

王传聪,唐修阳,项 杰,等. 罗氏沼虾转录组 SSR 标记信息分析[J]. 江苏农业科学,2018,46(22):56–59.
doi:10.15889/j.issn.1002–1302.2018.22.011

罗氏沼虾转录组 SSR 标记信息分析

王传聪, 唐修阳, 项 杰, 欧江涛
(盐城工学院海洋与生物工程学院, 江苏盐城 224051)

摘要:对罗氏沼虾肝胰腺组织转录组测序获得的 33 450 条 Unigenes,进行简单重复序列(simple sequence repeat, 简称 SSR)检测与分析。结果共检测出 15 356 个 SSR 位点,分布在 12 930 条 Unigenes 上,发生频率为 38.65%,平均距离为 1.91 kb。其中,优势重复基序为单核苷酸、二核苷酸、三核苷酸。SSR 序列重复长度主要为 12~16 bp 和 18 bp,占罗氏沼虾总 SSR 的 86.33%。单核苷酸重复基序中多数为 A/T,二核苷酸重复基序以 GA/TC、AG/CT 及 AT/TA 为主,三核苷酸重复基序主要为 AAT/TTA、AAG/TTC、ATT/TAA、CCT/GGA、CTC/GAG 和 CTT/GAA。最后,对 SSR 所在的 Unigene 进行了 KOG 分类注释。通过对罗氏沼虾转录组 SSR 位点信息分析,获得多态性较为丰富的 SSR 位点,具有极大的开发价值,可为罗氏沼虾遗传多样性的分析、遗传图谱的构建及分子标记辅助育种提供有效工具。

关键词:罗氏沼虾;转录组;SSR 序列;分子标记
中图分类号: S917.4 **文献标志码:** A **文章编号:** 1002–1302(2018)22–0056–04

罗氏沼虾别称马来西亚大虾,隶属节肢动物门、甲壳纲、十足目、长臂虾科、沼虾属。原产于南亚、东南亚以及大洋洲北部等地区,生活在淡水或咸淡水水域,自然栖息于受潮汐影响的河口区,是一种较大型的热带经济虾类^[1],具有生长快、肉质营养成分好以及养殖周期短等优点,素有“淡水虾王”之称^[2]。然而,随着集约化养殖的快速发展,一些严重的疾病已在罗氏沼虾中流行,包括由新型病原——螺原体引起的 2010 年罗氏沼虾螺原体病,引发大规模的死亡,给水产养殖业造成重大经济损失。这些水产动物病害在带来巨大经济效益损失的同时,也严重制约了罗氏沼虾产业的可持续发展^[3]。因此,及时开展罗氏沼虾的抗病研究是极为重要的。

微卫星标记是目前最常用的分子标记之一。由于与其他分子标记相比,它具有保守性高、多态性丰富、在基因组中分布广泛等特点^[3],近年来,在动物遗传连锁图谱的构建、种质鉴定、遗传多样性分析以及分子标记辅助育种等领域被广泛应用。目前在水生动物相关研究中,研究者已经对牙鲆、二长棘鲷、兴国红鲤、中国明对虾等品种进行了转录组测序,并基于测序结果对所得数据进行了深入分析^[4–7]。因此,开展罗氏沼虾转录组简单重复序列(simple sequence repeat,简称 SSR)的研究具有重要的理论和现实意义。本研究通过对罗氏沼虾转录组测序,筛选出 SSR 位点,对其进行数量分析和多态性评价等,从而为罗氏沼虾基因克隆、遗传图谱构建、遗传多样性分析以及为罗氏沼虾病害防治和选育种工作等提供基础数据。

收稿日期:2017–07–03
基金项目:国家自然科学基金面上项目(编号:31570176);国家自然科学基金青年科学基金(编号:31000072)。
作者简介:王传聪(1995—),男,安徽淮北人,硕士,主要从事海洋生物技术研究。E-mail:2582302990@qq.com。
通信作者:欧江涛,博士,副教授,主要从事分子遗传学和基因组学研究。E-mail:ojt110@126.com。

1 材料与方法

1.1 试验材料

罗氏沼虾:购于江苏省南京市仙林农贸市场,个体质量在 20~25 g,共计 100 尾,用 PCR 方法检测螺原体,结果均呈阴性,在 26~28 ℃水温条件下养殖备用。螺原体 MR-1008:分离于江苏省高邮市自然发病的罗氏沼虾,在 R2 液体培养基中于 30 ℃孵育 48 h,待其生长到最具有侵染能力的对数期备用。

1.2 总 RNA 提取

采用 TRIzol 提取法提取罗氏沼虾肝胰腺组织总 RNA,提取产物用 15 g/L 琼脂糖凝胶电泳进行质检。质检合格的样品保存在干冰中送往联川生物公司进行测序。

1.3 RNA 测序及组装

经 Illumina HiSeq2000/2500 进行测序,在健康罗氏沼虾和螺原体感染罗氏沼虾的肝胰腺组织中得到总 mRNA 原始数据(raw reads)分别为 53 070 612、61 244 504 条,对 raw reads 进行过滤得到有效数据(clean reads)分别为 52 757 004、60 719 728 条(表 1)。将得到的有效数据使用 Trinity 软件进行拼接,最终得到 33 450 条 Unigenes(总长度 29 327.225 kb)对应 43 405 个转录本,长度在 200~2 000 bp 之间的超过 85%,符合基因转录本的长度规律(表 2)。

表 1 测序数据预处理结果

罗氏沼虾 类别	原始数据 (条)	有效数据 (条)	Q ₂₀ (%)	Q ₃₀ (%)	GC (%)
健康	53 070 612	52 757 004	96.86	89.66	46.02
螺原体感染	61 244 504	60 719 728	97.05	90.29	43.17

注:Q₂₀和 Q₃₀分别代表碱基被测错的概率 P 为 1% 和 1‰,其计算公式为 $Q = -10\lg P$ 。GC 为 G 和 C 的数量总和占总碱基数量的百分比,用于检测有无 AT/GC 分离现象。

1.4 SSR 标记筛选

SSR 检测是以组装出来的 Unigene 作为参考序列,使用 MISA 软件批量识别和定位所有 SSR,然后对罗氏沼虾转录组

表 2 拼接结果统计

类别	总碱基数 (bp)	预测数 (条)	平均 GC 含量 (%)	N ₅₀ (bp)	不同长度 Unigenes 或转录本的数量(个)与占比			
					200 ~ 500 bp	500 ~ 1 000 bp	1 000 ~ 2 000 bp	≥ 2 000 bp
Unigenes	29 327 225	33 450	39.80	1 547	17 560(52.50%)	6 920(20.69%)	5 367(16.04%)	3 603(10.77%)
转录本	43 710 264	43 405	40.41	1 782	20 415(47.03%)	9 339(21.52%)	7 743(17.84%)	5 908(13.61%)

注: N₅₀表示将 Unigenes 从长到短排序,依次累加 Unigenes 碱基数后,Unigenes 总碱基数 50% 的 Unigenes 的长度,转录本同。

SSR 数据进行分类统计分析。

2 结果与分析

2.1 罗氏沼虾转录组中 SSR 位点的数量与分布

对组装所得的 33 450 条 unigenes 序列进行比对筛选,结果发现 15 356 个 SSR 位点,分布在 12 930 条 Unigenes 上,发生频率(含有 SSR 的 Unigenes 条数占总 Unigenes 条数的比例)为 38.65%,其中 10 504 条 Unigenes 序列只含有单个 SSR 位点,2 426 条 Unigenes 序列含有 2 个或 2 个以上的 SSR 位点。由表 3 可知,罗氏沼虾的 SSR 序列平均距离为 1.91 kb,总长度为 230 820 bp,平均长度为 15 bp。

罗氏沼虾转录组中各种 SSR 出现频率差异较大,各类型出现的频率不同,主要为单核苷酸、二核苷酸、三核苷酸重复(图 1)。单核苷酸重复的 SSR 数量最多,其分布频率为 17.70%,占罗氏沼虾总 SSR 的 38.56%,其次为二核苷酸重

表 3 SSR 在罗氏沼虾转录组中出现的频率

重复类型	SSR 位点数 (个)	频率 (%)	平均距离 (kb)	总长度 (bp)	平均长度 (bp)
单核苷酸	5 922	17.70	5.53	83 260	14
二核苷酸	5 528	16.53	5.31	81 192	15
三核苷酸	3 581	10.71	8.19	59 733	17
四核苷酸	176	0.53	166.63	3 656	21
五核苷酸	80	0.24	366.60	1 485	20
六核苷酸	69	0.21	425.03	1 494	26
总计	15 356	38.65	1.91	230 820	15

复,其分布频率为 16.53%,占罗氏沼虾总 SSR 的 36.00%,三核苷酸重复的 SSR 分布频率为 10.71%,占罗氏沼虾总 SSR 的 23.32%。四核苷酸、五核苷酸、六核苷酸重复类型的 SSR 类型较少,分布频率分别为 0.53%、0.24%、0.21%,分别占罗氏沼虾总 SSR 的 1.15%、0.52%、0.45%。

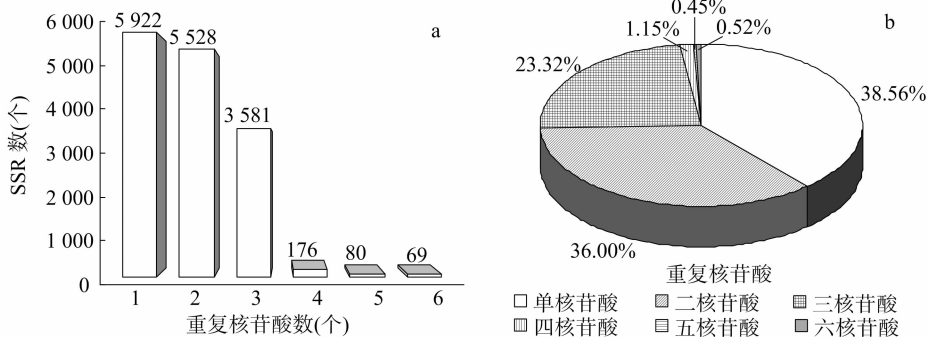


图 1 不同类型 SSR 数量统计(a)与重复单元分布(b)

由表 4 可知,罗氏沼虾 SSR 重复次数为 5、6、7、12 次的 SSR 位点较多,分别为 2 245、4 540、1 647、1 796 个,分别占罗氏沼虾总 SSR 的 14.62%、29.56%、10.73%、11.70%,其次为重复 8、13、14 次的 SSR 位点,分别为 938、1 216、925 个,分别占总 SSR 的 6.11%、7.92%、6.02%,重复次数 ≥ 16 次的 SSR 位点相对较少,总共为 1 277 个,占总 SSR 的 8.32%。

2.2 罗氏沼虾转录组中 SSR 的特点

罗氏沼虾转录组中共搜索到 512 种不同序列类型的 SSR,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重复序列类型分别有 40、66、201、89、60、56 种。单核苷酸中出现的重复类型多为 A/T,分别占罗氏沼虾总 SSR 的 39.33%,C/G 类型单核苷酸序列的 SSR 非常少。在二核苷酸重复中主要以 GA/TC、AG/CT 和 AT/TA 为主,分别为 1 184、1 501、1 300 个,分别占二核苷酸重复 SSR 的 21.42%、27.15%、23.52%,分别占罗氏沼虾总 SSR 的 7.71%、9.77%、8.47%。三核苷酸重复在罗氏沼虾总 SSR 中的占比较大,其重复类型见表 5。各种三核苷酸重复类型出现的数量差别较大,其中 AAT/TTA 出现的数量最多,为 216 个,出现频率为 0.65%,占罗氏沼虾总 SSR 的 1.41%;其次为

AAG/TTC、ATT/TAA、CCT/GGA、CTC/GAG、CTT/GAA,分别为 204、187、197、191、215 个,分别占罗氏沼虾总 SSR 的 1.33%、1.22%、1.28%、1.24%、1.40%;其中 CGG/GCC 出现的数量最少,仅为 8 个,占罗氏沼虾总 SSR 的 0.05%。

2.3 罗氏沼虾转录组 SSR 多态性评价

由图 2 可知,罗氏沼虾 SSR 重复长度分布中重复长度最小为 12 bp,最大为 25 bp,平均长度为 15 bp。其中重复长度主要为 12 ~ 16 bp 及 18 bp,占罗氏沼虾总 SSR 的 86.33%,其次是重复长度为 17 bp 及 20 ~ 22 bp,占罗氏沼虾总 SSR 的 11.91%,重复长度在 19 bp 和 ≥ 23 bp 的 SSR 位点数相对较少,分别仅占罗氏沼虾总 SSR 的 0.79%、0.96%。根据该结果,可推测该研究中 II 型 SSR(长度为 12 ~ < 20 bp)的比例为 89.07%,具有中等多态性,有较高的潜在可用性。而具有高度多态性的 I 型 SSR(长度 ≥ 20 bp)则较少。

2.4 含 SSR 位点 Unigene 的功能注释

利用软件对成功筛查到的 SSR 位点所在的 12 930 条 Unigene 序列进行功能注释。通过 Blast 比对,有 410 个 Unigene 被注释到 NCBI Nr 数据库的已知序列,并呈现较高的相似性和同源性,但有 96.83% 的 Unigene 未得到任何注释。

表 4 罗氏沼虾 SSR 重复次数分布

重复类型 (次)	位点数(个)					
	单核苷酸重复	二核苷酸重复	三核苷酸重复	四核苷酸重复	五核苷酸重复	六核苷酸重复
4	0	0	0	0	75	69
5	0	0	2 092	148	5	0
6	0	3 514	998	28	0	0
7	0	1 195	452	0	0	0
8	0	899	39	0	0	0
9	0	716	0	0	0	0
10	0	517	0	0	0	0
11	0	187	0	0	0	0
12	1 792	4	0	0	0	0
13	1 216	0	0	0	0	0
14	925	0	0	0	0	0
15	712	0	0	0	0	0
16	483	0	0	0	0	0
17	299	0	0	0	0	0
18	221	0	0	0	0	0
19	121	0	0	0	0	0
20	103	0	0	0	0	0
21	24	0	0	0	0	0
22	23	0	0	0	0	0
23	3	0	0	0	0	0
总和	5 922	5 528	3 581	176	80	69

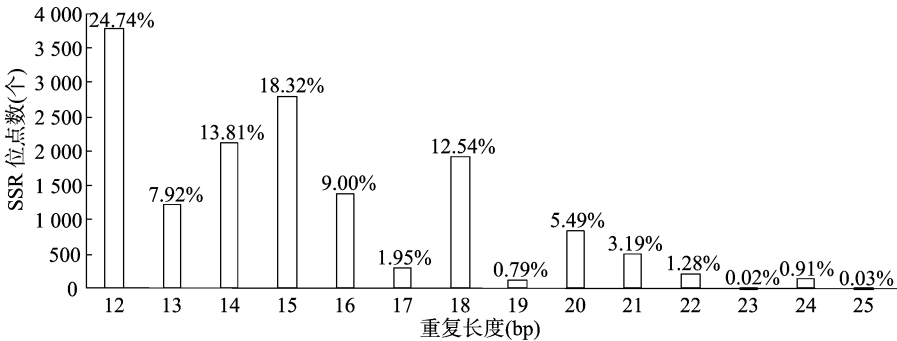


图2 罗氏沼虾 SSR 重复长度分布

将 15 356 条 SSR 位点所在的 12 930 条 Unigene 序列注释到 KOG 数据库中进行相关基因功能的预测和分类。结果显示其中的 275 条 Unigene(2.13%)能够在 KOG 找到相应的注释信息,根据其功能可以被分为 22 类(图 3),并对其进行数量统计。从分析结果可以看出,这 275 条被注释的 Unigene 功能种类较为全面,涉及大多数的生命活动过程或功能。“一般功能预测”是最大的一个分类,包含 57 条 Unigene。其次是“信号转导机制”“翻译后修饰,蛋白质周转,分子伴侣”和“RNA 加工和修饰”,分别包含 34、26 和 19 条 Unigene。“核结构”这个分类中包含的 Unigene 数最少,仅有 1 条。

3 讨论

本研究通过 IlluminaHiSeq2000/2500 高通量测序平台对罗氏沼虾肝胰腺组织转录组测序,从大量组装所得转录组 Unigene 序列中筛选获得 15 356 个 SSR 位点。相对于传统卫星标记的筛选方式,此方法筛选效率较高,工作量相对较小,适合大范围开拓 SSR 标记位点^[7]。

SSR 在整个基因组的不同位点都有分布,多态信息含量(polymorphism information content,简称 PIC)起衡量基因位点

多态性的作用,通常 PIC 能反映某个群体的遗传变异程度、位点多样性等^[8]。另外 SSR 片段长度也是判断其多态性的重要依据^[9]。从转录组数据筛选得到的 SSR 中,重复序列长度在 12 bp 以上的 SSR 标记位点数占总 SSR 数的 76.95%,多态性较丰富,基于该研究结果能够进行有针对性的引物设计。在数量性状位点(QTL)定位研究及遗传连锁图谱的构建中,SSR 多态性越高,所建立的图谱越精密和精确,基因的定位越精准^[10-11]。

随着微卫星辅助育种技术的广泛应用,通过遗传操作培养出罗氏沼虾抗病品种,是虾病控制中一个值得选择的方法。已有的研究表明,疾病发生时群体内会有显著的变异,通过微卫星等标记对发病动物中死亡群体和存活群体的遗传变异进行研究,有望获得与抗病性状相连锁的标记,进而对抗病基因进行标记和定位以培育出具有持久抗病能力的优良品种^[12]。

本研究对罗氏沼虾 SSR 的分布特征进行分析,不仅在 RNA 水平上体现了罗氏沼虾 SSR 的分布特点和规律,而且也作为开发罗氏沼虾功能基因奠定了 SSR 分子标记基础,同时也为罗氏沼虾 QTL 定位、基因克隆、遗传结构分析及其他遗传学研究提供了有效的分子标记。

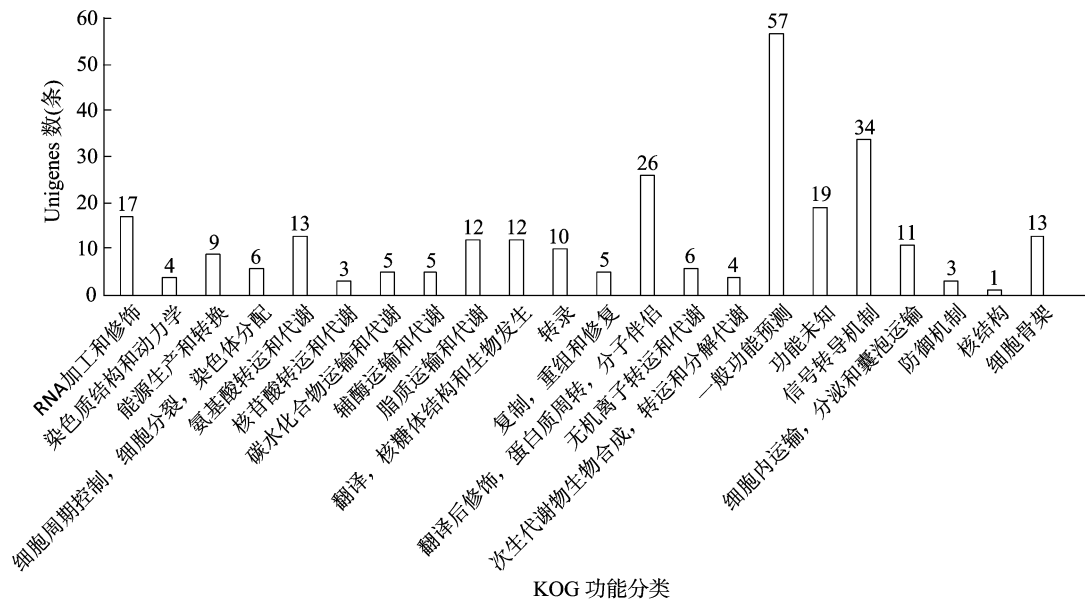


图3 含 SSR 位点 Unigene 的 KOG 分类

表 5 三核苷酸 SSR 不同重复单元的比例

重复单元	数量 (个)	频率 (%)	占比 (%)
AAC/TTG	69	0.21	0.45
AAG/TTC	204	0.61	1.33
AAT/TTA	216	0.65	1.41
ACA/TGT	71	0.21	0.46
ACC/TGG	97	0.29	0.63
ACG/TGC	92	0.28	0.60
ACT/TGA	126	0.38	0.82
AGA/TCT	156	0.47	1.02
AGC/TCG	96	0.29	0.63
AGT/TCA	160	0.48	1.04
AGG/TCC	152	0.45	0.99
ATA/TAT	167	0.50	1.09
ATC/TAG	86	0.26	0.56
ATG/TAC	143	0.43	0.93
ATT/TAA	187	0.56	1.22
CAA/GTT	68	0.20	0.44
CAC/GTG	91	0.27	0.59
CAG/GTC	135	0.40	0.88
CAT/GTA	96	0.29	0.63
CCA/GGT	92	0.28	0.60
CCG/GGC	24	0.07	0.16
CCT/GGA	197	0.59	1.28
CGA/GCT	87	0.26	0.57
CGC/GCG	28	0.08	0.18
CGG/GCC	8	0.02	0.05
CGT/GCA	87	0.26	0.57
CTA/GAT	114	0.34	0.74
CTC/GAG	191	0.57	1.33
CTG/GAC	126	0.38	0.82
CTT/GAA	215	0.64	1.40

注:频率为含有 SSR 的 Unigenes 数与总 Unigenes 数之比;占比为该重复单元 SSR 占总 SSR 的比例。

参考文献:

[1] 马 劳. 淡水虾养殖 罗氏沼虾养殖手册[M]. 罗马:联合国粮
食及农业组织,2005:1-11.

[2] 刘 恩,生万全. 罗氏沼虾的养殖现状与发展前景[J]. 安徽农
业大学学报,1997,24(2):189-192.

[3] 杨 明,丁福江. 罗氏沼虾产业健康发展措施初探[J]. 上海农
业科技,2014(3):62.

[4] 赵莹莹,朱晓琛,孙效文,等. 虾夷扇贝的多态性微卫星座位[J].
动物学报,2006,52(1):229-233.

[5] Song W T, Li Y Z, Zhao Y W, et al. Construction of a high density
microsatellite genetic linkage map and mapping of sexual and
growthrelated traits in halfsmooth tongue sole(*Cynoglossus semilaevis*)
[J]. PLoS One,2012,7(11):1-13.

[6] OZAKIA, Yoshida K, Fuji K, et al. Quantitative trait loci (QTL)
associated with resistance to a monogenean parasite (*Benedenia
seriolae*) in yellowtail owtail (*Seriola quinqueradiata*) through genome
wide analysis[J]. PLoS One,2013,8(6):1-14.

[7] 刘 磊,彭士明,高权新,等. 基于银鲈 RNA-seq 数据中 SSR 标
记的信息分析[J]. 安徽农业科学,2016,44(28):102-105.

[8] Shetes, Tiwari H, Elston R C. On estimating the eterozygosity and
polymorphism information content value[J]. Theoretical Population
Biology,2000,57(3):265-271.

[9] 孙蛟龙,方 扬,靳艳玲,等. 浮萍转录组数据 SSR 位点的生物信
息学分析[J]. 应用与环境生物学报,2015,21(3):401-405.

[10] Pinto L R, Oliveira K M, Marconi T, et al. Characterization of novel
sugarcane expressed sequence tag microsatellites and their
comparison with genomic SSRs[J]. Plant Breeding,2006,125(4):
378-384.

[11] 孙成飞,叶 星,董浚键,等. 罗氏沼虾 6 个养殖群体遗传多样
性的微卫星分析[J]. 南方水产科学,2015,11(2):20-26.

[12] 朱其建,戴习林,邹卫丽,等. 罗氏沼虾抗病选育群体的抗病性
能及其遗传多样性分析[J]. 水产学报,2013,37(10):1468-
1478.