

DOI: 10.19663/j.issn2095-9869.20180802001

http://www.yykxjz.cn/

于爱清, 施永海, 徐嘉波, 陆根海, 张海明, 谢永德, 刘永士. 长江刀鲚选育群体转录组EST-SSR的分布特征分析. 渔业科学进展, 2019, 40(5): 101-109

Yu AQ, Shi YG, Xu JB, Lu GH, Zhang HM, Xie YD, Liu YS. Characteristic analysis of microsatellites in selected *Coilia ectenes* using a transcriptome dataset. Progress in Fishery Sciences, 2019, 40(5): 101-109

长江刀鲚选育群体转录组 EST-SSR 的分布特征分析*

于爱清 施永海^① 徐嘉波 陆根海 张海明 谢永德 刘永士

(上海市水产研究所 上海市水产技术推广站 上海 200433)

摘要 本研究利用 MISA 软件挖掘长江刀鲚(*Coilia ectenes*)肌肉和肝脏转录组中的微卫星标记,为刀鲚选育群体的种质资源评估和分子标记辅助育种奠定基础。结果显示,从 71869 条 Unigenes 中共获得 33896 条重复单元长度为 1~6 碱基的微卫星序列;刀鲚转录组中不同类型微卫星的重复基序具有不同的分布特征,其中,单核苷酸重复、二核苷酸重复和三核苷酸重复为主要的微卫星重复类型,分别占总微卫星数目的 34.94%、49.47%和 13.34%;不同微卫星重复类型的优势重复基序亦有所不同,其中, A/T 为单核苷酸重复基序的优势重复基序占 86.25%, AC/GT 为二核苷酸重复基序的为优势重复基序占 75.25%, AGG/CCT 为三核苷酸重复基序的优势重复基序占 28.57%;不同微卫星重复基序核苷酸的数量和重复次数亦有所不同,重复次数伴随着重复单元中核苷酸数量的增加而呈现降低的趋势;从 100 对四核苷酸重复的 SSR 引物中筛选获得了 16 对多态性微卫星标记,并以此为基础,对长江刀鲚选育群体(F₃)的遗传学特征进行了初步评估,结果显示,长江刀鲚选育群体 F₃ 的平均有效等位基因数(N_e)、平均观测杂合度(H_o)、平均期望杂合度(H_e)和 Shannon 多样性指数 I 分别为 1.7580、0.3414、0.3977 和 0.6278。以上结果表明,基于刀鲚转录组数据批量开发微卫星是切实可行的,所开发的多态性微卫星标记能够应用于长江刀鲚选育群体的遗传背景评估和进一步的遗传育种研究。

关键词 刀鲚; 转录组; 微卫星; 分子标记

中图分类号 S937.3 文献标识码 A 文章编号 2095-9869(2019)05-0101-09

刀鲚(*Coilia ectenes*)隶属于鲱形目(Clupeiformes)、鲱科(Engraulidae)、鲚属(*Coilia*),又名长颌鲚,俗称刀鱼、毛刀鱼,主要分布于我国黄渤海、东海海域及

通海的江河,其中,又以长江下游的产量最高,因其鱼体丰腴肥满、肉质细嫩鲜美而享有“长江三鲜”之首的美誉(施永海等, 2014、2015; 邓平等, 2014)。历

* 上海市科技兴农重点攻关项目[沪农科攻字(2016)第 6-2-2 号]、上海市科学技术委员会重点科技攻关项目(17391900300)和上海市科学技术委员会重点科技攻关项目(11391901300)共同资助 [This work was supported by the Shanghai Key Scientific and Technological Project on Agriculture from Shanghai Municipal Agricultural Commission (2016-2-2), Shanghai Key Scientific Technological Project from Shanghai Science and Technology Commission (17391900300), and Shanghai Key Scientific Technological Project from Shanghai Science and Technology Commission (11391901300)]. 于爱清, E-mail: aiqingyu0125@163.com

① 通讯作者: 施永海, 教授级高级工程师, E-mail: yonghais@163.com

收稿日期: 2018-08-02, 收修改稿日期: 2018-08-29

史上,我国刀鲚的种质资源极为丰富。然而,近年来,由于酷捕滥渔、水域环境污染、水文条件改变、生态环境恶化及海岸工程建设等诸多因素,长江刀鲚自然种质资源急剧衰退,产量波动很大,且呈现出逐年降低且个体日趋小型化的趋势,甚至不能形成优势种群(施永海等, 2014、2015; 魏广莲等, 2013; 袁传宓, 1988),亟需对刀鲚的种质资源进行保护和恢复。而刀鲚基因组信息匮乏、遗传背景不清晰、繁殖周期较长等问题亦在一定程度上制约了其种质资源的开发和恢复。刀鲚种质资源的有效保护和恢复,除了政府相关部门采取相关行政干预(建立禁渔期、加强捕捞监管、适时增殖放流等)之外,还需相关部门适时地监测洄游型刀鲚的遗传多样性,从整体上把握刀鲚野生群体的遗传多样性水平和遗传背景,同时加快刀鲚良种的选育进程,特别是分子标记辅助良种选育的研究。

分子标记由于能够对动物整个基因组的遗传变异进行探究,且相对稳定,而被应用于水产动物的遗传育种研究,其中,微卫星标记(Microsatellite)的应用最为广泛。微卫星标记,又称为简单重复序列(Simple sequence repeat, SSR),是指由 1~6 个核苷酸组成的简单串联重复序列,这些序列的重复单位和重复次数的差异使其具有高度变异性和多态性,被广泛应用于水产动物种质资源评估和鉴定、遗传图谱的构建及分子标记辅助育种等方面(杨铭等, 2017; 管奥等, 2018; 白翠翠等, 2016)。近年来,学者利用线粒体基因 D-loop(张燕萍等, 2017; Cheng *et al.*, 2011)、线粒体 Cytb(魏广莲等, 2012; Ma *et al.*, 2010)、RAPD(马春艳等, 2004)、AFLP(葛家春等, 2008)等方法研究了刀鲚不同群体的遗传结构和遗传差异,而对刀鲚微卫星标记开发的报道相对较少。Rong 等(2013)利用 FIASO 法构建了微卫星富集文库方法,从 89 对刀鲚微卫星引物中获得了 20 对多态性微卫星标记。马春燕等(2011)利用 FIASO 法构建微卫星富集文库的方法,从 70 对刀鲚微卫星引物中获得了 12 对多态性微卫星标记。陈芳等(2012)利用磁珠富集法,从 198 对刀鲚微卫星引物中获得了 34 对多态性微卫星标记。邓平等(2014)利用磁珠富集法,从 59 对刀鲚微卫星引物中获得了 9 对多态性微卫星标记。利用磁珠富集法从水产动物基因组中开发的微卫星标记,多态性相对较高,但相对费时费力,效率相对较低,且标记大规模的开发成本相对较高,远远不能满足刀鲚高密度遗传图谱的构建、QTL 精细定位和分子标记辅助育种等对分子标记需求量较大的研究,亟需大规模开发分子标记。而高通量转录组测序技术所开发的微卫星,又被

称为表达序列标签 SSR (Expressed sequence tag SSR, EST-SSR),其多态性虽然相对低于基因组所开发的微卫星(G-SSR),但由于数量多、通量大、周期短、成本低,且直接与某些功能基因相关联(李东明等, 2017),而被广泛应用于水产动物分子标记大规模的开发研究。

本研究基于刀鲚肌肉和肝脏的高通量转录组数据,对长江刀鲚转录组中的微卫星位点进行筛选,分析刀鲚 EST-SSR 的序列特征、分布和组成等信息,并利用所筛选的多态性微卫星标记对长江刀鲚选育群体(F₃)的遗传学特征进行初步评估,以期能够为刀鲚遗传多样性的分析、种质资源的保护和利用、高密度遗传连锁图谱的构建及分子标记辅助良种选育奠定基础。

1 材料与方 法

1.1 实验材料与数据来源

实验所用长江刀鲚来自于上海市水产研究所核心选育群体(F₃)的 2 龄刀鲚,提取刀鲚肌肉和肝脏的 RNA 后,构建高通量转录组测序文库,利用 Illumina HiSeq4000 进行高通量转录组测序。刀鲚肌肉和肝脏的转录组测序总共获得 28.20 Gb 的数据量,原始测序数据经 Trinity 软件(<http://trinityrnaseq.sourceforge.net/>)组装并去冗余后,获得平均长度为 932.49 bp 的 Unigenes 71869 条, N50 达到 1364 bp,质量不低于 20 的碱基比例(Q20)为 97.31%,表明测序结果良好(未发表)。

1.2 EST-SSR 的筛查

以刀鲚转录组中的 Unigenes 作为参考序列,利用 MISA(MicroSAteellite identification tool)软件(<http://pgrc.ipk-gatersleben.de/misa/misa.html>)搜索刀鲚肝脏和肌肉转录组中的 SSR 位点,其中,单核苷酸、二核苷酸、三核苷酸、四核苷酸、五核苷酸和六核苷酸的最少重复次数分别设置 10、6、5、5、5 和 5。复合 SSR 两个位点间最大间隔碱基数设置为 100。

1.3 EST-SSR 多态性引物的设计、筛选和初步验证

利用 Primer Premier 3.0 软件对筛选出的微卫星序列批量设计引物,主要设置参数为:引物长度和 PCR 产物分别为 18~25 bp 和 100~500 bp, 55℃ ≤ 退火温度(T_m) ≤ 65℃, 40% ≤ GC 含量 ≤ 60%,上下游引物的 T_m ≤ 2℃。考虑到四碱基重复的微卫星标记相对二碱基重复和三碱基重复具有较长的演化历史和更加稳定的突变积累,本研究随机选取 100 对基于四核

苷酸重复微卫星序列所设计的 PCR 扩增引物交由生工生物工程(上海)股份有限公司合成, 并以刀鲚核心选育群体(F₃)为实验材料筛选具有多态性的微卫星引物。PCR 扩增产物经 10% 非变性聚丙烯酰胺凝胶电泳检测后, 利用银染法显色定影, 条带显色清晰后用数码相机拍照保存, 进行微卫星引物后续的多态性分析。利用 GenAlEx 6.0 软件(Peakall *et al*, 2006)计算长江刀鲚选育群体的有效等位基因数(N_e)、观测杂合度(H_o)、期望杂合度(H_e)、Shannon 多样性指数(I)等遗传多样性参数。参照 Bostein 等(1980)的计算公式计算多态信息含量(Polymorphism information content, PIC):

$$PIC = 1 - \sum_{i=1}^n p_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^n 2p_i p_j$$

式中, P_i 、 P_j 分别为群体中第 i 、 j 个等位基因频率, n 为等位基因数。

2 结果

2.1 刀鲚转录组中 EST-SSR 位点的数量与分布

利用 MISA 软件挖掘刀鲚转录组中的 EST-SSR 标记(搜索结果见表 1)。从刀鲚 71869 条 Unigenes 中共获得 33896 个 SSR 位点, 占所评估序列总数目的 47.16%; 22463 条 Unigenes 包含 SSR 位点, 占总 Unigenes 的 31.25%; 含有 1 个以上的 SSR 位点的 Unigenes 有 7701 个, 约占总 Unigenes 的 10.71%; 具有复合型 SSR 位点的 Unigenes 有 2989 个, 约占总 Unigenes 的 4.16%(表 1)。

2.2 刀鲚 EST-SSR 的基序类型和频率特征

基于微卫星重复单元的类型和重复次数, 将不同

表 1 刀鲚转录组中 EST-SSR 位点的搜索结果

Tab.1 EST-SSR search results in the transcriptome of *C. ectenes*

项目 Items	数据 Data
搜索序列总数目 Total number of sequences examined	71869
搜索序列总长度 Total size of examined sequences (bp)	67017196
识别的 SSR 总数 Total number of identified SSRs	33896
包含 SSR 的序列数目 Number of SSR containing sequences	22463
包含 1 个以上 SSR 的序列数目 Number of sequences containing more than 1 SSR	7701
包含复合型 SSR 的序列数目 Number of SSRs present in compound formation	2989

的 SSR 位点进行分类(表 2)。结果显示, 从单核苷酸重复 SSR 到六核苷酸重复 SSR 在刀鲚转录组中均有分布, 但数量分布上具有不同程度的差异。其中, 二核苷酸重复 SSR 数量最多, 占 SSR 位点总数的 49.47%; 单核苷酸重复 SSR 数量次之, 占 SSR 位点总数的 34.94%; 三核苷酸重复 SSR 的数量亦较为丰富, 占 SSR 总位点总数的 13.34%; 四核苷酸重复 SSR 的数量相对较少, 占 SSR 总位点总数的 2.11%; 五核苷酸和六核苷酸重复 SSR 的数量最少, 分别占 SSR 总位点总数的 0.09% 和 0.06%。

此外, 刀鲚转录组 EST-SSR 中共有 94 种核苷酸重复基序, 主要核苷酸重复基序亦有不同类型的优势重复基序。其中, 单核苷酸重复 SSR 的优势重复基序为 A/T, 占 86.25%, 这可能与 Poly A 被纳入计算结果有关; 二核苷酸重复 SSR 的优势重复基序为 AC/GT, 占 75.25%; 三核苷酸重复 SSR 的优势重复基序为 AGC/CTG 和 AGG/CCT, 分别占 27.16% 和 28.57%; 四核苷酸重复 SSR 的优势重复基序为 AAAC/GTTT 和 ACAG/CTGT, 分别占 17.51% 和 16.81%。

2.3 刀鲚 EST-SSR 不同基序重复次数分布

刀鲚 SSR 主要重复基序的重复次数的分布情况见表 3。刀鲚 SSR 基序的重复次数大部分为 5~24 次, 其他重复次数相对较少, 仅 ATC/ATG 出现了 45 次重复。低重复次数(5~9 次重复)的 SSR 位点最多, 约占 61.21%; 中等重复次数(10~11 次重复)的 SSR 位点次之, 约占 22.84%; 而 >11 次重复的 SSR 位点仅占 15.95%。其中, 单核苷酸基序重复次数均 ≥ 10, 且重复 11 次以上的最多, 占 45.60%; 二核苷酸基序以低重复次数为主, 占 92.26%, 重复 11 次以上的较少, 仅占 0.02%; 三核苷酸至六核苷酸基序均以重复次数为 5 的最多, 其余重复次数的重复基序相对较少。

2.4 刀鲚 EST-SSR 多态性引物的初步筛选和分析

以本单位核心选育群体 F₃ 随机选择的 30 尾鱼苗的基因组 DNA 为模板, 针对 100 对四核苷酸重复 SSR 位点的引物进行多态性检验, 最终筛选到 16 对具有多态性的微卫星引物(表 4)。

基于 16 对经筛选获得的多态性微卫星标记, 对长江刀鲚核心选育群体(F₃)进行遗传多样性评估(表 5)。结果显示, 长江刀鲚核心选育群体(F₃)的平均有效等位基因数(N_e)、平均观测杂合度(H_o)、平均期望杂合度(H_e)和 Shannon 信息指数(I)分别为 1.7580、0.3414、0.3977 和 0.6278。基于等位基因频率计算不同位点的多态信息含量, 结果显示, 16 对多态性微

表2 刀鲚转录组中 EST-SSR 的主要重复基序的类型和数量
Tab.2 The type and number of main repeat motif of EST-SSR in the transcriptome of *C. ectenes*

重复类型 Type of SSR	主要重复基序 Main repeat motif	数量 Number	重复类型 Type of SSR	主要重复基序 Main repeat motif	数量 Number
单核苷酸 Mononucleotide	A/T	10215	四核苷酸 Tetranucleotide	AAGG/CCCT	7
	C/G	1628		AAGT/ACTT	2
二核苷酸 Dinucleotide	AC/GT	12617		AATC/ATTG	7
	AG/CT	3460		AATG/ATTC	25
	AT/AT	666		AATT/AATT	3
	CG/CG	24		ACAG/CTGT	120
	三核苷酸 Trinucleotide	AAC/GTT		239	ACAT/ATGT
三核苷酸 Trinucleotide	AAG/CTT	277		ACCC/GGGT	5
	AAT/ATT	406		ACCG/CGGT	2
	ACC/GGT	595		ACCT/AGGT	3
	ACG/CGT	16		ACGC/CGTG	22
	ACT/AGT	56		ACGG/CCGT	3
	AGC/CTG	1228		ACGT/ACGT	1
	AGG/CCT	1292		ACTC/AGTG	22
	ATC/ATG	337		ACTG/AGTC	19
	CCG/CGG	76		AGAT/ATCT	45
	四核苷酸 Tetranucleotide	AAAC/GTTT	125	AGCC/CTGG	16
AAAG/CTTT		84	AGCG/CGCT	5	
AAAT/ATTT		40	AGCT/AGCT	4	
AACC/GGTT		13	AGGC/CCTG	23	
AACG/CGTT		3	AGGG/CCCT	34	
AACT/AGTT		3	ATCC/ATGG	45	
AAGC/CTTG		3	ATGC/ATGC	2	

表3 刀鲚转录组主要 EST-SSR 重复基序的重复次数
Tab.3 The repeat number of main repeat motif of EST-SSR in the transcriptome of *C. ectenes*

重复类型 Type of SSR	重复次数 Repeat number								总计 Total	比例 Ratio(%)
	5	6	7	8	9	10	11	>11		
单核苷酸 Mononucleotide	-	-	-	-	-	4158	2284	5401	11843	34.94
二核苷酸 Dinucleotide	-	5461	3626	3299	3083	1119	175	4	16767	49.47
三核苷酸 Trinucleotide	2452	1362	688	16	3	-	-	1	4522	13.34
四核苷酸 Tetranucleotide	657	55	2	-	-	-	-	-	714	2.11
五核苷酸 Pentanucleotide	15	6	1	1	2	2	2	-	29	0.09
六核苷酸 Hexanucleotide	8	3	4	3	-	-	1	2	21	0.06
总计 Total	3132	6887	4321	3319	3088	5279	2462	5408	33896	100
比例 Ratio (%)	9.24	20.32	12.75	9.79	9.11	15.57	7.26	15.95	100	

表 4 刀鲚多态性微卫星引物
Tab.4 Polymorphic SSR primers of *C. ectenes*

位点 Locus	引物序列 Primer sequence (5'~3')	重复基序 Repeat motif	产物长度 Product size (bp)	退火温度 Annealing temperature (°C)
CeP4-01	F: AGAATACACACAGGCCCAGG R: TGAAGTGGACGTCTCAACCA	(TGGC) ₅	200	70
CeP4-02	F: GCCACAACAGTTCTTCGACA R: TCTCTTCCCTTTTCCCATCC	(GGAT) ₅	202	62
CeP4-03	F: GTCAAGAGCTATGAAGCCGC R: GAGTCCAGCCATTATGAGCC	(CAAT) ₅	203	65
CeP4-04	F: GTGAAGACGACAAAGCGTGA R: GTGGGAATGCTGAAGACCAT	(TACA) ₅	204	65
CeP4-05	F: GGTCACCCATGAATCCTGAC R: TCAGGCATTGGTTTTTCCTC	(TGTC) ₅	204	60
CeP4-06	F: TTGGGAGGTAGCTAGAGGCA R: TGACAGGTGGCAGAAGCAG	(ACCA) ₅	205	62
CeP4-07	F: GTCCTGCTTGCTTGAAAAGG R: TCTCCCCTATGTTTTGCGAG	(TCAG) ₅	233	67
CeP4-08	F: ACTCCTGCTGGGTTGAGCTA R: TGTGTTTGCATACGGTTCGT	(GCAC) ₅	246	67
CeP4-09	F: GGGAGGGGTTATGCACTGTA R: CAAAACCGGAATAAGCCTCA	(TACT) ₅	247	60
CeP4-10	F: ATGGTGCAAAGGACAAGGAC R: AGCAAATAACGTCCGGATTG	(CCTC) ₅	249	62
CeP4-11	F: TCCTCACACTGATCACAGGC R: AAAAACGGCACAGAATGACC	(TCGT) ₅	252	60
CeP4-12	F: TGGATGGATGAATGAATGGA R: GCGGTTATGACTAGGCGTGT	(GATG) ₅	260	62
CeP4-13	F: GTGTTTCGATGGAATGGAGGT R: TGTTGCTGAAGAGGTCGATG	(TTGT) ₅	266	63
CeP4-14	F: TATCAGGTGCACCCTCAGTG R: GCAGCTGCTCAAGAAGGAGT	(CTGG) ₆	266	68
CeP4-15	F: TACATGAGAACGTCGGGTGA R: GAGAGAGGAGGTGGATGCTG	(TTTG) ₅	271	67
CeP4-16	F: AGGAAGTACCTGGGGAATGC R: TCCTATACCCTCTGGTGGCA	(TTTG) ₅	276	65

卫星标记中, 有 2 对属于高度多态, 10 对属于中度多态, 4 对属于低度多态。可见, 基于刀鲚转录组开发的多态性 EST-SSR 能够应用于刀鲚选育群体的种质资源评估, 具有较高实用性。

3 讨论

微卫星分子标记在动物进化过程中, 由于 DNA 分子的复制、缺失、滑移或错配及姐妹染色单体的不均交换等原因而具有高度变异性, 被广泛应用于水产动物群体遗传结构分析(滕爽爽等, 2018)、亲子鉴定

(张丹等, 2018)、遗传图谱构建(郭香等, 2013)、数量性状定位(顾颖等, 2016)和分子标记辅助育种(孙际佳等, 2017; 顾颖等, 2016)等方面的研究。对于具有基因组信息的物种而言, 微卫星标记的大规模开发相对方便; 而对于目前还没有基因组信息的物种, 则可利用高通量转录组测序技术大规模开发微卫星标记。刀鲚作为我国具有较高经济、营养和养殖价值的江海洄游性鱼类, 其基因组信息、转录组信息及可用分子标记资源均极为匮乏, 有鉴于全基因组测序的成本相对较高, 因此, 亟需开展刀鲚的转录学研究以获得大量的

表5 16个微卫星标记的遗传参数
Tab.5 Genetic parameters at 16 microsatellite markers

位点 Locus	有效等位基因数 N_e	观测杂合度 H_o	期望杂合度 H_e	Shannon 多样性指数 I	多态信息含量 PIC
CeP4-01	1.5327	0.3438	0.3466	0.5304	0.2872
CeP4-02	1.9255	0.5000	0.4805	0.6735	0.3660
CeP4-03	1.6431	0.3958	0.3841	0.5701	0.3456
CeP4-04	2.0080	0.1354	0.4759	0.7996	0.4421
CeP4-05	1.5063	0.2188	0.3296	0.5087	0.2798
CeP4-06	2.4890	0.5313	0.5879	0.9812	0.5384
CeP4-07	1.6704	0.3542	0.3980	0.5866	0.3224
CeP4-08	1.3524	0.1607	0.2462	0.4470	0.2299
CeP4-09	1.1886	0.0833	0.1419	0.2550	0.1406
CeP4-10	1.9074	0.3438	0.4746	0.6673	0.3670
CeP4-11	2.1001	0.4896	0.5204	0.8118	0.4421
CeP4-12	1.3509	0.1979	0.2303	0.3718	0.2235
CeP4-13	1.9161	0.2813	0.4681	0.8103	0.4287
CeP4-14	1.3320	0.2917	0.2487	0.4145	0.2182
CeP4-15	2.3696	0.7396	0.5753	0.9698	0.5136
CeP4-16	1.8356	0.3958	0.4544	0.6467	0.3528
平均值 Mean	1.7580	0.3414	0.3977	0.6278	0.3436

可用分子标记。此外,我国长江流域的刀鲚产量最为丰富,相应的野生群体遗传多样性较高和资源分布较广,是开展刀鲚遗传育种研究的理想种质资源库。而本研究的刀鲚最初的繁育亲本为长江流域的野生群体,历经3代高强度的人工选育,亟需适时地利用合适的分子标记检测其遗传变异情况,以制定科学有效的育种策略。

本研究基于刀鲚转录组数据,利用生物信息学方法挖掘刀鲚EST-SSR,获得了33896条刀鲚EST-SSR,这些EST-SSR的种类较为丰富,且1~6核苷酸重复基序均有不同程度的分布。其中,以出现频率为49.47%的二核苷酸重复基序为主,其次是单核苷酸重复基序(34.94%),三核苷酸重复基序亦较为丰富,占13.34%,其他类型的核苷酸重复基序相对较少。蔡磊等(2015)研究表明,多数物种EST-SSR中的重复基序均以二核苷酸重复基序为主,这与本研究结果一致;而其余类型的优势重复基序则有较大的差异,这可能与不同物种转录组文库构建的组织来源、不同物种的种间特异性、位点突变频率及选择性进化机制有关。在刀鲚所有EST-SSR中,不同的核苷酸重复基序具有不同的重复次数,其中,单核苷酸重复基序大部分属于中高等重复次数,其余类型重复基序的大部分重复次数在5~9之间,相对低于其他鱼类基因组中的微卫星重复次数,可能与编码区和非编码区所受到的选择压力差异有关。不同重复基序的碱基优势重复类型表现出不同程度的偏倚性,单核苷酸重复基序以A/T

类型为主,可能是由于富含A/T的微卫星退火温度相对较低,利于双链DNA的解链,通过DNA复制的重组和滑动机制而增加A/T重复基序的机率(倪守胜等,2018)。而二核苷酸重复基序以AC/GT为主,这与鱼类基因组微卫星中以AC/GT重复为主相符;其余重复基序的优势重复碱基类型在不同物种间差异极大,可能与不同物种的种间差异性有关。此外,本研究发现,刀鲚EST-SSR重复单元的拷贝次数随着重复单元长度的增加而下降,微卫星的数量亦呈现出逐步减少的趋势。

近年来,相关学者在鱼类中利用高通量转录组测序技术陆续开发出了多态性较高的微卫星标记。蔡磊等(2015)利用高通量转录组测序技术从诸氏鲮虾虎鱼(*Mugilogobius chulae*)肝脏转录组中获得6225个微卫星位点,并从76对引物中筛选获得32对具有多态性的微卫星标记。龚诗琦等(2016)利用高通量转录组测序技术从黄姑鱼(*Nibealalbiflora*)转录组获得12254个微卫星标记,并从80对引物中筛选获得18个具有多态性的微卫星标记。岳华梅等(2016)利用高通量转录组测序技术从兴国红鲤(*Cyprinus carpio var. singuonensis*)的垂体和性腺转录组中获得13652个微卫星标记,并从30对引物中筛选获得了9个具有多态性的微卫星标记。本研究从100对四碱基重复的微卫星引物中筛选获得了16对具有多态性的微卫星标记,引物多态率相对低于其他鱼类的研究结果。造成多态性较低的原因,一方面是由于本研究选取的四碱

基重复的微卫星相对于二碱基和三碱基重复的微卫星比较保守, 且变异率低; 另一方面是由于基于转录所获得的 EST-SSR 扩增产物可能跨越内含子或者 PCR 引物位于内含子和外显子的结合处, 而造成扩增失败导致多态率较低。此外, 筛选多态性引物所用样本的种类、数量、遗传差异程度以及不同物种的 DNA 序列的保守性对于微卫星位点多态性的筛选亦有不同程度的影响。

基于 Botstein 等(1980)对于多态信息含量(PIC)的评估标准, 当 $PIC > 0.5$ 、 $0.25 < PIC < 0.5$ 、 $PIC < 0.25$ 时分别代表高度多态、中度多态和低度多态性。Rong 等(2013)利用 FIASCO 法所开发的微卫星标记大部分为高度多态(70%), 中度多态占 25%, 而本研究所获得的多态性微卫星标记则大部分属于中度多态(62.5%), 高度多态仅占 12.5%, 表明不同的微卫星开发方法所获得的位点多态性有所不同, 而实验样本的来源、数量和位点的核苷酸的重复基序、次数的不同位点多态性亦有不同程度的影响。杂合度作为度量群体遗传变异的重要参考指标, 能够有效反映出群体遗传变异程度的高低。在本研究中, 基于刀鲚多态性 EST-SSR 所获得的刀鲚选育群体(F_3)的平均观测杂合度(H_o)和平均期望杂合度(H_e)分别为 0.3414 和 0.3977, 低于 Rong 等(2013)($H_o=0.49$, $H_e=0.68$)、Ma 等(2011)($H_o=0.675$, $H_e=0.650$)、Chen 等(2012)($H_o=0.53$, $H_e=0.77$)利用磁珠富集法所获的 G-SSR 的研究结果。究其原因, 一方面是由于本研究选用的四核苷酸重复微卫星序列在动物进化过程中相对保守, 相对磁珠富集法所获得的二核苷酸多态性和变异率较低, 另一方面, 本研究所选用的实验样本经过了 3 代高强度的人工选育, 而其他研究样本大部分为野生群体。总体而言, 基于转录组所获得的 EST-SSR 多态性虽然相对低于基于基因组所获得的 G-SSR, 但仍然能够有效区分个体或者群体的遗传差异, 加上所开发的微卫星数量相对较多, 具有相对较高的实用性。

本研究利用生物信息学方法筛选刀鲚高通量转录组 Unigenes 中的微卫星标记, 分析刀鲚 EST-SSR 分布和组成特征, 并利用所开发的多态性微卫星标记初步分析了刀鲚核心选育群体(F_3)的遗传特征。研究结果不仅丰富了刀鲚分子标记的数量, 而且对刀鲚后续选育群体的育种效果评估及分子标记辅助育种实践具有重要的指导意义。

参 考 文 献

Bai CC, Liu SF, Zhuang ZM. Characteristic analysis of

microsatellite DNA in the genome of Gobiidae. Progress in Fishery Sciences, 2016, 37(5): 9–15 [白翠翠, 柳淑芳, 庄志猛. 虾虎鱼科(Gobiidae)基因组微卫星 DNA 的分布特征. 渔业科学进展, 2016, 37(5): 9–15]

Botstein D, White RL, Skolnick M, et al. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American Journal of Human Genetics, 1980, 32(3): 314–331

Cai L, Yu LJ, Chen XQ, et al. A preliminary screening and characteristic analysis of microsatellite markers from transcriptome sequences in *Mugilogobius chulae*. Biotechnology Bulletin, 2015(9): 146–151 [蔡磊, 余露军, 陈小曲, 等. 诸氏鲮虾虎鱼转录组序列中微卫星标记的初步筛选及特征分析. 生物技术通报, 2015(9): 146–151]

Chen F, Zeng LY, Cheng QQ. Development of thirty-four novel polymorphic microsatellite markers in *Coilia ectenes* (Clupeiformes: Engraulidae) and cross-species amplification in two closely related taxa. Journal of Genetics, 2012, 91(1): 1–7

Cheng QQ, Zhang QY, Ma CY, et al. Genetic structure and differentiation of four lake populations of *Coilia ectenes* (Clupeiformes: Engraulidae) based on mtDNA control region sequences. Biochemical Systematics and Ecology, 2011, 39(4–6): 544–552

Deng PP, Shi YH, Zhang GY, et al. Isolation of microsatellite in *C. ectenes* by magnetic beads. Biotechnology Bulletin, 2014(6): 145–149 [邓平平, 施永海, 张根玉, 等. 磁珠富集法分离刀鲚微卫星标记. 生物技术通报, 2014(6): 145–149]

Ge GC, Cao T, Chen CJ, et al. Analysis of genetic diversity based on amplified fragment length polymorphism fingerprint of *Coilia nasus* from Yangtze River. Journal of Nanjing University (Natural Sciences), 2008, 44(3): 332–338 [葛家春, 曹廷, 陈婵娟, 等. 利用扩增片段长度多态性技术分析长江刀鲚的遗传多样性. 南京大学学报(自然科学版), 2008, 44(3): 332–338]

Gong SQ, Wang ZY, Xiao SJ, et al. Development and verification of SSR based on transcriptome of yellow drum, *Nibeia albiflora*. Journal of Jimei University (Natural Science), 2016, 21(4): 241–246 [龚诗琦, 王志勇, 肖世俊, 等. 黄姑鱼转录组 SSR 的开发与验证. 集美大学学报(自然科学版), 2016, 21(4): 241–246]

Gu Y, Lu CY, Zhang Q, et al. Analysis of genetic diversity and growth traits in *Cyprinus carpio* var. *haematopterus* using microsatellite QTL markers. Freshwater Fisheries, 2016, 46(4): 9–18 [顾颖, 鲁翠云, 张芹, 等. 微卫星 QTL 标记分析豫选黄河鲤群体遗传结构及生长性状相关性. 淡水渔业, 2016, 46(4): 9–18]

Guan A, Wu YT, Chen Y, et al. Deep sequence-based transcriptome analysis of microsatellites in the cuttlefish (*Sepiella japonica*). Progress in Fishery Sciences, 2018, 39(3): 144–151 [管奥, 毋玉婷, 陈宇, 等. 曼氏无针乌贼转录组微卫星特征分析. 渔业科学进展, 2018, 39(3):

- 144–151]
- Guo X, Li Q, Kong LF, *et al.* Microsatellite-based consensus map of the Pacific oyster (*Crassostrea gigas*). *Journal of Fisheries of China*, 2013, 37(6): 823–829 [郭香, 李琪, 孔令锋, 等. 基于微卫星标记整合长牡蛎遗传图谱. *水产学报*, 2013, 37(6): 823–829]
- Li DM, Yang AG, Wu B, *et al.* Development and application of the EST-SSR markers in *Atrina pectinata*. *Progress in Fishery Sciences*, 2017, 38(2): 137–142 [李东明, 杨爱国, 吴彪, 等. 栉江珧(*Atrina pectinata*)EST-SSR 标记的开发与应用. *渔业科学进展*, 2017, 38(2): 137–142]
- Ma CY, Cheng QQ, Zhang QY. Genetic variation of *Coilia ectenes* (Clupeiformes: Engraulidae) revealed by the complete cytochrome b sequences of mitochondrial DNA. *Journal of Experimental Marine Biology and Ecology*, 2010, 385(1): 14–19
- Ma CY, Cheng QQ, Zhang QY. Development of 12 polymorphic microsatellite markers in *Coilia ectenes* Jordan and Seale, 1905 (Clupeiformes: Engraulidae) and cross-species amplification in *Coilia mystus* Linnaeus, 1758. *Environmental Biology of Fishes*, 2011, 91(2): 243–249
- Ma CY, Liu M, Ma LB, *et al.* Genetic diversity in *Coilia ectenes* by RAPD analysis. *Marine Fisheries Research*, 2004, 25(5): 19–24 [马春艳, 刘敏, 马凌波, 等. 长江口刀鲚遗传多样性的随机扩增多态 DNA(RAPD)分析. *海洋水产研究*, 2004, 25(5): 19–24]
- Ni SS, Yang Y, Liu SF, *et al.* Microsatellite analysis of *Patinopecten yessoensis* using next-generation sequencing method. *Progress in Fishery Sciences*, 2018, 39(1): 107–113 [倪守胜, 杨钰, 柳淑芳, 等. 基于高通量测序的虾夷扇贝基因组微卫星特征分析. *渔业科学进展*, 2018, 39(1): 107–113]
- Peakall R, Smouse PE. GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Resources*, 2006, 6(1): 288–295
- Rong XJ, Xu YJ, Wang QY, *et al.* Isolation and characterization of polymorphic microsatellite markers from *Coilia ectenes*. *Genetics and Molecular Research*, 2013, 12(4): 6011–6017
- Shi YH, Zhang GY, Zhang HM, *et al.* Analysis and comparison of nutritional quality of muscle from *Coilia nasus* FED formulated feed and live feed. *Chinese Journal of Animal Nutrition*, 2014, 26(2): 427–436 [施永海, 张根玉, 张海明, 等. 配合饲料和活饵料喂养刀鲚肌肉营养品质分析与比较. *动物营养学报*, 2014, 26(2): 427–436]
- Shi YH, Zhang GY, Zhang HM, *et al.* Complete artificial propagation and embryonic development of *Coilia nasus*. *Journal of Shanghai Ocean University*, 2015, 24(1): 36–43 [施永海, 张根玉, 张海明, 等. 刀鲚的全人工繁殖及胚胎发育. *上海海洋大学学报*, 2015, 24(1): 36–43]
- Sun JJ, Li GF, Liu L, *et al.* Analysis of microsatellite DNA markers emphasis on correlation with some economically important traits in Mandarin fish *Siniperca chuatsi*. *Chinese Journal of Fisheries*, 2017, 30(1): 11–18 [孙际佳, 李桂峰, 刘丽, 等. 翘嘴鲌微卫星标记及其与主要经济性状的关联分析. *水产学杂志*, 2017, 30(1): 11–18]
- Teng SS, Fang J, Shao YQ, *et al.* Microsatellite analysis of genetic variation in the fast growth families of the second generation of *Tegillarca granosa* and correlation with growth traits. *Acta Hydrobiologica Sinica*, 2018, 42(4): 681–689 [滕爽爽, 方军, 邵艳卿, 等. 泥蚶 G2 代快速生长家系遗传结构的微卫星分析及其与生长性状的关联. *水生生物学报*, 2018, 42(4): 681–689]
- Wei GL, Xu GC, Gu RB, *et al.* Studies on the genetic diversity of farmed and wild populations of *Coilia nasus* by analysing mitochondrial DNA Cytb genes. *Acta Agriculturae Universitatis Jiangxiensis*, 2012, 34(6): 1216–1221 [魏广莲, 徐钢春, 顾若波, 等. 基于 mtDNA Cytb 序列分析养殖与野生刀鲚群体的遗传多样性. *江西农业大学学报*, 2012, 34(6): 1216–1221]
- Wei GL, Xu GC, Gu RB, *et al.* Biology and artificial breeding progress in *Coilia nasus*. *Journal of Yangtze University (Natural Science)*, 2013, 9(7): 31–35 [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. *长江大学学报(自然科学版)*, 2013, 9(7): 31–35]
- Yang M, Yu Y, Zhang XJ, *et al.* 2017. Development of microsatellite markers from the transcriptome sequences of Pacific white shrimp (*Litopenaus vannamei*). *Marine Sciences*, 41(2): 96–102 [杨铭, 于洋, 张晓军, 等. 基于转录组数据的凡纳滨对虾微卫星标记开发. *海洋科学*, 2017, 41(2): 96–102]
- Yuan CM. The resources of *Coilia ectenes* in middle and lower Yangtze River, and the status and the causation of the population formation changes. *Chinese Journal of Zoology*, 1988, 23(3): 12–15 [袁传必. 长江中下游刀鲚资源和种群组成变动状况及其原因. *动物学杂志*, 1988, 23(3): 12–15]
- Yue HM, Zhai Q, Song MY, *et al.* Development of microsatellite markers in *Cyprinus carpio* var. *singunensis* using next-generation sequencing. *Freshwater Fisheries*, 2016, 46(1): 24–28 [岳华梅, 翟晴, 宋明月, 等. 基于转录组测序的兴国红鲤微卫星标记筛选. *淡水渔业*, 2016, 46(1): 24–28]
- Zhang D, Fu JJ, Zhang LD, *et al.* The parentage analysis of bighead carp (*Hypophthalmichthys nobilis*) based on ten microsatellite markers. *Genomics and Applied Biology*, 2018, 1–11 [张丹, 傅建军, 张利德, 等. 鳊基于 10 个微卫星标记的亲缘鉴定分析. *基因组学与应用生物学*, 2018, 1–11]
- Zhang YP, He G, Wang S, *et al.* Analysis of genetic diversity of *Coilia nasus* and *Coilia brachygnathus* based on mitochondrial DNA D-loop sequence. *Jiangsu Agricultural Sciences*, 2017, 45(20): 59–64 [张燕萍, 贺刚, 王生, 等. 基于线粒体 DNA D-loop 序列分析刀鲚与短颌鲚的遗传多样性. *江苏农业科学*, 2017, 45(20): 59–64]

Characteristic Analysis of Microsatellites in Selected *Coilia ectenes* Using a Transcriptome Dataset

YU Aiqing, SHI Yonghai^①, XU Jiabo, LU Genhai, ZHANG Haiming, XIE Yongde, LIU Yongshi
(Shanghai Fisheries Research Institute, Shanghai Fisheries Technology Extension Station, Shanghai 200433)

Abstract The objective of this study was to identify microsatellites from the transcriptome sequences of *Coilia ectenes* using MISA software. These microsatellites could create very useful resources in the evaluation of germplasm resource and marker-assisted breeding of *C. ectenes*. The results showed that a total of 33,896 microsatellites were identified with repeating units, at lengths of 1~6 bases, from 71,869 unigenes. Different types of repeat SSRs had considerably different distribution characteristics. The majority of the microsatellite loci consisted of mono-, di-, and tri-nucleotide motifs (34.94%, 49.47%, and 13.34%, respectively). Dinucleotide microsatellite repeating units were the most abundant in the *C. ectenes* transcriptome, and the AC/GT repeating units were the most ascendant repeating unit (75.25%). The dominant repeating units for the mononucleotide and trinucleotide motifs were A/T (86.25%) and AGG/CCT (28.57%), respectively. Different nucleotide repeat motifs had different repetitions that had a reducing trend, with the increase in the number of nucleotides in repeat motifs. Among the 100 designed primer pairs, 16 pairs proved to be polymorphic microsatellite markers. In the present study, 16 polymorphic microsatellite loci were characterized to evaluate the genetic diversity of the selected *C. ectenes* (F₃). The results showed that the average number of effective alleles (N_e), the average observed heterozygosity (H_o), the average expected heterozygosity (H_e), and the average Shannon's information index (I) of the selected *C.ectenes* (F₃) were 1.7580, 0.3414, 0.3977, and 0.6278, respectively. These results indicate that it is feasible to develop microsatellite markers based on the *C. ectenes* transcriptome and polymorphic microsatellite loci obtained in this study will facilitate further studies on the population genetic management and conservation of *C. ectenes*.

Key words *Coilia ectenes*; Transcriptome; Microsatellites; Molecular markers

^① Corresponding author: SHI Yonghai, E-mail: yonghais@163.com