
- [5] Xu M X, Hong Y J, Dai Y G, *et al.* Preliminary study on disease resistance of pearl bivalves (*Hyriopsis schlegelii* L.) to disease-causing sources of the *Hyriopsis cumingii* L. [J]. *Journal of Nanchang University* (Natural Science), 2004, **28**(3): 262-265. [徐毛喜, 洪一江, 戴银根, 等. 池螺蚌(*Hyriopsis schlegelii*)对三角帆蚌病源的抗病性 [J]. 南昌大学学报(理科版), 2004, **28**(3): 262-265.]
- [6] Qiao Q, Wang Q, Han X, *et al.* Transcriptome sequencing of *Crucihimalaya himalaica* (Brassicaceae) reveals how Arabidopsis close relative adapt to the Qinghai-Tibet Plateau [J]. *Scientific Reports*, 2016(6): 21729.
- [7] Zheng P, Wang M, Li C, *et al.* Insights into deep-sea adaptations and host-symbiont interactions: A comparative transcriptome study on Bathymodiolus mussels and their coastal relatives [J]. *Molecular Ecology*, 2017(26): 5133-5148.
- [8] Bolger A M, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data [J]. *Bioinformatics*, 2014, **30**(15): 2114-2120.
- [9] Grabherr M G, Haas B J, Yassour M, *et al.* Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data [J]. *Nature Biotechnology*, 2013, **29**(7): 644-652.
- [10] Fu L, Niu B, Zhu Z, *et al.* CD-HIT: accelerated for clustering the next-generation sequencing data [J]. *Bioinformatics*, 2012, **28**(23): 3150-3152.
- [11] Davidson N M, Oshlack A. Corset: enabling differential gene expression analysis for de novo assembled transcriptomes [J]. *Genome Biology*, 2014(15): 410.
- [12] Huerta-cepas J, Forslund K, Coelho L P, *et al.* Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper [J]. *Molecular Biology and Evolution*, 2017, **34**(8): 2115-2122.
- [13] Huerta-Cepas J, Szklarczyk D, Forslund K. eggNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences [J]. *Nucleic Acids Research*, 2016(44): D286-D293.
- [14] Buchfink B, Xie C, Huson D H. Fast and sensitive protein alignment using DIAMOND [J]. *Nature Methods*, 2014, **12**(1): 59-60.
- [15] Emms D M, Kelly S. OrthoFinder2: fast and accurate phylogenomic orthology analysis from gene sequences [J]. *bioRxiv*, 2018, <https://doi.org/10.1101/466201>.
- [16] Kumar S, Stecher G, Li M, *et al.* MEGA X: molecular evolutionary genetics analysis across computing platforms [J]. *Molecular Biology and Evolution*, 2018(35): 1547-1549.
- [17] Katoh K, Kuma K, Toh H, *et al.* MAFFT version 5: improvement in accuracy of multiple sequence alignment [J]. *Nucleic Acids Research*, 2005, **33**(2): 511-518.
- [18] Katoh K, Toh H. Improved accuracy of multiple ncRNA alignment by incorporating structural information into a MAFFT-based framework [J]. *BMC Bioinformatics*, 2008, **9**(1): 212.
- [19] Price M N, Dehal P S, Arkin A P. FastTree 2-Approximately Maximum Likelihood Trees for Large Alignments [J]. *PLoS One*, 2010, **5**(5): e9490.
- [20] Sanderson M J. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock [J]. *Bioinformatics*, 2003, **19**(2): 301-302.
- [21] Carlson M, Pagès H. AnnotationForge: Code for Building Annotation Database Packages [CP]. *Bioconductor*, 2018, <https://doi.org/10.18129/B9.bioc.AnnotationForge>.
- [22] Yu G C, Wang L G, Han Y Y, *et al.* clusterProfiler: an R package for comparing biological themes among gene clusters [J]. *OMICS: A Journal of Integrative Biology*, 2012, **16**(5): 284-287.
- [23] Lin Q H. Sequence analysis of mitochondrial DNA on *Hyriopsis schlegelii* [D]. Nanchang: Nanchang University, 2010: 20-22. [林巧惠. 池螺蚌线粒体全基因组序列分析 [D]. 南昌: 南昌大学, 2010: 20-22.]
- [24] Yang Y, Smith S A. Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: improving accuracy and matrix occupancy for phylogenomics [J]. *Molecular Biology and Evolution*, 2014, **31**(11): 3081-3092.
- [25] Johnson B, Borowiec M, Chiu J, *et al.* Phylogenomics resolves evolutionary relationships among ants, bees, and wasps [J]. *Current Biology*, 2013, **23**(20): 2058-2062.
- [26] Morgan P H, Jacobs H G, Segrest J P, *et al.* A comparative study of glycopeptides derived from selected vertebrate collagens [J]. *Journal of Biological Chemistry*, 1970, **245**(19): 5042-5048.
- [27] Bauer R C, Laney A O, Smith R, *et al.* Jagged1 (JAG1) mutations in patients with tetralogy of fallot or pulmonic stenosis [J]. *Human Mutation*, 2010, **31**(5): 594-601.
- [28] Guo H J. Studies on the immunity-related enzymes activities and the genes expression characteristics in *Hyriopsis schlegelii* and *Hyriopsis cumingii* [D]. Nanchang: Nanchang University, 2008: 12-21. [郭红军. 池螺蚌与三角帆蚌抗菌酶活性及免疫相关基因表达特性的研究 [D].

- 南昌: 南昌大学, 2008: 12-21.]
- [29] Wolters J C, Abele R, Robert T. Selective and ATP-dependent translocation of peptides by the homodimeric ATP binding cassette transporter TAP-like (ABCB9) [J]. *Journal of Biological Chemistry*, 2005, **280**(25): 23631-23636.
- [30] Bonnon C, Wendeler M W, Paccaud J P, *et al.* Selective export of human GPI-anchored proteins from the endoplasmic reticulum [J]. *Journal of Cell Science*, 2010, **123**(10): 1705-1715.

THE PHYLOGENY AND GENETIC ANALYSIS OF THE *HYRIOPSIS SCHLEGELII* BASED ON TRANSCRIPTOME

ZHOU Wei, ZHANG Wan-Chang, ZHOU Ye, WANG Xiao-Min, HU Bei-Juan, WANG Jun-Hua and HONG Yi-Jiang
(Key Lab of Aquatic Resources and Utilization of Jiangxi, School of Life Sciences, Nanchang University, Nanchang 330031, China)

Abstract: There has been controversy about the genetic relationship between *Hyriopsis cumingii* and *Hyriopsis cumingii*. To further analyze the phylogeny of freshwater pearl mussels, we conducted comparative transcriptomic analysis between *H. schlegelii* and its relatives, including *Hyriopsis cumingii*, *Anodonta woodiana* and *Cristaria plicata*. The genetic distance between the four freshwater pearl mussels was calculated after the multiple sequence alignment (MSA) of the single-copy orthologs they shared. The results showed that the genetic distance between *H. schlegelii* and *H. cumingii* was only 0.00746 (<1%), while the genetic distance between other freshwater pearl mussels was almost 10 times bigger. Phylogenetic analysis showed that Ambloinae and Anodontinae diverged approximately 51.44 million years ago (Mya), while *H. schlegelii* and *H. cumingii* were divided about 4.24 million years ago. Further analysis identified the specific genes in the transcriptome of *H. schlegelii*. The enrichment of the KEGG pathway indicated that these genes were mainly enriched in pathways such as CD molecules and protein digestion and absorption. Moreover, GO annotation results showed that there were many specific genes related to *H. schlegelii* excellent characteristics, including development-related biological processes, such as developmental process, system development, animal organ development, and immune-related biological process, immune system process and antigen processing and presentation in GO classification. The above results showed that *H. schlegelii* and *H. cumingii* are closely related, and may be different subspecies or populations, which has important reference significance for the germplasm conservation and genetic breeding of freshwater pearl mussels. In addition, compared with *H. cumingii*, *H. schlegelii* may have several specific genes related to development and immunity, which provides some clues for exploring the related molecular mechanism.

Key words: *Hyriopsis schlegelii*; *Hyriopsis cumingii*; Transcriptome; Phylogenesis; Genetic differences