International Journal of Odonatology, 2018
Vol. 21, Nos. 3–4, 221–227, https://doi.org/10.1080/13887890.2018.1531066

Development of SSR markers based on the head transcriptome of *Pantala flavescens* (Fabricius, 1798) (Odonata: Libellulidae)

L.Z. Cao[a*] and Y.G. Yuan[b]

[a]*College of Life Science, Jiangxi Normal University, Nanchang, PR China;* [b]*College of Pharmacy, Guangdong Pharmaceutical University, Guangzhou, PR China*

(*Received 19 September 2017; accepted 28 September 2018*)

Pantala flavescens (Fabricius, 1798) is one of the most common species of dragonflies and has been found throughout from tropic to temperate zones worldwide. In this study, RNA-seq of *P. flavescens* was carried out through Illumina high-throughput sequencing technology. Approximately 37,868 unigenes and 47,188 transcripts were obtained. The average length of the assembled unigenes was 908.59 bp. We identified 1442 cDNA simple sequence repeats (SSRs) among the 37,868 unigenes, with 864 (59.91%) di-nucleotide repeats, 537 (37.32%) tri-nucleotide repeats, 32 (2.22%) complex-nucleotide repeats, and 9 (0.62%) with tetra-nucleotide repeats. Sixty microsatellite molecular markers were randomly selected to test amplification. Of the 60 markers, 32 (53.33%) produced clear amplicons of the expected size, 10 (16.67%) amplified nonspecific products, and 18 (30%) failed to amplify the DNA products. In order to assess their applicability, genetic diversity of the 32 SSR loci was tested in 32 individuals from Nanchang in China. Of these loci, 14 markers were highly polymorphic, with the observed (Ho) and expected (He) heterozygosities ranged 0.69 to 0.88 and from 0.96 to 0.98 respectively. PIC ranged from 0.52 to 0.83. These highly polymorphic loci will be valuable for the genetic analysis of distinct populations of *P. flavescens*.

Keywords: dragonfly; transcriptome sequencing; microsatellite; genetic diversity

Introduction

*Pantala flavesce*ns (Fabricius, 1798), commonly known as the wandering glider or the globe skimmer, may be the most widespread of any known dragonfly species (Russell, May, Soltesz, & Fitzpatrick, 1998). Previous studies indicated that *P. flavescens* was the most abundant species of Odonata, and possessed the capability to migrate several thousands of kilometers worldwide (Anderson, 2009; Artiss, 2004; Borisov, 2009; Feng, Wu, Ni, Cheng, & Guo, 2006). *P. flavescens* has a nearly cosmopolitan distribution; this species was thought to undertake extensive migration following the movement of the Inter-Tropical Convergence Zone (ITCZ; Hobson, Anderson, Soto, & Wassenaar, 2012). The larval stage of *P. flavescens* is very short, about 34–43 days, which allowed this species to breed in the ephemeral freshwater pools and floods produced by ITCZ rainfall (Hobson et al. 2012). Seasonal migration of *P. flavescens* occurs in many parts of its distribution region, with large swarms during long-distance migration (Srygley, 2003). In China, *P. flavescens* was first reported as migratory species (Feng et al., 2006) and had three

*Corresponding author. Email: clzclz1011@163.com

movement strategies: wandering around northern China, and north-bound (positive) and south-bound (negative) movements, with the majority engaged in wandering around northern China (Cao, Fu, Hu, & Wu, 2018). Catling (2005) used dragonflies to measure the efficiency of sewage lagoons, and stated that several species worked well as indicators for that purpose. Dragonflies have been used as bioindicators of habitat quality and environmental health (Gamboa, Reyes, & Arrivillaga, 2008; Martin & Maynou, 2016). Many dragonflies are also considered the important natural enemies of many insect pests, such as *Anopheles* mosquitoes, flies, and gnats (Russell et al., 1998).

Microsatellites or simple sequence repeats (SSRs) **are** tandem repeated motifs of 1–6 bp in length, and they were widely distributed throughout the eukaryotic genomes (Yang, Sun, Xue, Zhu, & Hong, 2012). Because of their ubiquity and high level of polymorphism, microsatellite markers have been utilized extensively to characterize genetic structure and diversity, to construct phylogenetic trees and to identify unique sources of allelic diversity (Sun, Li, Yang, & Hong, 2012). The traditional techniques for constructing enriched genomic libraries to develop SSR markers of *P. flavescens* were usually low efficiency, labor intensive and time consuming (Cao, Fu, & Wu, 2015). Next-generation sequencing technologies and the rapid development of high-throughput platforms have raised the bar regarding the sequencing of non-model organisms (Patnaik et al., 2016). These technologies offer a collated and comprehensive output for discovery of novel transcripts for functional studies and molecular marker development (Bräutigam, Mullick, Schliesky, & Weber, 2011). In this study, we processed a large number of high-quality transcriptome sequences and discovered a set of SSRs of *P. flavescens* using the Illumina HiSeq 4000 platform. The genetic characteristic of SSR markers of *P. flavescens* was analyzed and further application in genetic diversity analysis was validated. Finally, we obtained 14 highly polymorphic microsatellite markers that would allow researchers to investigate the genetic diversity and population genetic structure of *P. flavescens* in China and worldwide.

Materials and methods

Sample collection

In total, 32 *P. flavescens* were collected on 12 July 2016 from Nanchang, Jiangxi province, China. The species had been evaluated as least concern species by IUCN (www.iucnredlist.org/species/59971/65818523#threats). DNA was extracted from the thoracic muscle to analyze genetic diversity. Total RNA was extracted from 10 heads for RNA-seq.

cDNA library construction and Illumina sequencing

Total pooled RNA samples were sent to Shanghai Major Medical Laboratory Ltd (Shanghai, China) for mRNA cDNA-seq library construction and sequencing using Illumina HiSeq 4000 (Illumina Inc. San Diego, CA, USA) sequencing platform. In brief, the cDNA library of pooled RNA was obtained using Illumina TruSeqTM RNA sample preparation kit (Illumina Inc.), according to the manufacturer's instructions. Poly-T attached magnetic beads (Illumina, Inc.) were used to isolate poly-A mRNA from total RNA. The total mRNA was fractured into about 200 bp fragments randomly by adding fragmentation buffer (Ambion Inc, Austin, TX, USA). First-strand cDNA was synthesized from the fragmented mRNA using reverse transcriptase (Invitrogen Inc., Carlsbad, CA, USA) and random primers. In the following, the synthesis of the second strand of cDNA was accomplished using DNA polymerase I (New England Biolabs, Beijing, China) and RNase H (Invitrogen). Subsequently, the 3′ end of double-stranded cDNA

was repaired using T4 DNA polymerase, T4 polynucleotide kinase and Klenow fragment (New England BioLabs, Inc.). The end-repaired cDNA was ligated to Illumina paired-end (PE) adapter and then enriched by 15 cycles of PCR amplification. The suitable fragments (about 200 bp) were sequenced in paired-end pattern on an Illumina HiSeq 4000 sequencing platform. The sequencing data were automatically collected and generated into FASTQ format.

Simple sequence repeat markers discovery and primer design

The microsatellites in the unigene sequences of *P. flavescens* were localized by microsatellite identification tool (MISA; http://pgrc.ipk-gatersleben.de/misa/). All types of simple sequence repeat markers were identified from mononucleotide to tetranucleotide repeats. The selection criteria for SSR markers were as follows: the minimum repeat unit was 5 for trinucleotide to tetranucleotide and 6 for dinucleotide. Mononucleotides repeats were not regarded as the SSR analysis. The SSR loci that were used for genetic markers should include a perfect repeat motif and two unique flanking sequences with about 200 bp on each sides of the repeat.

The forward and reverse primers were designed based on unique flanking sequences using Batch Primer 3 (http://perlprimers.sourceforge.net). Primer lengths ranged from 15 to 25 bases with optimal sizes of 20 nt. Annealing temperature was between 45–60°C, with 55°C as the optimum annealing temperature. PCR product lengths ranged from 100 to 300 bp.

SSR markers validation and genetic diversity analysis

To verify the availability of SSR markers, 60 primer pairs were randomly chosen for PCR amplification. Genomic DNA was extracted from thoracic muscles of an individual *P. flavescens* using the hexadecyl trimethyl ammonium bromide (CTAB) method (Englen & Kelley, 2000). PCR amplification was performed in 25 μl volume containing 50 ng of genomic DNA, 0.4 μM each forward and reverse primer, 12.5 μl 2 $\times$ PCRMix using Bio-Rad Thermal Cycler (Bio-Rad Laboratories, Inc., Hercules, CA, USA). Cycling conditions were: initial denaturation at 95°C for 2 min, 35 cycles of 30 s at 94°C, the locus specific melting temperature (Tm) for 30 s, 72°C for 30 s, and a final extension at 72°C for 5 min. The amplification products were separated on a 3% agarose gel. The 50 bp DNA ladder (Sangon Biotech Co., Ltd, Shanghai, China) was used as the standard size marker. For these SSR markers with stable and clear bands, each of the forward primers was synthesized with a universal adapter (M13-21) on the 5′ end. PCR amplification was performed in the 25 μl volume containing three primers (0.3 μM fluorescent dye-labeled adapter, 0.1 μM forward primer added adapter and 0.4 μM reverse primer). The reaction conditions were identical to the above. PCR amplicons were separated by capillary electrophoresis on ABI a 3730 DNA Analyzer (Applied Biosystems, Foster City, CA, USA) along with the GeneScan-500 LIZ size standard.

Data analysis

Allele data were scored using GENMARKER version 4.0 (Applied Biosystems, Foster City, CA, USA). MICRO-CHECKER ver. 2.2.3 (Van Oosterhout, Hutchinson, Derek, & Shipley, 2004) was used to assess null alleles and scoring errors. The program PopGene 32 (Yeh, Yang, & Boyle, 1999) was used to estimate expected heterozygosities (He), observed heterozygosities (Ho) and the number of alleles (Na), departure from HWE. Polymorphism information content (PIC) values for each of the SSR markers was calculated using PIC_Calc 0.6.

Results

Illumina HiSeq 4000 and P. flavescens transcriptome

We obtained the head transcriptome of *P. flavescens*, with over 47,039,040 raw sequences of 7,102,895,040 bases (Q20 percentage of 95.28%). After stringent filtration, there were 45,908,046 clean reads (Q20 percentage of 97.34%) with an accumulated length of 6,694,354,464 bases representing 97.60% of the raw reads. The combined sequences of these reads were assembled into 37,868 unigenes and 47,188 individual transcripts. The average length of the assembled unigenes was 908.59 bp with N50 = 1836bp which was shorter than the average length of the assembled transcripts (1112.82 bp with N50 = 2223 bp). Among the unigenes, 18,441 (48.7%) were 100–400 bp: 5016 (13.25%) were 401–600 bp; 2710 (7.16%) were 600–800 bp; 1774 (4.68%) were 801–1000 bp; and 9927 (26.21%) were more than 1000 bp in length (Figure 1).

Frequency and types of SSR markers

Among the 37,868 unigenes, 1442 SSRs were identified. Of the unigenes, 1235 contained SSRs, and 32 (2.12%) sequences had more than one SSR. The types of repeat motif were not evenly distributed; the di-nucleotide repeat motifs were the most abundant at 864 (59.92%), followed by 537 (37.24%) tri-nucleotide repeat motifs, and 9 (0.62%) tetra-nucleotide repeat motifs (Figure 2).

The number of SSR repeats ranged from 5 to 12, with six repeats being the most abundant, followed by five and seven repeats as the next most abundant. Motifs with more than 10 repeats were rare 12 (0.85%). Among the nucleotide repeats, AT/TA was the most abundant. The distributed positions of SSRs were different with 438 SSRs in the coding regions, 537 SSRs in the 3′UTR, 117 SSRs in the 5′UTR and the others undetermined.

Development of polymorphic SSR markers

Sixty SSR markers were randomly selected to test amplification and genetic nature by analyzing 32 individuals from Nanchang, Jiangxi province. Of the 60 SSR markers, 32 (53.33%)

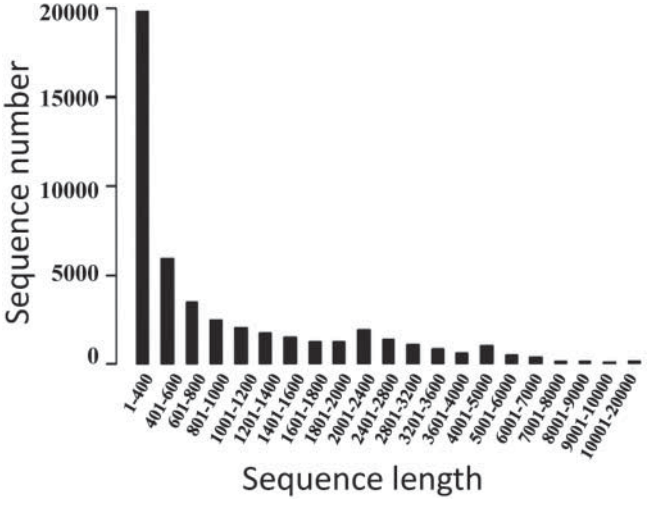

Figure 1. Frequency length distribution of Illumina read sequences.

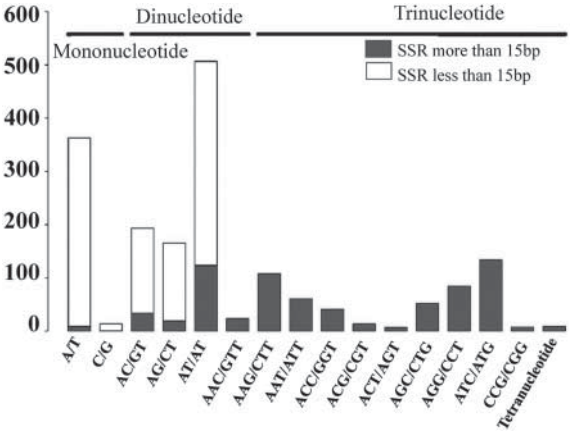

Figure 2. Summary of the number of repeat units in *P. flavescens* SSR loci.

produced clear bands of the expected size, 10 (16.67%) amplified non-specific products, and 18 (30%) failed to amplify the DNA product. Among the 32 SSR markers, 10 showed no polymorphisms, eight showed moderate polymorphisms and 14 showed high polymorphisms in the 32 individuals. These 14 high polymorphism loci (PIC > 0.5) were submitted to Genbank and their characteristics are outlined in Table 1. Among 32 individuals of *P. flavescens*, 126 alleles in total were detected for 14 high polymorphism loci with allele numbers varying between 5 and 14 with

Table 1. Characterization of 14 high polymorphic microsatellite loci of *P. flavescens*.

Locus	Primer sequences (5′–3′)	Repeat motif	Size range (bp)	Tm (°C)	Accession no.
SSR2	F:*TCAAAAACTGTAAACCCGCC R:GGGCAATGAATGCAGAAAAG	(CT)6	179–228	52	MF002465
SSR5	F:*CAAAATTGGAGCAGCACTGA R:GCCGCCGTCATACAGATATT	(ATG)6	225–238	52	MF002466
SSR7	F:*GGAGCCAGCTGTTTTTATCG R:ATCTCAGCCACCCCTACCTT	(AG)6	162–270	52	MF002467
SSR8	F:*CACTTACATAATGGGTGAATCCTG R:AGGCAACTACATTATGCCCG	(AT)6	152–173	52	MF002468
SSR10	F:*TCAGGTTCAGCCTAGGTGCT R:TGCAGCATTACCTTCAGCAT	(AT)9	128–322	52	MF002469
SSR11	F:*ACCGATCAAAACGAAGCAAT R:TGACGTCACCCTTTAGTGTCAT	(AT)9	136–156	52	MF002470
SSR14	F:*GAGCGATAGGAAAAGAGGGG R:GACTCTGTTGGGGTTCGTGT	(AG)8	205–213	52	MF002471
SSR16	F:*TTTTTCCAAAGGTTTTCCTGA R:TGCTACATGAAGGAGCAATGA	(CT)8	242–256	50	MF002472
SSR17	F:*TCACACCTTTCCACACTTGC R:CACTGTAAAAAGAAATGTTGACCA	(AT)8	236–245	50	MF002473
SSR18	F:*GATTGGGAAATGAGCTGATGA R:ACACACAACAATGCTGCCAT	(TA)10	216–285	50	MF002474
SSR19	F:*GGAGACACCTTTTGTGAGTCG R:TGTTGTTGCCTGTGTCACCT	(TA)7	116–164	55	MF002475
SSR20	F:*CCTCGACTCTGGATCTCCACT R:CCAATGTGGATTCTTCGCTT	(ATC)7	191–203	55	MF002476
SSR21	F:*GCTCAGAGCACTTCAGCATTT R:GAGTATAAATTGCACAGCCAAAAA	(TA)8	274–296	55	MF002477
SSR22	F:*CGGGAGAGAGGGCTTAAGAG R:CAGACTCTTGGCAGTGAGATG	(AC)7	105–193	50	MF002478

Notes: Tm, annealing temperature; accession no.: GenBank accession number.

Table 2. Summary of genetic diversity of 14 SSRs in 32 individuals of *P. flavescens*.

Loci	Na	Ho	He	PIC	I	H-W	Loci	Na	Ho	He	PIC	I	H-W
SSR2	14	0.72	0.97	0.83	2.15	0.59	SSR16	13	0.78	0.97	0.8	2.05	0.45
SSR5	6	0.75	0.98	0.64	1.36	0.07	SSR17	6	0.72	0.97	0.55	1.18	0.00
SSR7	11	0.81	0.98	0.62	1.57	0.04	SSR18	8	0.88	0.97	0.7	1.6	0.03
SSR8	10	0.78	0.98	0.7	1.67	0.73	SSR19	6	0.78	0.97	0.62	1.3	0
SSR10	12	0.81	0.98	0.66	1.67	1	SSR20	11	0.81	0.97	0.8	1.96	0.01
SSR11	9	0.69	0.96	0.62	1.47	0.99	SSR21	9	0.81	0.96	0.7	1.61	0.9
SSR14	5	0.72	0.97	0.68	1.43	0.00	SSR22	6	0.69	0.97	0.52	1.11	0.02

a mean of 9 per marker. The Ho and He ranged from 0.69 to 0.88 and from 0.96 to 0.98, respectively. Shannon's diversity index ranged from 1.11 to 2.15. Four SSRs (SSR7, SSR18, SSR20 and SSR22) deviated from Hardy–Weinberg equilibrium, and three SSRs (SSR14, SSR17 and SSR19) significantly deviated from Hardy–Weinberg equilibrium. PIC varies from 0.52 to 0.83 with average 0.57. The genetic diversities for these SSR markers are given in Table 2.

Discussion

With a nearly global distribution, *P. flavescens* may be the most widespread of any known dragonfly species (Hobson et al., 2012; Russell et al., 1998). Isotopic evidence suggested that the multigenerational journey may total over 18,000 km with single individuals traveling over 6000 km during the transoceanic trek from northern India to east Africa (Hobson et al., 2012). Migratory behavior of *P. flavescens* presented a unique opportunity to ask questions regarding the amount of gene flow that may be occurring on a global scale as well as its influence on both the population structure and genetic diversity of the species.

Previous studies suggested low genetic diversity and a high rate of gene flow among five geographically isolated populations of *P. flavescens* within India using randomly amplified polymorphic DNA (Christudhas & Mathai, 2014). Troast, Suhling, Jinguji, Sahlén, and Ware (2016) suggested high rates of gene flow were occurring among all included geographic regions and genes were being shared among individuals across the globe using PCR-amplified cytochrome oxidase one (CO1) mitochondrial DNA data.

Although the use of the mitochondrial CO1 gene has proven to be an effective marker for studying divergence within and among species (Hebert, Ratnasingham & DeWaard, 2003), microsatellites might help to evaluate any findings uncovered thus far. Genetic analyses with more loci over an even wider geographic range will be required to address this question.

In this study, we aim to develop a set of microsatellite loci for genetic analysis of distinct populations of *P. flavescens* from large geographical scales. Our result showed that transcriptome sequencing was a very useful tool for SSR development compared with screening partial genomic libraries enriched (Cao et al., 2015). Fourteen highly polymorphic SSR markers were obtained among 60 SSRs selected randomly from the unigenes. The set of reliable SSR markers might be used to research genetic structure and genetic diversity of migratory Odonata in a wider geographic region. In addition, the potential ecological impacts of the migratory behavior are worthy of further discussion.

Funding

This research was supported by the National Natural Science Funds [number 31460489].

References

Anderson, R. C. (2009). Do dragonflies migrate across the western Indian Ocean? *Journal of Tropical Ecology*, *25*, 347–358.

Artiss, T. (2004). Phylogeography of a facultatively migratory dragonfly, *Libellula quadrimaculata* (Odonata: Anisoptera). *Hydrobiologia*, *515*, 225–234.

Borisov, S. N. (2009). Study of dragonfly (Odonata) migrations in the Western Tien Shan Mountain using ornithological traps. *Entomological Review*, *89*, 1025–1029.

Bräutigam, A., Mullick, T., Schliesky, S., & Weber, A. P. M. (2011). Critical assessment of assembly strategies for non-model species mRNA-Seq data and application of next-generation sequencing to the comparison of C3 and C4 species. *Journal of Experimental Botany*, *62*(9), 3093–3102.

Cao, L. Z., Fu, X. W., & Wu, K. M. (2015). Development of 10 microsatellite markers from *Pantala flavescens* and their applicability in studying genetic diversity. *Molecular Biology Report*, *42*(8), 1275–1279.

Cao, L. Z., Fu, X. W., Hu, C. X., & Wu, K. M. (2018). Seasonal migration of *Pantala flavescens* across the Bohai Strait in Northern China. *Environmental Entomology*, *47*(2), 264–270.

Catling P. M. A. (2005). A potential for the use of dragonfly (Odonata) diversity as a bioindicator of the efficiency of sewage lagoons. *Canadian Field-naturalist*, *119*(2), 233–236.

Christudhas A., & Mathai M. T. (2014). Genetic variation of a migratory dragonfly characterized with random DNA markers. *Journal of Entomology and Zoology Studies*, *2*, 182–184.

Englen, M. D., & Kelley, L. C. (2000). A rapid DNA isolation procedure for the identification of *Campylobacter jejuni* by the polymerase chain reaction. *Letters in Applied Microbiology*, *31*(6), 421–426.

Feng, H. Q., Wu, K. M., Ni, Y. X., Cheng, D. F., & Guo, Y. Y. (2006). Nocturnal migration of dragonflies over the Bohai Sea in northern China. Journal Economic. *Entomology*, *31*, 511–520.

Gamboa, M., Reyes, R., & Arrivillaga, J. (2008). Benthic macroinvertebrates as bioindicators of environmental health. *Boletin De Malaiologia Y Salud Ambiental*, *48*, 109–120.

Hebert P. D. N., Ratnasingham S., & DeWaard J. R. (2003). Barcoding animal life: Cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings Biological Sciences*, *270*, 96–99.

Hobson, K. A., Anderson, R. C., Soto, D. S., & Wassenaar, L. I. (2012). Isotopic evidence that dragonflies (*Pantala flavescens*) migrating through the Maldives come from the Northern Indian subcontinent. *Plos One*, *7*(12), e52594.

Martin, R., & Maynou, X. (2016). Dragonflies (Insecta: Odonata) as indicators of habitat quality in Mediterranean streams and rivers in the province of Barcelona (Catalonia, Iberian Peninsula). *International Journal of Odonatology*, *19*(3), 107–124.

Patnaik, B. B., Park, S. Y., Kang, S. W., Hwang, H. J., Wang, T. H., Park, E. B., … Lee, Y. S. (2016). Transcriptome profile of the Asian giant hornet (*Vespa mandarinia*) using Illumina HiSeq 4000 sequencing: de novo assembly, functional annotation, and discovery of SSR markers. *International Journal of Genomics*, *4*, 1–15.

Russell, R. W., May, M. L., Soltesz, K. L., & Fitzpatrick, J. W. (1998). Massive swarm migrations of dragonflies (*Pantala flavescens*) in eastern North America. *American Midland Naturalist*, *140*, 325–342.

Sun, J. T., Li, J. B., Yang, X. M., & Hong, X. Y. (2012). Development and characterization of nine polymorphic microsatellites for the small brown planthopper *Laodelphax striatellus* (Hemiptera: Delphacidae). *Genetic Molecular Research*, *11*(2), 1526–1531.

Srygley, R. B. (2003). Wind drift compensation in migrating dragonflies *Pantala flavescens* (Odonata: Libellulidae). *Journal of Insect Behavior*, *16*, 217–232.

Troast, D., Suhling, F., Jinguji, H., Sahlén, G., & Ware, J. (2016). A global population genetic study of *Pantala flavescens*. *PLoS One*, *11*(3), e0148949.

Van Oosterhout, C. V., Hutchinson, W. F., Derek, P. M. W., & Shipley, P. (2004). MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, *4*, 535–538.

Yang, X. M., Sun, J. T., Xue, X. F., Zhu, W. C., & Hong, X. Y. (2012). Development and characterization of 18 novel EST-SSRs from the western flower thrips, *Frankliniella occidentalis* (Pergande). *International Journal of Molecular Sciences*, *13*, 2863–2876.

Yeh, F., Yang, R., & Boyle, T. (1999). *POPGENE version 1.32. Microsoft Window-based Freeware for population genetic analysis*. Canada: Molecular Biology and Biotechnology Center, University of Alberta.