

DOI: 10.19663/j.issn2095-9869.20171226002

http://www.yykxjz.cn/

高海钰, 李健, 王佳佳, 李吉涛. 十足目甲壳动物的 EST-SSR 分析. 渔业科学进展, 2019, 40(3): 94-102

Gao HY, Li J, Wang JJ, Li JT. Analysis of SSR information in EST resource of decapod crustaceans. Progress in Fishery Sciences, 2019, 40(3): 94-102

十足目甲壳动物的 EST-SSR 分析*

高海钰^{1,2,3} 李健^{1,2,3}① 王佳佳^{2,3} 李吉涛^{2,3}

(1. 上海海洋大学 上海 201306; 2. 青岛海洋科学与技术试点国家实验室海洋渔业科学与食品产出过程功能实验室 青岛 266071; 3. 农业农村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 本研究应用生物信息学方法,对 21 种十足目甲壳动物 EST-SSR 的特点进行了归纳总结和分析。结果显示,不同物种间的 EST-SSR 丰度存在明显差异,腹胚亚目中脊尾白虾(*Exopalaemon carinicauda*)的 EST-SSR 位点最为丰富(868.02 个/Mb),拟穴青蟹(*Scylla paramamosain*)位点丰度最低(286.48 个/Mb);枝鳃亚目中斑节对虾(*Penaeus monodon*) EST-SSR 位点最为丰富(641.19 个/Mb),白滨对虾(*Litopenaeus setiferus*)位点丰度最低(166.96 个/Mb);腹胚亚目和枝鳃亚目中均是二、三、四核苷酸基元的 SSR 较为常见,在腹胚亚目中占总数的 41.11%,在枝鳃亚目中占总数的 28.00%;复合(I)类型的 SSR 在腹胚亚目(51.38%)和枝鳃亚目(65.13%)中占很大比例;腹胚亚目二核苷酸基元的 SSR 中 AC/GT 分布频率最高,三核苷酸基元的 SSR 中 ACC/GGT 和 AAT/ATT 占优势;枝鳃亚目二核苷酸基元的 SSR 中 AG/CT 占优势,三核苷酸基元的 SSR 中 AAT/ATT 的分布频率高于其他三核苷酸基元。此外,对脊尾白虾含有 SSR 的 EST 进行 GO 分析后发现,细胞代谢过程、链接产物、细胞组分和细胞的比例分别在 3 个类型的注释中占优势。本研究加深了对十足目甲壳动物 SSR 分布规律的认识,可为甲壳动物 EST-SSR 标记的开发及实际应用提供参考。

关键词 十足目;甲壳动物;EST-SSR

中图分类号 Q959.223 **文献标识码** A **文章编号** 2095-9869(2019)03-0094-09

十足目动物是甲壳动物中种类最多(8000~10000 种)、进化位阶最高、经济价值最大的一个生物类群(堵南山,1993)。近年来,随着人们对优质水产品的需求增加,包括龙虾、螯虾、蟹和对虾等的腹胚亚目和枝鳃亚目在内的十足目动物在水产养殖中的地位日益突出,因此,有必要加强种质鉴定和群体遗传学研究。发展十足目甲壳动物基因组中大量有效的分子标记,可以帮助保护种质资源,改善影响重要性状基

因和分子标记辅助基因的识别和开发。作为优良遗传标记的微卫星(Microsatellite)与其他分子标记相比,具有较高稳定性、共显性表达、多态性信息丰富、易于检测等特点,被大量应用于构建遗传图谱(Guo *et al*, 2017)、基因定位(Zhou *et al*, 2017)、遗传多样性分析(王日芳等,2017)及亲缘关系鉴定(李东宇等,2016)等。

研究各物种基因组特征,第一步需要生成表达序列标签(Expressed sequence tag, EST)信息。EST在物

* 农业产业技术体系(CARS-48)、山东省泰山产业领军人才工程项目(LNJY2015002)、国家自然科学基金(31472275)和青岛市产业培育计划科技惠民专项(17-3-3-62-nsh)共同资助 [This work was supported by the Earmarked Fund for China Agriculture Research System (CARS-48), Program of Shandong Leading Talent (LNJY2015002), National Natural Science Foundation of China (31472275), and Qingdao Industrial Development Program (Science and Technology Benefit Special Project) (17-3-3-62-nsh)]. 高海钰, E-mail: hygao163@163.com

① 通讯作者: 李健, 研究员, E-mail: lijian@ysfri.ac.cn

收稿日期: 2017-12-26, 收修改稿日期: 2018-04-19

种间高度保守,具有良好的通用性(Varshney *et al.*, 2005),目前已广泛应用于多种植物(Yumurtaci *et al.*, 2017; Singh *et al.*, 2013)和动物(白俊艳等, 2016)。在甲壳类动物如中国明对虾(*Fenneropenaeus chinensis*) (Wang *et al.*, 2005)、日本沼虾(*Macrobrachium nipponense*) (周巧, 2013)等中也有报道,例如EST-SSR在EST中的分布频率、大小、引物开发等。

本研究运用生物信息学方法,从 GenBank 的 EST 数据库中收集 21 种十足目甲壳动物 EST 数据,发掘 EST-SSR 信息,通过比较不同十足目甲壳动物间 EST 的特点及不同基元 EST-SSR 的分布频率和类型,加深对十足目甲壳动物 SSR 分布规律的认识,为甲壳动物 EST-SSR 标记的开发及实际应用提供参考。此外,对脊尾白虾(*Exopalaemon carinicauda*)含有 SSR 的 EST 进行 GO 分析,进一步对脊尾白虾功能基因研究提供重要信息。

1 材料与方法

1.1 数据来源

EST数据来源于GenBank(www.ncbi.nih.gov)中的dbEST数据库,截止于2017年9月1日。包括21种十足目甲壳动物:美国龙虾(*Homarus americanus*)、淡水螯虾(*Pacifastacus leniusculus*)、日本龙虾(*Panulirus japonicus*)、拟穴青蟹(*Scylla paramamosain*)、中华绒螯蟹(*Eriocheir sinensis*)、三疣梭子蟹(*Portunus trituberculatus*)、克氏原螯虾(*Procambarus clarkii*)、招潮蟹(*Uca pugilator*)、红蟹(*Gecarcoidea natalis*)、珍宝蟹(*Metacarcinus magister*)、蓝蟹(*Callinectes sapidus*)、岸蟹(*Carcinus maenas*)、脆壳蟹(*Petrolisthes cinctipes*)、脊尾白虾、日本沼虾、罗氏沼虾(*Macrobrachium rosenbergii*)以及凡纳滨对虾(*Litopenaeus vannamei*)、斑节对虾(*Penaeus monodon*)、中国明对虾、日本囊对虾(*Marsupenaeus japonicus*)、白滨对虾(*Litopenaeus setiferus*)。物种及EST信息详列于表1。

1.2 EST 数据处理

本实验采用EMBOSS软件中的TRIMEST工具,去除尾部包含Poly-A或T的EST序列,利用CROSS_MATCH软件截去过短的序列片段;运行CAP3软件对EST序列进行聚类 and 拼接处理,获得非冗余EST,包括Contigs和Singlets两类序列(黄启星等, 2012)。

1.3 EST-SSR 分析

采用自行编写的Perl语言脚本程序对处理后的

EST序列进行SSR筛选。筛选标准为二核苷酸重复次数 ≥ 6 次,三、四、五和六核苷酸重复次数 ≥ 5 次。同时,筛选重复基元之间没有插入序列的复合型SSR位点(Compound microsatellite),即复合(I)型;筛选重复基元之间存在插入序列长度 ≤ 100 nt的复合型SSR,即复合(II)型。

1.4 脊尾白虾 EST 的功能注释

使用BLAST将Unigenes与NCBI的NR蛋白数据库和Uniprot蛋白数据库进行比对,设置E值为“ 10^{-5} ”(孙政等, 2013),利用Interproscan搜索得到的GO编号对Unigenes按照分子功能、细胞组分和生物学过程进行分类(Ye *et al.*, 2006)。

2 结果与分析

2.1 甲壳动物 EST 长度及 SSR 位点数

从GenBank数据库下载了21种十足目甲壳动物已发布的EST。由表1可见,16种腹胚亚目的EST共242178条,碱基总长度约为183.50 Mb。处理后获得85647条非冗余EST序列,包括30120条Contigs和55527条Singlets序列,碱基总长度约为66.72 Mb。从27536条Unigenes序列发掘34601个SSR位点,占EST总数的32.15%。其中,脊尾白虾含有SSR位点的EST最丰富(43.00%),分布频率为868.02个/Mb;拟穴青蟹含有SSR位点的EST所占百分比最小(12.70%),分布频率为286.48个/Mb。

由表2可见,5种枝鳃亚目的EST共231130条,碱基总长度约为141.28 Mb。获得的59605条Unigenes序列中包括20506条Contigs和39099条Singlets序列,碱基总长度约为37.89 Mb。从5种枝鳃亚目的18475条Unigenes序列中发掘出21911个SSR位点,占EST总数的31.00%。其中,斑节对虾含有SSR位点的EST占总数的百分比最大,为35.49%,分布频率为641.19个/Mb;白滨对虾含有SSR位点的EST所占百分比最小(10.56%),分布频率为166.96个/Mb。

2.2 甲壳动物不同基元 EST-SSR 的分布频率

表3和表4列出了十足目甲壳动物不同类型EST-SSR的数量和分布频率。在腹胚亚目和枝鳃亚目EST序列中,复合(I)型远高于其他类型,分别占总数的51.38%、65.13%;二核苷酸、三核苷酸和四核苷酸基元的SSR数目相差不多,多于五核苷酸和六核苷酸基元的SSR数目,三者腹胚亚目中占总数的41.11%,在枝鳃亚目中占总数的28.00%。在2种

表1 腹胚亚目 EST、Unigenes 及 SSR 在 Unigenes 序列中的丰度分布
Tab.1 The EST, Unigenes and abundance of SSR in Unigenes of Pleocyemata

物种 Species	EST 个数 EST number	EST 长度 EST length (Mb)	非冗余 EST Non-redundant EST			含有 SSR 位点的 Unigenes Unigenes containing SSR site		
			Congtigs 数量 Congtigs number	Singlets 数量 Singlets number	长度 Length (Mb)	数量 Number	SSR 数量 SSR Number	频率 Frequency (个/Mb)
美国龙虾 <i>H. americanus</i>	51833	42.59	8581	13338	17.78	5907	7097	399.13
淡水螯虾 <i>P. leniusculus</i>	1063	0.69	123	361	0.32	170	200	630.15
日本龙虾 <i>P. japonicus</i>	2679	1.78	240	706	0.61	221	249	408.62
拟穴青蟹 <i>S. paramamosain</i>	3975	1.85	34	3903	1.83	500	524	286.48
中华绒螯蟹 <i>E. sinensis</i>	17067	10.45	1339	4664	3.86	2302	2797	723.81
三疣梭子蟹 <i>P. trituberculatus</i>	14396	11.00	1016	3069	3.42	1316	1569	458.13
克氏原螯虾 <i>P. clarkii</i>	917	0.53	85	98	0.11	21	27	249.08
招潮蟹 <i>U. pugilator</i>	3646	2.14	452	1019	0.93	463	565	609.65
红蟹 <i>G. natalis</i>	2118	0.92	163	459	0.26	142	155	590.04
珍宝蟹 <i>M. magister</i>	1137	0.80	154	277	0.27	115	128	471.48
蓝蟹 <i>C. sapidus</i>	10930	9.51	865	1433	2.04	1057	1472	723.29
岸蟹 <i>C. maenas</i>	24570	20.91	4545	6708	8.87	2926	3471	391.23
脆壳蟹 <i>P. cinctipes</i>	97806	73.28	11593	15439	22.86	10908	14595	638.36
脊尾白虾 <i>P. carinicaude</i>	2900	1.57	218	1678	1.13	818	985	868.02
日本龙虾 <i>P. japonicus</i>	2679	1.78	240	706	0.61	221	249	408.62
罗氏沼虾 <i>M. rosenbergii</i>	4462	3.70	472	1669	1.81	449	518	286.72

表2 枝鳃亚目 EST、Unigenes 及 SSR 在 Unigenes 序列中的丰度分布
Tab.2 The EST, Unigenes and abundance of SSR in Unigenes of Dendrobranchiata

物种 Species	EST 个数 EST number	EST 长度 EST length (Mb)	非冗余 EST Non-redundant EST			含有 SSR 位点的 Unigenes Unigenes containing SSR site		
			Congtigs 数量 Congtigs number	Singlets 数量 Singlets number	长度 Length (Mb)	数量 Number	SSR 数量 SSR Number	频率 Frequency (个/Mb)
凡纳滨对虾 <i>L. vannamei</i>	163062	99.53	14048	19216	21.23	9791	11779	554.74
斑节对虾 <i>P. monodon</i>	52714	33.27	4946	15037	12.98	7091	8323	641.19
中国明对虾 <i>F. chinensis</i>	10512	5.25	1064	2431	1.77	813	891	504.08
日本囊对虾 <i>M. japonicus</i>	3783	2.54	374	1684	1.37	695	828	603.90
白滨对虾 <i>L. setiferus</i>	1059	0.68	74	731	0.54	85	90	166.96

亚目中,五核苷酸和六核苷酸基元的 SSR 数目占有很小的比例,在腹胚亚目中分别占总数的 3.67%、0.99%,在枝鳃亚目中分别占总数的 2.01%、1.33%。克氏原螯虾和淡水螯虾的 EST 序列中没有搜索到六核苷酸基元的 SSR 位点。

2.3 甲壳动物 EST-SSR 的主要类型

表 5 和表 6 列出了二、三核苷酸基元的 SSR 在各物种内的数量和分布频率。21 种十足目甲壳动物

中,各种类型的基元分布频率不一,有的基元甚至没有出现。不同基元的 SSR 在 21 种甲壳动物中的分布频率具有很大的差异。16 种腹胚亚目中,二核苷酸基元的 SSR 中分布频率最高的类型为 AC/GT,远高于 CG/CG 的分布频率,AG/CT 和 AT/AT 次之;三核苷酸基元的 SSR 中,分布频率最高的集中于 ACC/GGT 和 AAT/ATT,远高于 ACG/CGT 的分布频率。5 种枝鳃亚目中,二核苷酸基元 SSR 中,AG/CT 的分布频率最高;三核苷酸基元的 SSR 中,AAT/ATT

表 3 腹胚亚目不同重复单元 EST-SSR 的数量和分布频率

Tab.3 The number and frequency of SSR containing different size of repeat unit of EST sequences in Pleocyemata

物种 Species	复合(I) Compound SSRs(I)	复合(II) Compound SSRs(II)	二核苷酸 Dinucleotide	三核苷酸 Trinucleotide	四核苷酸 Tetranuc- leotide	五核苷酸 Pentanuc- leotide	六核苷酸 Hexnuc- leotide	合计 Total
美国龙虾 <i>H. americanus</i>	3441(193.52)	226(12.71)	882(49.60)	1330(74.80)	755(42.46)	402(22.61)	61(3.43)	7097(399.13)
淡水螯虾 <i>P. leniusculus</i>	65(204.80)	2(6.30)	17(53.56)	103(324.53)	12(37.81)	1(3.15)	0(0)	200(630.15)
日本龙虾 <i>P. japonicus</i>	140(229.74)	4(6.56)	38(62.36)	37(60.72)	27(44.31)	2(3.28)	1(1.64)	249(408.62)
拟穴青蟹 <i>S. paramamosain</i>	201(109.89)	15(8.20)	67(36.63)	121(66.15)	86(47.02)	29(15.85)	5(2.73)	524(286.48)
中华绒螯蟹 <i>E. sinensis</i>	1617(418.45)	66(17.08)	316(81.78)	439(113.61)	224(57.97)	62(16.04)	73(18.90)	2797(723.81)
三疣梭子蟹 <i>P. trituberculatus</i>	813(237.39)	37(10.80)	200(58.40)	361(105.41)	104(30.37)	31(9.05)	23(6.72)	1569(458.13)
克氏原螯虾 <i>P. clarkii</i>	9(83.03)	0(0)	3(27.68)	9(83.03)	3(27.68)	3(27.68)	0(0)	27(249.08)
招潮蟹 <i>U. pugilator</i>	294(317.23)	27(29.13)	79(85.24)	91(98.19)	61(65.82)	8(8.63)	5(5.40)	565(609.65)
红蟹 <i>G. natalis</i>	94(357.83)	4(15.23)	18(68.52)	23(87.55)	14(53.29)	1(3.81)	1(3.81)	155(590.04)
珍宝蟹 <i>M. magister</i>	68(250.47)	5(18.42)	7(25.78)	27(99.45)	16(58.94)	2(7.37)	3(11.05)	128(471.48)
蓝蟹 <i>C. sapidus</i>	850(417.66)	31(15.23)	134(65.84)	306(150.36)	105(51.59)	24(11.79)	22(10.81)	1472(723.29)
岸蟹 <i>C. maenas</i>	1655(186.54)	94(10.60)	419(47.23)	775(87.35)	350(39.45)	146(16.46)	32(3.61)	3471(391.23)
脆壳蟹 <i>P. cinctipes</i>	7505(328.26)	430(18.81)	1545(67.58)	2431(106.33)	2044(89.40)	553(24.19)	87(3.81)	14595(638.36)
脊尾白虾 <i>P. carinicaude</i>	682(601.01)	20(17.62)	189(166.55)	41(36.13)	38(33.49)	8(7.05)	7(6.17)	985(868.02)
日本沼虾 <i>M. nipponense</i>	678(124.60)	41(7.53)	185(34.00)	311(57.15)	162(29.77)	27(4.96)	30(5.51)	1434(263.53)
罗氏沼虾 <i>M. rosenbergii</i>	274(151.66)	18(9.96)	79(43.73)	67(37.09)	60(33.21)	14(7.75)	6(3.32)	518(286.72)
合计 Total	18386(275.57)	1020(15.29)	4178(62.62)	6472(97)	4061(60.87)	1313(19.68)	356(5.34)	35786(536.36)

表 4 枝鳃亚目不同重复单元 EST-SSR 的数量和分布频率

Tab.4 The number and frequency of SSR containing different size of repeat unit of EST sequences in Dendrobranchiata

物种 Species	复合(I) Compound SSRs(I)	复合(II) Compound SSRs(II)	二核苷酸 Di- nucleotide	三核苷酸 Tri- nucleotide	四核苷酸 Tetra- nucleotide	五核苷酸 Penta- nucleotide	六核苷酸 Hex- nucleotide	合计 Total
凡纳滨对虾 <i>L. vannamei</i>	7261(341.96)	362(17.05)	1135(53.45)	1183(55.71)	1357(63.91)	306(14.41)	175(8.24)	11779(554.74)
斑节对虾 <i>P. monodon</i>	5828(448.98)	357(27.50)	725(55.85)	620(47.76)	598(46.07)	110(8.47)	85(6.55)	8323(641.19)
中国明对虾 <i>F. chinensis</i>	619(350.20)	19(10.75)	75(42.43)	77(43.56)	73(41.30)	13(7.35)	15(8.49)	891(504.08)
日本囊对虾 <i>M. japonicus</i>	525(382.91)	34(24.80)	92(67.10)	74(53.97)	77(56.16)	10(7.29)	16(11.67)	828(603.90)
白滨对虾 <i>L. setiferus</i>	38(70.49)	1(1.86)	9(16.70)	25(46.38)	14(25.97)	2(3.71)	1(1.86)	90(166.96)
合计 Total	14271(376.64)	773(20.4)	2036(53.73)	1979(52.23)	2119(55.93)	441(11.64)	292(7.71)	21911(578.28)

表5 腹胚亚目中二、三核苷酸基元 SSR 的数量和频率分布
 Tab.5 The number and frequency of EST-SSR with di- and tri-nucleotide motifs in Pleocyemata

重复基元 Repeat motifs	AC/GT	AG/CT	AT/AT	CG/CG	AAC /GTT	AAG /CTT	AAT /ATT	ACC /GGT	ACG /CGT	ACT /AGT	AGC/ CTG	AGG /CCT	ATC /ATG	CCG /CGG	合计 Total
美国龙虾	982 (55.23)	375 (21.09)	1240 (69.74)	3 (0.17)	221 (12.43)	250 (14.06)	657 (36.95)	717 (40.33)	23 (1.29)	113 (6.36)	328 (18.45)	306 (17.21)	260 (14.62)	59 (3.32)	5534 (17.78)
<i>H. americanus</i>															
淡水螯虾	2 (6.25)	5 (15.63)	2 (6.25)	1 (3.13)	5 (15.63)	2 (6.25)	8 (25)	20 (62.5)	0 (0)	1 (3.13)	16 (50)	1 (3.13)	0 (0)	0 (0)	63 (0.32)
<i>P. leniusculus</i>															
日本龙虾	32 (52.46)	21 (34.43)	37 (60.66)	1 (1.64)	3 (4.92)	8 (13.11)	13 (21.31)	12 (19.67)	1 (1.64)	2 (3.28)	16 (26.23)	9 (14.75)	7 (11.48)	1 (1.64)	163 (0.61)
<i>P. japonicus</i>															
拟穴青蟹	53 (28.96)	39 (21.31)	5 (2.73)	1 (0.55)	8 (4.37)	17 (9.29)	13 (7.1)	39 (21.31)	1 (0.55)	20 (10.93)	33 (18.03)	38 (20.77)	17 (9.29)	2 (1.09)	286 (1.83)
<i>S. paramamosain</i>															
中华绒螯蟹	697 (180.57)	335 (86.79)	150 (38.86)	2 (0.52)	226 (58.55)	426 (110.36)	312 (80.83)	792 (205.18)	52 (13.47)	270 (69.95)	351 (90.93)	908 (235.23)	232 (60.1)	383 (99.22)	5136 (3.86)
<i>E. sinensis</i>															
三疣梭子蟹	251 (73.39)	170 (49.71)	57 (16.67)	5 (1.46)	35 (10.23)	48 (14.04)	84 (24.56)	167 (48.83)	2 (0.58)	60 (17.54)	103 (30.12)	194 (56.73)	36 (10.53)	14 (4.09)	1226 (3.42)
<i>P. trituberculatus</i>															
克氏原螯虾	1 (9.09)	0 (0)	0 (0)	0 (0)	2 (18.18)	0 (0)	0 (0)	13 (118.18)	0 (0)	0 (0)	1 (9.09)	4 (36.36)	0 (0)	2 (18.18)	23 (0.11)
<i>P. clarkii</i>															
招潮蟹	91 (97.85)	63 (67.74)	18 (19.35)	1 (1.08)	12 (12.9)	14 (15.05)	112 (120.43)	33 (35.48)	0 (0)	21 (22.58)	31 (33.33)	26 (27.96)	17 (18.28)	19 (20.43)	458 (0.93)
<i>U. pugnator</i>															
红蟹	19 (73.08)	23 (88.46)	3 (11.54)	0 (0)	3 (11.54)	10 (38.46)	14 (53.85)	16 (61.54)	4 (15.38)	1 (38.46)	10 (38.46)	32 (123.08)	6 (23.08)	3 (11.54)	153 (0.26)
<i>G. natalis</i>															
珍宝蟹	14 (51.85)	8 (29.63)	1 (3.7)	0 (0)	4 (14.81)	7 (25.93)	8 (29.63)	19 (70.37)	0 (0)	2 (7.41)	6 (22.22)	9 (33.33)	6 (22.22)	8 (29.63)	92 (0.27)
<i>M. magister</i>															
蓝蟹	308 (150.98)	104 (50.98)	34 (16.67)	4 (1.96)	34 (16.67)	32 (15.69)	55 (26.96)	282 (138.24)	3 (1.47)	47 (23.04)	194 (95.10)	136 (66.67)	35 (17.16)	121 (59.31)	1389 (2.04)
<i>C. sapidus</i>															
岸蟹	525 (59.19)	188 (21.2)	500 (56.37)	1 (0.11)	112 (12.63)	151 (17.02)	215 (24.24)	503 (56.71)	14 (1.58)	54 (6.09)	218 (24.58)	209 (23.56)	129 (14.54)	47 (5.3)	2866 (8.87)
<i>C. maenas</i>															
脆壳蟹	2822 (123.45)	1101 (48.16)	576 (25.2)	1 (0.04)	1882 (82.33)	340 (14.87)	1695 (74.15)	969 (42.39)	80 (3.5)	520 (22.75)	361 (15.79)	594 (25.98)	652 (28.52)	20 (0.87)	11613 (22.86)
<i>P. cinctipes</i>															
脊尾白虾	44 (38.94)	1269 (1123.01)	32 (28.32)	0 (0)	6 (5.31)	17 (15.04)	95 (84.07)	3 (2.65)	0 (0)	8 (7.08)	5 (4.42)	13 (11.5)	12 (10.62)	0 (0)	1504 (1.13)
<i>P. carinicauda</i>															
日本沼虾	73 (13.42)	269 (49.45)	74 (13.6)	1 (0.18)	58 (10.66)	92 (16.91)	47 (8.64)	58 (10.66)	12 (2.21)	45 (8.27)	86 (15.81)	92 (16.91)	89 (16.36)	14 (2.57)	1010 (5.44)
<i>M. nipponense</i>															
罗氏沼虾	27 (14.92)	69 (38.12)	100 (55.25)	3 (1.66)	8 (4.42)	28 (15.47)	39 (21.55)	11 (6.08)	13 (7.18)	7 (3.87)	18 (9.94)	14 (7.73)	28 (15.47)	0 (0)	365 (1.81)
<i>M. rosenbergii</i>															
合计 Total	5941 (89.04)	4039 (60.54)	2829 (42.4)	24 (0.36)	2619 (39.25)	1442 (21.61)	3367 (50.46)	3654 (54.77)	205 (3.07)	1180 (17.69)	1777 (26.63)	2585 (38.74)	1526 (22.87)	693 (10.39)	31881 (477.83)

表 6 枝鳃亚目中二、三核苷酸基元 SSR 的数量和频率分布
Tab.6 The number and frequency of EST-SSR with di- and tri-nucleotide motifs in Dendrobranchiata

重复基元 Repeat motifs	AC/GT	AG/CT	AT/AT	CG /CG	AAC /GTT	AAG /CTT	AAT /ATT	ACC /GGT	ACG /CGT	ACT /AGT	AGC /CTG	AG G/CC T	ATC /ATG	CCG /CGG	合计 Total
凡纳滨对虾 <i>L. vannamei</i>	1756 (82.71)	2186 (102.97)	647 (30.48)	18 (0.85)	282 (13.28)	792 (37.31)	1530 (72.07)	130 (6.12)	168 (7.91)	158 (7.44)	171 (8.05)	640 (30.15)	570 (26.85)	42 (1.98)	9090 (21.23)
斑节对虾 <i>P. monodon</i>	1283 (98.84)	2085 (160.63)	1560 (120.18)	126 (9.71)	167 (12.87)	566 (43.61)	1495 (115.18)	105 (8.09)	40 (3.08)	115 (8.86)	133 (10.25)	365 (28.12)	274 (21.11)	32 (2.47)	8346 (12.98)
中国明对虾 <i>F. chinensis</i>	85 (48.02)	81 (45.76)	161 (90.96)	0 (0)	7 (3.95)	32 (18.08)	120 (67.8)	12 (6.78)	8 (4.52)	7 (3.95)	16 (9.04)	20 (11.3)	31 (17.51)	0 (0)	580 (1.77)
日本囊对虾 <i>M. japonicus</i>	113 (82.48)	14 (7.107.3)	151 (110.22)	1 (0.73)	19 (13.87)	33 (24.09)	79 (57.66)	16 (11.68)	6 (4.38)	18 (13.14)	19 (13.87)	44 (32.12)	25 (18.25)	7 (5.11)	678 (1.37)
白滨对虾 <i>L. setiferus</i>	15 (27.78)	7 (12.96)	5 (9.26)	0 (0)	0 (0)	6 (11.11)	19 (35.19)	5 (9.26)	1 (1.85)	0 (0)	4 (7.41)	1 (1.85)	6 (11.11)	0 (0)	69 (0.54)
合计 Total	3252 (85.83)	4506 (118.92)	2524 (66.61)	145 (3.83)	475 (12.54)	1429 (37.71)	3243 (85.59)	268 (7.07)	223 (5.89)	298 (7.86)	343 (9.05)	1070 (28.24)	906 (23.91)	81 (2.14)	18763 (495.2)

的分布频率远高于其他类型的三核苷酸基元。从整体来看, 21 种十足目甲壳动物的 SSR 中多为富含 A/T 的类型。

2.4 脊尾白虾 EST 功能注释

通过 Blast2GO 功能注释工具, 比对脊尾白虾含 SSR 的序列与 NCBI 数据库中已知物种的基因(图 1), 脊尾白虾 EST-SSR 注释过程中引用已知注释基因的物种主要为端足虫(*Hyalella azteca*)(4.67%), 此外, 白滨对虾为 1.83%, 蓝蟹为 1.37%, 日本囊对虾和克氏原螯虾分别为 1.28%和 0.92%。

将脊尾白虾含有 SSR 的 EST 序列从生物学过程、分子功能和细胞组分 3 类进行 GO 分析。其中 47 条 (5.87%)序列有一个或多个已经预测的功能。由图 2 可以看出, 在 11 个生物学过程中细胞代谢过程所占比例最大, 达到 20%, 其次, 定位和代谢过程的比例为 18%。如图 3 所示, 5 种分子功能被记录 47 次, 其中, 链接产物的比例最大为 47%, 其次, 催化活性的比例为 28%。如图 4 所示, 9 个细胞组分中, 细胞组分和细胞比例为 24%。

3 讨论

本研究分析了 21 种十足目甲壳动物 EST 序列中 SSR 的分布频率和重复基元的特点, 发现腹胚亚目中含有 SSR 位点的非冗余 EST 序列占总数的 32.15%, 枝鳃亚目中为 31.00%, 二者比例均高于其他水产动物, 如牙鲆(*Paralichthys olivaceus*)中的比例为 7.95% (陈松波等, 2010), 在斑点叉尾鲷(*Ictalurus punctatus*)中的比例为 11.2%(Serapion *et al*, 2004), 在东部牡蛎

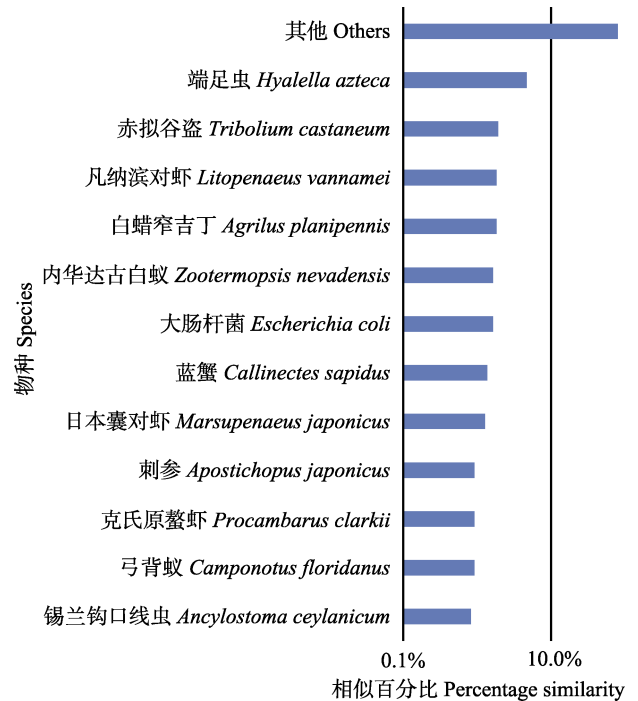


图 1 脊尾白虾 EST-SSR 与其他物种相似度统计图
Fig.1 Percentage similarity of Top-Blast hits for EST-SSR in *E. carinicauda* against other species

(*Crassostrea virginica*)中的比例为 1.4% (Wang *et al*, 2007), 在金头鲷(*Sparus aurata*)中的比例为 4.94% (Vogiatzi *et al*, 2011)。这些差异可能是因为 EST-SSR 在水产动物中的物种特异性较高(陈松波等, 2010), 也可能是使用不同的搜索软件或者搜索条件不同所引起的。

在腹胚亚目和枝鳃亚目中, 不同重复基元 SSR 的分布频率有着非常明显且一致的分布特征, 均是复

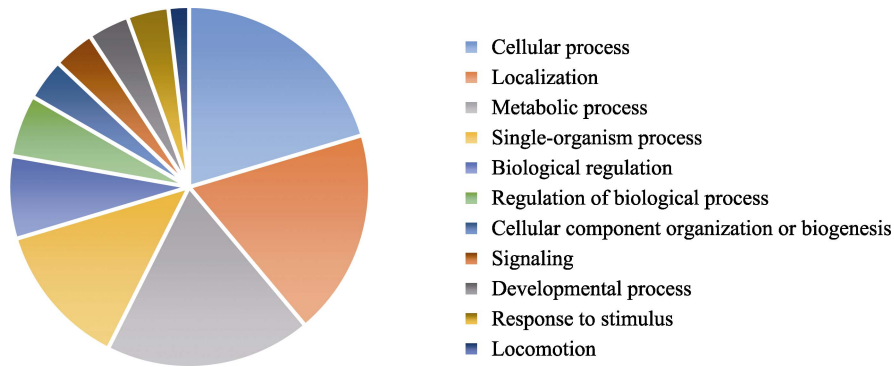


图 2 GO 分析-生物学过程
Fig.2 GO analysis-Biological process

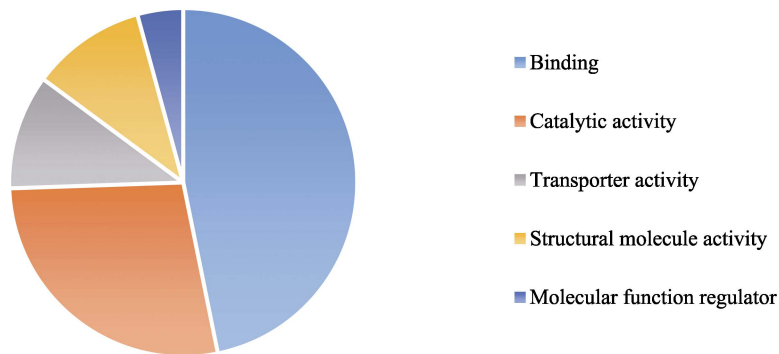


图 3 GO 分析-分子功能
Fig.3 GO analysis-Molecular function

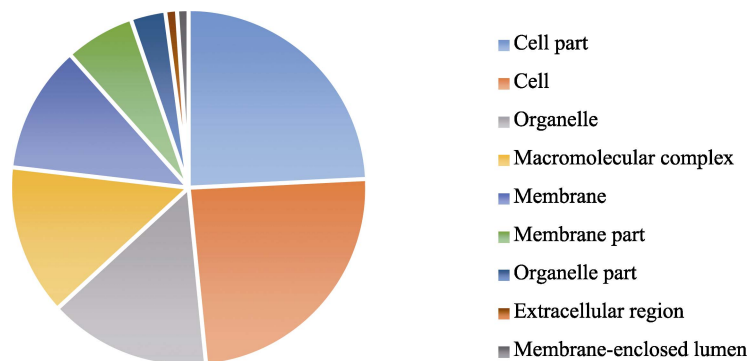


图 4 GO 分析-细胞组分
Fig.4 GO analysis-Cellular component

合(I)型远高于其他类型,斑鳢中复合 SSR 类型也占有相当的比例(占总数的 7.07%)(谢楠等, 2016)。二、三、四核苷酸基元的数目多于五、六核苷酸基元,在其他的水产动物如牙鲆(陈松波等, 2010)和金头鲷(Vogiatzi *et al.*, 2011)中,同样是二、三核苷酸基元的数目较多, Wang 等(2005)研究发现,中国对虾 EST-SSR 中三核苷酸基元的数目最多,与本研究对中国对虾的结果相一致;十足目甲壳动物中四核苷酸基元的数目也占很大比例,特别是在凡纳滨对虾中,四核苷酸数目最多,相似的报道较少。

腹胚亚目中二核苷酸重复中分布频率最高的类型为 AC/GT,同 Vogiatzi 等(2011)对金头鲷的研究中发现 AC/GT 占优势的结果相一致;枝鳃亚目中较高的类型为 AG/CT,同样在南亚野鲮(*Labeo rohita*)的研究中发现 AG/CT 占优势(Dinesh *et al.*, 2013)。2 种亚目中均为 CG/CG 的分布频率最低,其中,中国明对虾、脊尾白虾、白滨对虾、克氏原螯虾、红蟹和珍宝蟹中没有 CG/CG 类型,大部分鱼类的研究中,也没有 CG/CG 类型(Ju *et al.*, 2005)。不同三核苷酸基元的 SSR 的分布频率在不同物种中差异较大,整体上来

看,在枝鳃亚目中分布较腹胚亚目集中,枝鳃亚目中为 AAT/ATT 最多,腹胚亚目中较多的类型集中于 ACC/GGT、AAT/ATT,3 种核苷酸基元均为富含 A/T 的 SSR,在许氏平鲉(*Sebastes schlegelii*)的研究中,三核苷酸基元同样为富含 A/T 的 CTG/CAG 和 GAA/TTC 出现频率最高(薛蕊等,2015)。

比对脊尾白虾含 SSR 的序列与 NCBI 数据库中已知物种的基因,发现脊尾白虾同虾蟹类的基因相似度较低,Wang 等(2017)的研究发现,脊尾白虾与凡纳滨对虾和斑节对虾等物种的基因相似比例也较低,主要原因是 NCBI 数据库中有关十足目甲壳动物的数据资源较少。GO 分析后发现,细胞代谢过程、链接产物、细胞组分和细胞的比例分别在 3 个类型的注释中占优势,这与凡纳滨对虾(孙政等,2013)和梭鲈(*Sander lucioperca*)(Han *et al.*,2016)的研究结果相一致。

4 结论

本研究发现十足目甲壳动物中 EST-SSR 出现频率较高,2 种亚目中二、三、四核苷酸基元的 SSR 较多,且类型丰富。EST-SSR 信息的挖掘有利于加深人们对于十足目甲壳动物 SSR 分布规律的认识,为甲壳动物 EST-SSR 标记的开发及实际应用提供参考,推动甲壳动物种质资源的创新利用。针对脊尾白虾的基因功能注释分析发现,在 3 个类型的注释中,细胞代谢过程、链接产物、细胞组分和细胞的比例占有优势,为脊尾白虾在基因功能方面的深入研究提供参考资料。

参 考 文 献

- Bai JY, Pang YZ, Yang YB, *et al.* Development and application of EST-SSR markers in the black quail. *Animal Husbandry and Veterinary Medicine*, 2016, 48(11): 49–51 [白俊艳, 庞有志, 杨又兵, 等. 黄羽鹌鹑 EST-SSR 标记的开发与应用. *畜牧与兽医*, 2016, 48(11): 49–51]
- Chen SB, Gong L, Liu HJ. Analysis of SSR information in EST resources of Japanese flounder (*Paralichthys olivaceus*). *Journal of Northeast Agricultural University*, 2010, 42(42): 1215–1220 [陈松波, 龚丽, 刘海金. 牙鲆 EST 资源的 SSR 信息分析. *东北农业大学学报*, 2010, 42(42): 1215–1220]
- Dinesh KS, Soumya PP, Sujata P, *et al.* Identification of reproduction-related genes and SSR-markers through expressed sequence tags analysis of a monsoon breeding carp rohu, *Labeo rohita* (Hamilton). *Gene*, 2013, 524(1): 1–14
- Du NS. *Crustacean biology*. Beijing: Science Press, 1993, 675–760 [堵南山. 甲壳动物学. 北京: 科学出版社, 1993, 675–760]
- Guo Q, Guo LL, Zhang L, *et al.* Construction of a genetic linkage map in tree peony (*Paeonia Sect. Moutan*) using simple sequence repeat (SSR) markers. *Scientia Horticulturae*, 2017, 219: 294–301
- Han X, Ling Q, Li C, *et al.* Characterization of pikeperch (*Sander lucioperca*) transcriptome and development of SSR markers. *Biochemical Systematics and Ecology*, 2016, 66: 188–195
- Huang QX, Zuo J, Kong H, *et al.* Data-mining and diversity analysis of EST-SSRs from 11 tropical plant species. *Chinese Journal of Tropical Crops*, 2012, 33(7): 1208–1214 [黄启星, 左娇, 孔华, 等. 11 种热带植物 EST-SSR 标记的开发和多样性分析. *热带作物学报*, 2012, 33(7): 1208–1214]
- Ju Z, Wells MC, Martinez A, *et al.* An in silico mining for simple sequence repeats from expressed sequence tags of zebrafish, medaka, fundulus, and xiphophorus. *In Silico Biology*, 2005, 5: 439–463
- Li DY, Kong J, Meng XH, *et al.* Development of multiplex PCR systems of microsatellite markers for pacific white shrimp (*Litopenaeus vannamei*) and its application for parentage identification. *Progress in Fishery Sciences*, 2016, 37(3): 58–67 [李东宇, 孔杰, 孟宪红, 等. 凡纳滨对虾 (*Litopenaeus vannamei*) 微卫星多重 PCR 体系的建立及其在家系亲权鉴定中的应用. *渔业科学进展*, 2016, 37(3): 58–67]
- Serapion J, Kucuktas H, Feng JN, *et al.* Bioinformatic mining of type I microsatellites from expressed sequence tags of channel catfish (*Ictalurus punctatus*). *Marine Biotechnology*, 2004, 6(4): 364–377
- Singh RK, Jena SN, Khan S, *et al.* Development, cross-species/genera transferability of novel EST-SSR markers and their utility in revealing population structure and genetic diversity in sugarcane. *Gene*, 2013, 524(2): 309–329
- Sun Z, Liu CZ, Li FH, *et al.* Bioinformatic processing of a large number of *Litopenaeus vannamei* ESTs and analysis of tissue-specific gene expression. *Acta Oceanologica Sinica*, 2013, 35(5): 128–136 [孙政, 柳承璋, 李富花, 等. 凡纳滨对虾大量 EST 的生物信息学分析. *海洋学报*, 2013, 35(5): 128–136]
- Varshney RK, Graner A, Sorrells ME. Genic microsatellite markers in plants: Features and applications. *Trends in Biotechnology*, 2005, 23(1): 48–55
- Vogiatzi E, Lagnel J, Pakaki V, *et al.* In silico mining and characterization of simple sequence repeats from gilthead sea bream (*Sparus aurata*) expressed sequence tags (EST-SSRs): PCR amplification, polymorphism evaluation and multiplexing and cross-species assays. *Marine Genomics*, 2011, 4(2): 83–91
- Wang HX, Li FH, Xiang JH. Polymorphic EST-SSR markers and their mode of inheritance in *Fenneropenaeus chinensis*. *Aquaculture*, 2005, 249(1–4): 107–114
- Wang JJ, Ge QQ, Li JT, *et al.* Identification of novel EST-SSR markers by transcriptome sequencing in ridgetail white prawn *Exopalaemon carinicauda*. *Genes and Genomics*, 2017, 40(1): 1–9
- Wang RF, Li J, Li JT, *et al.* Genetic analysis of 33 microsatellite loci in 3 families of *Exopalaemon carinicauda* inbred line. *Progress in Fishery Sciences*, 2017, 38(4): 78–86 [王日芳, 李健, 李吉涛, 等. 脊尾白虾 (*Exopalaemon carinicauda*) 近交系 3 个家系 33 个微卫星座位的遗传分析. *渔业科学进展*, 2017, 38(4): 78–86]
- Wang YP, Guo XM. Development and characterization of EST-SSR markers in the eastern oyster *Crassostrea virginica*. *Marine Biotechnology*, 2007, 9(4): 500–511

- Xie N, Liu K, Feng XY, *et al.* Analysis of simple sequence repeats information in expressed sequence tags of *Channa argus* and *Channa maculatus*. *Freshwater Fisheries*, 2016, 44(3): 16–22 [谢楠, 刘凯, 冯晓宇, 等. 乌鳢和斑鳢 EST 序列微卫星信息分析. *淡水渔业*, 2016, 44(3): 16–22]
- Xue R. Development of EST-SSR markers and genetics assessment of fast-growing breeding line for the black rockfish, *Sebastes schlegelii*. Master's Thesis of Shanghai Ocean University, 2015 [薛蕊. 许氏平鲷 (*Sebastes schlegelii*) EST-SSR 标记开发及快速生长选育系遗传结构分析. 上海海洋大学硕士研究生学位论文, 2015]
- Ye J, Fang L, Zheng H, *et al.* WEGO: A web tool for plotting GO annotations. *Nucleic Acids Research*, 2006, 34: 293–297
- Yumurtaci A, Sipahi H, Al-Abdallat A, *et al.* Construction of new EST-SSRs for *Fusarium* resistant wheat breeding. *Computational Biology and Chemistry*, 2017, 68: 22–28
- Zhou L, Liu T, Cheng YK, *et al.* Molecular mapping of a stripe rust resistance gene in Chinese wheat landrace “Hejiangyizai” using SSR, RGAP, TRAP, and SRAP markers. *Crop Protection*, 2017, 94: 178–184
- Zhou Q. Screening of highly polymorphic EST-SSR markers and investigation of genetic diversity in *Macrobrachium nipponensis*. Master's Thesis of Nanjing Agricultural University, 2013 [周巧. 青虾高多态性 EST-SSR 标记的筛选及遗传多样性的研究. 南京农业大学硕士研究生学位论文, 2013]

(编辑 冯小花)

Analysis of SSR Information in EST Resource of Decapod Crustaceans

GAO Haiyu^{1,2,3}, LI Jian^{1,2,3}①, WANG Jiajia^{2,3}, LI Jitao^{2,3}

(1. Shanghai Ocean University, Shanghai 201306; 2. Laboratory for Marine Fisheries Science and Food Production Processes, Pilot National Laboratory for Marine Science and Technology (Qingdao), Qingdao 266071; 3. Key Laboratory of Sustainable Development of Marine Fisheries, Ministry of Agriculture and Rural Affairs, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao 266071)

Abstract The distribution frequency and characteristic of base repeats of EST-SSR (expressed sequence tag - simple sequence repeats) were induced and analyzed in 21 kinds of decapod crustaceans by applying bioinformatics methods. The results showed that EST-SSR abundances were different between various crustacean species. *Exopalaemon carinicauda* had the most abundant EST-SSR (868.02/Mb), whereas *Scylla paramamosain* had the least (286.48/Mb) in pleocyemata. *Penaeus monodon* had the most abundant EST-SSR (641.19/Mb), whereas *Litopenaeus setiferus* had the least (166.96/Mb) in dendrobranchiata. The abundance of EST-SSR with di-, tri-, and tetra-nucleotide motifs was more than that of those with penta- and hex- nucleotide motifs, which accounts for 41.11% in pleocyemata and 28.00% in dendrobranchiata. EST-SSR of compound (I) type occupied a large proportion in the pleocyemata (51.38%) and dendrobranchiata (65.13%). The frequency of the AC/GT repeat motif distribution was highest in dinucleotides, and ACC/GGT and AAT/ATT repeat motifs were the most abundant in the trinucleotide repeats in the pleocyemata. The frequency of AG/CT repeat motif distribution was highest in dinucleotides, and the distribution frequency of AAT/ATT motifs was significantly higher than other motifs of dinucleotide repeats in the dendrobranchiata. A total of 12,155 sequences containing SSRs were predicted; moreover, Gene Ontology (GO) Classification with the blast2go application was performed based on sequences containing SSRs of *Exopalaemon carinicauda*. The results demonstrated that ‘cellular process’ comprised the largest proportion in the biological process category, whereas ‘binding’ comprised the largest proportion in the molecular function category. Additionally, the cellular component category showed that many sequences likely possessed ‘cell parts’ and ‘cell’ by GO annotation. This study compared the characteristics of EST-SSR in different species of crustaceans and the diverse regions of the species genome. The consequences deepened our understanding of the distribution of SSRs and provided a reference for the development and practical applications of EST-SSR markers. The results also provided powerful information for future conservation and breeding research.

Key words Decapoda; Crustacean; EST-SSR

① Corresponding author: LI Jian, E-mail: lijian@ysfri.ac.cn