

我国不同地理群体栉江珧遗传多样性及系统发生分析

严加坤^{1,2} 杨爱国^{2*} 周丽青² 吴彪² 白临建^{1,2} 刘志鸿²

(¹上海海洋大学水产与生命学院, 201306)

(²农业部海洋渔业可持续发展重点实验室 中国水产科学研究院黄海水产研究所, 青岛 266071)

摘 要 利用通用引物对 5 个栉江珧野生群体(山东长岛、山东日照、山东文登、广东湛江和海南海口) 28S rRNA 和 COI 基因片段进行扩增测序, 分别得到 983bp 和 623bp 的片段, 基于 28S rRNA 序列分析系统发生的结果表明, 栉江珧与 *Atrina vexillumju* 具有较近的遗传关系。基于 COI 基因序列的遗传多样性分析结果显示, 山东文登群体具有最高的遗传多样性水平。AMOVA 分析显示, 群体遗传分化系数为 0.132 3 ($P < 0.001$), 说明栉江珧遗传变异主要来源于群体内的变异。由 28S rRNA 和 COI 基因聚类分析结果推测, 由于地理隔离, 我国栉江珧南北方群体可能早已分化为不同亚种。这些数据为我国栉江珧种质资源保护和利用补充了分子生物学资料。

关键词 栉江珧 28S rRNA COI 遗传多样性 系统进化

中图分类号 S917.4 **文献标识码** A **文章编号** 1000-7075(2013)02-0029-07

The genetic diversity and phylogenetic studies of five populations of *Atrina pectinta* in China

YAN Jia-kun^{1,2} YANG Ai-guo^{2*} ZHOU Li-qing²

WU Biao² BAI Lin-jian^{1,2} LIU Zhi-hong²

(¹College of Fisheries and Life, Shanghai Ocean University, 201306)

(²Key Laboratory of Sustainable Development of Marine Fisheries, Ministry of Agriculture, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao 266071)

ABSTRACT Mitochondrial COI gene fragments and ribosomal 28S rRNA of *Atrina pectinta* sampled from five locations including three in Shandong (Changdao, Wendeng, and Rizhao), one Guangdong (Zhanjiang) and one in Hainan (Haikou) were amplified with universal primers, and 623bp and 983bp nucleotide sequences were obtained, respectively. Phylogenetic studies of Pinnidae based on 28S rRNA partial sequences showed that *A. pectinta* had a close genetic relationship with *Atrina vexillumju*. Genetic diversity of the five populations based on COI partial sequences demonstrated that the Wendeng population had the highest level of genetic diversity. The fixation indices (F_{st}) analyzed by AMOVA was 0.132 3 ($P < 0.001$), indicating that the genetic variation of *A. pectinta* was mainly from genetic fixation inter-populations.

山东省科技发展计划项目(2010GHY10513)、山东省自然科学基金项目(ZR2009DQ006)、青岛市成果转化项目(10-3-4-17-chg)和市南区 2011 年第二批科技发展基金项目(2011-5-023-QT)共同资助

* 通讯作者。E-mail: yangag@ysfri.ac.cn; Tel: (0532) 85811982

收稿日期: 2012-05-04; 接受日期: 2012-05-30

作者简介: 严加坤(1988-), 男, 硕士研究生, 主要从事贝类遗传育种研究。E-mail: himingse@163.com

Cluster analysis based on 28S rRNA and COI showed that *A. pectinta* in China might have differentiated into different subspecies. These findings would provide molecular biological data for the protection and utilization of *A. pectinta* resources in China.

KEY WORDS *Atrina pectinta* 28S rRNA COI Genetic diversity Phylogeny

栉江珧 *Atrina pectinta*, 俗称大海红、带子, 隶属于软体动物门 Mollusca、瓣鳃纲 Lamellibranchia、贻贝目 Mytiloida、江珧科 Pinnidae、江珧属 *Atrina*, 是一种广温、广盐性贝类, 主要分布在温带、热带海洋, 在我国广泛分布于各大海区。栉江珧闭壳肌大而圆, 味道鲜美, 是一种具有较高经济价值的大型海产贝类。但是, 在某些海区的过度捕捞和人工育苗工作的相对滞后导致栉江珧资源量下降、种质资源退化和遗传多样性水平降低等问题。因此, 开展有关栉江珧的遗传结构和遗传变异水平的研究是非常必要的。

有关栉江珧的研究主要集中在繁殖生物学、人工育苗以及同工酶等方面(余祥勇等 1999、2000; 于瑞海等 2007; 张红云 2010), 分子遗传学方面的研究较少。线粒体 COI 基因具有母系遗传、进化速度快、结构简单、序列短、易于扩增等特点, 被广泛应用于不同地理生物种群之间的亲缘关系分析(Canapa *et al.* 2000; Martínez-Navarro *et al.* 2005; Shih *et al.* 2008)。真核生物核糖体基因组依次由 18S、ITS1、5.8S、ITS2 和 28S 组成, 其中 5.8S、18S 和 28S 区域进化速度较慢、序列高度保守, 适用于种间分子系统发育研究(Hedgecock *et al.* 1999; 董长永等 2008; 陈诚等 1997)。本研究对分布于中国沿海 5 个自然群体的栉江珧的 mtDNA COI 及核糖体 28S 基因序列片段进行了扩增和测序, 结合 GenBank 中搜索到的江珧科其他种的相同序列, 对栉江珧群体遗传特性及系统发育进行分析, 以期为栉江珧种质资源的研究和保护提供背景基础资料, 为今后开展栉江珧家系选育、亲本选择及分子辅助育种等相关研究工作的展开奠定分子生物学基础。

1 材料与方法

1.1 材料

2011 年 3 月~2012 年 2 月共收集我国北方山东长岛 CD(37°54'49.66"N, 120°47'36.94"E)、山东文登 WD(36°54'23.31"N, 122°02'13.35"E)、山东日照 RZ(35°23'17.76"N, 119°38'41.15"E)和南方广东湛江 ZJ(21°10'33.04"N, 110°46'42.41"E)、海南海口 HN(20°07'32.10"N, 110°27'11.99"E) 5 个海区的野生种群, 活体运回实验室。测量相关生物学数据后, 取肌肉存于-80℃备用。

1.2 DNA 提取

采用盐析法提取肌肉组织基因组 DNA, 方法参照 Aljanabi 等(1997), 有改动。具体改动为在 DNA 析出时将异丙醇更换为 1/10 体积的 3mol/L NaCl 和两倍体积的无水乙醇。基因组 DNA 经 1% 琼脂糖凝胶电泳检测、NanoDrop 2000 生物分光光度计检测浓度后, ddH₂O 稀释至 100ng/μl, -20℃保存备用。

1.3 目的基因扩增

基因片段扩增所通用引物的核苷酸序列为: 28S F: GGGACTACCCCTGAATTTAAGCAT; 28S R: CCAGCTATCCTGAGGGAACTTCG; COI F: GGTCACAAATCATAAAGATATTGG; COI R: TAACTTCAGGGTGACCAAAAATCA。PCR 反应体系为 1.5U *Taq* 酶、10×Buffer 3μl、dNTP(2.5mol/L each) 3.6 μl、Primer(10 μmol/L)0.5μl、基因组 DNA 100ng, 双蒸水补足 30μl 体系。PCR 反应程序为: 94℃ 预变性 5min, 之后 94℃ 30s、57℃ (28S)/52℃ (COI) 30s、72℃ 1min、35 个循环, 最后 72℃ 延伸 10min。PCR 产物经 1% 琼脂糖凝胶电泳检测后由华大基因纯化测序。

1.4 数据处理

测序结果经 EditSeq、Seqman 根据测序峰图人工编辑、整理。经 ClustalW(Thompson *et al.* 1994) 比对

确定序列长度,用 DnaSP 5.01(Librado *et al.* 2009)软件计算各个群体的单倍型、单倍型多态性、多态位点数、平均核苷酸差异数、核苷酸多样性指数等。用 MEGA 5.01(Tamura *et al.* 2011)计算不同序列间的碱基组成、变异位点、简约信息位点和不同地理种群的 Kimura2-paramter 遗传距离。采用 Kimura2-paramter 距离矩阵采用邻接法(NJ)以及非加权配对算数平均法(UPGMA)构建单倍型分子系统树,系统树中节点的自举置信水平应用自引导(Bootstrap)估计,共 1 000 次循环。用 Arlequin 3.11(Excoffier *et al.* 2005)中的分子变异分析(AMOVA)分析方法估算遗传变异在群体内和群体间的分布及遗传分化系数(Fixation indices, F_{st}),并用排列测验法(Permutation test)检验 F_{st} 的显著性(重复次数为 1 000)。基因流 N_m 由公式 $N_m = (1/F_{st}) - 1$ 计算得出。

江珧科其他物种(表 1)选自 GenBank(<http://www.ncbi.nlm.nih.gov/genbank/>),以紫贻贝 *Mytilus edulis* 为外群,利用 PAUP 4.0 beta 10(Swofford 2002)构建 Maximum parsimony tree(MP)和 Neighbor-joining tree(NJ)进化树。

表 1 GenBank 所选物种名称及登录号

Table 1 Name of selected species and accession number in GenBank

物种	Species	GenBook Locus
雙色江珧蛤	<i>Pinna bicolor</i>	AB594398
黑旗江珧蛤	<i>Atrina vexillum</i>	AB594404
尖角江珧蛤	<i>Pinna muricata</i>	AB594400
* * * *	<i>Streptopinna saccata</i>	AB594402
紫贻贝	<i>Mytilus edulis</i>	Z29550

注:所选江珧科物种命名规则依据 <http://taibnet.sinica.edu.tw>,选取原则为在我国海域有分布

Note:The selected species distributes in China, and are named according to <http://taibnet.sinica.edu.tw>

2 结果与分析

2.1 核糖体 28S 序列片段分析

测序结果经过人工校对、编辑之后共得到 5 个群体 25 条 983bp 序列片段。ClustalW 比对显示,北方 3 个群体 28S 序列完全一致,同样的现象也出现在南方两群体上。利用 DnaSP 做单体型检测,25 条序列共定义两种碱基颠换形成的单体型。其中北方群体和南方群体分别属于单体型 1、单体型 2。

依据 28S rRNA 序列对两种单体型及所选的江珧科物种构建系统进化树,基于 MP 和 NJ 方法构建的进化树聚类结果相同,NJ 进化树如图 1 所示。可以看出所选的江珧科 5 种江珧分为两支,本研究的栉江珧与来自 GenBank 的 *A. vexillum* 首先聚为一支,*P. bicolor* 和 *P. muricata* 聚在一起之后再与 *S. saccata* 聚在一起形成另一支。

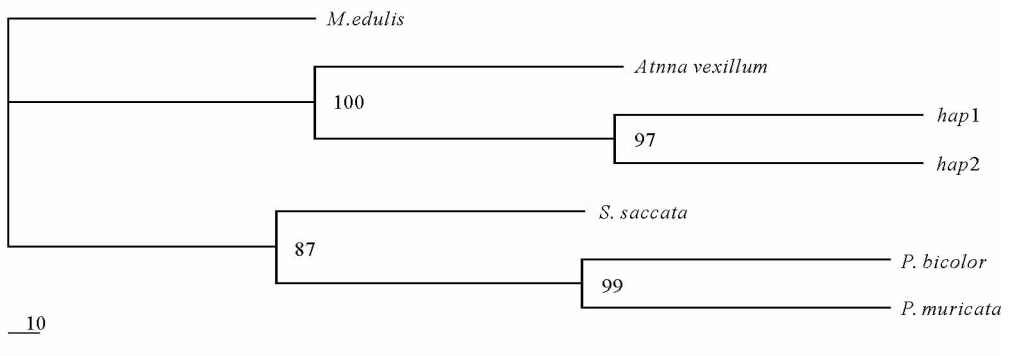


图 1 基于 28S rRNA 基因序列构建的江珧科贝类的 NJ 树

Fig. 1 NJ tree constructed based on 28S rRNA gene sequences

2.2 线粒体 COI 序列分析

2.2.1 碱基组成及序列分析

5 个群体共得到碱基序列 55 条,碱基组成如表 2 所示,可以看出 AT 含量在 5 个群体中均大于 61%,显著

高于 GC 含量。5 个群体中共检测到多态位点 77 个,其中 Parsimony informative sites(简约位点)72 个、Singleton variable sites(单变异位点)5 个。77 个位点共定义 19 种单体型,单体型分布如表 3 所示;单体型 1、3、4 为北方群体所共享,南方群体共享单体型 7,其余为个别群体所有。DnaSP 分析群体遗传多样性参数结果如表 4;文登群体在单倍体多态性(Hd)、核苷酸多样性指数(Pi)等方面均明显高于其他群体,湛江群体最低;此外在北方 3 个群体在各项遗传参数上均高于南方群体。

表 2 栉江珧群体 COI 序列片段碱基组成

Table 2 Base compositions of COI sequences in five *A. pectinta* populations

碱基 Base	日照 RZ	文登 WD	长岛 CD	海南 HN	湛江 ZJ	均值 Avg
T(%)	43.06	42.94	42.99	42.04	42.05	42.61
C(%)	14.89	14.99	14.95	15.91	15.87	15.33
A(%)	19.10	19.02	19.07	19.26	19.26	19.14
G(%)	22.95	23.05	22.98	22.79	22.81	22.92
A+T(%)	62.2	62.0	62.1	61.3	61.3	61.8

表 3 栉江珧 COI 基因单体型(H1~H19)在不同群体中的分布

Table 3 Distribution of COI gene haplotypes in five *A. pectinta* populations

群体 Population	单体型 Haplotype																			
	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13	H14	H15	H16	H17	H18	H19	
日照 RZ	1	1	1	3	1	1														
文登 WD	1		1	2		1						1	1	1	1	1				
长岛 CD	2	2	1	3	1								1	1		1				
海南 HN							6	1	1	1	1									
湛江 ZJ																		1	1	1

表 4 栉江珧 COI 遗传多样性参数

Table 4 Summary of genetic diversity of different *A. pectinta* populations

遗传多样性参数 Items	日照 RZ	文登 WD	长岛 CD	海南 HN	湛江 ZJ	总计 Total
单倍体型数 Number of haplotypes(<i>H</i>)	6	9	8	5	4	19
单倍体多态性 Haplotype diversity(<i>Hd</i>)	0.892 9	0.977 8	0.924 2	0.666 7	0.533 3	0.899 6
多态位点 Segregating sites(<i>S</i>)	6	14	11	6	3	77
平均核苷酸差异数 Average number of differences(<i>K</i>)	2.642 9	4.044 4	2.575 8	1.355 6	0.600 0	34.193 5
核苷酸多样性指数 Nucleotide diversity(<i>Pi</i>)	0.004 24	0.006 49	0.004 13	0.002 18	0.000 96	0.054 89

2.2.2 群体间遗传及聚类分析

栉江珧群体间平均核苷酸差异数(*K*)及遗传距离(*D*)(表 5)可以看出在南方和北方群体间遗传参数存在显著差异,而南北方群体各自之间则差异较小;平均核苷酸差异数 *K*、遗传距离 *D* 的极大值均出现在与文登群体相关的组合中,而南方的海南和湛江群体之间的两个遗传参数在比较中均表现为最低。

表 5 栉江珧群体间平均核苷酸差异数 K 及遗传距离 D Table 5 Average number of nucleotide difference between populations (K) and the pairwise distances matrix for the COI gene fragment (D)

群体 Population	日照 RZ	文登 WD	长岛 CD	海南 HN	湛江 ZJ
日照 RZ	* * *	0.000 18	0.000 11	0.051 83	0.052 10
文登 WD	3.200	* * *	0.000 13	0.052 22	0.052 50
长岛 CD	2.479	3.158	* * *	0.052 15	0.052 43
海南 HN	67.150	67.500	67.317	* * *	0.000 10
湛江 ZJ	67.150	67.500	67.317	1.000	* * *

注: 对角线左下方为 K , 右上方为 D

Note: K (below diagonal), D (above diagonal)

遗传分化系数(F_{st})和基因流(N_m)(表 6)分析显示,基因流最大值(22.002 25)出现在海南与湛江群体之间,文登与日照群体(11.638 3)之间次之,南北两方群体之间基因流较小,遗传分化程度较高。5 个地理群体的分子方差分析显示,群体间遗传分化系数 $F_{st}=0.1323(P<0.001)$ (表 7),说明整个遗传变异群体间的占 13.23%,而 86.77%的变异来自群体内。因此,栉江珧 COI 基因遗传变异的主要来源是群体内变异。

表 6 基因流(N_m)和遗传分化系数(F_{st})Table 6 Gene flow (N_m) and genetic fixations index (F_{st}) between and within the five populations of *A. pectinata*

群体 Population	日照 RZ	文登 WD	长岛 CD	海南 HN	湛江 ZJ
日照 RZ	* * *	11.638 3	10.025 6	0.015 342	0.012 374
文登 WD	-0.044 89	* * *	10.905 8	0.020 833	0.017 813
长岛 CD	-0.052 49	-0.048 05	* * *	0.015 039	0.012 08
海南 HN	0.970 23	0.960 00	0.970 80	* * *	22.002 25
湛江 ZJ	0.975 85	0.965 60	0.976 41	0.022 22	* * *

注: 对角线左下方为 F_{st} , 右上方为 N_m

Note: F_{st} (below diagonal), N_m (above diagonal)

表 7 栉江珧群体间遗传差异的分子方差分析

Table 7 Analysis of molecular variance (AMOVA) among populations of *A. pectinata*

变异来源 Source of variation	自由度 df	平方和 Sum of squares	方差 Variance components	百分比 Percentage(%)
群体间 Among populations	4	4.032	0.061 02 ^{Va}	13.23
群体内 Within populations	45	18.008	0.400 19 ^{Vb}	86.77
总计 Total	49	22.040	0.461 20	

注: 上标 V_a 和 V_b 表示差异显著程度, $P<0.001$, 其通过将样本经过 1023 次随机排列改变计算而得到

Note: V_a and V_b indicate the difference level. Probabilities were calculated by 1023 random permutations of individuals across populations.

利用 MEGA 5.01 软件对得到的 COI 基因序列构建 NJ(图 2)和 UPGMA 进化树,两种构建方法得到的结果基本一致。表明 5 个栉江珧群体分为两支,即北方 3 群体一支、南方 2 群体一支,两支上各群体之间存在广泛的交叉。

3 讨论

28S rRNA 序列在生物体内分布广泛、含量大而且进化缓慢序列保守性高,近些年来已经被广泛地应用在水生生物属间水平的系统进化研究中(潘颖瑛等 2009)。为了进一步弄清栉江珧在分布于我国的江珧科生物中的进化地位,对栉江珧 5 个地理群体的 28S rRNA 序列进行了测序比对。结合 GenBank 的 4 种江珧科生物的 28S rRNA 序列,构建系统进化树。结果表明,栉江珧与 *A. vexillum* 在系统进化中关系最近,遗传距离最小;而与 *P. bicolor* 及 *P. munitata* 遗传距离最远。

COI 序列进化速度较快,在检测群体遗传变异和分析不同地理群体生物遗传多样性方面具有广泛的应用。本研究扩增、测序得到 5 个地理群体栉江珧 COI 序列 50 条,平均 AT 含量(61.8%)高于 GC 含量,这与之前在脊尾白虾(马朋等 2011)、长蛸(孙宝超等 2010)、缢蛏(牛东红等 2008)中得到的结果相似。遗传多样性分析显示,文登群体的遗传多样性水平最高,长岛次之,湛江最低。文登群体遗传多样性水平高可能由于深入黄海海域,靠近日韩海域,因此得到更多的与日韩群体间的基因交流机会,从而增加了自身的多样性水平;而日照和长岛群体则分别位于内海的渤海及黄海边缘,从而进行基因交流的机会减少。另外作者在采样过程中走访发现,栉江珧在山东文登、荣成等地并不是很受欢迎,在海洋捕捞时并不是主要的捕捞对象,这也许是野生资源得到保护的原因之一。

群体间遗传分析结果显示,文登群体与海南、湛江群体的平均核苷酸差异数及遗传距离非常大,说明它们之间存在着巨大的遗传差异,基因流极小。遗传分化系数(F_{st})是表观群体间遗传分化程度的重要参数,本研究中北方 3 个群体之间的 F_{st} 为 $-0.044\ 89$ 、 $-0.052\ 49$ 及 $-0.048\ 05$,都非常接近于 0,说明北方群体之间几乎不存在遗传分化;而南方两群体之间的 $F_{st}=0.022\ 22$,同样非常小。南北方群体之间作比较,发现 F_{st} 在 $0.960\ 0\sim 0.976\ 41$ 之间,说明南北方群体之间存在极大的遗传分化。并且 28S rRNA 显示,南北方群体只构成两种单体型,而南北方群体分别隶属于单体型 1 和单体型 2,COI 基因单体型也区分为北方群体共享和南方群体共享两大类型。综上所述,作者

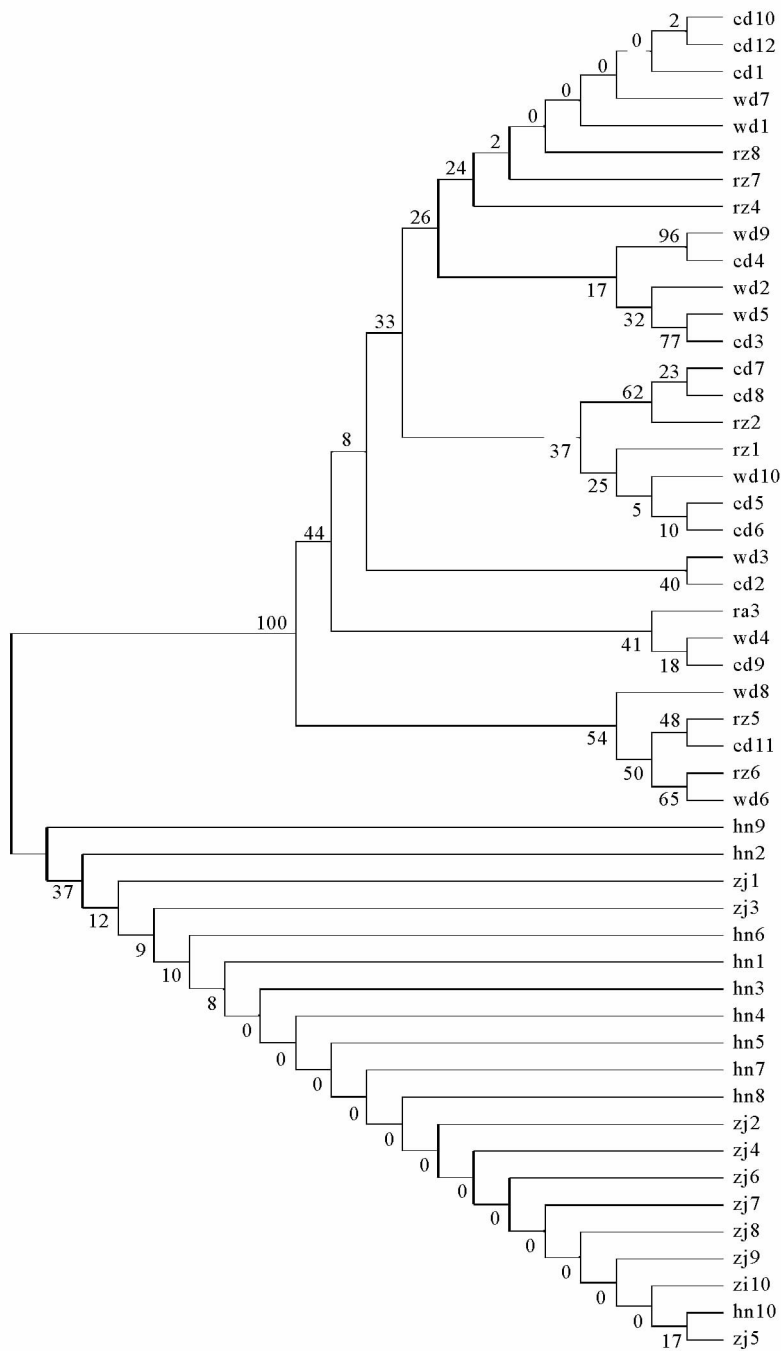


图 2 基于 COI 序列构建的 NJ 分子系统树

Fig. 2 NJ tree of COI gene based on genetic distance among the *A. pectinata* populations

认为南北方群体之间的遗传分化程度已经达到将南北群体分为两种栉江珧亚种的程度,课题组将对南北方各群体进行形态差异与判别分析,从形态学方面进一步证明种群分化程度。

就栉江珧遗传育种而言,应该尽可能选用遗传多样性水平高和遗传距离大的群体,虽然海南、湛江群里的遗传多样性水平不高,但由于地理隔离造成的基因交流的匮乏,南北方之间的遗传差异在选择育种亲本时不容忽视。因此选用北方群体和南方群体的亲本组合例如文登与海南群体组合,将会增加亲本的遗传多样性水平,进而增加子代产生优良性状的几率。

目前 GenBank 中只收录了存在的 8 种江珧科生物中的 4 种的 28S rRNA 序列,因此要更为系统地分析江珧科进化关系还需要进一步的采样、实验。COI 数据分析得到的变异适中,较好地阐明了 5 个地理群体野生栉江珧的遗传现状。希望这些工作为我国栉江珧优良种质的筛选提供一定的理论基础。

参 考 文 献

- 于瑞海,王昭萍,王世党,李琪,郑小东,周维武. 2007. 栉江珧工厂化育苗技术研究. 中国海洋大学学报, 37(5): 704-708
- 马朋,刘萍,李健,李吉涛,陈萍. 2011. 脊尾白虾 3 个野生群体线粒体 COI 基因的遗传多样性及其系统发育分析. 渔业科学进展, 32(6): 50-56
- 牛东红,李家乐,沈和定,姜志勇. 2008. 缢蛏六群体线粒体 DNA-COI 基因序列变异及群体遗传结构分析. 海洋学报(中文版), 30(3): 109-111
- 王梅芳,余祥勇,杨书婷,桂建芳. 2000. 无裂栉江珧种内同工酶表型差异的比较研究. 热带海洋, 19(4): 45-50
- 王玲,陈诚,沈和定,吴文健,魏峦峦. 2010. 基于 28SrDNA 部分序列的石磺科系统发育研究. 生物技术通报, 6: 172-179
- 王明昌,张筠,李妍,侯林,隋娜,董长永. 2008. 中国沿海蛾螺科 5 属 10 种 28S rRNA 基因的系统学分析. 动物学报, 54(5): 814-821
- 孙宝超,孙国华,刘相全,刘丽娟,王卫军,杨建敏,郑小东. 2010. 中国沿海长蛸(*Octopus variabilis*) 自然群体线粒体 COI 基因遗传多样性研究. 海洋与湖沼, 41(2): 259-265
- 余祥勇,王梅芳,杨书婷,桂建芳. 1999. 有棘和无棘两种表型栉江珧同工酶差异的比较. 湛江海洋大学学报, 19(2): 6-8
- 张红云. 2010. 栉江珧繁殖生物学及人工苗种繁育技术研究. 见:集美大学硕士研究生学位论文
- 俞晓平,董胜张,潘颖瑛. 2009. 福寿螺 18S rRNA 和 28S rRNA 基因片段的克隆与进化分析. 植物保护, 35(1): 33-36
- Aljanabi SM, Martinez I. 1997. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. Nucleic Acids Research 25(22): 4692-4693
- Canapa A, Barucca M, Marinelli A, Olmo E. 2000. Molecular data from the 16S rRNA gene for the phylogeny of Pectinidae (Mollusca: Bivalvia). Journal of Molecular Evolution 50(1): 93-97
- Excoffier L, Laval G, Schneider S. 2005. Arlequin ver. 3. 0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 47-50
- Hedgecock D, Li G, Banks MA, Kain Z. 1999. Occurrence of the Kumamoto oyster *Crassostrea sikamea* in the Ariake Sea, Japan. Marine Biology 133(1): 65-68
- Librado P, Rozas J. 2009. DnaSPv5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451-1452
- Martinez-Navarro EM, Galián J, Serrano J. 2005. Phylogeny and molecular evolution of the tribe Harpalini (Coleoptera, Carabidae) inferred from mitochondrial cytochrome-oxidase I. Molecular Phylogenetics and Evolution 35(1): 127-146
- Shih HT, Suzuki H. 2008. Taxonomy, phylogeny and biogeography of the endemic mudflat crab *Helicel/Chasmagnathus* complex (Crustacea: Brachyura: Varunidae) from East Asia. Zoological Studies 47(1): 114-125
- Swofford DL. 2002. PAUP* Phylogenetic Analysis Using Parsimony and other methods Version 4.0 beta version
- Tamura K. 2011. MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution 28(10): 2731-2739
- Thompson JD, Higgins DG, Gibson TJ. 1994. ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research 22(22): 4673-4680