昆 虫 学 报 *Acta Entomologica Sinica*, July 2013, 56(7): 724 – 728 ISSN 0454-6296

High-throughput discovery of SSR genetic markers in the yellow mealworm beetle, *Tenebrio molitor* (Coleoptera: Tenebrionidae), from its transcriptome database

ZHU Jia-Ying[1,*], WU Guo-Xing[2], YANG Bin[1]

(1. Key Laboratory of Forest Disaster Warning and Control of Yunnan Province, Southwest Forestry University, Kunming 650224, China; 2. College of Plant Protection, Yunnan Agricultural University, Kunming 650201, China)

Abstract: *Tenebrio molitor* is a well known model insect. Although a lot of achievements have been made in many research aspects related to this insect, only very few molecular/genetic resources are available. In this study, a high-throughput method was used for discovering the simple sequence repeat (SSR) genetic markers from this beetle. In total, 1 249 SSR genetic markers were developed from the previously constructed trancriptome database. The majority of them contained mono- and trinucleotide motifs (44.44% and 41.15%, respectively), and A/T (42.70%) was the most abundant motif. Except for mononucleotide, the SSRs with five repeat units were the most common, with the frequency of 30.90%. Base on the identified SSRs, 1 004 pairs of primers were designed, of which a maximum of 5 pairs of alternative primers were designed from a single SSR. The SSRs identified here will constitute an important resource for marker-assisted investigation in functional and comparative genomics of *T. molitor*.

Key words: *Tenebrio molitor*; transcriptome; microsatellite; genetic marker; primer

1 INTRODUCTION

Simple sequence repeats (SSRs) or microsatellites, ubiquitously distributed throughout eukaryotic and prokaryotic genomes, are tandem repeat sequences of 1 – 6 base pairs of DNA (Goldstein and Schlötterer, 1999). With the advantages of harboring high levels of polymorphism, being stable, PCR-based and relatively low-cost, SSRs have been developed and become one of the most popular genetic markers widely used in many areas of molecular biology such as genome characterization, genome mapping, comparative genomics, phylogenetic studies and population genetics (Li *et al.*, 2002, 2004). Traditionally, SSR marker development has typically involved in constructing and screening recombinant libraries, which is generally laborious, time consuming and expensive (Zane *et al.*, 2002; Tang *et al.*, 2008). With the advent of next generation sequencing technologies (a powerful alternative for generating a tremendous number of DNA sequences), high-throughput transcriptome sequencing dramatically expedites the traditional methods of developing SSR markers (Wei *et al.*, 2011; Churbanov *et al.*, 2012; Gao *et al.*, 2012).

The yellow mealworm beetle, *Tenebrio molitor* (Coleoptera: Tenebrionidae), is considered scavengers that infest stored products including cereals, baking flour, and livestock feed. It is thought to have originated in Europe and now widespread over the world. It is one of excellently ideal model organisms among the higher eukaryotes for studies in biology, biochemistry, evolution, immunology and physiology because of the reasons such as its relatively large size, ease of rearing and handling, and genetic diversity (Pursall and Rolff, 2011; Dobson *et al.*, 2012). Since Arenssen Hein started experiments with *T. molitor* on a large scale in 1915, a large amount of information related to this insect has been gathered in numerous research fields (Arendsen Hein, 1920; Andersen *et al.*, 1997; Pölkki *et al.*, 2012). Despite the focused investigation of *T. molitor* has a long history, only very few molecular genetic/genomic resources are available. The transcript dada set of this insect has been generated using Illumina HiSeq™ 2000 platform and a large amount of assembled unigenes are available (Zhu *et al.*, 2012). Based on this database, a mega-identification of SSRs in *T. molitor* was reported in this study.

基金项目：云南省应用基础研究面上项目(2010CD063)；国家自然科学基金项目(31260449)；教育部科学技术研究重点项目(211171)
作者简介：朱家颖，男，1984 年 10 月生，云南宣威人，博士，副教授，研究方向为昆虫生理生化与分子生物学，E-mail: jyzhu001@gmail.com
* 通讯作者 Corresponding author, E-mail: jyzhu001@gmail.com
收稿日期 Received: 2013-03-10；接受日期 Accepted: 2013-06-13

2 MATERIALS AND METHODS

2.1 Transcriptome database

The transcriptome of *T. molitor* pupae was sequenced using Illumina paired-end sequencing (Zhu *et al.*, 2012). Raw data were deposited in DDBJ database under accession no. DRA000603. The unigenes assembled using SOAPdenovo software were used for SSR identification.

2.2 SSR identification

A perl script, MIcroSAtelitte identification tool (MISA, http://pgrc.ipk-gatersleben.de/misa/) was used to identify SSRs in all these unigene sequences. The parameters were designed for identification of mono-, di-, tri-, tetra-, penta-, and hexa-nucleotide SSR motifs with the minimum of repeat numbers of 12, 6, 5, 5, 4 and 4, respectively. The maximum length of interruption between two adjacent SSR repeat units was set to 100 bp.

2.3 SSR primer design and validation

With respect to the unigenes containing SSRs, only those with sequences not less than 150 bp in both back and forth of the repeat units were screened out for primer design. The primers were designed using the Primer 3 v2.2.2 software with length of 18 – 23 bp (optimum length: 23 bp), annealing temperature of 55 – 65℃ (optimum annealing temperature: 60℃; maximum difference in temperature of forward and reverse primer: 2℃), and the product size ranging from 80 to 300 bp. Then, the following selection criteria were applied: primers can not contain an SSR (2 – 6 bases repeat more than four times), and primers have only one matched unigene when mapped against all unigenes (allow three-base mismatch at the 5′ end and one-base mismatch at the 3′ end). To validate the success of the SSR primer design, 12 primers were randomly selected and their products were amplified by PCR using *T. molitor* genomic DNA as the template. The PCR reaction profile was: 94℃ for 2 min, followed by 30 cycles of 94℃ for 20 s, 55℃ for 30 s, 72℃ for 1 min, and a final extension step at 72℃ for 10 min. The PCR amplification products were analyzed by electrophoresis in 4% agarose gels.

3 RESULTS

A total of 71 514 unigenes composed of 30 319 407 nucleotides were examined for identifying SSRs. Among these sequences, a subset of 1 190 unigenes containing 1 249 SSRs were found, suggesting that merely 1.67% of sequences contained SSRs. These motifs included mono-, di-, tri-, tetra-, penta- and hexa-nucleotides with the lengths ranging from 1 to 6 bp. Fifty five unigenes contained more than one SSR. The majority of SSRs was in perfect formation. There were only 30 SSRs present in compound formation, accounting for 2.4% of the total. The numbers of SSRs with different numbers of tandem repeats were calculated. Of the 1 249 SSR motifs identified, mono-, di-, tri-, tetra-, penta- and hexanucleotide repeats were 555, 140, 514, 21, 12 and 7 in total, respectively. Based on the distribution of the SSRs, mononucleotide repeats were found to be the most abundant (44.44%), followed by trinucleotide (41.15%), dinucleotide (11.21%), tetranucleotide (1.68%), pentanucleotide (0.96%) and hexanucleotide repeats (0.56%) (Table 1). Within mononucleotide, the major SSRs were with 12 and > 15 repeat units. Except for mononucleotide, SSRs with five repeat units (30.90%) were the most common, followed by six (15.21%), seven (5.12%), eight (1.84%), and four (1.04%) repeat units. Among all nucleotide repeats, A/T (42.70%) was the most abundant motif (Table 2). Regarding to di- and tri- and tetranucleotide repeats, only CG/CG, ACT/AGT, AGG/CCT, and AAAT/ATTT displayed relatively low abundance. The other types were nearly equal. Using the Primer 3 v2.2.2 software, a total of 1 480 primer pairs were designed. After filtration, 1 004 primer pairs were successfully obtained, of which a maximum of 5 pairs of alternative primers were designed from a single SSR. Twelve primer pairs were randomly selected for validation of the predicted SSRs. Of them, all successfully amplifed PCR products (Table 3).

4 DISCUSSION

As the powerful applications of SSRs, using microsatellite markers to understand the evolution process and diversification of insects may help us protect useful insects and control pests (Wang *et al.*, 2009). Large numbers of SSRs have been isolated and characterized in different insect orders. Abundance of SSRs, with which they occur differ greatly between taxa, was known to be dependent on the SSR search criteria, the size of the dataset, and the database-mining tools (Varshney *et al.*, 2005). In this study, 1.67% sequences of the transcriptome contained the SSRs . The percent number is similar to the SSR contents in the genomes of *Bombyx mori*, *Drosophila melanogaster*, *Anopheles gambiae*, *Apis*

Table 1 Frequency of SSRs based on the number of repeat units in *Tenebrio molitor* transcriptome

SSR type	Number of repeats													Total	%
	4	5	6	7	8	9	10	11	12	13	14	15	>15		
Mononucleotide	–	–	–	–	–	–	–	–	183	97	67	38	170	555	44.44
Dinucleotide	–	–	78	38	14	3	4	–	1	2	–	–	–	140	11.21
Trinucleotide	–	364	107	26	9	5	2	1	–	–	–	–	–	514	41.15
Tetranucleotide	–	16	5	–	–	–	–	–	–	–	–	–	–	21	1.68
Pentanucleotide	8	4	–	–	–	–	–	–	–	–	–	–	–	12	0.96
Hexanucleotide	5	2	–	–	–	–	–	–	–	–	–	–	–	7	0.56
Total	13	386	190	64	23	8	6	1	184	99	67	38	170		
%	1.04	30.90	15.21	5.12	1.84	0.64	0.48	0.08	14.73	7.93	5.36	3.04	13.61		

Table 2 Frequency distribution of SSRs based on motif types in *Tenebrio molitor* transcriptome

SSR motif	Number of repeats													Total	%
	4	5	6	7	8	9	10	11	12	13	14	15	>15		
A/T	–	–	–	–	–	–	–	–	178	90	63	37	161	529	42.70
C/G	–	–	–	–	–	–	–	–	5	7	4	1	9	26	2.10
AC/GT	–	–	20	14	11	1	–	–	1	2	–	–	–	49	3.95
AG/CT	–	–	28	14	3	1	4	–	–	–	–	–	–	50	4.04
AT/AT	–	–	28	10	–	1	–	–	–	–	–	–	–	39	3.15
CG/CG	–	–	2	–	–	–	–	–	–	–	–	–	–	2	0.16
AAC/GTT	–	48	16	2	1	2	–	1	–	–	–	–	–	70	5.65
AAG/CTT	–	52	15	3	–	–	1	–	–	–	–	–	–	71	5.73
AAT/ATT	–	60	31	8	5	1	–	–	–	–	–	–	–	105	8.47
ACC/GGT	–	23	7	4	–	–	–	–	–	–	–	–	–	34	2.74
ACG/CGT	–	56	8	3	–	–	–	–	–	–	–	–	–	67	5.41
ACT/AGT	–	4	4	–	–	1	–	–	–	–	–	–	–	9	0.73
AGC/CTG	–	30	4	1	1	–	–	–	–	–	–	–	–	36	2.91
AGG/CCT	–	13	4	–	1	1	–	–	–	–	–	–	–	19	1.53
ATC/ATG	–	29	8	2	–	–	1	–	–	–	–	–	–	40	3.23
CCG/CGG	–	49	10	3	1	–	–	–	–	–	–	–	–	63	5.08
AAAT/ATTT	–	9	2	–	–	–	–	–	–	–	–	–	–	11	0.89
Others	13	6	–	–	–	–	–	–	–	–	–	–	–	19	1.53

The sequence complementary was considered.

mellifera, and *Tribolium castaneum*, which are 0.72%, 1.56%, 1.58%, 3.41%, and 0.41%, respectively (Archak *et al*., 2007). With regard to the SSR formation of *T. molitor*, it is similar to the above five insect genomes that account for nearly 3.2% compound SSRs (Archak *et al*., 2007). In general, trinucleotide repeats are the most abundant microsatellites in coding ESTs besides mononucleotide repeats (Xu *et al*., 2012). Trinucleotide repeats were predominant in *T. molitor* transcriptome sequences. On the basis of the detected SSR marker sequences, the number of large designed SSR primer pairs indicated that transcriptome data is a fast and cost-effective way for high-throughput discovery of SSR markers in non-model organisms that lack a reference genome. The randomly selected primer pairs all successfully amplified PCR products, which highly validated the utility of predicted SSRs. The SSR data constructed here will facilitate the development of suitable SSR genetic markers for investigating the functional and comparative genomics of *T. molitor*.

Table 3 The characterization of selected SSR primers

Number	Unigene ID	Primer sequences	Product size (bp)	Motif and repeat number	PCR amplification
1	Unigene10146	F: CCTTCGTCTTCGCCTTGAG	145	(GGC)5	S
		R: AAGGAACTGTCCCAGAAGCTC			
2	Unigene11122	F: ACGAAACAAAAACCACAG	144	(TTA)7	S
		R: TTTTAACATGAATGTTGGCTTCA			
3	Unigene1309	F: GATCACGATCCTGGTCTTCC	142	(GGC)5	S
		R: TCTGATTATTAGACAACGCACGA			
4	Unigene13708	F: AACGAGATGGTGCTGGAATC	135	(GCG)5	S
		R: CATCCCTAGCAGACTCGCA			
5	Unigene14044	F: GTACCTGTATTGCTGAAACCGTC	154	(AGA)10	S
		R: CGTAACATGGAGTGTCGTCGTAT			
6	Unigene17249	F: ACTCGTATGTCCCTTTGTCCTTT	124	(TCG)5	S
		R: GACAACGAAATCACTCTGGAGAT			
7	Unigene17273	F: CTTCTCTGTCCATTTAGCACCAT	148	(AC)8	S
		R: GAAGTCCGTGTTGTGTGTTGTAA			
8	Unigene1769	F: AGCTGGTGAACGACAGCAAC	155	(GAC)5	S
		R: GTTTTGGTTGAAGTACGTCGTG			
9	Unigene63254	F: CTTCTTCTCTTTCTCCGGTTTTT	100	(TTC)6	S
		R: GACGAACTCGATGAGCAACCT			
10	Unigene5676	F: ACAATCTCAACAAGAGGAAGTGC	121	(TCG)5	S
		R: GTGGAATTCTCCGAAGGTTCTT			
11	Unigene64074	F: CACGCCGAAGTACCTACAAATAA	144	(ATTT)5	S
		R: TGGGAATAAACCATAAAAACCAA			
12	Unigene8285	F: GTACCTATGGCGTTCTTTACACG	127	(AT)6	S
		R: TACACTCTTAGGGTCTGGGCATA			

S: Successful amplification.

ACKNOWLEDGEMENTS We are grateful to YANG Yu-Zhi at Southwest Forestry University, China for conducting the SSR primers validation.

References

Andersen SO, Rafn K, Roepstorff P, 1997. Sequence studies of proteins from larval and pupal cuticle of the yellow meal worm, *Tenebrio molitor*. *Insect Biochem. Mol. Biol.*, 27(2): 121 – 131.

Archak S, Meduri E, Kumar PS, Nagaraju J, 2007. InSatDb: a microsatellite database of fully sequenced insect genomes. *Nucl. Acids Res.*, 35: D36 – 39.

Arendsen I, 1920. Technical experiences in the breeding of *Tenebrio molitor*. *Proc. Kon. Ned. Akad. Wet. Amst.*, 23: 193.

Churbanov A, Ryan R, Hasan N, Bailey D, Chen H, Milligan B, Houde P, 2012. HighSSR: high-throughput SSR characterization and locus development from next-gen sequencing data. *Bioinformatics*, 28(21): 2797 – 2803.

Dobson AJ, Johnston PR, Vilcinskas A, Rolff J, 2012. Identification of immunological expressed sequence tags in the mealworm beetle *Tenebrio molitor*. *J. Insect Physiol.*, 58(12): 1556 – 1561.

Gao X, Han J, Lu Z, Li Y, He C, 2012. Characterization of the spotted seal *Phoca largha* transcriptome using Illumina paired-end sequencing and development of SSR markers. *Comp. Biochem. Physiol.*, 7D(3): 277 – 284.

Goldstein DB, Schlötterer C, 1999. Microsatellites: Evolution and Applications. Oxford University Press, Oxford.

Li YC, Korol AB, Fahima T, Beiles A, Nevo E, 2002. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol. Ecol.*, 11(12): 2453 – 2465.

Li YC, Korol AB, Fahima T, Nevo E, 2004. Microsatellites within genes: structure, function, and evolution. *Mol. Biol. Evol.*, 21(6): 991 – 1007.

Pölkki M, Krams I, Kangassalo K, Rantala MJ, 2012. Inbreeding affects sexual signalling in males but not females of *Tenebrio molitor*. *Biol. Lett.*, 8(3): 423 – 425.

Pursall ER, Rolff J, 2011. Immune responses accelerate ageing: proof-of-principle in an insect model. *PLoS ONE*, 6(5): e19972.

Tang J, Baldwin SJ, Jacobs JM, Linden CG, Voorrips RE, Leunissen JA, van Eck H, Vosman B, 2008. Large-scale identification of polymorphic microsatellites using an *in silico* approach. *BMC Bioinform.*, 9: 374.

Varshney RK, Graner A, Sorrells ME, 2005. Genetic microsatellite markers in plants: features and applications. *Trends Biotechnol.*, 23(1): 48 – 55.

Wang ML, Barkley NA, Jenkins TM, 2009. Microsatellite markers in plants and insects. Part I: applications of biotechnology. *Genes, Genomes and Genomics*, 3(S1): 54 – 67.

Wei W, Qi X, Wang L, Zhang Y, Hua W, Li D, Lv H, Zhang X, 2011. Characterization of the sesame (*Sesamum indicum* L.) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers. *BMC Genomics*, 12: 451.

Xu Y, Zhou W, Zhou Y, Wu J, Zhou X, 2012. Transcriptome and comparative gene expression analysis of *Sogatella furcifera* (Horváth) in response to southern rice black-streaked dwarf virus. *PLoS ONE*, 7(4): e36238.
Zane L, Bargelloni L, Patarnello T, 2002. Strategies for microsatellite isolation: a review. *Mol. Ecol.*, 11(1): 1-16.
Zhu JY, Yang P, Zhang Z, Wu GX, Yang B, 2012. Transcriptomic immune response of *Tenebrio molitor* pupae to parasitization by *Scleroderma guani*. *PLoS ONE*, 8(1): e54411.

基于转录组数据高通量发掘黄粉甲微卫星引物

朱家颖[1,*], 吴国星[2], 杨　斌[1]
(1. 西南林业大学云南省森林灾害预警与控制重点实验室, 昆明 650224; 2. 云南农业大学植物保护学院, 昆明 650201)

摘要: 黄粉虫 *Tenebrio molitor* 作为理想的模式研究生物, 虽然已围绕该昆虫在多个研究领域开展了诸多研究, 但是有关其分子和遗传方面的研究仍知之甚少。为此, 本研究基于前期构建的黄粉甲转录组数据库, 成功发掘获得 1 249 个微卫星序列。其中, 单碱基或三碱基序重复列最多, 分别占 44.44% 和 41.15%; A/T 型重复序列出现频率最高, 占 42.70%。除单核苷酸重复序列外, 重复单元的重复次数以 5 次最多, 占 30.90%。基于鉴定获得的微卫星序列, 共设计获得 1 004 对微卫星引物, 而且每对引物还设计了 5 对替代引物。研究获得的微卫星引物将有助于今后开展黄粉甲功能和比较基因组学方面的研究。

关键词: 黄粉甲; 转录组; 微卫星; 遗传标记; 引物
中图分类号: Q966　　**文献标志码**: A　　**文章编号**: 0454-6296(2013)07-0724-05

(责任编辑: 赵利辉)