The Distribution Characteristics of Microsatellites in the Transcriptome of *Periplaneta americana*

Lifang Zhang[a, *, **] and Shangeng He[a]

[a] *College of Food and Biological Engineering/Institute of Food Science and Engineering Technology, Hezhou University, Hezhou, 542899 China*

*e-mail: zlfdlnu@163.com

**e-mail: 1718944622@qq.com

Abstract—To obtain information of *Periplaneta americana*, we analyzed the distribution characteristics of microsatellite sequences in the *P. americana* transcriptome (229 MB) by using MSDBv2.4. The total number of perfect microsatellite sequences was 38 082 and covered about 0.3% of *P. americana* transcriptome. The cumulative length of microsatellites was 618 138 bp, and the density of microsatellites was 2978.54 bp/Mb. In the different repeat types of the microsatellites, the number of the mononucleotide repeats was 20 002 (accounting for 52.52%), which obviously was the most abundant type. While the trinucleotide, tetranucleotide, dinucleotide, pentanucleotide and hexanucleotide repeats accounted for 24.51, 12.97, 8.13, 1.61 and 0.26%, respectively. The kind of different repeat copy categories in each repeat type was also quite different, such as the A in mononucleotide repeat type, the AG in dinucleotide, the AAT in trinucleotide, AAAT in tetranucleotide, the AAGAA in pentanucleotide, and the CAGTAG in hexanucleotide were the most of each category. The A, T, AC, AG, AT, GT, AAG, AAT, ATC, ATG, ATT, CTT, AAAG and AAAT were the dominant repeat copy categories, the total number of all these types was 29 933, accounting for 78.6% in the total number of microsatellite sequences. These results based on a foundation for developing high polymorphic microsatellites to research the functional genomics, population genetic structure and genetic diversity of *P. americana*.

INTRODUCTION

Cockroaches (Blattodea) are the well-known primitive winged insects, with an extremely high diversity of about 4000 species worldwide (Chung et al., 2005). American cockroach is a synanthropic pest that generally inhabits from cosmopolitan to urban areas (Jaramillo-Ramirez et al., 2010). The American cockroach is normally named *Periplaneta americana* which belongs to Blattoidea Blattidae Periplaneta. *P. americana* is an important medicinal insect. Its series of products, such as "Kang Fu Xin" and "*P. americana* edible dried worm power", developed with *P. americana* as raw material, had the functions of promoting blood circulation, repairing wounds, regulating immune function, etc. These products had been put into clinical usage and had a good clinical effect. However, the reports on *P. americana* were mainly focused on biology, ecology, chemical composition analysis and medicinal value. There were few studies on the genetic background and molecular biology of *P. americana*.

Transcriptome is the complete set and quantity of transcripts in a cell at a specific developmental stage or under a physiological condition. The transcriptome provides information on gene expression, gene regulation, and amino acid content of proteins. Therefore, transcriptome analysis is essential to explain the functional elements of the genome and reveal the molecular constituents of cells and tissues (Gao et al., 2012). Transcriptome or expressed sequence tag (EST) sequencing is an efficient way to generate functional genomic-level data for non-model organisms (Emrich et al., 2006; Gavery and Roberts, 2012; Li et al., 2011; Sun et al., 2011). Large collections of EST sequences are valuable for comparative genomics, development of molecular markers (Li et al., 2010; Li et al., 2011; Gao et al., 2007; Shen et al., 2012), and population genomics studies (Santure et al., 2011).

Microsatellites were invented by Hamada et al. in 1982. This invention was verified by Tautz and Renz in 1984, who hybridized different microsatellite sequences to genomic DNA from a variety of organisms and found that many types of simple sequences were present. Microsatellites were a special class of repetitive DNA sequences that had many merits (Tautz and Renz, 1984; Schlötterer, 2004; Sharma

et al., 2007). Microsatellites are defined as highly variable DNA sequences composed of tandem repeats of 1−6 base pair (bp) long units with codominant inheritance (Simbaqueba et al., 2011), also called Simple Sequence Repeats (SSR) or Short Tandem Repeats (STR). SSR were commonly regarded as genomic "junk" with no significant role as genomic information in a long time until the more utilizing of SSR repeat-number variation and accumulating evidence to support the hypothesis that SSR could play a positive role in adaptive evolution (Kashi et al., 2006; Li et al., 2010; Li et al., 2004).

SSR have many advantages, such as simplicity, effectiveness, abundance, hypervariability, reproducibility, codominant inheritance, and extensive genomic coverage (Powell et al., 1996). According to the origin of the sequences used for the initial identification of simple repeats, SSR are divided into two categories: genomic SSR which are derived from random genomic sequences and EST-SSR derived from expressed sequence tags or from coding sequences. EST-SSR are more evolutionary conserved than noncoding sequences (Wei et al., 2011). EST had been a valuable resource to develop SSR-markers for many animals (Marta et al, 2011), but so far, only a few SSR of P.americana were available in GenBank. EST-SSR are tightly linked with functional genes that may influence certain important agronomic characters (Simbaqueba et al., 2011). The development of next generation sequencing technologies have enabled rapid isolation and identification of EST-SSR. Compared with genomic-SSRs detected in noncoding sequences, EST-SSR are more efficient for QTL mapping, gene targeting, and MAS (Varshney et al., 2005). As transcribed sequences are more conserved than noncoding sequences, the transferability of EST-SSR is better than that of g-SSR (Eujayl et al., 2004; Zhang et al., 2005; Saha et al., 2006).

In this study, we sequenced the transcriptome of *P. americana* by Illumina sequencing method, studied the microsatellite sequences by the MSDBv2.4 developed in our laboratory, and analyzed the microsatellite repeat sequences. To understand the characteristics and compositions of microsatellite repeats in the transcriptome of *P. americana,* we studied and analyzed the transcriptional microsatellites of *P. americana*, which were used to develop the microsatellite markers of the transcriptome of *P. americana*. The results provided data based on studying the functional genomics, population genetic structure and population genetic diversity of *P. americana*.

MATERIALS AND METHODS

Sample Selection

The samples of *P. americana* were provided by Sichuan Good Doctor Pharmaceutical Group Co. Ltd. We selected five groups of *P. americana* samples, they were 3−4-instar-old nymphs (3-instar-old nymphs), 7−8-instar-old nymphs (8-instar-old nymphs), newly emerging adults, female adults and male adults, and separated them.

RNA Extraction

We used animal tissue total RNA extraction Kit (Foregene (Chengdu) Biotechnology Co., Ltd) to extract RNA. Three individuals were selected from each group, intestine and wing were removed, liquid nitrogen was added to grind, and then used about 20 mg powder to extract RNA. The integrity of RNA was detected by 1.5% agarose gel electrophoresis and the purity of RNA was detected by nucleic acid analyzer, then RNA was stored at −80°C condition.

Transcriptome Determination

The *P. americana* RNA samples were sent to Novogene (Beijing) Technology Co., Ltd., and sequenced the transcriptome on Hiseq 2000 platform of Illumina company. The basic steps for sequencing were as follows: The mRNA was enriched with magnetic beads with Oligo (dT). The Fragmentation Buffer was added to the mRNA to break it into the short fragments, and then the mRNA after the fragmentation was used as the template. The cDNA first strand was synthesized from random hexamers, and the cDNA second strand was synthesized by adding the buffer solution, dNTPs, RNase H, DNA polymerase I. After the purification of RNA kit, elution of EB buffer, terminal repair, addition of base A and sequencing connector, recovery of target fragment by agarose gel electrophoresis, PCR amplification, then construction of RNA library, and finally sequence of the constructed library. After quality control, a pair of high quality PE reads of 35 718 099, 31 505 768, 58 573 112, 34 455 335, 40 364 085 were obtained from each group of samples, and the number of each read was more than 30 million pairs. Since *P. americana* had no reference genome, de novo assembly was used to assemble the transcriptome. The common reference transcripts of the 5 sets of samples was assembled by Trinity software. There were intergrowth bacteria of 250 transcripts of the assembled transcripts, 291 000 transcripts of *P. americana* were obtained. The 229 Mb transcriptome sequence of *P. americana* was obtained by splicing and assembling, and the sequence was preserved in FASTA format.

Microsatellite Search and Statistical Analysis

MSDBv2.4 was an interface friendly analysis software developed in Perl language, which could quickly identify and search the whole genome microsatellites. The microsatellite search software MSDBv2.4 was used to search and statistics the transcriptional sequence of *P. americana*. The search criteria of this

Table 1. The numbers, percent and abundance of microsatellites in different types of repeats

Repeat type	Number	Total length, bp	Frequency, loci/Mb	Density, bp/Mb	Percent, %
Mono-	20002	301188	96.38	1451.295	52.52
Di-	3096	51922	14.92	250.19	8.13
Tri-	9334	162462	44.98	782.834	24.51
Tetra-	4939	87008	23.8	419.254	12.97
Pentra-	612	12810	2.95	61.726	1.61
Hexa-	99	2748	0.48	13.241	0.26
Total	38082	618138	183.51	2978.54	100

The suffixes of Mono-, Di-, Tri-, Tetra-, Pentra-, Hexa- are nucleotide.

study was set by software default, the mode was perfect search, the minimum repeats was set to above 12 repeats of mononucleotide, above 7 repeats of dinucleotide, above 5 repeats of trinucleotide, above 4 repeats of tetranucleotide, pentanucleotide, hexanucleotide (Du et al., 2013). According to the arrangement difference between the principle of base complementary pairing and the initial base sequence of statistical copy number, the Perl language script was used to merge the same kind as a duplicate copy type, such as three base repeats AAC, which can be merged with ACA, CAA, TTG, TGT, GTT.

RESULTS

Distribution Characteristics of Six Base Repeats in Microsatellites Abundance and Density of Repeat Types

MSDBv2.4 software was used to search for the microsatellite sequences in *P. americana* transcriptome. The distribution characteristics of the perfect microsatellites with respect to motif length were shown in Table 1. The transcriptome size of *P. americana* was 229 Mb, the total number of microsatellites was 38082, the total length was 618138 bp (accounting for 0.3% of the transcriptome size), and the total frequency was 183.51 loci/Mb. The total density was 2978.54 bp/Mb, in other words, it was 2978.54 bp of microsatellite sequences in 1 Mb transcriptome sequence.

In the transcriptional microsatellites of *P. americana*, the proportions of 1−6-base pair repeats was are shown in Table 1. The number of mononucleotide repeats was the highest (20002), accounting for 52.52% of the total number of microsatellites, the density was 1451.295 bp/Mb and the frequency was 96.38 loci/Mb. The number of trinucleotide repeats was 9.334, accounting for 24.51% of the total number of microsatellites, and the frequency was 44.98 loci/Mb. Tetranucleotide was 4939, accounting for 12.97% of the total microsatellites, and the frequency was 23.8 loci/Mb. Dinucleotide was 3,096, accounting for 8.13% of the total microsatellites, and

the frequency was 14.92 loci/Mb. Pentranucleotide was 612, accounting for 1.61%, and the frequency was 2.95 loci/Mb. The number of hexanucleotide repeats was the least (99), accounting for only 0.26% of the total number of microsatellites, and the frequency was 0.48 loci/Mb.

Duplicate Copy Number Distribution

The number of microsatellites included in different base repeats is shown in Table 2. The distribution of mononucleotide repeats ranged from 12 to 24 times, mainly distributed in 12 to 21 times, accounting for 99.38%; in the mononucleotide repeats, the number of microsatellites with 12−19 times was 4828, 3022, 2187, 1664, 1865, 1910, 1505 respectively. The distribution of dinucleotide repeats ranged from 7 to 12 times, mainly distributed in 7 to 11 times, accounting for 99.52%; in the dinucleotide repeats, the number of microsatellites with 7−11 times was 1116, 698, 534, 484, 249 respectively. The distribution of trinucleotide repeats ranged from 5 to 9 times, mainly distributed in 5 to 7 times, accounting for 99.60%; in the trinucleotide repeats, the number of microsatellites with 5−7 times was 3847, 3553, 1897 respectively.

The distribution of tetranucleotide repeats ranged from 4−8, 13, 25 times, mainly distributed in 4−6 times, accounting for 99.83%; in the tetranucleotide repeats, the number of microsatellites with 4−5 times was 3064, 1790 respectively. The distribution of pentranucleotide repeats ranged from 4−7, 9−11, 15−17 times, mainly distributed in 4, 5 times, accounting for 98.76%; in the pentranucleotide repeats, the number of microsatellites with 4 times was 579. The distribution of hexanucleotide repeats ranged from 4−7, 9, 11, 13, 15−17, 25 times, mainly distributed in 4 times, accounting for 98.29%; in the hexanucleotide repeats, the number of microsatellites with 4 times was 86. The number of microsatellites in the pentranucleotide and hexanucleotide repeats was small.

Among the microsatellites with mononucleotide to hexanucleotide, the maximum repeats number of microsatellites was 5, accounting for 14.84% of the

Table 2. Distribution of SSRs based on the number of repeat units

Number of repeats	Mono-	Di-	Tri-	Tetra-	Pentra-	Hexa-	Percent, %
4	0	0	0	3064	579	86	9.79
5	0	0	3847	1790	12	2	14.84
6	0	0	3553	75	7	2	9.55
7	0	1116	1897	6	5	1	7.94
8	0	698	1	2	0	0	1.84
9	0	534	4	0	1	5	1.43
10	0	484	5	0	3	0	1.29
11	0	249	6	0	2	1	0.68
12	4828	15	7	0	0	0	12.74
13	3022	0	8	1	0	1	7.96
14	2187	0	9	0	0	0	5.77
15	1792	0	10	0	1	0	4.73
16	1664	0	11	0	1	1	4.40
17	1865	0	12	0	1	0	4.93
18	1910	0	13	0	0	0	5.05
19	1505	0	14	0	0	0	3.99
≥20	1229	0	81	1	0	0	3.44
Total	2002	3096	9334	4939	612	99	100

The suffixes of Mono-, Di-, Tri-, Tetra-, Pentra-, Hexa- are nucleotide.

total number of transcriptional microsatellites. The distribution range of repeat copy number in mononucleotide, dinucleotide and trinucleotide was continuous and concentrated; while the distribution range of repeat copy number in tetranucleotide, pentranucleotide and hexanucleotide was not continuous, and it was a fault phenomenon, and there was a big difference in the number of repeat copies.

Distribution of Duplicate Copy Categories

In the same base repeat type, the number of repeat types was varied, and the 1−6 repeats were described as below. They were showed in Fig. 1.

In the mononucleotide repeats, the number of repeats type A was 10323, accounting for 51.61% of the total number of mononucleotide microsatellites (20 002), which was the most. Followed by 8837 copies of T and G of 479 copies, accounting for 44.18 and 2.39%, respectively. The number of C duplicates was the least (363), only accounting for 1.81%.

In the dinucleotide repeats, there were 6 repeats, in which the first was AG repeats (783), accounting for 25.29% of the total number of dinucleotide microsatellites (3096), The second was AT, 702 (22.67%); GT, 616 (19.90%). The third was AC, 519 (16.76%); CT, 462 (14.92%), respectively. The number of CG repeat categories was the lowest, only 14, accounting for 0.45%. It can be seen that the number of the five copy types of AT, AC, AG, CT, GT in the dinucleotide repeats was close to each other, the CG copy types were relatively small.

In the trinucleotide repeats, there were 20 repeat types, the total number was 9334, the main repeats were AAG, AAT, ATC, ATG, ATT, CTT, the number of all of them was more than 500. Among them, the number of AAT repeats was 1713 (36.53%), followed by ATT, 1444 (15.47%) and ATG, 1256 (13.46%), the lowest was CCG, 33 (0.35%). The number of AAT, ATT, ATG, AAG in the trinucleotide repeats was close to each other, all of them were above 1,000, and the other copy types were very small.

There were 59 types in the tetranucleotide repeats, the number of AAAT repeats was 645, accounting for 23.22% of the tetranucleotide repeats (4939), followed by AAAG, 564 (11.42%) and ATTT, 470 (9.52%). Due to the large number of repeat types of tetranucleotide microsatellites, only 9 copy types (more than 100) were listed as below, accounting for more than 2% of the tetranucleotide microsatellites, they were AATG, 342(6.92%); ATGT, 333 (6.74%); ACAT, 271 (5.49%); AAGT, 258 (5.22%); ATTG, 224 (4.54%); ACTT, 210 (4.25%); CTTT, 207 (4.19%); ATTG, 120 (2.43%); ATGG, 115 (2.33%).

The pentranucleotide repeats contain a large number of repeat copies (109 types), however, only 10 or more repeat units were listed below. AAGAG was 31, accounting for 5.07% of the total number of pentranu-
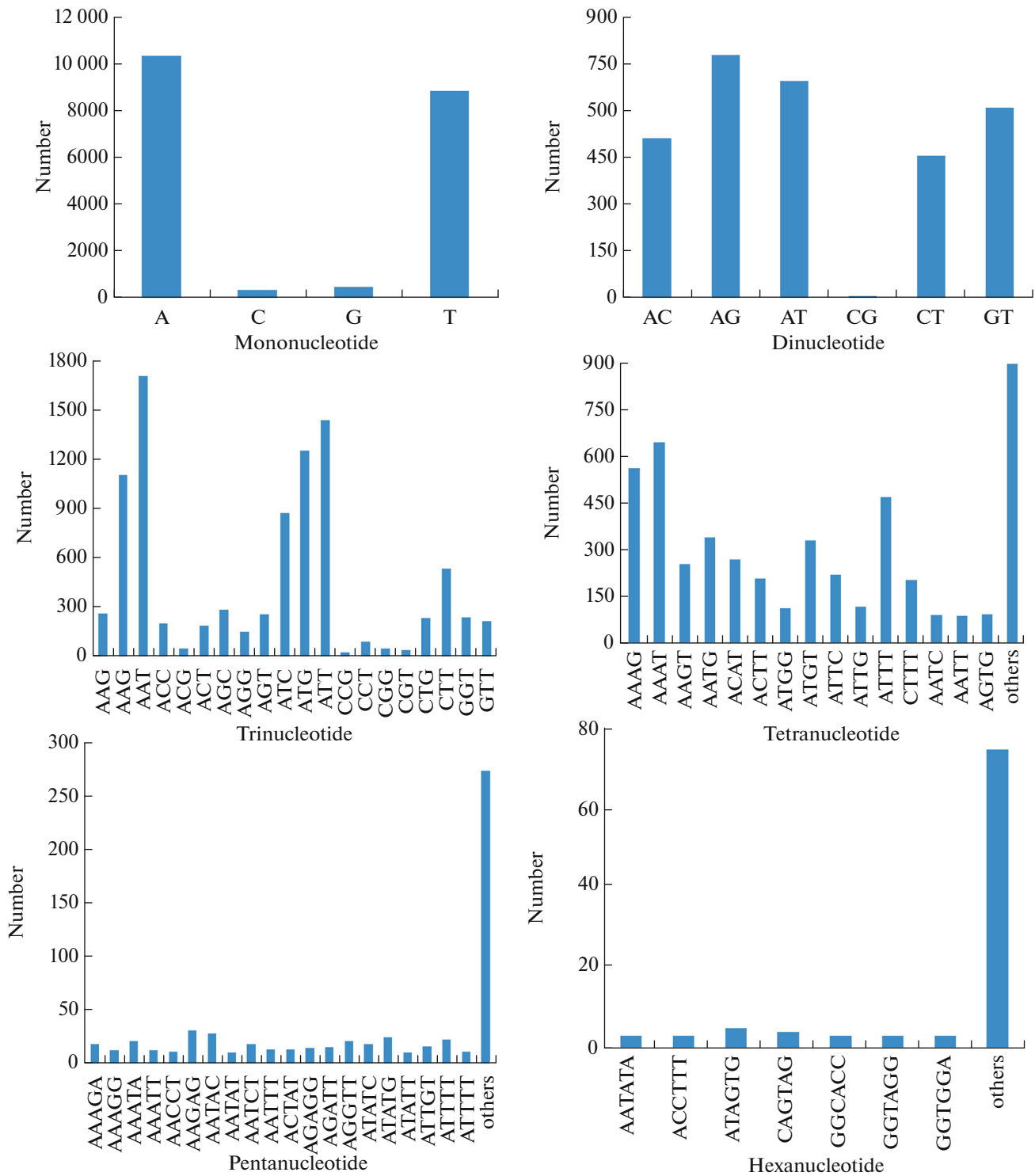
Fig. 1 The number distributions of each repeat copy categories in 1−6 bp repeats.

cleotide repeats (612), followed by AATAC, 28 (4.58%), both of which accounted for more than 4% of the total number of repeats. There were also 9 copy categories, which were ATATG, 24 (3.92%); ATTTT, 22 (3.59%); AAATA, 21 (3.43%); AGGTT, 21 (3.43%); AAAGA, 18 (2.94%); AATCT, 18 (2.94%); ATATC, 18 (2.94%); ATTGT, 16(2.61%); AGATT, 15 (2.45%) in order.

There were 66 repeat copy categories in hexanucle-otide microsatellites, the total number was 99, but the
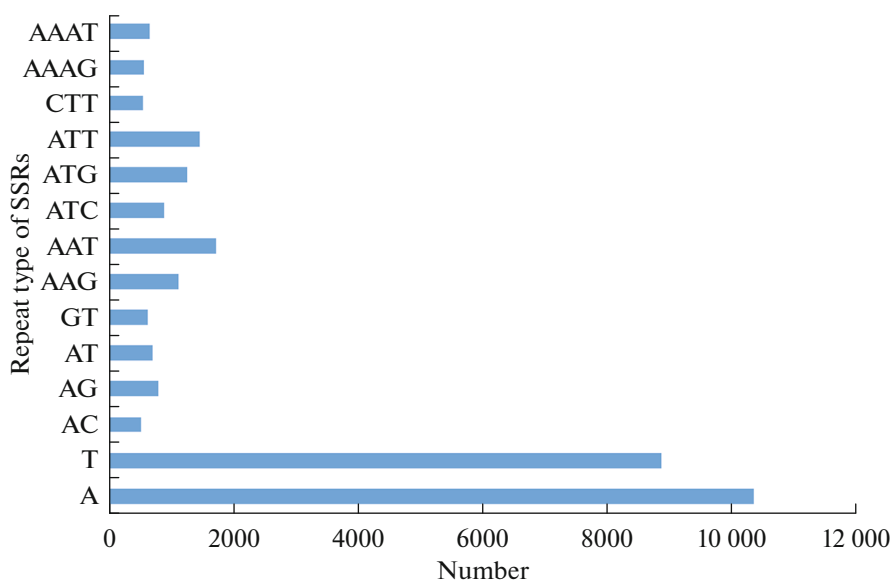
Fig. 2 The most frequency of microsatellite repeats copy.

number of copy categories contained in them was even rarer, and only more than 3 copies was listed. ATAGTG duplicated copy category was the largest, which was 5, accounting for 5.05% of the total number of hexanucleotide microsatellite types, followed by CAGTAG (4), accounting for 4.04%. The number of these two microsatellites was very similar, both of which were more than 4%. AATATA, 3 (3.03%); ACCTTT, 3 (3.03%); GGCACC, 3 (3.03%); GGTAGG, 3 (3.03%); GGTGGA, 3 (3.03%). The total number of the remaining 59 duplicate copy categories was only 75, accounting for 75.76%.

The Dominant Repeat Motif2

According to the descending order, the largest number of the 14 microsatellite repeats in the transcriptome of *P. americana*, they were A, T, AC, AG, GT, AAG, AAT, ATC, ATG, ATT, CTT, AAAG, AAAT, and shown in Fig. 2. The number of these copy types in the transcriptional microsatellites of *P. americana* was more than 500, and they accounted for 29 933, accounting for 78.60% of the total number of transcriptional microsatellites (38 082), while the other microsatellite types accounted for 21.40% of the total transcriptional microsatellites.

DISCUSSION

In the study, the EST-SSR frequency in the *P. americana* was 183.51 loci/Mb, and the density was 2978.54 bp/Mb, accounting for 0.3% of the transcriptome size. The SSR frequency in this study was lower than 8.93% of the EST-SSR in *Sesamum indicum* L. (Wei et al., 2011), *Odontotermes formosanus* (9.98%) (Huang et al., 2012), *Bemisia tabaci* (5.07%) (Xie

et al., 2012), *Rhyacionia leptotubula* (3.09%) (Zhu et al., 2013), *Tenebrio molitor* (1.79%) (Zhu et al., 2013). The percentage of the transcriptomic microsatellites was lower compared with some published projects using genome (rather than transcriptome) sequencing. This may be because the transcriptome sequences can contain introns that do not appear in the transcribed DNA (Wang et al., 2016). The EST-SSR frequency is dependent on several factors such as transcriptome structure or composition (Toth et al., 2000), arithmetical method for SSR detection, and the parameters of the search for microsatellites.

In the microsatellite sequences of the transcriptome of *P. americana*, the number of mononucleotide repeats was the largest (20 002), accounting for 52.52% of the total number of microsatellites. The result was the same as *Catalpa bungei* (Bignoniaceae) (Wang et al., 2016). While dinucleotide repeats were also the most frequent SSR motif type, This finding was consistent with results reported for *Sesamum indicum* L. (Wei et al., 2011), *Phlebotomus papatasi* (67%) (Omar H. and Ahmad A., 2011). The trinucleotide repeats had been observed to have the highest frequency, such as *Rhyacionia leptotubula* (Zhu et al., 2013), 60.7% of *Blattella germanica* (Zhou et al., 2014), *Jatropha Curcas* L. (Wen et al., 2010), *Sogatella furcifera* (Horväth) (Xu et al., 2012), *Nilaparvata lugens* Stál (Jing et al., 2012), *Tetrao tetrix* (Wang et al., 2012). However the tetranucleotide, pentranucleotide and hexanucleotide had very rare dominant repeats.

In general, trinucleotide repeats have been observed to have the highest frequency. It is likely that trinucleotides were so prevalent because they can remain in coding regions without causing reading frame shifts (Wang et al., 2012). This may be the result

of triplet codon selection, as other several types of repeat units (excluding hexanucleotide repeats) can cause change in reading frames, resulting in frameshift mutations, and the gene expression product is to produce completely different protein or is to become shorter. Since the number of repeats of trinucleotide and hexanucleotide repeats does not change the reading frame of the gene, the effect on the gene expression product is relatively small, the coding region sequence has better tolerance to trinucleotide and hexanucleotide microsatellite repeats. Under the selective action, trinucleotide and hexanucleotide microsatellites will be enriched. So the trinucleotide microsatellites have the largest number. In any case, this study and previous studies showed that some species had the most mononucleotide repeats, some had dinucleotide and some had trinucleotide, Among them, it was in the majority that the trinucleotide was the dominant repeats. It could be seen that there are different dominant repeats in different species. The bias of the number of repeats and the differences in types may be related to the evolution or functional specificity of the species themselves.

The number of copy repeats in different microsatellites was quite different. Among the mononucleotide repeats of *P. americana*, A (10323) was the most frequent motif in our data sheet, and this result was consistent with the data reported for *Phlebotomus papatasi* (A, 28.7%) (Omar H. and Ahmad A., 2011). Among the dinucleotide repeats, AG (783) was the most frequent motif, whereas CG (14) motifs were the rarest. This was similar to that AG/CT (46.29%) of *Sesamum indicum* L. (Wei et al., 2011). It was different from that in AC/GT (8.34%) of *Rhyacionia leptotubula* (Zhu et al., 2013). The majority of the trinucleotide repeats was AAT (1713), our results are consistent with those for other species, such as *Blattella germanica* (Zhou et al., 2014), *Nilaparvata lugens* Stäl (AAT, 15.1%) (Jing et al., 2012). It was different from that in ATC/ATG (13.59%) of *Rhyacionia leptotubula* (Zhu et al., 2013). The number of different copy types in the same type of repeats varies greatly, which may be related to the frequency of use of the corresponding encoded proteins in different species.

In the transcriptomic microsatellites, A was the dominant repeats in the mononucleotide, AG, AAT, AAAT, AAGAGG, ATAGTG was the dominant repeats in the dinucleotide, trinucleotide, tetranucleotide, pentranucleotide, hexanucleotide, respectively. This phenomenon may be caused by the fact that transcriptome microsatellites are mainly coming from the coding region sequence. Compared with the genome non-coding sequence, the coding region sequence is under more selection pressure and is relatively difficult to mutate. It is more helpful to study the gene function of species.

The number of microsatellites in the transcriptome of *P. americana* was analyzed. It was found that when the repeat number of microsatellites of the same motif increased, the number of microsatellites decreased. While the number of microsatellites with the same number of repeats tended to decrease with the increase of the motof size. And with the increasing of the length of different repeat motifs, the range of repeats decreases gradually. For example, dinucleotide repeats were divided into six types according to different number of repeats, in which the number of dinucleotide repeats with repeats number of 7 was 1116, gradually decreasing with the number of repeats increased, when the number of repeats increased to 10, the number of repeat motifs reduced to 15; when the length of repeats increased to hexanucleotide repeats, there were 8 repeat types and only 1 in repeats number of 7. It can be seen that the number of microsatellites and the range of repeats decreased with the increasing of the length of repeats. It was inferred that the variation rate of shorter microsatellites in the transcriptome of *P. americana* was faster than that of the longer ones.

Microsatellites are widely distributed in the coding and non-coding regions of eukaryotic genome. It can provide important information resources for the development of a large number of highly efficient microsatellite molecular markers by analyzing the characteristics of microsatellites in the transcriptome of *P. americana*. In addition, microsatellites in the transcriptome are gene-functional sequences, which are helpful for the development of microsatellite markers associated with important genes of *P. americana*. The transcriptome microsatellites of *P. americana* laid a foundation for genetic diversity, germplasm resources identification and molecular marker-assisted breeding.

CONCLUSIONS

Using Illumina sequencing technology, a large of *P. americana* transcriptome (229 MB) was achieved. Of these, mononucleotide SSRs were the dominant repeat type. A, AG, AAT, AAAT, AAGAGG, ATAGTG were the dominant repeat motives. The maximum repeats number of microsatellites was 5, accounting for 14.84% of the total number of transcriptional microsatellites. These results based on a foundation for developing high polymorphic microsatellites. The platform constructed in this study is beneficial for us to have a better understanding of the fundamental molecular knowledge of this pest. It is also valuable for further research of gene expression, genomics, and functional genomics on this species.

ACKNOWLEDGMENTS

ceutical Group Co., Ltd. Wenqi Ma assisted in the experiment, Chen Wang and Wujiao Li supported in data analysis.

COMPLIANCE WITH ETHICAL STANDARDS

Conflict of interest. The authors declare that they have no conflicts of interest.

Statement on the welfare of animals. All applicable international, national, and/or institutional guidelines for the care and use of animals were followed.

REFERENCES

Chung, H.S., Yu, T.H., Kim, B.J., et al., Expressed sequence tags analysis of *Blattella germanica, Korean J. Parasitol.*, 2005, vol. 43, no. 4, p. 149.

Du, L., Li, Y., Zhang, X., et al., MSDB: a user-friendly program for reporting distribution and building databases of microsatellites from genome sequences, *J. Hered.*, 2013, vol. 104, no. 1, p. 154.

Emrich, S.J., Barbazuk, W.B., Li, L., et al., Gene discovery and annotation using LCM-454 transcriptome sequencing, *Genome Res.*, 2007, vol. 17, no. 1, p. 69.

Eujayl, I., Sledge, M.K., Wang, L., et al., *Medicago truncatula* EST-SSRs reveal cross-species genetic markers for *Medicago* spp., *Theor. Appl. Genet.*, 2004, vol. 108, no. 3, pp. 414–422.

Gao, Z., Wang, W., and Zhou, X., DNA marker technologies and their applications in aquaculture genetics, *Biotechnol. Bull.*, 2007, vol. 238, no. 1, pp. 1–37.

Gao, X., Han, J., Lu, Z., et al., RETRACTED: Characterization of the spotted seal *Phoca largha* transcriptome using Illumina paired-end sequencing and development of SSR markers, *Comp. Biochem. Physiol., Part D: Genomics Proteomics*, 2012, vol. 7, no. 3, pp. 277–284.

Gavery, M.R. and Roberts, S.B., Characterizing short read sequencing for gene discovery and RNA-Seq analysis in *Crassostrea gigas, Comp. Biochem. Physiol., Part D: Genomics Proteomics*, 2012, vol. 7, no. 2, pp. 94–99.

Hamada, H., Petrino, M.G., and Kakunaga, T., A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes, *Proc. Natl. Acad. Sci. U. S. A.*, 1982, vol. 79, no. 21, pp. 6465–6469.

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), *PloS One*, 2012, vol. 7, no. 11, p. e50383.

Jaramillo-Ramirez, G.I., Cardenas-Henao, H., Gonzalez-Obando, R., et al., Genetic variability of five *Periplaneta Americana* L. (Dyctioptera: Blattidae) populations in southwestern Colombia using the AFLP molecular marker technique, *Neotrop. Entomol.*, 2010, vol. 39, no. 3, p. 371.

Jing, S., Liu, B., Peng, L., et al., Development and use of EST-SSR markers for assessing genetic diversity in the brown planthopper (*Nilaparvata lugens* Stål), *Bull. Entomol. Res.*, 2012, vol. 102, no. 1, p. 113.

Kashi, Y. and King, D.G., Simple sequence repeats as advantageous mutators in evolution, *Trends Genet.*, 2006, vol. 22, no. 5, pp. 253–259.

Li, Y.C., Korol, A.B., Fahima, T., et al., Microsatellites within genes: structure, function, and evolution, *Mol. Biol. Evol.*, 2004, vol. 21, no. 6, pp. 991–1007.

Li, H., He, C., Yang, Q., et al., Characterization of single nucleotide polymorphisms from expressed sequence tags of Chinese mitten crab *Eriocheir sinensis, Aquat. Biol.*, 2010, vol. 11, pp. 193–199.

Li, Y., Korol, A.B., Fahima, T., et al., Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review, *Mol. Ecol.*, 2010, vol. 11, no. 12, pp. 2453–2465.

Li, H., Liang, Y., Sui, L., et al., Characterization of 10 polymorphic microsatellite markers for Mediterranean blue mussel *Mytilus galloprovincialis* by EST database mining and cross-species amplification, *J. Genet.*, 2011, vol. 90, no. 1, p. e30.

Li, H.J., Liu, W.D., Gao, X.G., et al., Identification of host-defense genes and development of microsatellite markers from ESTs of hard clam *Meretrix meretrix, Mol. Biol. Rep.*, 2011, vol. 38, no. 2, pp. 769–775.

Marta, M., Cavagnaro, P.F., Megan, B., et al., De novo assembly and characterization of the carrot transcriptome reveals novel genes, new markers, and genetic diversity, *BMC Genomics*, 2011, vol. 12, no. 1, p. 389.

Omar, H. and Ahmad, A., Characterization of simple sequence repeats (SSRs) from *Phlebotomus papatasi* (Diptera: Psychodidae) expressed sequence tags (ESTs), *Parasites Vectors*, 2011, vol. 4, no. 1, p. 189.

Powell, W., Morgante, M., Andre, C., et al., The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis, *Mol. Breed.*, 1996, vol. 2, no. 3, pp. 225–238.

Saha, M.C., Cooper, J.D., Mian, M.A., et al., Tall fescue genomic SSR markers: development and transferability across multiple grass species, *Theor. Appl. Genet.*, 2006, vol. 113, no. 8, pp. 1449–1458.

Santure, A.W., Gratten, J., Mossman, J.A., et al., Characterisation of the transcriptome of a wild great tit *Parus major* population by next generation sequencing, *BMC Genomics*, 2011, vol. 12, no. 1, p. 283.

Schlötterer, C., The evolution of molecular markers–just a matter of fashion, *Nat. Rev. Genet.*, 2004, vol. 5, no. 1, pp. 63–69.

Sharma, P.C., Grover, A., and Kahl, G., Mining microsatellites in eukaryotic genomes, *Trends Biotechnol.*, 2007, vol. 25, no. 11, pp. 490–498.

Shen, Y., Catchen, J., Garcia, T., et al., Identification of transcriptome SNPs between *Xiphophorus*, lines and species for assessing allele specific gene expression within F_1, interspecies hybrids, *Comp. Biochem. Physiol., Part C: Pharmacol., Toxicol. Endocrinol.*, 2012, vol. 155, no. 1, pp. 102–108.

Simbaqueba, J., Sánchez, P., Sanchez, E., et al., Development and characterization of microsatellite markers for the cape gooseberry *Physalis peruviana, PloS One*, 2011, vol. 6, no. 10, p. e26719.

Sun, L., Chen, M., Yang, H, et al., Large scale gene expression profiling during intestine and body wall regeneration in the sea cucumber *Apostichopus japonicas, Comp.*

Biochem. Physiol., Part D: Genomics Proteomics, 2015, vol. 6, no. 2, pp. 195−205.

Tautz, D. and Renz, M., Simple sequences are ubiquitous repetitive components of eukaryotic genomes, *Nucleic Acids Res.*, 1984, vol. 12, no. 10, pp. 4127−4138.

Tóth, G., Gáspári, Z., and Jurka, J., Microsatellites in different eukaryotic genomes: survey and analysis, *Genome Res.*, 2000, vol. 10, no. 7, p. 967.

Varshney, R.K., Graner, A., et al., Genic microsatellite markers in plants: features and applications, *Trends Biotechnol.*, 2005, vol. 23, no. 1, pp. 48−55.

Wang, B., Ekblom, R., Castoe, T.A., et al., Transcriptome sequencing of black grouse (*Tetrao tetrix*) for immune gene discovery and microsatellite development, *Open Biol.*, 2012, vol. 2, no. 4, p. 120054.

Wang, P., Ma, Y., Ma, L., et al., Development and characterization of EST-SSR markers for *Catalpa bungei* (Bignoniaceae), *Appl. Plant Sci.*, 2016, vol. 4, no. 4, p. 1500117.

Wei, W., Qi, X., Wang, L., et al., Characterization of the sesame (*Sesamum indicum* L.) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers, *BMC Genomics*, 2011, vol. 12, no. 1, p. 451.

Wen, M., Wang, H., Xia, Z., et al., Developmenrt of EST-SSR and genomic-SSR markers to assess genetic diversity in *Jatropha Curcas* L., *BMC Res. Notes*, 2010, vol. 3, no. 1, p. 42.

Xie, W., Meng, Q.S., Wu, Q.J., et al., Pyrosequencing the *Bemisia tabaci* transcriptome reveals a highly diverse bacterial community and a robust system for insecticide resistance, *PloS One*, 2012, vol. 7, no. 4, p. e35181.

Xu, Y., Zhou, W., Zhou, Y., et al., Transcriptome and comparative gene expression analysis of *Sogatella furcifera* (Horváth) in response to southern rice black-streaked dwarf virus, *PloS One*, 2012, vol. 7, no. 4, p. e36238.

Zhang, L.Y., Bernard, M., Leroy, P., et al., High transferability of bread wheat EST-derived SSRs to other cereals, *Theor. Appl. Genet.*, 2005, vol. 111, no. 4, pp. 677−687.

Zhou, X., Qian, K., Tong, Y., et al., *De Novo* transcriptome of the hemimetabolous German cockroach (*Blattella germanica*), *PloS One*, 2014, vol. 9, no. 9, p. e106932.

Zhu, J.Y., Li, Y.H., Yang, S., et al., De novo assembly and characterization of the global transcriptome for *Rhyacionia leptotubula* using Illumina paired-end sequencing, *PloS One*, 2013, vol. 8, no. 11, p. e81096.

Zhu, J.Y., Wu, G.X., and Yang, B., High-throughput discovery of SSR genetic markers in the yellow mealworm beetle, *Tenebrio molitor* (Coleoptera: Tenebrionidae), from its transcriptome database, *Acta Entomol. Sin.*, 2013, vol. 56, no. 7, pp. 724−728.