proceedings

MDPI

Proceedings

Analysis of Simple Sequence Repeats (SSRs) Dynamics in *Reticulitermes chinensis* †

Haroon [1E*], Yu-Xin Li [1E], Chen-Xu Ye [1], Xiao-Qin Ma [1], Jian Su[1], Xiao-Hong Su [1,2,3] and Lian-Xi Xing [1,2,3*]

[1] College of Life Sciences, Northwest University, No. 229, North Taibai Rd., Xi'an, Shaanxi Province, 710069, P. R. China
[2] Shaanxi Key Laboratory for Animal Conservation (Northwest University), Xi'an 710069, China
[3] Key Laboratory of Resource Biology and Biotechnology in Western China (Northwest University), Ministry of Education, Xi'an 710069, China
* Corresponding author: lxxing@nwu.edu.cn; Tel.: +86-153-9809-5144; haroon@stumail.nwu.edu.cn; Tel.: +86-155-2962-7930
† Presented at the 1st International Electronic Conference on Entomology (IECE 2021), 1–15 July 2021; Available online: https://iece.sciforum.net/.

Abstract: An abundance of potential inherent variation in SSR or microsatellite repeats has resulted in a valuable genetic marker in eukaryotes. We describe how SSRs in *Reticulitermes chinensis* castes are organized and abundant. Collectively we sequence 184,436 unigenes through the Trinity system, ranging from 201-43, 214 bp length from transcriptome data. Using MISA to find out SSRs in unigenes and a total of 10740 SSRs loci were identified as di-, tri-, tetra-, penta-, and hexanucleotide. Among them, trinucleotide SSRs were the most abundant in *R. chinensis* genome 2702, followed by chromosome six (tri-) 1110. Pentanucleotide repeats were found at frequency 1 from each chromosome 8, 9 and 13, respectively. The frequency of AC/GT motif was (21.91%) reported, followed by others motif (16.6%), AAG/CTT (8.49%) and AGC/CTG (8.2%). The minimum motif types were AATG/ATTC (1.27%), followed by ACG/CGT (1.32%) and AAT/ATT (1.77%). Thes abundance and inherent variations in SSRs provide valuable information for taxonomics, phylogenetic, genome mapping, and population genetic research studies. SSR-based markers have high degrees of allelic variability and codominant legacy (inheritance), and analytical ease.

Keywords: *Reticulitermes chinensis;* SSRs; microsatellites; unigenes

Citation: Haroon; Li, Y.-X.; Ye, C.-X.; Ma, X.-Q.; Su, J.; Su, X.-H.; Xing, L.X. Analysis of Simple Sequence Repeats (SSRs) Dynamics in *Reticulitermes chinensis,* in Proceedings of the 1st International Electronic Conference on Entomology, 1–15 July 2021, MDPI: Basel, Switzerland, doi:10.3390/IECE-10566

Published: 2 July 2021

1. Introduction

Termites are well-known eusocial insects classified into nine families, with 14 subfamilies, 280 types, and more than 2600 species [1]. Approximately 4,000 species are estimated (only 2,600 are taxonomically known). It should be noted that only about 10% are considered pests. Termites are significant detritivores, especially in the subtropical and tropical regions, and it is considered ecological important to recycle wood and other plant materials [2]. Termite colonies depended entirely on tissues intact or partially decayed of wood, whether living or dead or mass of plants. It is becoming an economic pest when lumber products, construction materials, forests, and other commercial products begin to be destroyed [3]. Termite has diverse feeding behavior consuming wood cellulose from tree farming, structural lumber, and causes enormous economic losses up to $40 billion worldwide/year [4,5]. In China, more than 1.8 billion RMB is used to control *Reticulitermes aculabialis, R. chinensis, R. labralis, R. speratus*, and *R. flaviceps* termites species [3]. Annual losses were estimated to 0.3 billion dollars in 2004, and approximately 217 million dollars a year cost the wood damage alone by this pest. Termites affect more than 90% of homes south of the Yangtze River. The damage and economic losses caused by termites are

steadily increasing, even with management strategies, and cause enormous economic causes to the Chinese economy annually [6,7].

Microsatellite markers or simple sequence repeats (SSRs) are the best-known and influential tools that have emerged for determining genetic divergence among populations, structure, and genetic diversity estimates, among different taxa [8-10], which is critical to the development of efficient conservation strategies. In estimating possible annihilation caused in individual natural populations, microsatellites as an advanced tool are fundamentally important. In addition to the structural situation of genetic composition, microsatellites can also assess reproductive behavior, social structure, and dispersal of endangered insect species [11,12]. However, this approach is desirable to the generation of microsatellite markers because of low costs of in silico mining and large abundance of microsatellites in different sequence resources. The negligible costs of in silico mining and many microsatellites in various sequence resources make this approach particularly attractive for creating microsatellite markers. Due to its very polymorphic nature, SSR offers a powerful taxonomic, phylogenetic, and population genetic studies tool. Polymorphism in SSRs generally considered slippage and unequal recombination of DNA polymerase [12-14]. The abundance and distribution of SSRs helped to understand the relevance of genes or genome development. The current study aimed to analyze transcriptomes of *Reticulitermes chinensis*, and distribution of various SSR loci were in silico identified from transcriptome unigenes, which could help to understand the evolutionary consequences and analysis of diversity.

2. Material and Methods

2.1. Transcriptome Sequencing

Reticulitermes chinensis colonies were collected from Qinling mountains and reared at room temperature during April 2014-May 2018, Northwest University, Xian, China (http://english.nwu.edu.cn/). An adequate amount of RNA was obtained from PKs (primary king), PQs (primary queen) [15], ergatoid (SWRK "secondary worker reproductive king and SWRQ "queen"), WM (male workers) and WF (female workers), for Illumina sequencing using TRIzol reagent and an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA) [16]. The total amount of RNA was extracted and stored at −80°C for further experiments. Later, NEB-Next Prep-Kit for Illumina (NEB) sequences [17] was used to reverted cDNA according to the manufacturer protocol. A spectrophotometer was used for the quantity and quality of RNA and cDNA. Three technical and five biological replicates were pooled for each caste and stored at −20 °C for further experiments.

2.2. Simple sequence repeat (SSR) analysis

We annotated sequence reads through Illumina HiSeqTM 4000 platform and Trinity system (trinityrnaseq r2012–04–27), ranging from various bp length at transcriptome data [15]. The estimated GC performance evaluation and assessment were estimated as exemplary quality assemblies for all unigenes in the transcriptome data utilized MISA (http:/pgrc.ipk-gatersleben.de/misa/) to find out SSR in unigenes [18]. The predicted and identified SSRs were di-, tri-, tetra-, penta-, and hexanucleotide motifs with amplification products containing repeat units ≥6 times.

3. Results

We annotated a total of 103589264 (Figure 1) sequence reads (averages length of 561 bp) (Table S1) through Illumina HiSeqTM 4000 platform and 184,436 unigenes (184413) (Figure 2) viva Trinity system (trinityrnaseq r2012–04–27), ranging from 201 to 43,214 bp length from transcriptome data (7G) (Table S2). The estimated GC performance evaluation and assessment were estimated 43.02%, suggesting suitable quality assemblies. We use MISA (http:/pgrc.ipk-gatersleben.de/misa/) to search all unigenes in the transcriptome and find the SSR in unigenes. A total of 10740 SSRs were identified from the transcriptome

data of *R. chinensis* (Table S3). The SSRs motif was identified as di-, tri-, tetra-, penta-, and hexanucleotide. Among them, the fifth chromosome (tri-) possessed the highest number of SSRs (2702) followed by chromosome six (tri-) (1110), while the least number of SSRs (penta- 1) were reported from chromosomes 8, 9 and 13 (Figure 3). To understand more about the frequency and different types of motif AC/GT put (21.91%), others motif (16.6%), followed by AAG/CTT (8.49%) and AGC/CTG (8.2%) (Table S4). The minimum number of motif types are AATG/ATTC (1.27%) followed by ACG/CGT (1.32%) and AAT/ATT (1.77%) (Figure 4).

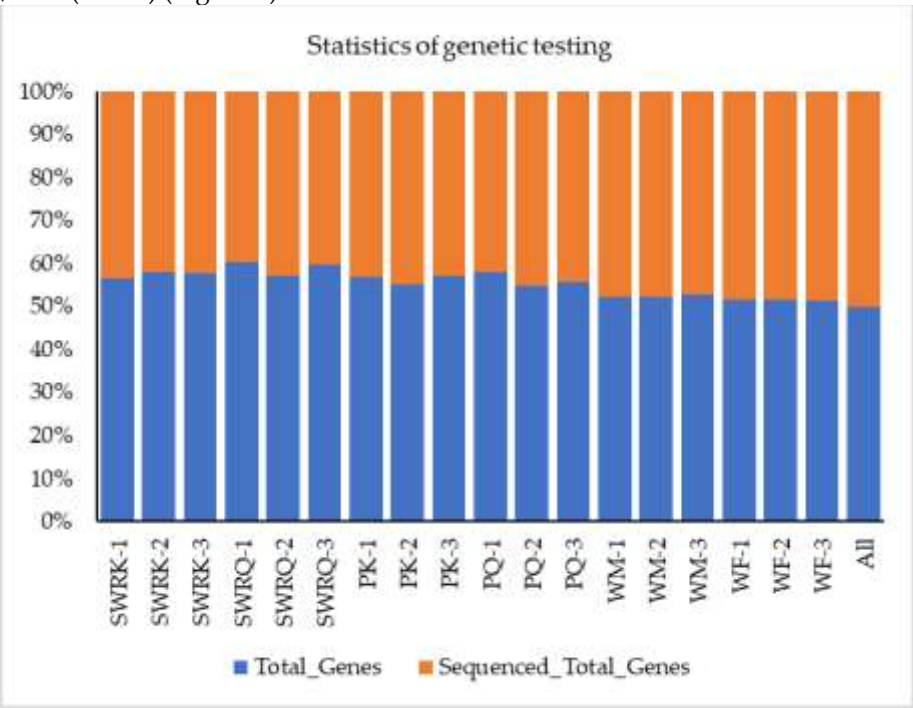

Figure 1. The total number of genes and sequence genes from *R. chinensis*.

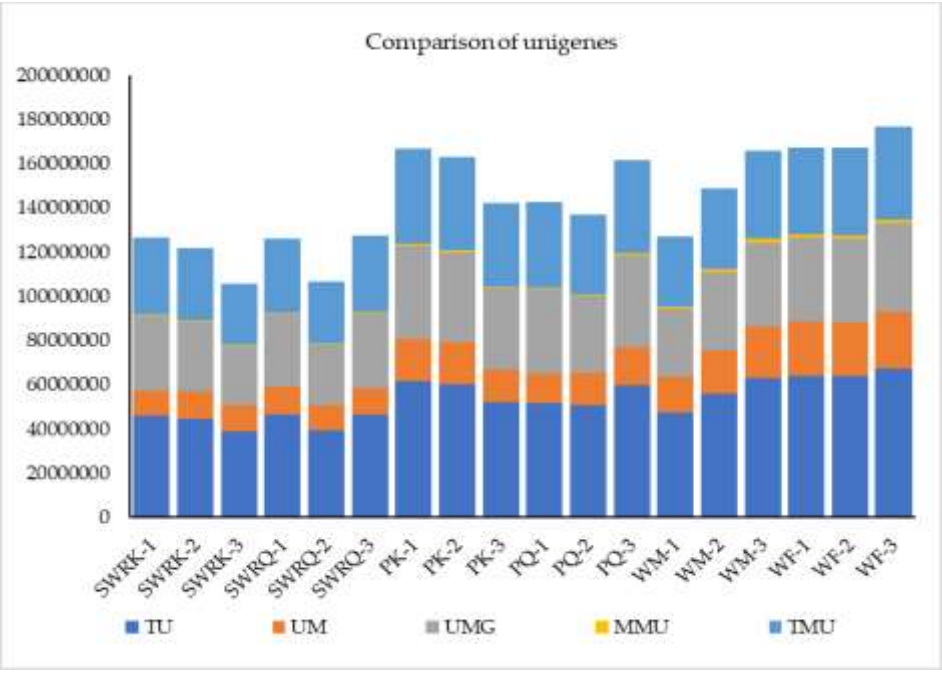

Figure 2. Total number of unigenes (TU), unmapped unigenes (UU), unique mapped unigenes (UMU), multiple mapped unigenes (MMU) and total mapped unigenes (TMU) in different caste of *R. chinensis*.

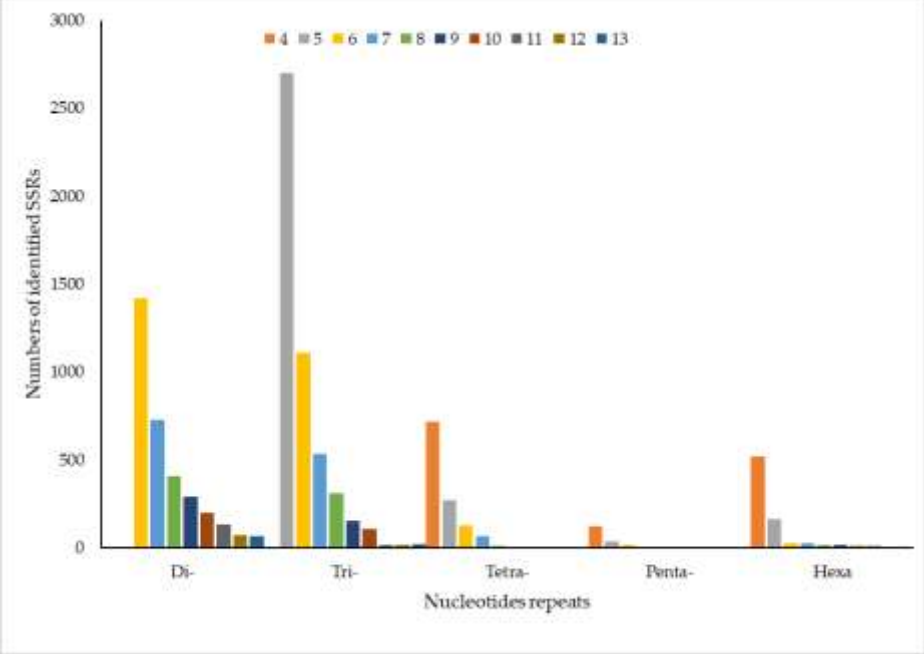

Figure 3. The number of identified SSRs and motif types di-, tri-, tetra- penta- and hexanucleotides repeats.

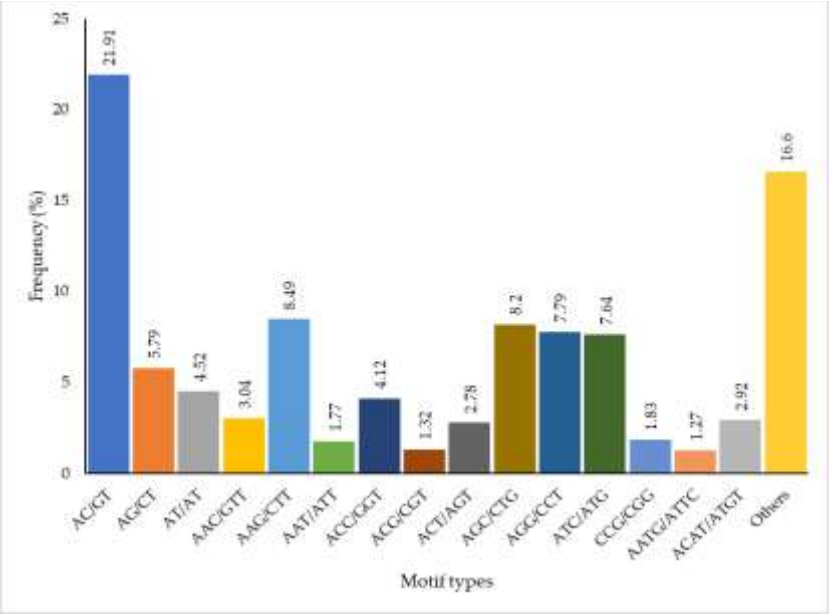

Figure 4. Frequency distribution of SSRs according to motif and sequence types.

4. Discussions

In research on genetic diversity, genetic mapping, comparative genomics, and marker-assisted selection breeding, polymorphic SSRs play an essential role. Transcriptomics provides a rich discovery source of SSR because it creates many sequences [19]. The transcriptomes analysis of *R. chinensis* was reported annotated 103589264 clean assembled sequence reads (561 average bp length) via the Illumina HiSeqTM 4000 platform. A total of 184,436 unigenes were assembled (201 to 43,214bp) from accurate transcriptome data via the Trinity system (7G). We estimated that 43.02% of GC performance evaluation and assessment suggests that quality and development in sequence are appropriate for homogenous studies.

The transcriptome-SSR frequency has several factors. The first factor of the transcriptome-SSR frequency, e.g., repeat length threshold and several repetition motifs, are called microsatellite criteria [20,21]. Most studies excluded repeat motifs of mononucleotides, as these could result from sequencing errors. In some studies, the number of dinucleotide reproductions based on three repeat units, while others are less [22-24]. Moreover, SSRs were mainly identified from unigenes above 1000 bp, reducing the frequency to some extent. Second, SSRs frequency may also be affected by genome structure or composition [25]. For instance, *Reticulitermes* small genome size was reported as the cause of the high frequency of transcriptomes-SSR sequences [15,26]. Finally, the various SSR software can also affect the frequency of SSRs loci. In addition to genomic stability, the effects of excess numbers of short-iterate repeats might significantly develop genomic characteristics, such as codon usage [13]. Microsatellites generally show an increase in repeat length [27] to decrease in abundance. However, there were also reports of more than expected long microsatellite replicates [28]. For the development of specific genome markers, the microsatellites identified in this study could be used in evolutionary studies in *Reticulitermes* castes.

5. Conclusion

In addition to the advanced computational field, the comprehensive review has shown development in identifying an SSRs motif in the *R. chinensis*. With the development of sequencing methodologies, massive sequence data for identification and quantification of SSRs motif in the *R. chinensis* genome has been generated in transcriptome data. The results of reliable next-generation sequencing (NGS) technologies with fast and cost-effective methods will continue to provide additional information on SSR markers in insects. Thus, in the new-model *R. chinensis* field, including evolutionary consequences and the analysis of diversity to improve genetically, the NGS area added advantage in the SSR marker development. The overall conclusion of this study is to provide valuable information for the SSR motif of *R. chinensis* and future genetic or genomic studies in economically significant (detritivores the plants and buildings) valuable.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Table S1: Total of annotated genes with sequenced genes in different castes of *R. chinensis*. Table S2: The total number of unigenes, unmapped, unique, multiple, and total mapped unigenes in *R. chinensis* castes. Table S3: Distribution of SSRs in different chromosomes identified from transcriptomes of *R. chinensis.* Table S4: Motif and their frequency (%) from transcriptomes data in *R. chinensis*.

Author Contributions: Conceptualization, H. and L.Y.X; Data curation, H., L.Y.X. and Y.C.; Formal analysis, Q.X.M. and J.S; Investigation, H., L.Y.X. and X.L.X.; Methodology, H. and L.Y.X.; Resources, H. and X.L.X.; Software, Y.C.; L.Q. and Q.X.M; Validation, H.; Visualization, S.X.H.; Writing—original draft, H., X.L.X. and L.Y.X.; Writing—review and editing, H., S.X.H. and X.L.X. All authors have read and agreed to publish the current version of the manuscript.

Funding: Financial support was provided by the National Natural Science Foundation of China (31870389) and Key Laboratory of Resource Biology and Biotechnology in Western China, Ministry of Education (ZSK2017002).

Informed Consent Statement: Not applicable.

Data Availability Statement: All data and transcriptomic sequences will be available under BioProject accession number PRJNA592596 at the NCBI. The corresponding author will answer any reasonable requests.

Conflicts of Interest: The authors declare no conflict of interest.

Ethical approval: *Reticulitermes chinensis* species is a common species; therefore, no ethical approval required.

Table S1. Total of annotated genes with sequenced genes in different castes of *R. chinensis*.

S.No	Sample	Total_Genes	Sequenced_Total_Genes
1	SWRK-1	184436	140939
2	SWRK-2	184436	134023
3	SWRK-3	184436	135852
4	SWRQ-1	184436	121130
5	SWRQ-2	184436	138925
6	SWRQ-3	184436	123895
7	PK-1	184436	140368
8	PK-2	184436	151112
9	PK-3	184436	138891
10	PQ-1	184436	133861
11	PQ-2	184436	151322
12	PQ-3	184436	147181
13	WM-1	184436	169058
14	WM-2	184436	169392
15	WM-3	184436	164677
16	WF-1	184436	172488
18	WF-2	184436	173035
19	WF-3	184436	174693
20	All	184436	184413

Table S2. The total number of unigenes, unmapped, unique, multiple, and total mapped unigenes caste of *R. chinensis*.

Sample	TU	UU	UMU	MMU	TMU
SWRK-1	46038870	11586945	33782988	668937	34451925
SWRK-2	44657610	12356581	31813518	487511	32301029
SWRK-3	39095042	11595066	26918618	581358	27499976
SWRQ-1	46331166	12989163	32848349	493654	33342003
SWRQ-2	39376280	11594529	27277503	504248	27781751
SWRQ-3	46545376	11960048	34109970	475358	34585328
PK-1	61875302	18635074	42503761	736467	43240228
PK-2	60491486	18546087	41141899	803500	41945399
PK-3	52211218	14661080	36965780	584358	37550138
PQ-1	51946682	13105283	38302564	538835	38841399
PQ-2	50594180	14769492	35197605	627083	35824688
PQ-3	59718954	17357458	41636743	724753	42361496
WM-1	47709742	15887884	30853450	968408	31821858
WM-2	56089922	19493776	35388793	1207353	36596146
WM-3	63032988	23340170	38047161	1645657	39692818
WF-1	64019360	24704908	37664936	1649516	39314452

WF-2	63868690	24231782	37948858	1688050	39636908
WF-3	67410876	25462040	40269037	1679799	41948836

TU (total unigenes), UU (unmapped unigenes), UMU (unique mapped unigenes), MMU (multiple mapped unigenes) and TMU (total mapped unigenes) in diffirent caste of *R. chinensis*.

Table S3. Distribution of SSRs in different chromosomes identified from transcriptomes of *R. chinensis*.

Number of repeat units	Di-	Tri-	Tetra-	Penta-	Hexa
4	0	0	716	124	517
5	0	2702	271	34	164
6	1420	1110	128	17	24
7	726	537	67	3	26
8	410	313	10	1	17
9	291	154	4	1	15
10	201	106	5	2	11
11	131	15	6	0	10
12	72	14	3	0	4
13	67	20	4	1	4

Table S4. Motif and their frequency (%) from transcriptomes data in *R. chinensis*.

	Motif	Frequency (%)
1	AC/GT	21.91
2	AG/CT	5.79
3	AT/AT	4.52
4	AAC/GTT	3.04
5	AAG/CTT	8.49
6	AAT/ATT	1.77
7	ACC/GGT	4.12
8	ACG/CGT	1.32
9	ACT/AGT	2.78
10	AGC/CTG	8.2
11	AGG/CCT	7.79
12	ATC/ATG	7.64
13	CCG/CGG	1.83
14	AATG/ATTC	1.27
15	ACAT/ATGT	2.92
16	Others	16.6

References

1. Pranesh, M.; Harini, B.J.T.E. Diversity and Distribution Pattern of Termites in Relation with Human Interference: A Study at Jnanabharathi Campus, Bangalore, India. **2015**, *9*, 671-676.
2. Ganesan, T.; Rajarajan, D.; Kumaresan, V.J.J.o.M.; Pathology, P. Microorganisms in termite mound soil and surface soil adjacent to mound. **2010**, *40*, 408.
3. Li, H.; Xu, Z.; Deng, T.; Chen, L.; Li, J.; Wei, J.; Mo, J.J.S. Species of termites (Isoptera) attacking trees in China. **2010**, *56*, 109-120.
4. Abd-Elkareem, E.; Fouad, H.J.E.J.o.A.; Studies, R. Termites, their role in the damaged mud buildings, and prevention methods: application on the ruins of the White Monastery, Sohag, Egypt. **2016**, *6*, 85.
5. Adam, R.; Mitchell, J.J.I.s. Energetics and development of incipient colonies of the harvester termite, Trinervitermes trinervoides (Sjöstedt)(Termitidae, Nasutitermitinae). **2009**, *56*, 21-27.
6. Ahmad, F.; Fouad, H.; Liang, S.Y.; Hu, Y.; Mo, J.C.J.I.s. Termites and Chinese agricultural system: applications and advances in integrated termite management and chemical control. **2019**.
7. Jasmi, A.H. Termite Assembledges, Niche Partitioning and Infestation in Natural, Disturbed and Plantation Forest in Northern Peninsular Malaysia. Universiti Sains Malaysia, 2016.
8. de Jesús May-Itzá, W.; Quezada-Euán, J.J.G.; Medina, L.A.M.; Enríquez, E.; De la Rúa, P.J.C.G. Morphometric and genetic differentiation in isolated populations of the endangered Mesoamerican stingless bee Melipona yucatanica (Hymenoptera: Apoidea) suggest the existence of a two species complex. **2010**, *11*, 2079-2084.
9. Lopes, D.M.; Campos, L.A.d.O.; Salomão, T.M.F.; Tavares, M.G.J.G.; biology, m. Comparative study on the use of specific and heterologous microsatellite primers in the stingless bees Melipona rufiventris and M. mondury (Hymenoptera, Apidae). **2010**, *33*, 390-393.
10. Tavares, M.G.; Dias, L.A.d.S.; Borges, A.A.; Lopes, D.M.; Busse, A.H.P.; Costa, R.G.; Salomão, T.M.F.; Campos, L.A.d.O.J.G.; Biology, M. Genetic divergence between populations of the stingless bee uruçu amarela (Melipona rufiventris group, Hymenoptera, Meliponini): Is there a new Melipona species in the Brazilian state of Minas Gerais? **2007**, *30*, 667-675.
11. Beaumont, M.A.J.M.E.; applications. Microsatellite in conservation genetics. **1999**, 165-182.
12. Negreiros, A.B.; Silva, G.R.; Oliveira, F.A.; Resende, H.C.; Fernandes-Salomão, T.M.; Maggioni, R.; Pereira, F.M.; Souza, B.A.; Lopes, M.T.; Diniz, F.M.J.I. Microsatellite Marker Discovery in the Stingless Bee Uruçu-Amarela (Melipona rufiventris Group, Hymenoptera, Meliponini) for Population Genetic Analysis. **2019**, *10*, 450.
13. Singh, R.; Sheoran, S.; Sharma, P.; Chatrath, R.J.B. Analysis of simple sequence repeats (SSRs) dynamics in fungus Fusarium graminearum. **2011**, *5*, 402.
14. Yin, H.; Ben-Abu, Y.; Wang, H.; Li, A.; Nevo, E.; Kong, L.J.P.o. Natural selection causes adaptive genetic resistance in wild emmer wheat against powdery mildew at "evolution canyon" microsite, mt. Carmel, Israel. **2015**, *10*, e0122344.
15. Haroon; Ma, X.M.; Li, Y.X.; Zhang, H.X.; Liu, Q.; Su, X.H.; Xing, L.X. Transcriptomic evidence that insulin signalling pathway regulates the ageing of subterranean termite castes. *Scientific Reports* **2020**, *10*, 1-13, doi:10.1038/s41598-020-64890-9.
16. Su, X.; Liu, H.; Yang, X.; Chen, J.; Zhang, H.; Xing, L.; Zhang, X.J.S.r. Characterization of the transcriptomes and cuticular protein gene expression of alate adult, brachypterous neotenic and adultoid reproductives of Reticulitermes labralis. **2016**, *6*, 1-9.
17. Jia, Z.-Q.; Liu, D.; Peng, Y.-C.; Han, Z.-J.; Zhao, C.-Q.; Tang, T.J.B.g. Identification of transcriptome and fluralaner responsive genes in the common cutworm Spodoptera litura Fabricius, based on RNA-seq. **2020**, *21*, 1-11.
18. Beier, S.; Thiel, T.; Münch, T.; Scholz, U.; Mascher, M.J.B. MISA-web: a web server for microsatellite prediction. **2017**, *33*, 2583-2585.
19. Luikart, G.; England, P.R.; Tallmon, D.; Jordan, S.; Taberlet, P.J.N.r.g. The power and promise of population genomics: from genotyping to genome typing. **2003**, *4*, 981-994.
20. Kumpatla, S.P.; Mukhopadhyay, S.J.G. Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. **2005**, *48*, 985-998.
21. Sharifova, S.; Mehdiyeva, S.; Shahmuradov, I.J.A.J.o.A.S. Mining and survey of simple sequence repeats in expressed sequence tags of tomato species. **2016**, *15*, 65.
22. Cloutier, S.; Niu, Z.; Datla, R.; Duguid, S.J.T.; Genetics, A. Development and analysis of EST-SSRs for flax (Linum usitatissimum L.). **2009**, *119*, 53-63.
23. Huang, D.; Zhang, Y.; Jin, M.; Li, H.; Song, Z.; Wang, Y.; Chen, J.J.M.E.R. Characterization and high cross-species transferability of microsatellite markers from the floral transcriptome of Aspidistra saxicola (Asparagaceae). **2014**, *14*, 569-577.
24. Jhanwar, S.; Priya, P.; Garg, R.; Parida, S.K.; Tyagi, A.K.; Jain, M.J.P.B.J. Transcriptome sequencing of wild chickpea as a rich resource for marker development. **2012**, *10*, 690-702.
25. Tóth, G.; Gáspári, Z.; Jurka, J.J.G.r. Microsatellites in different eukaryotic genomes: survey and analysis. **2000**, *10*, 967-981.
26. Varshney, R.K.; Thiel, T.; Stein, N.; Langridge, P.; Graner, A.J.C.; Letters, M.B. In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. **2002**, *7*, 537-546.
27. Khajuria, Y.P.; Saxena, M.S.; Gaur, R.; Chattopadhyay, D.; Jain, M.; Parida, S.K.; Bhatia, S.J.P.O. Development and integration of genome-wide polymorphic microsatellite markers onto a reference linkage map for constructing a high-density genetic map of chickpea. **2015**, *10*, e0125583.
28. Qi, W.-H.; Jiang, X.-M.; Du, L.-M.; Xiao, G.-S.; Hu, T.-Z.; Yue, B.-S.; Quan, Q.-M.J.P.O. Genome-wide survey and analysis of microsatellite sequences in bovid species. **2015**, *10*, e0133667.