

doi: 10.3969/j.issn.1674-0858.2016.05.15

沟眶象转录组微卫星特征分析

武政梅, 高朋, 温俊宝*

(北京林业大学林木有害生物防治北京市重点实验室, 北京 100083)

摘要: 本研究基于高通量测序获得的沟眶象 *Eucryptorrhynchus chinensis* (Olivier) 转录组数据, 采用 MISA (Micro Satellite) 软件对其简单序列重复 (simple sequence repeat, SSR) 位点进行了高通量发掘。对筛选得到的 1 kb 以上的 Unigene (17590 条, 占 Unigene 总数的 26.98%) 进行了 SSR 分析, 结果发现含 SSR 的序列有 1665 条, 共发现 1823 个 SSR, 平均 32.52 kb 出现一个 SSR。沟眶象微卫星主要以三碱基重复类型为主 (637, 34.94%), 其次为单碱基重复 (576, 31.60%)。共发现 104 种碱基重复基元, 所占比例最高的为 (A/T)_n (30.39%), 其次为 (AT/AT)_n (6.91%)。沟眶象 SSR 的平均长度为 15.74 bp, 长度大于 20 bp 的 SSR 仅占总数的 12.70%。研究还发现沟眶象微卫星出现的频率和其长度呈极显著负相关 ($P < 0.01$), 相关系数为 -0.572。研究结果为开发高多态性微卫星引物来进行沟眶象功能基因组学、种群遗传结构、种群遗传多样性等研究奠定了基础。

关键词: 沟眶象; 微卫星; 转录组; 重复类型; 重复基元

中图分类号: Q965; S433.5

文献标识码: A

文章编号: 1674-0858 (2016) 05-0979-05

Characteristic analysis of microsatellite in *Eucryptorrhynchus chinensis* transcriptome

WU Zheng-Mei, GAO Peng, WEN Jun-Bao* (Beijing Key Laboratory for Forest Pest Control, Beijing Forestry University, Beijing 100083, China)

Abstract: According to the constructed transcriptome database of *Eucryptorrhynchus chinensis*, the microsatellites have been explored at a high-throughput scale by using Micro Satellite identification tool. The microsatellites in the unigenes (the length ≥ 1 kb) (17590, 26.89%) were analyzed and a total of 1823 SSR were identified in 1665 unigenes, with one SSR per 32.52 kb. Tri-nucleotide repeats were the most abundant (637, 34.94%), followed by mono nucleotide repeats (576, 31.60%). Among all the 104 SSR motifs, (A/T)_n was the most frequent repeat motif (30.39%), followed by (AT/AT)_n (6.91%). The average length of microsatellites was 15.74 bp, the microsatellites over 20 bp were only 12.70%. There was significant negative correlation ($P < 0.01$) between the frequency and the length of microsatellites in *e. chinensis* transcriptome and the correlation coefficient was -0.572. These results lay a foundation for developing high polymorphic microsatellites to research the functional genomics, population genetic structure and genetic diversity of *e. chinensis*.

Key words: *Eucryptorrhynchus chinensis*; microsatellite; transcriptome; repeat type; repeat motif

微卫星 (microsatellite) 又称简单序列重复 (simple sequence repeat, SSR), 一般以 1–6 个碱基为核心序列, 其长度通常较短 (何平, 1998)。

由于这些序列广泛存在于真核细胞的基因组 (史洁等, 2012), 同时具有可重复性高、多态性强、数量丰富、共显性、对基因组有很好覆盖性等特

基金项目: 林业公益性行业科研专项 (201304412; 201204501)

作者简介: 武政梅, 女, 1991 年生, 山东泰安人, 硕士, 研究方向为昆虫分子生物学, E-mail: zhengmeibest@163.com

* 通讯作者 Author for correspondence, E-mail: wenjb@bjfu.edu.cn

收稿日期 Received: 2015-10-29; 接受日期 Accepted: 2015-12-29

点 (Liu *et al.*, 2013), 因此微卫星已经成为不同物种遗传研究中使用最广泛的分子标记 (王丽鸳等, 2014), 已被广泛应用于种群遗传多样性分析、动植物分类和进化、遗传图谱的构建、基因定位和克隆、分子标记辅助育种、品种鉴定等领域 (Li *et al.*, 2002; Varshney *et al.*, 2005)。SSR 按来源分为基因组 SSR (Genomic SSR, gSSR) 和表达序列标签 SSR (Expressed SSR, EST-SSR) (王东等, 2014)。两种 SSR 相比较, 因 EST-SSR 源于基因的转录区, 其多态性与基因功能可能有直接关系 (Eujayl *et al.*, 2002), 因此通用性更高。新一代高通量测序技术的快速发展为利用转录组序列开发 SSR 标记带来了福音, 它不仅具有 EST-SSR 标记的优点, 而且能产生更为海量的转录组数据, 因此为 SSR 标记的开发提供了更全面的信息, 从而提高了遗传多样性研究的准确性。目前, 研究者们已经对很多物种转录组中的微卫星进行了研究分析, 鞘翅目昆虫中就有云南切梢小蠹 *Tomicus yunnanensis* (袁远等, 2014)、绿豆象 *Callosobruchus chinensis* (Duan *et al.*, 2014)、黄粉虫 *Tenebrio molitor* (Zhu *et al.*, 2013) 等, 结果显示在不同物种中微卫星的分布特征差异很大。

沟眶象 *Eucryptorrhynchus chinensis* Olivier 隶属鞘翅目 Coleoptera 象甲科 Curculionidae 隐喙象亚科 Cryptorrhynchinae 沟眶象属 *Eucryptorrhynchus* Hdlr, 主要以幼虫蛀食臭椿根部为害。由于近 10 多年来检疫不严, 沟眶象及其近缘种臭椿沟眶象 *Eucryptorrhynchus brandti* 已传播扩散至西北多个省区并成片严重发生, 整个宁夏二代农田林网中臭椿受害株率达 15% 以上, 致死率达 7%, 严重影响了臭椿的环境生态效益和观赏经济价值 (杨贵军等, 2008a)。目前有关沟眶象的报道, 主要集中在生物学、生态学以及防治措施等领域, 几乎没有有关其遗传背景和分子生物学方面的研究。本文根据已构建的沟眶象转录组数据库, 对其微卫星进行了高通量发掘, 旨在为今后开发高多态性微卫星引物, 对沟眶象进行功能基因组学、种群遗传结构、种群遗传多样性等研究打下基础。

1 材料与方法

1.1 试虫来源和转录组数据

沟眶象样品采自宁夏回族自治区银川市灵武

市梧桐树乡一队, 经液氮瞬时冷冻后放于 -80°C 的冰箱中保存备用。沟眶象总 RNA 样品检测合格后, 利用 Illumina HiSeq 2500 测序平台进行高通量深度测序, 共获得 65186 条 Unigenes, 转录组数据未发表。

1.2 微卫星的筛选

利用 MISA 软件 (Micro Satellite identification tool) (<http://www.pgrc.ipk-gatersleben.de/misa/>) 对沟眶象转录组中 1 kb 以上的 Unigene 进行 SSR 位点分析, 筛选标准为: 单碱基、二碱基、三碱基、四碱基、五碱基、六碱基的最少重复次数分别为 12、6、5、5、5、5。复合 SSR 两个位点间最大间隔碱基数为 (Maximal number of bases interrupting 2 SSRs in a compound microsatellite): 100。

2 结果与分析

2.1 沟眶象转录组序列中 SSR 重复的分布

本研究利用 MISA 软件对筛选得到的 1 kb 以上的 Unigene 进行了 SSR 分析, 评估序列拼接总长度为 59281851 bp, 获得 SSR 数目为 1823 个 (完整型 SSR 1709 个), 这些 SSR 分布于 1665 条序列中, 平均相隔 32.52 kb 出现一个 SSR 序列, SSR 的出现频率为 10.36%。本研究仅对沟眶象转录组中 1-6 碱基完整型 SSR 进行分析, 部分 SSR 信息见表 1。

2.2 沟眶象转录组重复基元的分布

在沟眶象所有碱基重复类型中, 三碱基重复 SSR 含量最多, 约占总数的 37.27%, 其次为单碱基重复 SSR (33.70%)、二碱基重复 SSR (26.04%), 没有发现六碱基重复 SSR。研究还发现在沟眶象 104 种重复基元中, (A/T)_n、(AT/AT)_n、(TTA/TAA)_n、(ATAA/TTAT)_n、(AGGTT/AACCT)_n 和 (AAGTA/TACTT)_n 分别在单、二、三、四、五碱基中出现频率最多, 它们在各自重复基元类型中的比例分别是 96.18%、28.31%、14.76%、23.26%、50.00% (表 2)。

在沟眶象的 104 种碱基重复基元中, 单、二、三、四、五碱基重复基元分别有 4、12、58、23、7 种, 出现频率最高的是 (A/T)_n (30.39%), 其次是 (AT/AT)_n (6.91%)、(TA/TA)_n (6.80%)、(TTA/TAA)_n (5.16%)、(AAT/ATT)_n (4.44%) 等。不同类型重复基元的 SSR 分布见图 1。

表 1 沟眶象转录组中微卫星数据库的部分结果

Table 1 SSR database of *Eucryptorrhynchus chinensis* transcriptome

Unigene 编号 Unigene ID	SSR 类型 SSR type	SSR 序列 SSR sequence	SSR 长度 SSR size	SSR 在 Unigene 上的开始位置 Start	SSR 在 Unigene 上的结束位置 End
CL1Contig11	Trinucleotide	(CTT) ₅	15	1069	1083
CL75Contig1	Mononucleotide	(T) ₁₃	13	817	829
CL872Contig1	Mononucleotide	(A) ₂₂	22	1134	1155
CL5367Contig1	Trinucleotide	(AAT) ₆	18	349	366
CL7045Contig1	Tetranucleotide	(ATAA) ₅	20	812	831
CL7130Contig1	Pentanucleotide	(AACCA) ₅	25	3730	3754
CL18646Contig1	Dinucleotide	(TC) ₉	18	108	125
CL1293Contig1	Trinucleotide	(GAG) ₁₀	30	2968	2997

表 2 沟眶象转录组中不同微卫星优势重复基元出现的频率

Table 2 Occurrence frequency of different advantage repeat motifs of *Eucryptorrhynchus chinensis* transcriptome

重复类型 Repeat	数量 Number	频率 (%) Frequency	最多重复基元 Maximum repeat motif	数量 Number	比例 (%) Proportion
单碱基重复 Mononucleotide	576	33.70	A/T	554	96.18
二碱基重复 Dinucleotide	445	26.04	AT/AT	126	28.31
三碱基重复 Trinucleotide	637	37.27	TTA/TAA	94	14.76
四碱基重复 Tetranucleotide	43	2.52	ATAA/TTAT	10	23.26
五碱基重复 Pentanucleotide	8	0.47	AGGTT/AACCT	2	25.00
			AAGTA/TACTT	2	25.00

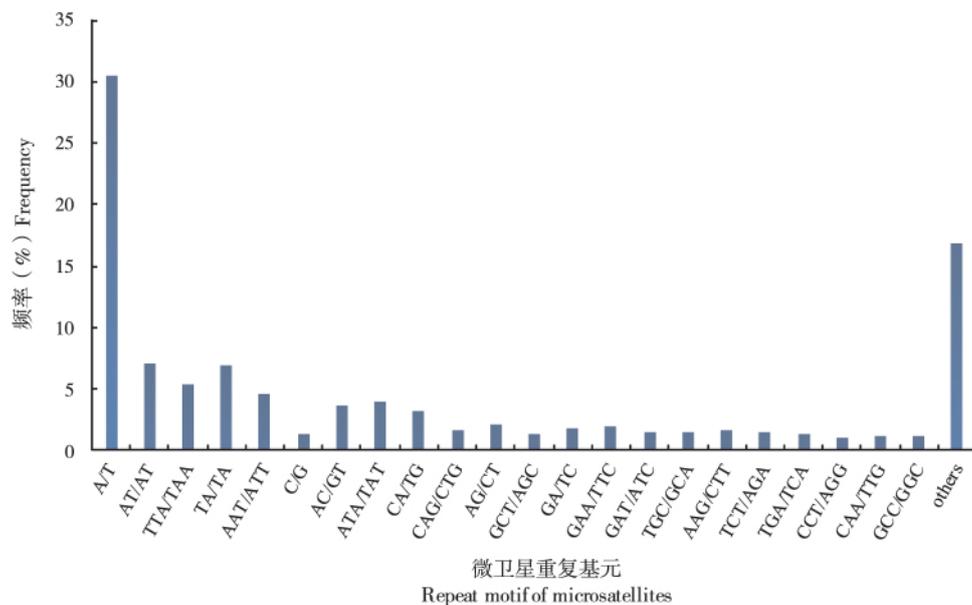


图 1 基于重复基元类型的微卫星分布 (考虑到碱基互补)

Fig. 1 Microsatellites distribution on different repeat motifs (considering sequence complementary)

注: Other motifs 表示频率小于 0.80% 的重复基元类型。Note: Other motifs denote the repeat motifs with frequency below 0.80%.

2.3 沟眶象转录组序列中微卫星的长度分布

沟眶象转录组中所发现的 1709 个完整型 SSR 中, 微卫星长度存在显著变异, 从 12 - 48 个碱基不等, 平均长度为 15.74 个碱基。如图 2 显示, 沟眶象微卫星以重复长度小于 20 bp 的短重复序列最多, 其中 12 - 15 bp 的数量最多 (1098, 64.25%), 其次为 16 - 20 bp (394, 23.05%), 长度大于 20 bp 的长序列重复仅占微卫星总数的 12.70%。本研究进一步利用 SPSS 软件进行 Person 相关性分析, 结果表明沟眶象微卫星出现的频率和其长度呈极显著负相关 ($P < 0.01$), 相关系数为 -0.572。

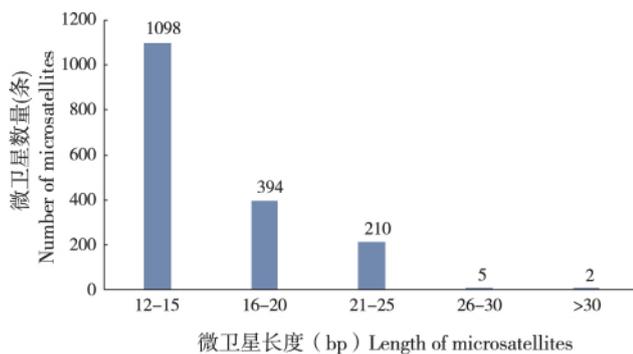


图 2 沟眶象转录组中微卫星的长度分布

Fig. 2 Length distribution of microsatellites in *Eucryptorrhynchus chinensis* transcriptome

3 结论与讨论

本研究从沟眶象转录组中发掘到 1823 个 SSR 位点, 分布于 1665 条 Unigene 序列中。沟眶象 SSR 的出现频率为 10.36%, 这与黑翅土白蚁 *Odontotermes formosanus* (9.98%) (Huang *et al.*, 2012) 中 SSR 的出现频率差不多, 高于烟粉虱 *Bemisia tabaci* (5.07%) (Xie *et al.*, 2012)、细梢小卷蛾 *Rhyacionia leptotubula* (3.09%) (Zhu *et al.*, 2013)、粘虫 *Mythimna separate* (1.93%) (胡艳华等, 2015) 中 SSR 的出现频率。鞘翅目昆虫中, 云南切梢小蠹 *Tomicus yunnanensis* (袁远等, 2014) 和黄粉虫 *Tenebrio molitor* (Zhu *et al.*, 2013) 转录组中检测出的 SSR 的出现频率分别为 1.29%、1.79%, 明显低于沟眶象转录组中 SSR 的出现频率, 这说明沟眶象转录组中 SSR 数量相对来说十分丰富。分析其中的原因, 除了 SSR 位点的搜索方法或计算标准有差异外, 最根本的原

因可能是物种特异性 (黄海燕等, 2013)。

在鉴定得到的沟眶象转录组 SSR 中, 三碱基重复类型的数量最高, 约占总数的 33.70%。这与绿豆象 *Callosobruchus chinensis* (Duan *et al.*, 2014)、二点委夜蛾 *Athetis lepigone* (Li *et al.*, 2013)、白背飞虱 *Sogatella furcifera* (Xu *et al.*, 2012)、云南切梢小蠹 *Tomicus yunnanensis* (袁远等, 2014) 等昆虫的 SSR 以三碱基重复为主类似。大部分昆虫中的微卫星都是以三碱基重复为主导重复类型, 这有可能是为了防止蛋白质编码时发生移码突变现象 (Sun *et al.*, 2011)。本研究中沟眶象微卫星重复基元中所占比例最高的为 (A/T)_n, 达到 30.39%, 这与黄粉虫、云南切梢小蠹、扶桑绵粉蚧 *Phenacoccus solenopsis* (罗梅等, 2014) 的转录组中 SSR 的分析结果一致。沟眶象三碱基重复中, TTA/TAA 为优势重复基元, 而鞘翅目昆虫中的黄粉甲和云南切梢小蠹三碱基重复中都是以 AAT/ATT 为优势重复基元。这种差异可能与昆虫自身 SSR 特点以及转录组数据来源等紧密相关。

SSR 分子标记的多态性是判断其可用性的重要依据, 而其长度又是影响其多态性高低的重要因素。Temnykh 等 (2001) 的研究表明, 当 SSR 长度 ≥ 20 bp 时多态性较高, 长度在 12 - 20 bp 时多态性中等, 而长度在 12 bp 以下时 SSR 多态性极低。本研究在 SSR 筛选过程中已经将长度在 12 bp 以下的 SSR 过滤掉, 最终发现沟眶象转录组中 SSR 长度在 12 - 48 个碱基不等, 平均长度为 15.74 个碱基, 具有高多态性的 SSR 所占比例仅为 12.70%, 这部分多态性潜能高的 SSR 在高效开发沟眶象微卫星标记方面应该具有较高的利用价值 (杨华等, 2011)。本研究还发现沟眶象微卫星出现的频率和其长度呈极显著负相关 ($P < 0.01$), 相关系数为 -0.572。根据 Samadi 等 (1998) 的研究可以得出, 基因组中长度较短的 SSR 变异速率较快, 而长度较长的 SSR 变异速率则较慢, 相对来说较稳定。虽然基因序列中的 SSR 增加了基因的不稳定性, 但正是这样的不稳定性为基因变异提供了动力。

本研究基于构建获得的沟眶象转录组数据库, 对其 SSR 进行了高通量发掘, 并对沟眶象 SSR 的特征进行了分析, 表明利用转录组数据发掘 SSR 是高效可行的。研究结果为开发高多态性微卫星引物来进行沟眶象功能基因组学、种群遗传结构、

种群遗传多样性等研究奠定了基础, 这对于从分子角度探究沟眶象的扩散机制及综合防治等都具有非常重要的意义, 同时为探索沟眶象和其近缘种臭椿沟眶象的起源进化关系打下基础。

参考文献 (References)

- Duan CX, Li DD, Sun SL, *et al.* Rapid development of microsatellite markers for *Callosobruchus chinensis* using Illumina paired-end sequencing [J]. *PLoS ONE*, 2014, 9 (5): e95458.
- Eujayl I, Sorrells M, Banm M, *et al.* Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat [J]. *Theoretical and Applied Genetics*, 2002, 104 (2): 399-407.
- He P. Abundance, polymorphism and applications of microsatellite in eukaryote [J]. *Hereditas*, 1998, 20 (4): 42-47. [何平. 真核生物中的微卫星及其应用[J]. 遗传, 1998, 20 (4): 42-47]
- Huang HY, Du HY, Wuyun TN, *et al.* Development of SSR molecular markers based on transcriptome sequencing of *Eucommia ulmoides* [J]. *Scientia Silvae Sinicae*, 2013, 49 (5): 176-181. [黄海燕, 杜红岩, 乌云塔娜, 等. 基于杜仲转录组序列的 SSR 分子标记的开发[J]. 林业科学, 2013, 49 (5): 176-181]
- Huang Q, Sun P, Zhou X, *et al.* Characterization of head transcriptome and analysis of gene expression involved in caste differentiation and aggression in *Odontotermes formosanus* (Shiraki) [J]. *PLoS ONE*, 2012, 7 (11): e50383.
- Hu YH, Li M, Zhang HF, *et al.* The information analysis of SSR loci in the *Mythimna separate* (Walker) transcriptome [J]. *Journal of Shanxi Agricultural University*, 2015, 35 (5): 484-489. [胡艳华, 李敏, 张虎芳, 等. 粘虫转录组中 SSR 位点的信息分析[J]. 山西农业大学学报, 2015, 35 (5): 484-489]
- Li LT, Zhu YB, Ma JF, *et al.* An analysis of the *Athetis lepigone* transcriptome from four developmental stages [J]. *PLoS ONE*, 2013, 8 (9): e73911.
- Li YC, Korol AB, Fahima T, *et al.* Microsatellites: Genomic distribution, putative functions and mutational mechanisms: A review [J]. *Molecular Ecology*, 2002, 11 (12): 2453-2465.
- Liu T, Zhu S, Fu L, *et al.* Development and characterization of 1827 expressed sequence tag-derived simple sequence repeat markers for ramie (*Boehmeria nivea* L. Gaud.) [J]. *PLoS ONE*, 2013, 8 (4): e60346.
- Luo M, Zhang H, Bin SY, *et al.* High-throughput discovery of SSR genetic markers in the mealybug, *Phenacoccus solenopsis* (Hemiptera: Pseudococcidae), from its transcriptome database [J]. *Acta Entomologica Sinica*, 2014, 57 (4): 395-400. [罗梅, 张鹤, 宾淑英, 等. 基于转录组数据高通量发掘扶桑绵粉蚧微卫星引物[J]. 昆虫学报, 2014, 57 (4): 395-400]
- Samadi S, Artiguelbielle E, Estoup A, *et al.* Density and variability of dinucleotide microsatellites in the parthenogenetic polyploid snail *Melanoides tuberculata* [J]. *Molecular Ecology*, 1998, 7 (9): 1233-1236.
- Shi J, Yin TM, Guan HW, *et al.* Characteristic analysis of microsatellites *Camellia* spp. [J]. *Journal of Nanjing Forestry University (Natural Sciences Edition)*, 2012, 36: 47-52. [史洁, 尹佟明, 管宏伟, 等. 油茶基因组微卫星特征分析[J]. 南京林业大学学报 (自然科学版), 2012, 36: 47-52]
- Temnykh S, DeClerck G, Lukashova A, *et al.* Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.) [J]. *Genome Research*, 2001, 11: 1441-1452.
- Varshney RK, Graner A, Sorrells ME. Genic microsatellite markers in plants: Features and applications [J]. *Trends in Biotechnology*, 2005, 23 (1): 48-55.
- Wang D, Cao LY, Gao JP. Data mining of simple sequence repeats in *Codonopsis pilosula* transcriptome [J]. *Chinese Traditional and Herbal Drugs*, 2014, 46 (8): 2390-2394. [王东, 曹玲亚, 高建平. 党参转录组中 SSR 位点信息分析[J]. 中草药, 2014, 46 (8): 2390-2394]
- Wang LY, Wei K, Zhang CC, *et al.* Characterization of microsatellites in tea (*Camellia sinensis*) floral transcriptome [J]. *Acta Agronomica Sinica*, 2014, 40 (1): 80-85. [王丽鸾, 韦康, 张成才, 等. 茶树花转录组微卫星分布特征[J]. 作物学报, 2014, 40 (1): 80-85]
- Xie W, Meng QS, Wu QJ, *et al.* Pyrosequencing the *Bemisia tabaci* transcriptome reveals a highly diverse bacterial community and a robust system for insecticide resistance [J]. *PLoS ONE*, 2012, 7 (4): e35181.
- Xu Y, Zhou W, Zhou Y, *et al.* Transcriptome and comparative gene expression analysis of *Sogatella furcifera* (Horvath) in response to southern rice black-streaked dwarf virus [J]. *PLoS ONE*, 2012, 7 (4): e36238.
- Yang GJ, Yong HL, Wang XP. The biological characters and behavior of *Eucryptorrhynchus chinensis* [J]. *Chinese Bulletin of Entomology*, 2008, 45 (1): 65-69. [杨贵军, 雍惠莉, 王新谱. 沟眶象的生物学特性及行为观察[J]. 昆虫知识, 2008, 45 (1): 65-69]
- Yang H, Chen Q, Wei ZL, *et al.* Analysis on SSR information in *Camellia sinensis* transcriptome [J]. *Journal of Anhui Agricultural University*, 2011, 38: 882-886. [杨华, 陈琪, 韦朝领, 等. 茶树转录组中 SSR 位点的信息分析[J]. 安徽农业大学学报, 2011, 38: 882-886]
- Yuan Y, Zhang LF, Wu GX, *et al.* High-throughput discovery microsatellites in *Tomicus yunnanensis* (Coleoptera: Scolytinae) [J]. *Journal of Environmental Entomology*, 2014, 36 (2): 166-170. [袁远, 张丽芳, 吴国星, 等. 云南切梢小蠹微卫星的高通量发掘[J]. 环境昆虫学报, 2014, 36 (2): 166-170]
- Zhu JY, Li YH, Yang S, *et al.* De novo assembly and characterization of the global transcriptome for *Rhyacionia leptotubula* using Illumina paired-end sequencing [J]. *PLoS ONE*, 2013, 8 (11): e81096.
- Zhu JY, Wu GX, Yang B. High-throughput discovery of SSR genetic markers in the yellow mealworm beetle, *Tenebrio molitor* (Coleoptera: Tenebrionidae), from its transcriptome database [J]. *Acta Entomologica Sinica*, 2013, 56 (7): 724-728.