

基于微卫星标记的地理群体仿刺参 遗传结构判别分析模型的建立

高银雪, 常亚青*, 庞震国, 白雪秋, 姬南京

(大连海洋大学农业部北方海水增养殖重点实验室, 辽宁 大连 116023)

摘要:通过主成分判别分析法(discriminant analysis of principal component, DAPC)建立仿刺参个体水平地理亚群体来源的判别模型。实验运用了 20 个微卫星引物,对俄罗斯海参崴(RS)和中国大连两个亚群(CN1、CN2)3 个仿刺参地理群体进行了产物扩增和标记分型,获得 3 个地理群体分子标记表型数据。开创性地运用主成分分析和判别分析法建立判别模型,以研究不同群体仿刺参个体水平上遗传背景及其群体来源。结果显示,运用 20 对微卫星引物在仿刺参不同地理群体中稳定扩增的 136 个等位基因位点数据建立的判别模型可以准确判别刺参个体的群体来源。用 75% 的原有数据集建立模型,剩余的 25% 数据对模型的准确度进行验证,3 个群体时模型准确率大于 80%,2 个群体时模型准确性超过 90%。传统的遗传多样性和遗传结构分析显示,刺参的 3 个地理群体间存在中度的遗传分化,且遗传距离较近($D_1 = 0.18$, $D_2 = 0.159$)。传统分析方法和新方法结果良好的一致性表明,利用 DAPC 和微卫星等位基因信息建立的刺参地理群体判别模型,可以很好地用于未知遗传背景刺参种质的来源鉴别。

关键词:仿刺参;杂交育种;遗传结构;微卫星标记;判别模型;主成分判别分析

中图分类号: S 966.1

文献标志码: A

刺参(*Apostichopus japonicus* Selenka)又称仿刺参,隶属于棘皮动物门(Echinodermata)、海参纲(Holothuroidea)、楯手目(Aspidochirota),是一种经济价值较高的海洋无脊椎动物,广泛分布于太平洋西部的中国、俄罗斯、韩国与日本沿海,在中国产于辽宁、山东、河北等北方沿海^[1]。目前随着刺参增养殖技术的成熟和产业规模化,种质遗传改良已经成为促使刺参养殖产业可持续性发展的重要手段和任务。其中地理远源杂交育种因为更容易使后代个体获得杂种优势,已在许多水产养殖品种改良中得到广泛应用^[2-3]。Chang 等^[4]培育的第一个刺参养殖品种“水院 1 号”具有生长速度快,刺多等杂种优势。因此,充分了解刺参不同地理群体的遗传结构差异,对于刺参种质资源评估和杂交育种工作具有重要意义。近几年,许多学者分别使用 ISSR、同工酶、微卫星遗传

标记、线粒体 DNA 等技术结合传统遗传学分析对刺参的不同群体进行了遗传多样性分析^[5-7]。但这些研究仅仅是分析了群体水平的遗传多样性,并未涉及到个体水平遗传距离的评估和个体遗传背景的推断。

主成分分析是一种采用“降维”思想,将多个变量结合成几个综合变量的分析方法。将原始变量重新转化成线性无关的新变量,按照协方差成分的大小排序,选取能够解释大部分协方差的变量作为主成分变量^[8]。主成分分析广泛应用于医学、化学和生态学等领域^[9-10]。将主成分分析后获得的新变量作为因变量进行线性判别分析的方法称为主成分线性判别分析,近年来主成分线性判别分析逐渐成为遗传学领域研究的有利工具。Paul 等^[11]利用主成分判别分析对芦苇的本土原种、引进种及其杂交种的

收稿日期:2014-06-24 修回日期:2014-07-18

资助项目:国家“八六三”高技术研究发展计划(2012AA10A412);渔业新品种引进、开发(2011203003);辽宁省农业攻关及成果转化重大项目(201401249)

通信作者:常亚青, E-mail: yqchang@dlou.edu.cn

<http://www.scxuebao.cn>

遗传结构进行了研究,为杂交种的遗传背景提供了有力的依据;Jombart 等^[8]利用流感病毒 H3N2 病原菌的 SNP 分子标记表型数据,运用该方法对病原菌在时间上的遗传变化规律进行了研究。

微卫星标记(microsatellite markers)具有多态性丰富、重复性好、共显性等特点,已被广泛地应用于水产动物的亲本识别、遗传育种计划、遗传图谱构建等研究领域^[12-13]。本研究运用开发的 20 对微卫星引物产生的多态性信息对刺参的两个地理群体的遗传多样性进行相关分析,开创性地运用主成因分析和线性判别法建立个体种质来源的判别模型,以期对刺参人工养殖及良种选育提供相关的遗传背景资料,同时也为刺参种质资源持续利用提供相关的依据。

1 材料与方法

1.1 实验材料

实验所用的刺参地理群体样品分别取自俄罗斯海参崴(131.55°E, 43.08°N)和中国大连(121.73°E, 39.03°N; 121.23°E, 38.76°N),数量分别为 85、25、30 头。分别取个体管足,液氮速冻后保存于 -80℃ 冰箱。

1.2 基因组 DNA 的提取

按照 TIANamp Marine Animals DNA Kit(北京天根)试剂盒说明书提取刺参基因组总 DNA,1% 琼脂糖凝胶电泳检测, -20℃ 保存备用。

1.3 微卫星扩增及产物大小判定

实验所用的微卫星引物来自本实验室通过 Illumina 测序构建的刺参转录组数据库,利用 Primer 3.0 设计引物,送至上海生工生物公司合成(附表 1)。PCR 体系为 20 μL: DNA 模板 1.6 μL, 10 mmol/L 上下游引物各 1.6 μL, 2.5 mmol/L dNTP 1.6 μL, 10 × PCR Buffer(含 Mg^{2+}) 3.2 μL, 5 U/μL Taq 酶(TaKaRa) 0.15 μL, ddH₂O 补充至 20 μL。PCR 反应条件为 94℃ 预变性 3 min; 94℃ 变性 30 s, 60℃ 退火 30 s, 72℃ 延伸 30 s, 29 个循环;最后 72℃ 终延伸 10 min。PCR 产物用 8% 非变性聚丙烯酰胺凝胶检测,硝酸银染色,拍照。以 TaKaRa 公司的 50 bp DNA ladder 为标准,利用 Gel-pro32 凝胶图像处理系统处理电泳图片,获得各个泳道中 PCR 产物的片段大小并输出数据。

1.4 数据分析

基本遗传信息分析(传统遗传结构分析)

利用 Microchecker v2.2.3^[14]检测 20 个微卫星位点在刺参不同地理群体中是否存在无效等位基因,GenAlEx^[15]计算 20 个位点在 3 个群体中的 Hardy-Weinberg 平衡指数、观测杂合度和期望杂合度等遗传信息以及群体间的(Nei's)遗传距离,并进行了 AMOVA 分析。

数据点转换 以 20 个微卫星位点产生的 136 个等位基因作为基本变量向量,将全部 140 个个体在变量中出现条带的情况作为数据点。未出现条带记为 0,将某等位基因位点出现条带且为杂合的记为 0.5,将等位基因位点出现条带且为纯合记为 1,共产生 136 × 140 约 1.5 万个数据点。

DAPC 分析 采用 R 软件平台,运行 adegenet 数据分析模块,调用 inbreeding 和 seppop 命令对不同群体进行近交系数正态性检验;运用贝叶斯信息参数(Bayesian information criterion, BIC)得到最佳聚类数目与实际群体数目比较,用以评定原始数据的可靠性。使用主成因判别分析法(discriminant analysis of principal component, DAPC)进行刺参群体的判别分析^[16]。利用 dapc 命令进行主成因线性判别分析,设置参数为主成因因子数 $n \leq N/3$ (N 为总的个体数),使用 scatter 命令得到判别分析的结果。最后对建立的模型进行准确性评估。

2 结果

2.1 传统遗传多样性分析

刺参不同地理群体的遗传多样性指数显示,在 3 个刺参群体中,20 个位点共检测出了 136 个等位基因。每个位点的等位基因数从 3 ~ 13 不等,其中 C1_Unigene_BMK.22951 位点等位基因数最多,为 13 个,表现出高度多态性(附表 2)。俄罗斯刺参群体平均等位基因数与中国两个群体的有所差别,分别为 6.4(RS)、5.2(CN1)、5.35(CN2),总的平均等位基因数为 5.65;RS 群体的平均观测杂合度($H_o = 0.541$)与平均期望杂合度($H_e = 0.651$)最大,3 个群体总的平均观测杂合度 H_o 为 0.532,平均期望杂合度 H_e 为 0.633。结果显示,俄罗斯海参崴群体的遗传多样性在 3 个群体中最高。

2.2 群体遗传结构分析

78% 的遗传变异存于个体间,6% 存在群体间,16% 变异来自群体内,平均的 F_{st} 值为 0.062 ($P < 0.001$) (表 1)。基于 20 个微卫星位点的等位基因频率计算刺参地理群体间的 Nei's 遗传距

离和遗传相似性系数(表 2)。RS 刺参群体与 CN1、CN2 群体的遗传距离相近($D_1 = 0.18, D_2 = 0.159$),遗传相似性系数分别为 0.835 和 0.853; CN1 与 CN2 遗传距离最小($D = 0.074$),遗传相似性系数最大($I = 0.929$)。

表 1 刺参群体差异的 AMOVA 分析
Tab.1 Analysis of molecular variance (AMOVA) among populations of *A. japonicus*

变异来源 source of variation	自由度 df	平方和 sum of squares	方差组分 variation of components	变异百分比/% variation	固化指数 fixation indices	P
种群间 among populations	2	81.740	0.429	6		
种群内 within populations	137	1 039.945	1.103	16		
个体间 within individuals	140	753.900	5.385	78		
总计 total	279	1 875.585	6.917	100	$F_{st}: 0.062$	0.001

表 2 刺参不同群体的 Nei's 遗传相似性
(对角线上方)和遗传距离(对角线下方)

Tab.2 Nei's genetic identity (above diagonal) and genetic distance (below diagonal) in *A. japonicus*

群体 population	俄罗斯海参崴 RS	中国大连 1 CN1	中国大连 2 CN2
RS		0.853	0.835
CN1	0.159		0.929
CN2	0.180	0.074	

2.3 判别分析

Microchecker v2.2.3 软件检测显示 20 个微卫星标记在三个刺参群体中均不存在无效等位基因。利用 R 软件对刺参不同群体的近交系数进行评估(图 1),俄罗斯群体近交系数符合正态分布,中国大连群体偏正态严重,其中,中国大连群体 2 呈现明显两极分化,表现出养殖行为对近海刺参群体近交系数的影响。

判别分析时使用的主成分变量数目为 45,所对应的累计方差变量组分为 89.3% (图 2-a)。应用 DAPC 内置 EM 算法计算贝叶斯信息参数 (bayesian information criterion, BIC),结果表明,最佳的聚类数为 2~3 ($BIC = 288.398\ 4, BIC = 289.068\ 9$,图 2-b)。

20 个等位基因位点在不同群体间的频率分析显示,位点 C1_Unigene_BMK. 15555、C1_Unigene_BMK. 6427、T1_Unigene_BMK. 12964、C1_Unigene_BMK. 7484、C1_Unigene_BMK. 54725 和 C1_Unigene_BMK. 6860 的贡献率较大,其中 T1_Unigene_BMK. 12964、C1_Unigene_BMK. 7484 及 C1_Unigene_BMK. 6860 的贡献率

均在 0.06 以上,为模型的建立提供了重要的理论基础(图 3)。

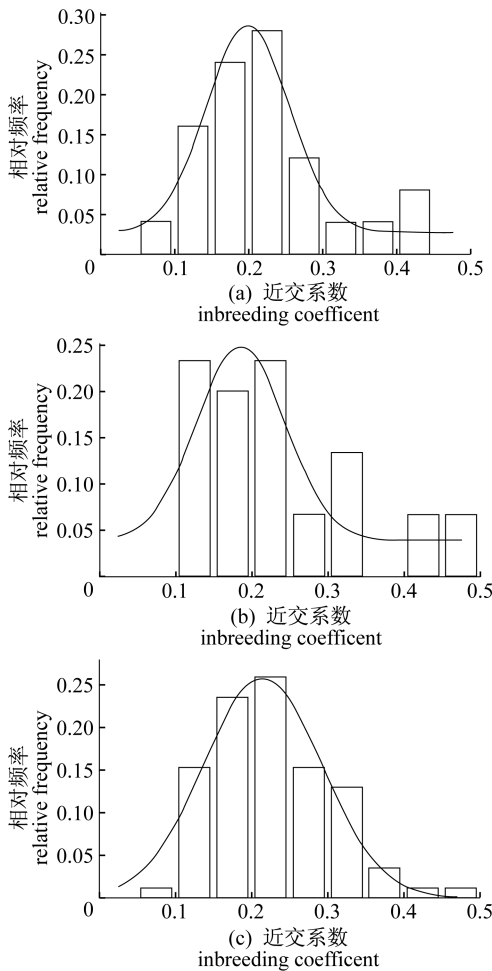


图 1 刺参不同群体近交系数分布图
(a) 中国群体 1, (b) 中国群体 2, (c) 俄罗斯海参崴群体
Fig.1 Average inbreeding coefficients in different populations of *A. japonicus*
(a) CN1, (b) CN2, (c) RS

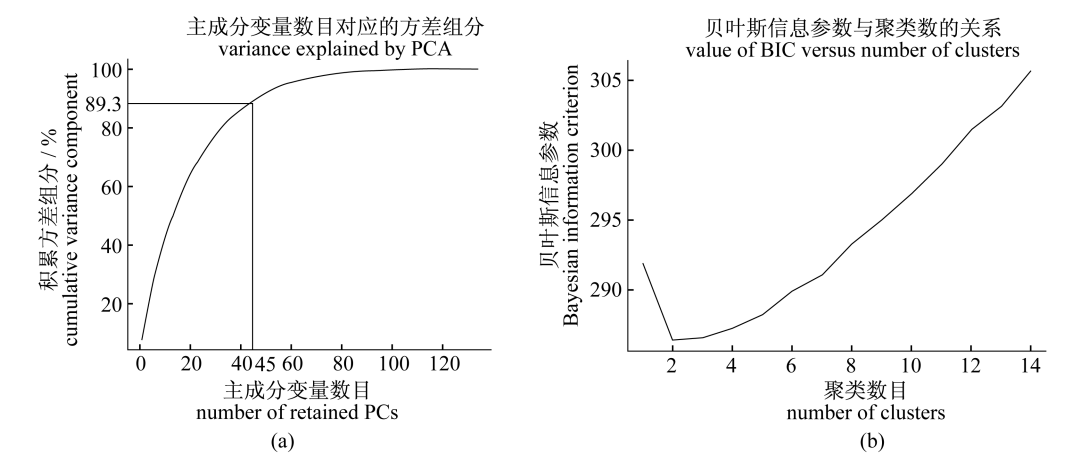


图 2 主成分判别分析遗传结构的主成分和聚类结果

(a) 累计方差组分变化趋势表明最佳的主成分变量为 45 个, (b) 贝叶斯标准值显示遗传聚类数目为 2 ~ 3

Fig. 2 PCA and BIC results using the DAPC to investigate genetic structure

(a) The optimal principal component amount found for the analysis was 45 based on cumulative variance component curve, (b) The BIC (Bayesian Information Criterion) values corresponding to different genetic clusters are shown

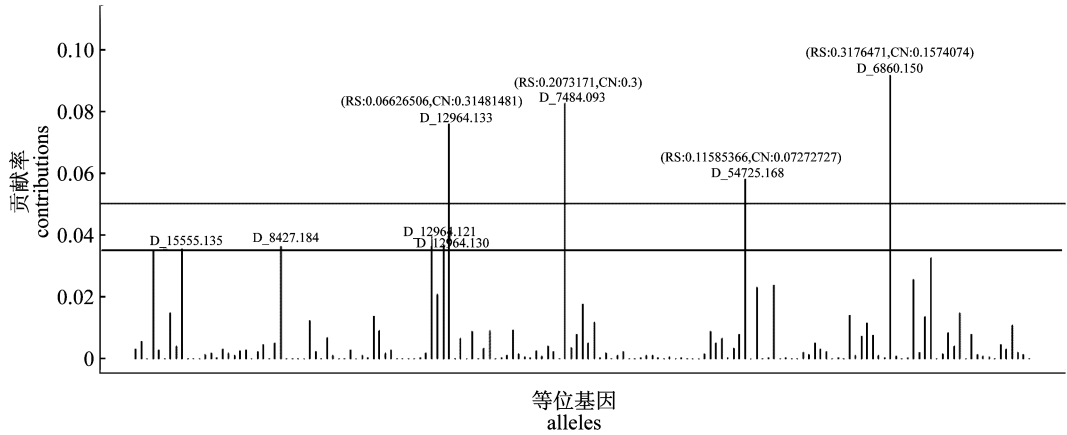


图 3 20 个等位基因在仿刺参不同群体中的贡献率

贡献率大于 0.05 的等位基因在不同群体间的频率已在图中标出

Fig. 3 Contributions of 20 alleles of *A. japonicus* in different populations

The alleles that contributions above 0.05 have been marked

DAPC 分析结果表明,将遗传聚类数设为 3 时,绝大部分刺参个体通过 DAPC 方法都可以被准确分类到原始群体(图 4-a)。特别是在俄罗斯群体中,100% 的个体被正确地判别到俄罗斯海参

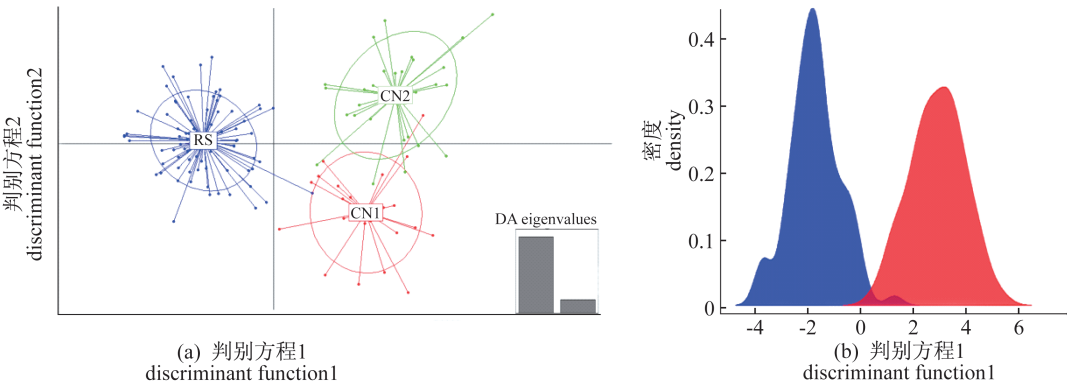


图 4 仿刺参群体主成分判别分析散点图 (a: n = 3, b: n = 2)

Fig. 4 Scatter plots of the DAPC for *A. japonicus* (a: n = 3, b: n = 2)

崴群体;中国大连两个群体分析结果显示,有部分刺参被聚类到了与其遗传距离较近的群体;而遗传聚类数为 2 时,该方法几乎可以对所有刺参个体进行完全的判别(图 4-b)。

2.4 判别模型验证

随机抽取原有数据集的 75% 建立判别模型,利用剩余的 25% 数据对模型的准确度进行验证。将原有数据分为 3 个群体的模型,验证准确率大于 80%;而对两个群体的模型准确性达到了 90% 以上,这说明了本研究建立的模型具备较高的稳定性和准确性。

3 讨论

作为种间遗传研究的一种标记方法,微卫星分子标记已在刺参种间研究中具有一定的适用性。在本研究之前,Kanno 等^[5]在仿刺参基因组中筛选出 20 个微卫星分子标记并用于分析日本沿海不同颜色仿刺参群体的遗传结构;孙国华等^[17]建立了从 EST 文库中筛选仿刺参微卫星分子标记的方法。本实验设计并筛选了 20 对微卫星引物,在 3 个刺参群体中均可稳定地扩增,产生了 136 个等位基因位点。这 20 个微卫星位点在刺参种内具有很好的遗传多样性和可信度。

等位基因数目(N_a)、遗传杂合度(H_e)等遗传参数均可反映群体的遗传多样性^[18]。种群内遗传多样性程度是进行种质资源分析的基础。本实验中 RS、CN1 和 CN2 刺参群体的平均等位基因数为 5.2~6.4,平均期望杂合度为 0.62~0.651。

基因固定指数(F_{st})常用于评价群体遗传分化程度,其标准为低度分化: $F_{st} < 0.05$,中等分化: $0.05 < F_{st} < 0.15$,高度分化: $0.15 < F_{st} < 0.25$,极高度分化: $0.25 < F_{st} < 1.00$ ^[19]。本实验利用 20 个微卫星多态位点的数据分析了 3 个不同刺参地理群体的 F_{st} ,其平均值为 0.062,即刺参的 3 个地理群体间存在中度的遗传分化。利用 AMOVA 对刺参不同地理群体的遗传分化分析表明,有 6% 的遗传变异来自种群间,同时俄罗斯海参崴群体与中国两个群体间的遗传距离较近。这与潘传燕等^[20]利用微卫星标记对中国、韩国、俄罗斯等不同地理仿刺参群体进行遗传多样性水平和遗传结构分析的研究结果一致。

判别分析与主成分分析都是通过多个变量转化为几个综合变量的统计方法,已广泛地应用

到水产动物的形态差异、生长体征等方面的研究中^[21-22]。主成分分析法在群体遗传结构分析中主要分析个体间的遗传变异,而判别分析则能够对群体间变异进行区分,结合以上两种分析的优点产生了主成分判别分析。DAPC 和微卫星 DNA 多态性,即以微卫星 DNA 在不同群体中等位基因频率的不同为基础,通过将不同位点的等位基因频率转化成若干代表各等位基因非线性组合的主成分,从而通过分析个体的基因频率便可进行个体遗传背景的判别。DAPC 从遗传学角度来分析群体遗传结构,是一种快速灵活的方法。近几年,已有很多研究把 DAPC 作为一种分析群体遗传学和系统发育学的工具,并且在海洋生物中已有应用。如 Morgan 等^[23]利用 DAPC 分析了在澳大利亚塔斯马尼亚海区分布的龙虾(*Jasus edwardsii*)的群体遗传结构;Vignaud 等^[24]通过微卫星标记对黑鳍礁鲨(*Carcharhinus melanopterus*)进行了遗传结构分析,该研究表明利用 DAPC 进行群体鉴别效率较高。

本实验使用 20 对微卫星引物对刺参不同地理群体进行 PCR 扩增,获得 3 个群体的分子标记表型数据,开创性地应用 DAPC 方法对群体进行判别分析,从而对个体遗传背景进行分析,进而建立对其群体来源进行判别的有效方法。DAPC 判别分析显示,刺参不同地理群体分为 2 个群体,其中中国群体又可分为 2 个亚群,两个中国群体与俄罗斯群体的遗传距离几乎相同。模型判定准确度验证结果也证明,DAPC 对刺参个体的群体来源判别具有很高的准确性和可靠性。

综上所述,本实验利用 DAPC 和 20 个微卫星标记数据对刺参群体的等位基因信息进行分析,可用于未知遗传背景的刺参群体鉴别。该方法的建立可以为刺参的种质资源保护和养殖种类的遗传改良提供重要的理论基础和实际应用价值。

参考文献:

- [1] Chang Y Q, Ding J, Song J, et al. Biology and aquaculture of sea cucumbers and sea urchins[M]. Beijing: China Ocean Press, 2004. [常亚青,丁君,宋坚,等. 海参,海胆生物学研究与养殖. 北京: 海洋出版社, 2004.]
- [2] Ding J, Chang Y, Wang C, et al. Evaluation of the growth and heterosis of hybrids among three commercially important sea urchins in China:

<http://www.sxuebao.cn>

- Strongylocentrotus nudus*, *S. intermedius* and *Anthocardia crassispina* [J]. *Aquaculture*, 2007, 272 (1): 273 – 280.
- [3] Yu F, Zhang Q W, Kong J, *et al.* Growth of juvenile hybrids from different crossing combinations of the imported *Scophthalmus maximus* L [J]. *Journal of Fisheries of China*, 2008, 32 (1): 58 – 64. [于飞, 张庆文, 孔杰, 等. 大菱鲆不同进口群体杂交后代的早期生长差异. *水产学报*, 2008, 32 (1): 58 – 64.]
- [4] Chang Y Q, Shi S B, Zhao C, *et al.* Characteristics of papillae in wild, cultivated and hybrid sea cucumbers (*Apostichopus japonicus*) [J]. *African Journal of Biotechnology*, 2011, 10 (63): 13780 – 13788.
- [5] Kanno M, Li Q, Kijima A, *et al.* Isolation and characterization of twenty microsatellite loci in Japanese Sea cucumber (*Stichopus japonicus*) [J]. *Marine Biotechnology*, 2005, 7 (3): 179 – 183.
- [6] Yao B, Hu X L, Bao Z M, *et al.* Genetic variation in two sea cucumber (*Apostichopus japonicus*) stocks revealed by ISSR markers [J]. *Chinese Journal of Oceanology and Limnology*, 2007, 25 (1): 91 – 96.
- [7] Ji N J, Chang Y Q, Zhao C, *et al.* D-loop sequences variation and phylogenetic analysis in three geographical populations of the sea cucumber *Apostichopus japonicus* [J]. *Journal of Fisheries of China*, 2014, 38 (4): 476 – 481. [姬南京, 常亚青, 赵冲, 等. 3 个地理群体仿刺参 D-loop 序列的变异及系统发生分析. *水产学报*, 2014, 38 (4): 476 – 481.]
- [8] Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations [J]. *BMC Genetics*, 2010, 11 (1): 94.
- [9] Segil J L, Weir R F F. Design and validation of a morphing myoelectric hand posture controller based on principal component analysis of human grasping [J]. *Neural Systems and Rehabilitation Engineering*, *IEEE Transactions on*, 2014, 22 (2): 249 – 257.
- [10] Tafarroj M M, Kalani H, Moavenian M, *et al.* An application of principal component analysis method in wood defects identification [J]. *Journal of the Indian Academy of Wood Science*, 2014: 1 – 6.
- [11] Paul J, Vachon N, Garroway C J, *et al.* Molecular data provide strong evidence of natural hybridization between native and introduced lineages of *Phragmites australis* in North America [J]. *Biological Invasions*, 2010, 12 (9): 2967 – 2973.
- [12] Li X Y, Sun G H, Yang J M, *et al.* Parentage identification using microsatellite DNA in artificial breeding of *Apostichopus japonicus* [J]. *Oceanologia et Limnologia Sinica*, 2013, 44 (5): 1263 – 1269. [李雪燕, 孙国华, 杨建敏, 等. 微卫星标记在刺参 (*Apostichopus japonicus*) 人工繁育中亲本识别的应用. *海洋与湖沼*, 2013, 44 (5): 1263 – 1269.]
- [13] Chang Y, Feng Z, Yu J, *et al.* Genetic variability analysis in five populations of the sea cucumber *Stichopus* (*Apostichopus japonicus*) from China, Russia, South Korea and Japan as revealed by microsatellite markers [J]. *Marine Ecology*, 2009, 30 (4): 455 – 461.
- [14] van Oosterhooft C, Hutchinson W, Wills D, *et al.* MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data [J]. *Molecular Ecology Notes*, 2004, 4 (3): 535 – 538.
- [15] Peakall R O D, Smouse P E. GENALEX 6: genetic analysis in Excel Population genetic software for teaching and research [J]. *Molecular Ecology Notes*, 2006, 6 (1): 288 – 295.
- [16] Jombart T. adegenet: a R package for the multivariate analysis of genetic markers [J]. *Bioinformatics*, 2008, 24 (11): 1403 – 1405.
- [17] Sun G H, Yang J M, Song Z L, *et al.* Isolation of microsatellite markers from *Apostichopus japonicus* ESTs [J]. *Oceanologia Et Limnologia Sinica*, 2010, 41 (1): 133 – 139. [孙国华, 杨建敏, 宋志乐, 等. 刺参 (*Apostichopus japonicus*) EST 序列中微卫星分布分析及其标记的筛选. *海洋与湖沼*, 2010, 41 (1): 133 – 139.]
- [18] Yang H, Li D Y, Cao X, *et al.* Genetic potential analysis of six *Tilapia* populations by microsatellite DNA markers [J]. *Hereditas*, 2011, 33 (7): 768 – 775. [杨弘, 李大宇, 曹祥, 等. 微卫星标记分析罗非鱼群体的遗传潜力. *遗传*, 2011, 33 (7): 768 – 775.]
- [19] Balloux F, Lugon-Moulin N. The estimation of population differentiation with microsatellite markers [J]. *Molecular Ecology*, 2002, 11 (2): 155 – 165.
- [20] Pan C Y, Zang Y P, Liao M J, *et al.* Development of microsatellite DNA markers of sea cucumber *Apostichopus japonicus* and their application in population genetic variation and structure analysis [J]. *Progress in Fishery Sciences*, 2012, 33 (4): 72 – 82. [潘传燕, 臧云鹏, 廖梅杰, 等. 仿刺参微卫星标记的筛选及群体遗传结构分析. *渔业科学进展*, 2012, 33 (4): 72 – 82.]

- [21] Tang Z Y, Xiao J, Li L P, *et al.* Principal component and discriminant analyses of traits of Nile tilapia (*Oreochromis niloticus*) at different ages [J]. *Oceanologia Et Limnologia Sinica*, 2013, 43 (2): 288 – 293. [唐瞻杨, 肖俊, 李莉萍, 等. 尼罗罗非鱼 (*Oreochromis niloticus*) 不同月龄性状的主成分与判别分析. 海洋与湖沼, 2013, 43(2): 288 – 293.]
- [22] Zhao X Q, Ni J, Chen L Q, *et al.* Analysis of morphological variations among four populations of *Macrobrachium nipponensis* [J]. *Journal of Fishery Sciences of China*, 2006, 13 (2): 224 – 229. [赵晓勤, 倪娟, 陈立侨, 等. 日本沼虾 4 种群的形态差异分析. 中国水产科学, 2006, 13(2): 224 – 229.]
- [23] Morgan E M J, Green B S, Murphy N P, *et al.* Investigation of genetic structure between deep and shallow populations of the southern rock lobster, *Jasus edwardsii* in Tasmania, Australia [J]. *PLoS One*, 2013, 8(10): e77978.
- [24] Vignaud T, Clua E, Mourier J, *et al.* Microsatellite analyses of blacktip reef sharks (*Carcharhinus melanopterus*) in a fragmented environment show structured clusters [J]. *PLoS One*, 2013, 8(4): e61067.

Linear discriminant model construction based on microsatellite alleles for geographic distant populations of sea cucumber (*Apostichopus japonicus*)

GAO Yinxue, CHANG Yaqing*, PANG Zhenguo, BAI Xueqiu, JI Nanjing

(Key Laboratory of Mariculture & Stock Enhancement in North China's Sea, Ministry of Agriculture, Dalian Ocean University, Dalian 116023, China)

Abstract: A Discriminant Analysis of Principal Components (DAPC) method was adopted to construct a germplasm resource prediction model at individual level, for the sea cucumber *Apostichopus japonicus*. One hundred and thirty six alleles generated by twenty microsatellite markers were used as the raw data input. Three populations were sampled from Vladivostok, Russia (RS, 131.55°E, 43.08°N) and Dalian, China (CN1, CN2). Principal Component Analysis (PCA) and Discriminant Analysis (DA) were carried out sequentially. Linear Discriminant functions were constructed with the two China subpopulations as one overall CN population and two separated populations respectively. The results showed that with 75% individuals as training set and the remaining 25% as test set, the accuracy of constructed model reached up to about 80% for the three-pop model and over 90% for the two-pop model. Conventional genetic diversity and genetic structure analysis showed moderate genetic differentiations among the three geographic populations of sea cucumber, and relatively close genetic distances ($D_1 = 0.18$, $D_2 = 0.159$). Nice accordance of results of conventional analysis and our novel analysis indicated that the current method can be practically well applied on prediction and discrimination on germplasm resource at individual level. To our knowledge, this is the first application of DAPC with molecule marker information to genetic study. The novel method built in the current study can be well applied to the practical breeding project in future.

Key words: *Apostichopus japonicus*; hybridization; genetic structure; microsatellite markers; prediction model; DAPC

Corresponding author: CHANG Yaqing. E-mail: yqchang@dlou.edu.cn

附表 1 微卫星引物的相关信息

位点 locus	引物序列 primer sequence(5'-3')	重复序列 repeat motif	基因片段大小/bp size range	温度/℃ <i>T_m</i>
C1_Unigene_BMK. 15611	F:GAATGATGTTCCAGCCCTTG R:GGGCCTTTCACACAGAGAGA	(GA)6	135 ~ 213	60
C1_Unigene_BMK. 15555	F:AACCTCTCGAGTCATTCTCCA R:GAGGGAAGAGACCACCGATA	(CT)10	127 ~ 199	60
C1_Unigene_BMK. 6545	F:GGGTGTTGTAAGCCAAGGA R:GGGAGAGAGAAGCACATTCG	(AAG)5	105 ~ 147	60
C1_Unigene_BMK. 8427	F:CCTCCTCCTCCACTTCTTCA R:TGGTGGTGCTTTTTCATCAG	(TTC)5	163 ~ 184	60
C1_Unigene_BMK. 32064	F:GACCGGTAAGGGGATGATTT R:CTCGTCCTTCTCTCACCAC	(GA)6	114 ~ 134	60
T1_Unigene_BMK. 45838	F:AAGGCCTGGAAGGGTAGAC R:TCCCAGGTTCCGAACACTAA	(GA)6	144 ~ 192	60
T1_Unigene_BMK. 9323	F:TCCTTGGTGGGGTTGTTTTA R:GAATGTTTTGGTCGGCAAGT	(CAA)5	117 ~ 189	60
C1_Unigene_BMK. 25927	F:ACTCGGGAAGCTTGATCTGA R:ATGTAGGACTGGAGCCGTCA	(TTC)5	186 ~ 264	60
T1_Unigene_BMK. 12964	F:AAATCGAAAGAAGTGACCGAAT R:CATGTCAGCTCCATTTGCTG	(AAG)5	106 ~ 139	60
T1_Unigene_BMK. 14571	F:CGGAATGGCAACTTGAGAAT R:TGTGTGCTCAGTCAAGTGTGA	(GAA)5	156 ~ 171	60
C1_Unigene_BMK. 41767	F:CACGGGTATGTTAACTGCTT R:ATTTGCGCCATGAACATTCT	(TAA)8	82 ~ 109	60
C1_Unigene_BMK. 7874	F:ACTCCGCTTTTGAGGAGGAC R:AACCACAATAGACCGCATCC	(GT)7	85 ~ 109	60
C1_Unigene_BMK. 22951	F:TCGACTGTTCATTGGATTACCA R:CATGTTTCGGGGCAATACTT	(TG)6	126 ~ 172	60
C1_Unigene_BMK. 54725	F:ATCCCCATCCCCTCTTTACA R:CGGATCGCTCTTTTGATTCT	(AAG)5	138 ~ 175	60
C1_Unigene_BMK. 10620	F:GTGCCCAGTTGAGATGGTTT R:GAGCAAGCGGACTCATCAAC	(GA)6	88 ~ 98	60
C1_Unigene_BMK. 3141	F:TTCTACCCAGATTCGCCACT R:AAACAGGGCTCTGCTTTGAA	(TAA)6	94 ~ 121	60
T1_Unigene_BMK. 3789	F:CACAGAGCCGATATTTGCTT R:AACAGCTTGTTACGGGGCTA	(AG)7	83 ~ 99	60
C1_Unigene_BMK. 6860	F:ACCGTTAGAAAGGCGTGATG R:ATGGGTTGTCCAAACTCGTC	(GA)6	148 ~ 180	60
T1_Unigene_BMK. 5796	F:CATGGATCCACCTGGGTAGT R:CAACATTGGTCCAAGCCTTT	(AT)8	170 ~ 184	60
T1_Unigene_BMK. 3177	F:CACGTTTACATGCTGGTTGG R:GGCTCCTACCCCTAGACTTT	(TTG)5	139 ~ 154	60

附表 2 刺参三个野生地理群体的遗传多样性指数

位点 locus	N_m	RS						CN1						CN2					
		N_a	N_e	I	H_o	H_e	P	N_a	N_e	I	H_o	H_e	P	N_a	N_e	I	H_o	H_e	P
CI_Unigene_BMK.15611	7.277	4	2.462	1.078	0.500	0.594	0.000	4	2.700	1.134	0.320	0.630	0.001	5	1.972	0.999	0.138	0.493	0.000
CI_Unigene_BMK.15555	4.846	4	2.059	0.991	0.325	0.514	0.000	4	2.654	1.134	0.560	0.623	0.329	5	3.292	1.301	0.345	0.696	0.000
CI_Unigene_BMK.6545	1.527	6	2.805	1.228	0.584	0.643	0.043	5	2.341	1.174	0.625	0.573	0.650	6	1.804	0.983	0.333	0.446	0.078
CI_Unigene_BMK.8427	4.189	6	3.718	1.495	0.614	0.731	0.000	4	2.465	1.080	0.400	0.594	0.008	4	1.960	0.949	0.517	0.490	0.636
CI_Unigene_BMK.32064	4.162	3	1.964	0.710	0.554	0.491	0.400	2	1.724	0.611	0.440	0.420	0.812	2	1.998	0.693	0.700	0.499	0.028
TI_Unigene_BMK.45838	6.309	6	3.701	1.483	0.771	0.730	0.401	6	2.677	1.277	0.680	0.626	0.838	5	2.713	1.193	0.724	0.631	0.906
TI_Unigene_BMK.9323	9.291	6	2.017	0.945	0.321	0.504	0.000	3	1.576	0.671	0.440	0.366	0.575	3	1.850	0.762	0.433	0.459	0.627
CI_Unigene_BMK.25927	4.335	4	2.033	0.944	0.361	0.508	0.001	4	3.307	1.253	0.520	0.698	0.186	4	3.049	1.163	0.560	0.672	0.204
TI_Unigene_BMK.12964	2.469	7	3.645	1.565	0.530	0.726	0.000	6	3.264	1.385	0.320	0.694	0.000	6	4.324	1.565	0.414	0.769	0.001
TI_Unigene_BMK.14571	63.823	4	1.664	0.767	0.333	0.399	0.344	3	1.615	0.659	0.440	0.381	0.496	3	1.510	0.628	0.267	0.338	0.107
CI_Unigene_BMK.41767	7.512	6	4.074	1.497	0.688	0.755	0.017	8	5.342	1.808	0.960	0.813	0.551	9	5.240	1.892	0.621	0.809	0.000
CI_Unigene_BMK.7874	4.645	10	6.744	2.045	0.659	0.852	0.002	8	3.894	1.717	0.440	0.743	0.000	10	6.545	2.034	0.800	0.847	0.071
CI_Unigene_BMK.22951	23.127	13	5.044	1.943	0.798	0.802	0.994	7	3.846	1.616	0.800	0.740	0.901	9	3.490	1.568	0.621	0.713	0.996
CI_Unigene_BMK.54725	4.887	8	3.146	1.454	0.646	0.682	0.030	9	4.579	1.759	0.840	0.782	0.017	5	4.157	1.509	0.733	0.759	0.828
CI_Unigene_BMK.10620	17.644	5	3.025	1.275	0.603	0.669	0.008	4	3.071	1.231	0.640	0.674	0.865	4	3.814	1.364	0.793	0.738	0.269
CI_Unigene_BMK.3141	15.720	9	2.538	1.360	0.469	0.606	0.086	6	1.806	0.957	0.500	0.446	0.995	4	1.899	0.909	0.400	0.473	0.265
TI_Unigene_BMK.3789	10.178	7	4.981	1.691	0.590	0.799	0.000	7	3.027	1.316	0.640	0.670	0.000	6	3.885	1.524	0.517	0.743	0.000
CI_Unigene_BMK.6860	3.376	7	4.207	1.605	0.647	0.762	0.000	5	3.834	1.413	0.520	0.739	0.005	9	6.675	2.032	0.586	0.850	0.000
TI_Unigene_BMK.5796	8.911	7	3.501	1.450	0.588	0.714	0.321	4	2.197	0.940	0.120	0.545	0.000	3	2.198	0.856	0.533	0.545	0.004
TI_Unigene_BMK.3177	3.574	6	2.158	1.056	0.235	0.537	0.000	5	2.747	1.211	0.480	0.636	0.185	5	2.483	1.076	0.367	0.597	0.044
平均 mean	10.390	6.4	3.274	1.329	0.541	0.651		5.2	2.933	1.217	0.534	0.620		5.35	3.243	1.250	0.520	0.628	