

# 黄河口与长江口近江牡蛎的比较转录组分析\*

涂康,于红,孔令锋,李琪\*

(中国海洋大学海水养殖教育部重点实验室,山东青岛266003)

**摘要:**为进一步开发近江牡蛎 *Crassostrea ariakensis* 的基因资源,探究黄河口与长江口近江牡蛎的遗传差异,采用 Illumina HiSeq 4000 测序平台对东营垦利和南通海门的近江牡蛎进行高通量转录组测序。对测序数据进行拼接后分别得到 83680 条和 71269 条 unigene,并对 unigene 进行了基因功能注释和正选择基因的富集分析。结果表明,同源基因经筛选得到正选择基因 259 个,GO(Gene Ontology)功能分析中共涉及到 967 个相关功能类别,其中有关到生物学过程的分类条目占比最多,为 590 项。总体上看,与细胞器、代谢过程、结合以及催化活性相关的类别占主导地位。KEGG(Kyoto Encyclopedia of Genes and Genomes)富集分析表明,共涉及 48 个代谢通路,其中与核糖体、细胞凋亡相关的通路占比最大。上述结果不仅加深了对黄河口以及长江口近江牡蛎基因表达水平差异的认识,同时也为今后进行近江牡蛎各类遗传学分析和一些关键基因的克隆以及具体功能分析提供了基础数据。

**关键词:**近江牡蛎;黄河口;长江口;转录组;正选择基因

中图分类号:S917.4; Q786 文献标志码:A 文章编号:1003-6482(2019)04-091-07

DOI:10.13984/j.cnki.cn37-1141.2019.04.011

## 引言

近江牡蛎(*Crassostrea ariakensis*),是世界性广分布种类,一般栖息于江河入海口的低盐海区,并由此而得名。广泛的分布性使近江牡蛎表现出了高度的遗传多样性,研究表明可能是由于地理隔离造成了基因流的阻塞,各地理群体为适应不同环境而产生新的性状<sup>[1]</sup>。

河口生态系统处于河流与海洋的交汇处,是两种不同生态系统物质交换的重要场所,占全球沿海水域的 15%<sup>[2]</sup>,是一类非常重要的生态系统。黄河口位于山东省东营垦利县境内,地处渤海湾与莱州湾的交汇处,其水文因素较为单一,冲淡水受渤海水团影响较大<sup>[3]</sup>,故而水温与同纬度其他海域相比较低。长江口是我国最大的入海河口,其外海域受台湾暖流影响较大,其在夏季可为长江口带来高温暖流,水温较高。基于两地之间生态环境的差异,本实验将黄河口与长江口作为近江牡蛎研究区域具有一定的代表性。

随着分子生物学的发展,近江牡蛎与香港巨牡蛎之间混杂的分类关系以及近江牡蛎自身大跨度的维度适应性越来越引起研究人员的关注。本实验通过转录组测序技术对黄河口和长江口两海域近江牡蛎进行测序与分析,首次从转录组学角度探讨长江口与黄河口近江牡蛎以及渤海湾长牡蛎之间在基因表达水平的差异,同时也为有关近江牡蛎基因组学研究提供了基础数据。

## 1 材料与方法

### 1.1 样品采集

本研究所用近江牡蛎分别于 2016 年取自山东东营垦利和江苏南通海门,均为健康成贝。每个取样点取一个个体,解剖后立即采集外套膜、鳃、闭壳肌、唇瓣和性腺组织。所有样品储藏在 -80℃ 冰箱,以便后续使用。

\* 基金项目:中央高校基本科研业务费专项(201762014);山东省科技发展计划项目(2016ZDJS06A06);青岛市产业培育计划项目(17-3-3-64-nsh)资助

第一作者简介:涂康(1991-),男,硕士生,主要从事贝类遗传育种研究。E-mail: 976825895@qq.com

\* 通讯作者:李琪(1966-),男,教授,主要从事贝类遗传育种研究。E-mail: qili66@ouc.edu.cn

收稿日期:2018-02-01

## 1.2 样品鉴定

每个样品取少量闭壳肌,以苯酚氯方法提取 DNA。通过 PCR 扩增后对 COI 基因进行测序,并与 NCBI 基因库进行比对,确定两个个体均为近江牡蛎。

## 1.3 总 RNA 的提取与处理

每种组织取 50—100mg 加入 1mL TRIzol 试剂(Ambion 公司,美国),置于匀浆机中匀浆。按照操作说明分别提取垦利和海门近江牡蛎五种组织的总 RNA,使用 DNase I 处理去除总 RNA 样品中残留的 DNA。NanoDrop 2000 分光光度计(Thermo Scientific 公司,美国)检测其质量和浓度均符合要求,琼脂糖凝胶电泳检测其完整性符合要求。

## 1.4 转录组文库制备

同一个体不同组织 RNA 提取检测合格后等量混合,用带有 Oligo(dT)的磁珠富集 mRNA。随后加入 fragmentation buffer 将 mRNA 打断成短片段,以 mRNA 为模板,用六碱基随机引物合成单链 cDNA,然后加入缓冲液、dNTPs 和 DNA polymerase I 和 RNase H 合成双链 cDNA,再用 AMPure XP beads 纯化双链 cDNA。纯化的双链 cDNA 先进行末端修复、加 poly(A)并连接测序接头,再用 AMPure XP beads 进行片段大小选择。最后进行 PCR 扩增,并用 AMPure XP beads 纯化 PCR 产物,得到最终的文库。文库构建完成后,先使用 Qubit2.0 进行初步定量,稀释文库至 1.5ng/ul,随后使用 Agilent 2100 (Agilent Technologies 公司,美国)对文库的 insert size 进行检测,insert size 符合预期后,使用 Q-PCR 方法对文库的有效浓度进行准确定量(文库有效浓度 > 2nM),以保证文库质量。建好的转录组文库用 Illumina HiSeq 4000 测序平台进行测序。目的样品的转录组建库及测序由北京诺禾致源公司完成。

## 1.5 数据处理与分析

首先对 Illumina HiSeq 4000 平台测序产生的原始数据进行处理,去除带接头的 reads、N(N 表示无法确定碱基信息)的比例大于 10% 的 reads 以及低质量 reads(质量值 Qphred  $\leq$  5 的碱基数占整个 reads 的 50% 以上的 reads),从而得到相应的 clean reads 数据库,使用 Trinity 软件对序列进行拼接,取每条基因中最长的转录本作为 unigene。

## 1.6 基因功能注释

本实验将拼接得到的两组参考序列分别比对到七大数据库:Nr, Nt, Pfam, KOG/COG, Swiss-prot, KEGG(Kyoto encyclopedia of genes and genomes)和 GO(Gene ontology)数据库,以获得全面的基因功能信息。GO 注释按照生物学过程、细胞组分和分子功能三个大类的下一层级进行描述。用 Blast2GO 提取 GO 注释信息,并与 KEGG 数据库进行比对,对这些基因所涉及的信号通路或代谢途径进行分析。

## 1.7 直系同源基因分析

分别取垦利和海门近江牡蛎的 unigene 代表该基因进行 CDS 预测,按照 Nr 蛋白库、Swiss-prot 蛋白库的优先级顺序进行比对,若比对上则提取转录本的 ORF 编码框信息并将其翻译为氨基酸序列(按 5'-3' 的顺序);若没有比对上则采用 estscan(v3.0.3)软件预测其 ORF,从而得到这部分基因编码的核酸序列和氨基酸序列。使用 OrthoMCL(v2.0.9)软件对垦利近江牡蛎-海门近江牡蛎的 CDS 全长序列进行直系同源基因搜索分析,并筛选 1:1 直系同源基因数据集用于后续分析。

## 1.8 正选择基因分析

Ka 为非同义替换率;Ks 为同义替换率。Ka/Ks 常用于判定是否有选择压力作用于该蛋白编码基因。若 Ka/Ks > 1,则认为有正选择效应;若 Ka/Ks = 1,则认为存在中性选择;如果 Ka/Ks < 1,则认为有纯化选择作用。本研究采用 paml-codeml 软件来计算 Ka、Ks 以及 Ka/Ks 值,筛选 Ka/Ks > 1 的表达基因作为正选择基因。对所得正选择基因运用 GOseq 和 KOBAS 软件分别进行 GO 和 KEGG 富集分析,以完成对相关基因功能的注释。

## 2 结果

### 2.1 转录组数据拼接与注释

垦利和海门近江牡蛎经 Illumina HiSeq 测序后分别获得 49987778 和 40109420 条 raw reads, 去除接头数据和低质量数据后分别得到 49258106 和 39345510 条 clean reads。数据已上传至 NCBI SRA 数据库, 检索号 PRJNA427660。利用 Trinity 软件对所有的 clean data 进行从头组装, 由此得到的 transcripts 数量依次为 109975 和 89262 条。最终获得 unigenes 分别为 83680 和 71269 条, 平均长度为 951bp 和 841bp, 其具体长度分布见图 1。

将拼接后得到的垦利与海门近江牡蛎 unigenes 与七大数据库进行比对, 至少注释到一个数据库的 unigene 数量分别为 51875 (61.99%) 条和 48910 (68.62%) 条, 具体注释情况见表 1。对基因进行 GO 分析后, 垦利与海门近江牡蛎分别有 25320 个和 24448 个基因成功注释, 注释基因涉及细胞组分、生物学过程及分子功能 3 个大类 54 个小类(图 2)。根据数量统计结果显示, 在细胞组分大类中, 富集较多的有细胞类别、细胞部分类别以及细胞膜类别; 在生物学过程大类中, 富集较多的有细胞过程类别、代谢过程类别和单一有机体过程类别; 在分子功能大类中, 富集最多的是绑定类别, 其次是催化活性类别和转运活性类别。在东营近江牡蛎和南通近江牡蛎基因中分别有 10508 和 12674 个基因映射到 231 个通路, 其中基因数量占比较高的代谢通路有核糖体通路、PI3K-Akt 信号通路、内质网蛋白加工通路和粘着斑通路等。

表 1 基因功能注释(括号内的值是给定物种中注释成功基因占所有参加注释基因的百分比)

Table 1 Functional annotation of unigenes (values in parentheses are the percentages of all assembled unigenes in a given species)

数据库	注释数量	
	垦利近江牡蛎	海门近江牡蛎
NR	39963 (47.75%)	39459 (55.36%)
NT	44063 (52.65%)	41344 (58.01%)
KO	10508 (12.55%)	12674 (17.78%)
Swiss-Prot	21238 (25.38%)	22072 (30.96%)
PFAM	25206 (30.12%)	24325 (34.13%)
GO	25320 (30.25%)	24448 (34.3%)
KOG	14335 (17.13%)	14399 (20.2%)
Annotated in all Databases	6946 (8.3%)	7876 (11.05%)
Annotated in at least one Database	51875 (61.99%)	48910 (68.62%)
Total Unigenes	83680 (100%)	71269 (100%)

Annotated in all Databases: 在以上 7 个数据库中都注释成功的 Unigene 数目及其占总 Unigene 数的比例

Annotated in at least one Database: 在以上 7 个数据库中至少 1 个数据库注释成功的 Unigene 数目及其占总 Unigene 数的比例

### 2.2 同源基因分析

对垦利近江牡蛎与海门近江牡蛎两组转录组数据进行同源分析, 初始一共得到假定的同源基因 18806 对, 筛除模糊比对结果以及包含终止密码子的结果后得到直系同源基因 16501 对。对直系同源基因计算  $K_a$ ,  $K_s$  以及  $K_a/K_s$  值, 得到  $K_a/K_s$  大于 1 的值共有 259 个。

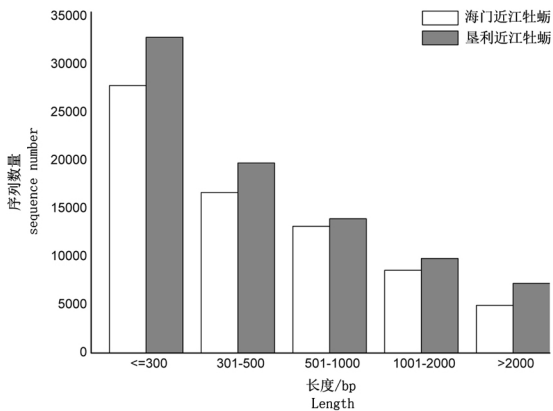


图 1 海门近江牡蛎与垦利近江牡蛎 unigene 长度分布

Fig. 1 Length distribution of unigene sequence in *C. ariakensis* from Haimen and Kenli

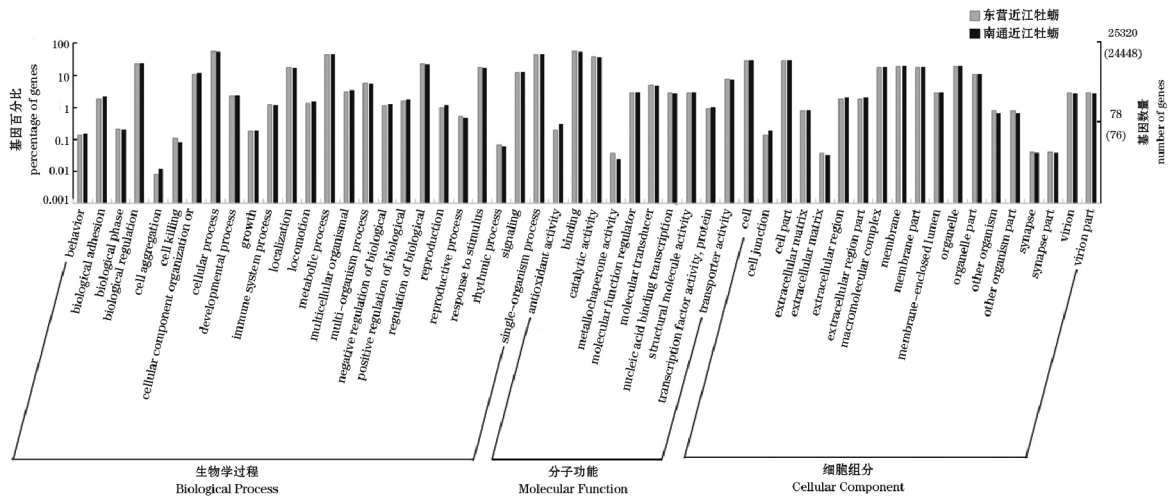


图2 垦利近江牡蛎与海门近江牡蛎基因 GO 功能分类

Fig. 2 Gene ontology classification of putative functions of *C. ariakensis* in Kenli and Haimen

2.3 GO 功能分类与 KEGG 富集分析

对正选择基因进行 GO 功能分析,大致可以将基因归类到三大类别:细胞组分、生物学过程和分子功能。对垦利近江牡蛎与海门近江牡蛎正选择基因进行分析可以发现(表 2):细胞组分分类中共有 153 个分类条目,代表类型为细胞成分、胞内受体和胞内组分等;富集到生物学过程的正选择基因最多,共 590 个分类条目,占主导地位的分类有细胞代谢过程、有机物生物合成过程等;分子功能分类涉及到 224 个分类条目,其中催化活性和结构分子活性占据优势地位。垦利近江牡蛎与海门近江牡蛎正选择基因 KEGG 富集分析结果显示一共涉及到 48 个代谢通路,其中核糖体通路最为富集,其次是过氧化物酶体通路(表 3)。

表 2 垦利近江牡蛎与海门近江牡蛎正选择基因 GO 注释中富集前 10 组分

Table 2 Top ten GO categories in the three terms of positive selected genes in *C. ariakensis* in Kenli compared to *C. ariakensis* in Haimen

GO 类别 Gene ontology term	功能描述 description	差异基因数目 DEG items	p 值 p-value
细胞组分 cellular components	细胞成分 Cell part	61	3.76E-03
	胞内受体 Intracellular	60	9.63E-04
	胞内组分 Intracellular part	59	2.39E-4
	细胞器 Organelle	47	2.36E-2
	细胞质 Cytoplasm	32	2.29E-05
	细胞质组分 Cytoplasmic part	26	5.45E-05
	无膜细胞器 Non-membrane-bounded organelle	22	5.59E-03
	核糖核蛋白复合体 Ribonucleoprotein complex	18	2.55E-08
	核糖体 Ribosome	15	1.17E-08
	类囊体 Thylakoid	4	5.36E-11
生物学过程 biological processes	代谢过程 Metabolic process	71	5056E-09
	细胞代谢过程 Cellular metabolic process	61	1.01E-03
	有机物生物合成过程 Organic substance biosynthetic process	41	1.14E-04
	生物合成过程 Biosynthetic process	41	1.55E-04
	细胞生物合成过程 Cellular biosynthetic process	39	4.67E-04
	基因表达 Gene expression	38	2.86E-03
	蛋白质代谢过程 Protein metabolic process	25	7.14E-04
	细胞组分起源 Cellular component biogenesis	20	7.53E-04
	细胞蛋白质代谢过程 Cellular protein metabolic process	20	1.9E-03
	单个有机体代谢过程 Single-organism metabolic process	18	2.04E-05

分子功能 molecular function	催化活性 Catalytic activity	33	4.72E-02
	结构分子活性 Structural molecule activity	18	8.64E-05
	核糖体的结构组成 Structural constituent of ribosome	14	3.22E-07
	氧化还原酶 Oxidoreductase activity	8	9.16E-10
	酶活性抑制剂 Enzyme inhibitor activity	7	3.53E-02
	双氧化活性 Dioxygenase activity	2	2.05E-06
	顺式还原酮加双氧酶活性 Acireductone dioxygenase [iron(II)-requiring] activity	2	1.35E-03
	翻译因子活性 Translation factor activity	2	2.23E-02
	辅因子结合 Cofactor binding	2	1.99 E-02
	半胱氨酸型肽链内切酶抑制剂活性 Cysteine-type endopeptidase inhibitor activity	2	5.57E-03

表 3 垦利近江牡蛎与海门近江牡蛎正选择基因 KEGG 通路富集分析

Table 3 The KEGG enrichment pathways of positive selected genes in *C. ariakensis* in Kenli compared to *C. ariakensis* in Haimen

通路 ID pathway ID	信号通路 pathway	基因数目 gene number	p 值 p-value
ko03010	核糖体 Ribosome	9	3.19E-11
ko04146	过氧化物酶体 Peroxisome	2	2.27E-02
ko00270	半胱氨酸和蛋氨酸代谢 Cysteine and methionine metabolism	1	1.06E-05
ko04520	阿尔茨海默病 Alzheimer's disease	1	3.96E-03
Ko03320	过氧化物增生物激活受体信号通路 PPAR signaling pathway	1	4.95E-02
ko04915	雌激素受体信号传导通路 Estrogen signaling pathway	1	5.11E-02

### 3 讨论

作为比较基因组学的一个热门分支,高通量比较转录组技术在水产动物遗传与进化中得到广泛运用<sup>[4-6]</sup>。相对于线粒体基因测序和核基因测序来说,转录组测序能更精准定位蛋白编码基因,在不借助参考基因组的情况下对所研究物种进行测序分析以研究生物在不同条件下基因的表达情况,进而高效而精确地定位表型相关基因。本研究通过对位于黄河口的垦利近江牡蛎和位于长江口的海门近江牡蛎的 5 种组织分别进行转录组文库构建,并在没有参考基因组的情况下使用 Illumina Hiseq 4000 进行测序,两组原始数据经过质控之后分别得到 49258106 和 39345510 条 clean reads。经过组装和拼接后,最终获得 83680 和 71269 条 unigenes。将从头组装得到的 unigenes 与七大数据库进行比对,分别有 38.01% (83680 条)和 31.38% (71269 条)的 unigenes 能够匹配到已知的七大数据库中,其余基因与所选数据库中已知基因同源性较低,未被注释。推测其原因主要有两种,首先测序所得转录组数据中本身有一部分是属于非编码序列,不能编码蛋白质;其次是注释情况受限于数据库本身收录的相近物种的基因注释数据量,现有的牡蛎基因组注释完善程度不够充分,无法提供全面的注释信息。

GO 功能分析是基因功能国际标准分类体系<sup>[7]</sup>,对 divergent( $Ka/Ks > 1$ )基因进行 GO 分析可以发掘出正选择基因在机体中所参与的生物过程、分子功能以及相关细胞组分。本研究运用 GOseq<sup>[8]</sup>对垦利近江牡蛎和海门近江牡蛎的正选择基因进行 GO 功能分析表明,细胞组分类别中与细胞器相关的分类条目(细胞器、无膜细胞器)、与胞内活动相关的分类条目(胞内受体、胞内组分)以及与细胞质相关的分类条目(细胞质、细胞质组分)所占比例最大;生物学过程类别中与代谢过程以及生物合成相关的分类条目占据主导地位;在分子功能类别中,与各类结构分子活性作用相关的分类条目富集最为显著。这些富集结果可能与两地近江牡蛎的生存环境有关,相较于海门近江牡蛎,垦利地区近江牡蛎显然更能适应低温环境<sup>[9]</sup>。研究表明,在一定的温度范围内,贝类代谢速率与外界温度变化成正相关<sup>[10]</sup>。在两地差异温度条

件下机体为了维持生命活动而调节代谢速率,该过程主要在细胞质内进行,同时需要调节膜内外各项活动以及胞内各种细胞器的共同作用才能完成。低温或高温刺激会使生物体内各项合成、代谢过程发生紊乱,垦利近江牡蛎和海门近江牡蛎为响应不同温度条件,与能量代谢相关的通路差异化发展,同时有机体生物合成过程又需要各类催化因子的辅助作用,低温下酶活性抑制剂可以帮助降低特定酶活性以降低能量消耗,对抗外界环境因子作用。

KEGG 是有关 Pathway 的主要公共数据库<sup>[11]</sup>,可以用于获取差异表达基因所涉及的信号通路,通过对关键信号通路的分析进一步掌握其生物学功能。本研究通过 KOBAS 软件<sup>[12]</sup>对垦利近江牡蛎和海门近江牡蛎的正选择基因进行 KEGG 富集分析可以发现,富集最显著的是核糖体通路,核糖体是蛋白合成场所<sup>[13]</sup>,RNA 转运是核糖体与蛋白之间的桥梁,而蛋白则是机体各项生命活动的主要承担者,由此可知,核糖体合成蛋白在两地近江牡蛎适应不同环境时发挥了重要作用。除了核糖体以外,还有一些基因富集到了与氨基酸和糖类有关的代谢通路上(半胱氨酸和蛋氨酸代谢),这类代谢通路受环境影响作用较大<sup>[14]</sup>。推测两地近江牡蛎在不同环境下,氨基酸与糖类的合成和代谢产生了不一样的变化。此外还有两项与过氧化物相关的通路(过氧化物酶体和过氧化物增生物激活受体信号通路),过氧化物酶体与脂质分解、生物合成相关,参与大部分代谢过程;过氧化物增生物激活受体信号通路与特定基因的转录相关,推测可能参与到近江牡蛎在不同环境下的基因表达。富集通路中还有一项是阿尔茨海默病,看似与近江牡蛎的生理活动没有直接联系,其实该通路所导致的生理现象和细胞凋亡有关。细胞凋亡对于机体稳态的维持有重要作用<sup>[15]</sup>,内在或外在因素都会触发该程序,受到外界刺激作用时细胞会自发启动凋亡程序<sup>[16]</sup>,由此清除衰老或突变细胞,我们推测这可能与两地近江牡蛎在适应各自环境过程中的自我保护机制有关。

综上所述,对黄河口和长江口近江牡蛎的比较转录组分析可以发现,正选择基因主要集中于与细胞器相关、与代谢相关的功能上,以及与核糖体、细胞凋亡等相关的通路上。相关功能与通路可能与两地近江牡蛎在不同环境条件下应激表达有关。本研究从分子水平解释了黄河口与长江口近江牡蛎特定适应性的遗传机制,补充和完善了近江牡蛎基因数据库资源,为今后开展与近江牡蛎相关遗传学以及组学方面的研究提供了基础资料。

## 参考文献

- [1] Ren J, Hou Z, Wang H, et al. Intraspecific variation in mitogenomes of five *Crassostrea* species provides insight into oyster diversification and speciation[J]. *Mar Biotechnol*, 2016, 18(2): 242-254.
- [2] Shen G Y, Shi B Z. *Marine Ecology*. 2nd Ed. Beijing: Science Press, 2002 (in Chinese).
- [3] 翟璐, 徐宾铎, 纪毓鹏, 等. 黄河口及其邻近水域夏季鱼类群落空间格局及其与环境因子的关系[J]. *应用生态学报*, 2015, 26(9): 2852-2858.
- [4] Wang Y, Yang L, Zhou K, et al. Evidence for adaptation to the Tibetan Plateau inferred from Tibetan loach transcriptomes[J]. *Genome BiolEvol*, 2015, 7(11): 2970-2982.
- [5] Mu X, Hou G, Song H, et al. Transcriptome analysis between invasive *Pomaceacaniculata* and indigenous *Cipangopaludinacahayensis* reveals genomic divergence and diagnostic microsatellite/SSR markers[J]. *BMC Genet*, 2015, 16(1): 12.
- [6] Ma X, Dai W, Kang J, et al. Comprehensive transcriptome analysis of six catfish species from an altitude gradient reveals adaptive evolution in Tibetan fishes[J]. *G3: Genes Genom Genet*, 2016, 6(1): 141-148.
- [7] Ashburner M, Ball C A, Blake J A. Gene ontology: Tool for the unification of biology[J]. *Nat Genet*, 2000, 25(1): 25-29.
- [8] Young M D, Wakefield M J, Smyth G K, et al. Gene ontology analysis for RNA-seq: accounting for selection bias[J]. *Genome Biol*, 2010, 11:1-12.
- [9] 张松, 于非, 刁新源, 等. 渤、黄、东海海表面温度年际变化特征分析[J]. *青岛: 海洋科学*. 2009, (8): 76-81.
- [10] 王俊, 姜祖辉, 唐启升, 2002. 栉孔扇贝耗氧率和排氨率的研究. *沈阳: 应用生态学报*, 13(9): 1157-1160.
- [11] Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes[J]. *Nucleic Acids Res*, 2000, 28(1): 27-30.
- [12] Xie C, Mao X, Huang J, et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases[J]. *Nucleic acids Res*, 2011, 39(suppl\_2): W316-W322.
- [13] Ramakrishnan V. Ribosome structure and the mechanism of translation[J]. *Cell*, 2002, 108(4): 557-572.

- [14] Kaplan F, Kopka J, Haskell D W, et al. Exploring the temperature-stress metabolome of Arabidopsis[J]. *Plant Physiol*, 2004, 136(4): 4159-4168.
- [15] Meier P, Finch A, Evan G, et al. Apoptosis in development[J]. *Nature*, 2000, 407(6805): 796-801.
- [16] Moon K, Zhou SY, Lee SH, et al. Hsp70 and XIAP are potential key molecular mechanisms causing impaired apoptosis and severe acute pancreatitis (AP) in CF mice[J]. *Pancreas*, 2007, 35: 417.

## A Comparative Transcriptional Analysis of *Crassostrea ariakensis* from Yellow River and Yangtze River Estuaries

TU Kang, YU Hong, KONG Lingfeng, and LI Qi\*

(Key Laboratory of Mariculture of Ministry of Education; Ocean University of China, Qingdao 266003, China)

**Abstract:** *Crassostrea ariakensis* inhabits a wide range of latitudes. In order to further develop the genetic resource of *C. ariakensis*, and explore the genetic difference of *C. ariakensis* between Yellow River and Yangtze River Estuaries, RNA-seq and bioinformatics analysis were conducted to compare the transcriptome of *C. ariakensis* from Kenli, Dongying and Haimen, Nantong. A total of 49,258,106 and 39,345,510 clean reads were generated from *C. ariakensis* in Kenli and Haimen, respectively. After de novo assembly, we received 83,680 and 71,269 unigenes, respectively. The complete sets of unigenes were queried against seven public annotation databases. Then the orthologous analysis and positive selection analysis were conducted to identify the positively selected genes. In total, 259 pairs of unigenes were identified with the Ka/Ks ratio  $> 1$ . The functional description of the GO (Gene ontology) and analysis of metabolic pathways of KEGG (Kyoto encyclopedia of genes and genomes) revealed that the positively selected genes were mainly enriched in the pathways relating to organelle, metabolism, ribosome and apoptosis. This study not only enhanced our understanding to gene differential expression of *C. ariakensis* in Yellow River Estuary and Yangtze River Estuary, but also provided basic gene resources for cloning, functional analysis and molecular genetic marker development of *C. ariakensis*.

**Key words:** *Crassostrea ariakensis*; Yellow River Estuary; Yangtze River Estuary; transcriptome; positively selected gene