

黑鲷、真鲷及其杂交子代基因编码区微卫星 序列及密码子偏好性分析*

曹广勇^{1,2} 张志勇² 张志伟² 陈淑吟² 祝 斐² 贾超峰²
陈自强^{1,2} 曾海峰² 汤晓建²

(1. 上海海洋大学水产与生命学院 上海 201306; 2. 江苏省海洋水产研究所 江苏省海水鱼类遗传育种
重点实验室 南通 226007)

摘要 本研究采用生物信息学方法分析比较了黑鲷、真鲷及其两种杂交子代基因编码区(Coding sequence, CDS)中微卫星序列(简单重复序列, SSR)的分布规律及密码子偏好性。结果表明: (1) 黑鲷、真鲷及其两种杂交子代微卫星序列均以三碱基重复类型为主, 其次为二碱基重复, 片段长度大多数在 12—20bp 之间。(2) 黑鲷与反交(黑鲷♀×真鲷♂)在三、五碱基中优势基元类型相同, 真鲷与正交(真鲷♀×黑鲷♂)在三碱基中优势基元类型相同, 而黑鲷与正交在四碱基中优势基元类型相同。(3) UUC、UAC 等 28 个密码子为黑鲷、真鲷及其两种杂交子代基因编码区的偏好性密码子, 偏爱使用以 C 或 G 结尾的密码子, 且第三位密码子碱基的 GC 含量(GC3s)均高于基因编码区平均 GC 含量。(4) 真鲷与正交的密码子使用的偏好性相似, 一定程度上说明黑鲷、真鲷杂交子代偏母系遗传。分析结果将为研究鲷科鱼类微卫星标记、研究其群体遗传多样性、构建转基因表达系统和开展分子遗传育种提供理论基础。

关键词 黑鲷; 真鲷; CDS 序列; 微卫星; 密码子偏好性

中图分类号 Q953; S931 **doi:** 10.11693/hyh20190200038

黑鲷 (*Acanthopagrus schlegelii*) 和真鲷 (*Pagrus major*) 均隶属鲷形目(Percoiformes)、鲷科 (Sparidae), 是我国沿海重要的海水养殖经济鱼类。真鲷个体大且生长快, 而黑鲷营养价值高、抗逆性强。杂交选育可以使杂种后代增加变异性和异质性, 综合双亲的优良性状, 进而获得优良性状稳定的新品种, 是鱼类育种的有效途径之一。目前, 黑鲷、真鲷及其杂交子代对比分析已有相关报道, 例如: 黑鲷和真鲷杂交子代脂肪含量明显高于亲本黑鲷, 具有较好的温度、盐度及低氧抗逆性(邱新志等, 2008; 蒋宏雷等, 2009; 姜景腾等, 2010; 王思婷等, 2016); 真鲷(♀)×黑鲷(♂)杂交子一代偏母系遗传, 且与真鲷的 Calmodulin 功能

基因的组织表达有所差异(林勉等, 2014; 陈淑吟等, 2017), 而在基因编码序列(Coding sequence, CDS)生物信息分析方面尚未见报道。

微卫星特征及密码子偏好性分析是在基因组水平进行生物信息分析的两种常见方法。在真核生物基因组中, 高等生物基因组倾向于单碱基微卫星类型占主导地位(黄杰等, 2015), 将单碱基类型除外, 则以二碱基微卫星序列最为丰富(Ma *et al.*, 1996); 而在水生生物中, 出现以二碱基重复和三碱基重复为主的两种现象(崔建洲等, 2006; 李偲等, 2011; 刘博等, 2012; 袁文成等, 2015)。此外, 生物体内编码某种氨基酸会出现倾向特定密码子的现象, 即密码子偏好

* 江苏省重点研发计划(现代农业)项目, BE2016326 号; 江苏省农业科技自主创新资金项目, CX(17)2021 号; 江苏省农业重大新品种创制项目, PZCZ201744 号; 江苏省自然科学基金项目, BK20181202 号。曹广勇, 硕士研究生, E-mail: 1584996121@qq.com

通信作者: 张志伟, 博士, 研究员, E-mail: zhzhwei2005@126.com

收稿日期: 2019-02-22, 收修改稿日期: 2019-05-05

性(codon bias), 一般相近物种的密码子使用模式相似(Lynn *et al.*, 2002), 而不同物种的密码子偏好性也有所差异。密码子的这种偏好性是物种进化过程中基因突变和选择的结果, 基因中的 GC 含量以及密码子第 3 位上 GC 含量(GC3s)可能为影响密码子偏性的主要因素(Adams *et al.*, 2003; Romero *et al.*, 2003)。本文采用生物信息学工具在基因编码序列水平上研究黑鲷、真鲷及其两种杂交子代微卫星序列的分布规律及密码子偏好性, 为鲷科鱼类构建转基因表达系统、提高特定基因的表达效率、分子遗传育种、种质资源保护和开发利用提供参考。

1 材料与方法

1.1 数据来源

本课题组于 2016 年对黑鲷、真鲷及其两种杂交子代(正交: 真鲷♀×黑鲷♂; 反交: 黑鲷♀×真鲷♂)进行转录组测序, 实验用鱼来源于吕四养殖基地, 取样均为 1 月龄仔鱼, 以全鱼进行测序比对并获得该 4 种鱼 CDS 序列数据。所有序列数据均以 FASTA 文件格式保存。

1.2 SSR 位点搜索

利用 Krait (Du *et al.*, 2018)搜索和统计黑鲷、真鲷及其两种杂交子代 CDS 序列中的 SSR 位点。搜索参数分别设置为: 单核苷酸、二核苷酸、三核苷酸、四核苷酸、五核苷酸、六核苷酸的最少重复次数分别为 10、6、5、4、4、4。Perl 语言操作系统下 MISA 软件(MISA, MicroSatellite identification tool) (Lu *et al.*, 2013)检验。

1.3 密码子分析指标和方法

相对密码子使用度 RSCU (Relative Synonymous Codon Usage), 是指对于某一特定的密码子在编码对

应氨基酸的同义密码子间的相对概率, 可直观地反映密码子使用的偏好性。当某一密码子 RSCU=1, 代表该密码子的使用没有偏好性; RSCU>1 时, 表明该密码子为使用相对较多的密码子, 反之亦然。第 i 个氨基酸的第 j 个密码子的 RSCU 的计算公式(吴宪明等, 2007)如下:

$$RSCU_{ij} = \frac{X_{ij}}{\bar{X}} = \frac{X_{ij}}{\frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}},$$

式中, X_{ij} 是编码第 i 个氨基酸的第 j 个密码子的出现次数, n_i 是编码第 i 个氨基酸的同义密码子的数量(值为 1—6)。

用 CodonW 对黑鲷、真鲷及其两种杂交子代 CDS 序列分别进行分析, 得出每条序列中各密码子的 RSCU 以及密码子的第 3 位碱基中 A、U、G、C 的含量在第 3 位碱基总量中所占的比率(分别为 A3s、U3s、G3s、C3s); 并使用 Microsoft Excel 2003 和 SPSS16.0 对黑鲷、真鲷及其两种杂交子代密码子偏好性数据进行统计及聚类分析。

2 结果与分析

2.1 微卫星特征分析

2.1.1 CDS 序列中 SSR 的数量和分布规律 黑鲷、真鲷及其两种杂交子代中共检索到 185532 条 CDS 序列, GC 含量约为 53%, 14225 个完整型 SSR 位点(黑鲷: 3613, 真鲷: 3597, 正交: 3315, 反交: 3700), 分布在 12709 条 CDS 序列(黑鲷: 3215, 真鲷: 3172, 正交: 2996, 反交: 3326), SSR 发生频率为 6.9%(黑鲷: 8.5%, 真鲷: 7.8%, 正交: 6.8%, 反交: 5.3%)。经对黑鲷、真鲷及其两种杂交子代 CDS 序列中不同碱基完整型 SSR 进行分析, 每种鱼的具体 SSR 信息见表 1。

表 1 黑鲷、真鲷及其两种杂交子代 CDS 序列中 SSR 位点的搜索结果

Tab.1 Results of SSR search in coding sequence of four sea-bream

序号	项目	黑鲷	真鲷	正交	反交
1	序列总数量	37887	40641	44152	62852
2	序列总长度(bp)	36848430	37735593	37773645	48081141
3	GC 含量(%)	53.92	53.49	53.7	52.53
4	检测到的完美型 SSR 数量	3613	3597	3315	3700
5	完美型 SSR 总长度(bp)	57689	58025	52856	58509
6	SSRs 的平均长度(bp)	15.97	16.14	15.95	15.82
7	SSR 占比(%)	0.16	0.16	0.14	0.13
8	相对丰度(loci/Mb)	98.05	95.32	87.76	76.95
9	相对密度(bp/Mb)	1565.58	1537.67	1399.28	1216.88

续表

序号	项目	黑鲷	真鲷	正交	反交
10	单碱基重复 SSR 数目(%)	257(7.11%)	202(5.62%)	142(4.28%)	226(6.11%)
11	二碱基重复 SSR 数目(%)	550(15.22%)	514(14.29%)	515(15.54%)	572(15.46%)
12	三碱基重复 SSR 数目(%)	2666(73.79%)	2724(75.73%)	2529(76.29%)	2748(74.27%)
13	四碱基重复 SSR 数目(%)	69(1.91%)	66(1.83%)	62(1.87%)	86(2.32%)
14	五碱基重复 SSR 数目(%)	18(0.5%)	19(0.53%)	15(0.45%)	23(0.62%)
15	六碱基重复 SSR 数目(%)	53(1.47%)	72(2.0%)	52(1.57%)	45(1.22%)

2.1.2 优势重复碱基分析 在 SSR 位点各碱基重复数量中,三碱基重复在黑鲷、真鲷及其两种杂交子代 CDS 序列中占比均为最大(约 75%),二碱基重复其次(约 15%),五碱基重复占比最小,约为 0.5%(表 1)。

黑鲷微卫星 81 种基元(motif)类型,真鲷微卫星 94 种基元类型,正交 80 种基元类型,反交 79 种基元类型。黑鲷、真鲷及其两种杂交子代 CDS 序列中二、三碱基重复数量约占总 SSR 位点 90%,其中二、三碱基重复类型分别以 AC、AG、AAC、ACC、ACG、AAG 重复基元分布最多,这表明 CDS 序列中微卫星的分布对 A/T 碱基不具有偏好性。

黑鲷、真鲷及其两种杂交子代在单碱基和二碱基重复的优势基元类型上完全一致(单碱基为 A/T,二碱基为 AC/GT)。黑鲷与反交在三、五碱基中优势基元类型相同,真鲷与正交在三碱基中优势基元类型相同,而黑鲷与正交在四碱基中优势基元类型相同。该 4 种鱼优势基元类型在单碱基和二碱基重复中的占比较为接近,黑鲷与反交在三碱基重复类型中的优势基元占比较为接近,而黑鲷与正交在四碱基中优势基元类型相同。黑鲷、真鲷及其两种杂交子代在四碱基重复的优势基元类型及占比差异较大。该 4 种鱼 CDS 序列中优势重复碱基具体信息见表 2。

表 2 黑鲷、真鲷及其两种杂交子代 CDS 序列中优势重复碱基的组成和频率统计(考虑到碱基互补作用)
Tab.2 Composition and frequency of the dominant repeat motifs in the coding sequence of four sea-bream (in consideration of sequence complementarity)

重复类型	黑鲷	真鲷	正交	反交
单碱基重复	225 (87.5%) [A/T]	173 (85.6%) [A/T]	119 (83.8%) [A/T]	192 (85.0%) [A/T]
二碱基重复	309(56.2%) [AC/GT]	304(59.1%) [AC/GT]	297(57.7%) [AC/GT]	316(55.2%) [AC/GT]
三碱基重复	760(28.5%) [ACG/CGT]	280(10.3%) [AAG/CTT]	295(11.7%) [AAG/CTT]	767(27.9%) [ACG/CGT]
四碱基重复	20(29.0%) [ACAG/CTGT]	4(6.1%) [AAAG/CTTT]	14(22.6%) [ACAG/CTGT]	12(14.0%) [AAAC/GTTT]
五碱基重复	4(22.2%) [AAAAC/GTTTT]	3(15.8%) [ACCGC/GCGGT]	/	3(13.0%) [AAAAC/GTTTT]
六碱基重复	/	2(2.8%) [AACCAG/CTGGTT]	3(5.8%) [AACAGG/CCTGTT]	/

2.1.3 微卫星的长度分布 黑鲷、真鲷及其两种杂交子代 CDS 序列中微卫星长度分布和数量与片段长度的分布关系结果相似。以黑鲷为例,微卫星长度分布规律和数量与片段长度的分布关系如图 1、图 2 所示。黑鲷 CDS 序列中所发现的 3613 个完整型微卫星中,微卫星长度差异不大,片段长度从 10—30bp 不等,平均长度为 15.97bp。黑鲷微卫星序列中,主要是重复长度大于 15bp 的重复序列,其中长度为 15bp 的三碱基重复数量最多(1585 条),占微卫星(完整型)总

数的 43.9%;长度为 18bp 的三碱基重复次之(807 条),占微卫星(完整型)总数的 22.3%(图 1)。

从图 2 可以看出,黑鲷 SSR 数量随着重复次数的增加呈明显下降趋势。三碱基重复随重复次数的增加,SSR 数量呈直线下降,当重复次数从 5 次增加到 8 次时,黑鲷 SSR 数量下降至 0,之后稳定。三碱基重复 SSR 数量则随重复次数增加而减少。当二碱基重复次数达到 10 次、单碱基重复次数达到 12 次、其他碱基重复曲线均较为平缓。

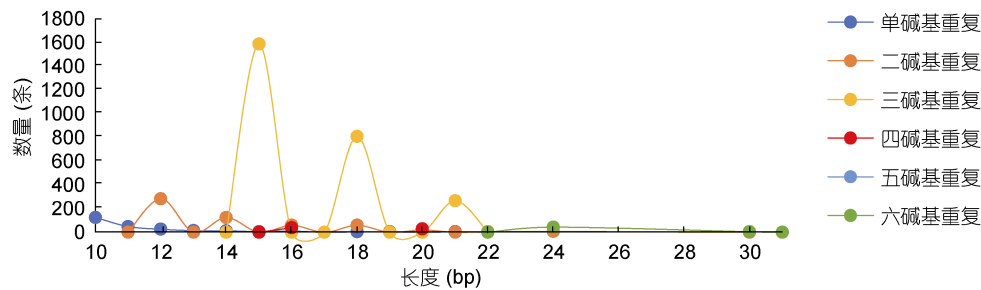


图 1 黑鲷 CDS 序列中微卫星的长度分布
Fig.1 Length distribution of microsatellites in the coding sequence of *A. schlegelii*

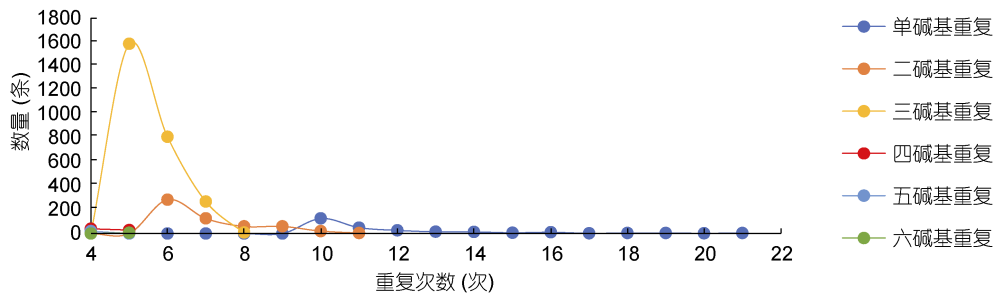


图 2 黑鲷 CDS 序列中 SSR 数量随重复次数的变化
Fig.2 Variation of repeat times with the number of microsatellites in the coding sequence of *A. schlegelii*

2.2 密码子偏好性

2.2.1 黑鲷、真鲷及其两种杂交子代基因编码区碱基组成 黑鲷基因编码区第三位密码子的碱基 GC3s 含量最高(60.88%), 反交最低(58.32%), 可见黑鲷、真鲷及其两种杂交子代的编码蛋白基因序列偏爱使用碱基以 GC 结尾的密码子。从第三位密码子碱基组成看, 黑鲷等 4 种鱼基因编码区 A、U、G 和 C 四种碱基结尾的密码子以 C 结尾的密码子最多, 其他依次为 G3s、U3s 和 A3s(表 3)。由表 1 与表 3 对比还可以发现, 黑鲷、真鲷及其两种杂交子代第三位密码子碱基的 GC 含量(GC3s)均高于基因编码区平均 GC 含量, 而第三位密码子也是导致密码子使用发生偏向性的重要因素之一。

表 3 黑鲷、真鲷及其两种杂交子代 CDS 序列密码子的 A3s、U3s、G3s、C3s
Tab.3 Values of A3s, U3s, G3s, and C3s in the coding sequences of four sea-bream

物种	A3s	U3s	G3s	C3s
黑鲷	18.75	20.37	27.66	33.22
真鲷	19.19	20.97	27.12	32.72
正交	19.01	20.65	27.41	32.92
反交	19.86	21.82	26.88	31.44

注: A3s, U3s, G3s 和 C3s 分别为 A、U、G 和 C 四种碱基结尾的密码子含量(百分比%)

2.2.2 密码子使用偏好性 UUC、UAC、CAC、CAG 等 28 个密码子的 RSCU 均大于 1, 为黑鲷编码蛋白质基因的偏好性密码子(表 4), 黑鲷与真鲷、正

表 4 黑鲷、真鲷及其两种杂交子代 CDS 序列偏好密码子的 RSCU
Tab.4 RSCU of preferable codons of coding sequences in four sea-bream

氨基酸	密码子	黑鲷	真鲷	正交	反交
苯丙氨酸	UUC	1.25	1.22	1.23	1.18
酪氨酸	UAC	1.39	1.37	1.38	1.31
组氨酸	CAC	1.31	1.29	1.31	1.25
谷氨酰胺	CAG	1.52	1.51	1.52	1.49
天冬酰胺	AAC	1.39	1.37	1.38	1.30
赖氨酸	AAG	1.17	1.15	1.16	1.16
天冬氨酸	GAC	1.27	1.25	1.26	1.20
谷氨酸	GAG	1.41	1.40	1.40	1.35
半胱氨酸	UGC	1.06	1.03	1.04	1.01
亮氨酸	CUC	1.33	1.32	1.32	1.26
	CUG	2.70	2.65	2.67	2.50
丝氨酸	UCU	1.01	1.04	1.03	1.08
	UCC	1.32	1.30	1.31	1.31
	AGC	1.46	1.43	1.44	1.36
精氨酸	AGA	1.46	1.50	1.48	1.61
	AGG	1.57	1.56	1.57	1.60
缬氨酸	GUC	1.17	1.15	1.15	1.10

续表					
氨基酸	密码子	黑鲷	真鲷	正交	反交
脯氨酸	GUG	1.77	1.76	1.76	1.70
	CCU	1.14	1.18	1.16	1.18
	CCC	1.13	1.12	1.12	1.11
	CCA	1.09	1.11	1.10	1.10
苏氨酸	ACC	1.36	1.34	1.35	1.32
	ACA	1.15	1.19	1.17	1.18
丙氨酸	GCU	1.14	1.18	1.16	1.17
	GCC	1.40	1.38	1.39	1.36
甘氨酸	GGC	1.20	1.18	1.19	1.10
	GGA	1.22	1.23	1.23	1.29
异亮氨酸	AUC	1.76	1.73	1.74	1.58

注: 色氨酸和甲硫氨酸的 RSCU 为 1, 终止密码子未列入表中

交、反交的编码蛋白基因的偏好性密码子完全相同, 其中尤其偏爱使用以 C 结尾的密码子(除 CGC 外, 其余的都为偏好性密码子), 其次为 G 结尾的密码子, 特别是 CUG, 它们的 RSCU 均达到 2.5 以上, 而以 A 或 U 碱基结尾的仅有 7 个。虽然在黑鲷等 4 种鱼基因编码区偏爱以 C 或 G 结尾的同义密码子, 但 NCG 形式的四种密码子(N 代表 A, U, G, C 四种碱基中的任一种), 没有一个为偏好性密码子, 且均在其同义密码子中为最小。此外, 以密码子编码氨基酸的原理计算, 黑鲷、真鲷及其两种杂交子代的编码 20 种氨基酸中丝氨酸(Ser)和亮氨酸(Leu)含量最高(如图 3 所示), 该两种氨基酸的偏好密码子均以 G/C 结尾。

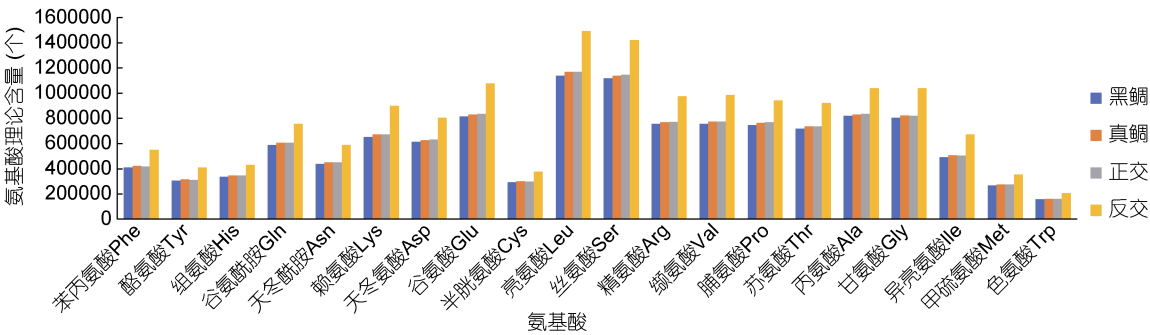


图 3 黑鲷、真鲷及其两种杂交子代各种氨基酸含量

Fig.3 The content of a variety of amino acid residues of coding sequences in four sea-bream

注: 氨基酸理论含量以密码子编码氨基酸的原理计算

2.2.3 聚类分析 在对基因密码子使用概率进行聚类分析的过程中, 将每种鱼的基因编码区序列作为一个对象, 将偏好性密码子的 RSCU 作为变量。除去 RSCU 始终为 1 的编码色氨酸(Trp)和甲硫氨酸(Met)的密码子 UGG 和 AUG, 再除去 3 个不编码氨基酸的

终止密码子, 取余下的 59 个密码子的 RSCU 对基因的密码子使用偏好性进行分析。聚类结果见图 4: 真鲷与正交首先聚为一类, 然后与黑鲷聚为一类。说明真鲷与正交的密码子使用的偏好性相似, 也一定程度表明黑鲷和真鲷杂交子代偏母系遗传。

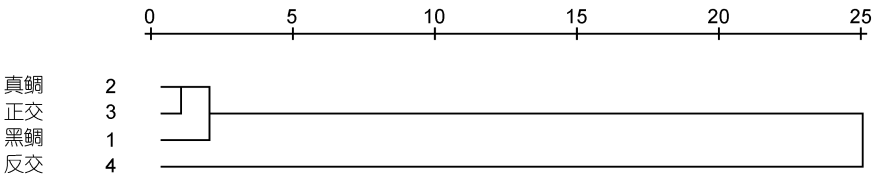


图 4 黑鲷、真鲷及其两种杂交子代的 CDS 序列的基因密码子偏好性聚类

Fig.4 Cluster tree based on the RSCU values of coding sequences in four sea-bream

3 讨论

3.1 微卫星特征分析

黑鲷、真鲷、正交和反交 CDS 序列 SSR 的发生频率(含 SSR 的 Unigene 数/Unigene 总数)分别为 8.5%、7.8%、6.8%、5.3%, 平均间隔分别 10.2kb、

10.5kb、11.4kb、13.0kb 出现 1 个 SSR 序列。这 4 种鱼微卫星序列的重复类型中, 均为三碱基重复类型的数量最高(约 75%); 二碱基重复类型的数量次之(约 15%)。这与翘嘴鲌(*Siniperca chuatsi*)、缙蛭(*Sinonovacula constricta*)等以三碱基重复为主的水生生物相同(刘博等, 2012; 袁文成等, 2015), 与草鱼

(*Ctenopharyngodon idellus*)、红鳍东方鲟(*Takifugu rubripes*)等水生生物以二碱基重复为主不同(崔建洲等, 2006; 李偲等, 2011)。

在所有的微卫星重复类型中, 二核苷酸重复类型最为活跃, 变异水平最高, 拷贝数分布范围最广(白翠翠等, 2016)。两碱基重复类型中, 黑鲟、真鲟及其两种杂交子代均以 AC 重复拷贝类别最为丰富, 这与真菌类生物的内含子和红鳍东方鲟的研究结果一致(Tóth *et al*, 2000; 崔建洲等, 2006), 但与中国明对虾(*Fenneropenaeus chinensis*)的 AT、三疣梭子蟹(*Portunus trituberculatus*)的 AG 重复拷贝多不同(高焕等, 2004; 宋来鹏等, 2008), 与禾本科植物也不同(郑燕等, 2011)。三碱基重复类型中, 黑鲟和反交以 ACG 最多, 其次是 AAG 和 ACC, 真鲟和正交以 AAG 最多, 其次是 ACC 和 ACG。有研究表明, 三碱基重复类型在水生生物中分布特征差异较大, 例如: 三疣梭子蟹基因组其中以 ACT 多, 其次是 AGG 和 AAT(宋来鹏等, 2008); 中国明对虾以 AAT 多, 其次是 AAG 和 ATC(高焕等, 2004); 红鳍东方鲟中以 AAT 多, 其次是 AGG 和 ATC, GCC 少(崔建洲等, 2006)。同时有研究发现, 二碱基重复引物的多态性较高, 多应用于种群间遗传多样性分析(房祖业等, 2018); 一些三碱基重复与某些遗传性病变密切相关(Pearson *et al*, 1998)。这 4 种鱼与上述几种物种在二碱基重复类型中均缺少 GC 重复, 可能由于 DNA CpG 的甲基化, 胞苷酸(C)很容易经过脱氨基作用转变成胸腺嘧啶(T), 导致 GC 重复减少, 剩余少量 GC 维持 DNA 热力学稳定性所必需, 而突变后的序列 TG(即 AC 类型)相应增加(Schorderet *et al*, 1992)。

SSR 分子标记多态性的高低取决于其长度大小, 可以判断其潜在利用价值(李珊等, 2010)。当 SSR 长度为 12—20bp 时, 呈中度多态性; 长度为 20bp 以上时, 表示多态性较高; 长度为 12bp 以下时, 则呈低多态性(Temnykh *et al*, 2001)。本研究在 SSR 位点搜索时已将长度在 10bp 以下的 SSR 序列剔除, 发现黑鲟、真鲟及其两种杂交子代 CDS 序列中完整型 SSR 长度均在 10—85 个碱基之间, 片段长度在 12—20bp 之间的 SSR 序列最多(约占各 SSR 总数的 85%), 表明该 4 种鱼 CDS 序列来源的 SSR 具有中等的多态性潜能, 在开发黑鲟等微卫星分子标记研究方面将具有一定的利用价值。

3.2 密码子偏好性分析

黑鲟、真鲟、正交及反交 4 种鱼的基因编码区在

密码子使用上呈现高度一致性, 均偏爱 C/G 结尾的密码子, 确定了 28 种偏好密码子(RSCU>1), 其中 21 种以 G/C 结尾, 偏好性较强的有 CUG (RSCU>2), 表明该 4 种鱼具有较近的亲缘关系。而在聚类分析中, 发现真鲟与正交的密码子使用的偏好性相似, 一定程度上说明黑鲟和真鲟杂交子代偏母系遗传, 这与染色体构成和 AFLP 分析结果一致(林勉等, 2014)。此外, 黑鲟、真鲟及其两种杂交子代的编码 20 种氨基酸中丝氨酸(Ser)和亮氨酸(Leu)含量最高, 且两种氨基酸的偏好密码子均以 G/C 结尾。该两种氨基酸分别在维持免疫系统和控制血糖具有重要作用, 间接表明鲟科鱼类具有较高营养价值。密码子偏好性分析也一定程度上反映了黑鲟等 4 种鱼具有较高的营养价值。

在进化过程中物种的密码子使用特点具有保守性, 不同的物种对密码子的使用偏好性有所差异, 相近物种的密码子使用模式相似(Lynn *et al*, 2002)。本研究发现, 黑鲟、真鲟、正交及反交 4 种鱼的基因编码区的 GC 含量, 尤其是第 3 位密码子的 GC 含量都较高, 这一特点与金鱼(*Carassius auratus*)的基因组编码区密码子组成相似(张欣等, 2017), 与鲤(*Cyprinus carpio*)和斑马鱼(*Danio rerio*)HOX 基因家族密码子组成存在差异(周丹等, 2013)。通过优化与改造密码子, 可促进翻译效率及提高生物中外源基因的表达, 例: 黑鲟偏好密码子的基因作为外源基因, 而在真鲟、金鱼等物种中获得高效表达; 若依据黑鲟的密码子偏好性信息进行密码子改造, 可提高以黑鲟偏好密码子的外源基因在其他模式生物中的表达效率。对于今后鲟科鱼类转基因过程中对构建合适的转基因表达系统具有重要的指导意义, 从而提高特定蛋白的表达量, 增强转基因育种的效果。然而, 关于黑鲟、真鲟及其两种杂交子代 4 种鱼基因编码区密码子偏好使用对功能基因保守结构域的影响及其蛋白所携带信息的影响还有待进一步研究。

4 结论

本研究利用生物信息学软件分析比较了黑鲟、真鲟及其两种杂交子代 4 种鱼基因编码区(CDS 序列)中微卫星序列的分布特征及密码子偏好性, 结果显示: 该 4 种鱼基因编码区均为以三碱基重复为主的微卫星序列并使用相同的偏好性密码子, 各碱基重复中优势基元类型及偏好性密码子的相对密码子使用度有所差异, 进而推断黑鲟和真鲟杂交子代一定程度

偏母系遗传。本研究结果将为研究鲴科鱼类微卫星标记、研究其群体遗传多样性、构建转基因表达系统和开展分子遗传育种提供理论基础。

参 考 文 献

- 王思婷, 李 鹏, 陈淑吟等, 2016. 黑鲟与真鲴及其杂交子一代脂肪含量测定与脂肪酸成分分析. 食品科学, 37(20): 81—85
- 白翠翠, 柳淑芳, 庄志猛, 2016. 虾虎鱼科(Gobiidae)基因组微卫星 DNA 的分布特征. 渔业科学进展, 37(5): 9—15
- 刘 博, 邵艳卿, 滕爽爽等, 2012. 缙蛭(*Sinonovacula constricta*)EST-SSR 分布特征及引物开发利用. 海洋与湖沼, 43(1): 132—137
- 李 珊, 周天华, 赵桂仿等, 2010. 马蹄香表达序列标签资源的 SSR 信息分析. 中草药, 41(3): 464—468
- 李 偲, 刘 航, 黄 容等, 2011. 草鱼 型微卫星标记的发掘及其多态性检测. 水生生物学报, 35(4): 681—687
- 吴宪明, 吴松峰, 朱云平等, 2007. 密码子偏性的分析方法及相关研究进展. 遗传, 29(4): 420—426
- 邱新志, 吴雄飞, 2008. 盐度对真鲴(♀)×黑鲴(♂)杂交子一代仔鱼活力的影响. 渔业现代化, 35(4): 41—43
- 宋来鹏, 刘 萍, 李 健等, 2008. 三疣梭子蟹基因组微卫星特征分析. 中国水产科学, 15(5): 738—744
- 张 欣, 尤宏争, 李荣妮等, 2017. 金鱼密码子用法特点分析. 华北农学报, 32(S1): 30—36
- 陈淑吟, 李 鹏, 张志勇等, 2017. 黑鲴♀×真鲴♂反交子代与黑鲴的 *CaM* 基因克隆与 mRNA 表达分析. 中国水产科学, 24(6): 1193—1202
- 林 勉, 苗 亮, 李明云等, 2014. 真鲴、黑鲴及其杂交子代的染色体构成与 AFLP 分析. 海洋学报, 36(8): 117—121
- 周 丹, 薛仁余, 张晓峰等, 2013. 鲤和斑马鱼 *HOX* 基因家族同义密码子使用偏性的分析. 水产学杂志, 26(2): 19—25
- 郑 燕, 张 耿, 吴为人, 2011. 禾本科植物微卫星序列的特征分析和比较. 基因组学与应用生物学, 30(5): 513—520
- 房祖业, 陈晓东, 吴咏诗等, 2018. 大刺鲃(*Mastacembelus armatus*)二、三、四碱基重复微卫星标记的筛选和特征分析. 海洋与湖沼, 49(1): 174—182
- 姜景腾, 吴雄飞, 蒋宏雷, 2010. 低氧胁迫对真鲴(♀)与黑鲴(♂)杂交子一代体内酶活力的影响. 宁波大学学报(理工版), 23(4): 10—14
- 袁文成, 黄鹤忠, 李文龙等, 2015. 翘嘴鲈(*Siniperca chuatsi*)转录组 EST-SSR 位点的信息分析及其多态性检测. 海洋与湖沼, 46(2): 403—409
- 高 焕, 刘 萍, 孟宪红等, 2004. 中国对虾(*Fenneropenaeus chinensis*)基因组微卫星特征分析. 海洋与湖沼, 35(5): 424—431
- 黄 杰, 周 瑜, 刘与之等, 2015. 基于 454 GS FLX 高通量测序的四川山鹧鸪基因组微卫星特征分析. 四川动物, 34(1): 8—14
- 崔建洲, 申雪艳, 杨官品等, 2006. 红鳍东方鲀基因组微卫星特征分析. 中国海洋大学学报, 36(2): 249—254, 272
- 蒋宏雷, 吴雄飞, 2009. 真鲴与黑鲴及杂交 F₁ 耐受温度、生长比较. 现代渔业信息, 24(2): 26—28
- Adams M J, Antoniw J F, 2003. Codon usage bias amongst plant viruses. Archives of Virology, 149(1): 113—135
- Du L M, Zhang C, Liu Q *et al*, 2018. Krait: an ultrafast tool for genome-wide survey of microsatellites and primer design. Bioinformatics, 34(4): 681—683
- Lu X, Wang H X, Liu B Z *et al*, 2013. Three EST-SSR markers associated with QTL for the growth of the clam *Meretrix meretrix* revealed by selective genotyping. Marine Biotechnology, 15(1): 16—25
- Lynn D J, Singer G A C, Hickey D A, 2002. Synonymous codon usage is subject to selection in thermophilic bacteria. Nucleic Acids Research, 30(19): 4272—4277
- Ma Z Q, Röder M, Sorrells M E, 1996. Frequencies and sequence characteristics of di-, tri-, and tetra-nucleotide microsatellites in wheat. Genome, 39(1): 123—130
- Pearson C E, Sinden R, 1998. Trinucleotide repeat DNA structures: dynamic mutations from dynamic DNA. Current Opinion in Structural Biology, 8(3): 321—330
- Romero H, Zavala A, Musto H *et al*, 2003. The influence of translational selection on codon usage in fishes from the family Cyprinidae. Gene, 317: 141—147
- Schorderet D F, Gartler S M, 1992. Analysis of CpG suppression in methylated and nonmethylated species. Proceedings of the National Academy of Sciences of the United States of America, 89(3): 957—961
- Tóth, Gáspári Z, Jurka J, 2000. Microsatellites in different eukaryotic genomes: survey and analysis. Genome Research, 10(7): 967—981
- Temnykh S, DeClerck G, Lukashova A *et al*, 2001. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Research, 11(8): 1441—1452

ANALYSIS OF THE MICROSATELLITE SEQUENCES AND CODON BIAS OF THE CODING SEQUENCE IN *ACANTHOPAGRUS SCHLEGELII*, *PAGRUS MAJOR* AND THEIR HYBRID PROGENIES

CAO Guang-Yong^{1,2}, ZHANG Zhi-Yong², ZHANG Zhi-Wei², CHEN Shu-Yin², ZHU Fei², JIA Chao-Feng²,
CHEN Zi-Qiang^{1,2}, ZENG Hai-Feng², TANG Xiao-Jian²

(1. Fishery and Life Science College, Shanghai Ocean University, Shanghai 201306, China; 2. Jiangsu Marine Fisheries Research Institute, Nantong 226007, China)

Abstract The microsatellite sequences (simple sequence repeats, SSR) and codon bias in the coding regions (CDS) of black porgy (*Acanthopagrus schlegelii*), red porgy (*Pagrus major*), and their offspring (orthogonal and reciprocal ones) were analyzed and compared using the method of bio-informatics. The results are followed. The microsatellite sequence is mainly composed of the three-base repeat type, followed by the two-base repeat, and the most fragment length of repeat sequences is between 12bp and 20bp. The black porgy and the reciprocal cross (black porgy♀× red porgy♂) have the same dominant primitive type in the three and five bases, and the red porgy and the orthogonal (red porgy♀× black porgy♂) dominant base types in the three bases as well, while the black porgy and the orthogonal are the same in the four bases. Twenty-eight pieces of genetic codons such as UUC and UAC are the preferred codons in gene coding regions of the four sea bream species, preferring to use the codons ended in C or G, and the GC content of the third codon base (GC3s). Both are higher than the average GC content of the coding region of the gene. The preference of codon use of the red porgy and the orthogonal is similar, which to some degree shows that black porgy and red porgy hybrids are of maternal hereditary characteristics. This research may provide a theoretical basis for the study of microsatellite markers of sea bream, their population genetic diversity, the construction of a transgenic expression system, and molecular genetic breeding.

Key words *Acanthopagrus schlegelii*; *Pagrus major*; coding sequence; microsatellite; codon usage bias