Available online freely at www.isisn.org

Bioscience Research

Print ISSN: 1811-9506 Online ISSN: 2218-3973

Journal by Innovative Scientific Information & Services Network

RESEARCH ARTICLE BIOSCIENCE RESEARCH, 2020 17(2): 753-758. **OPEN ACCESS**

Identification of SNPs and SSRs markers in *Lantana camara* L. An important medicinal and aromatic plant

Muzammil Shah[1*], Hesham F. Alharby[1], Khalid Rehman Hakeem[1], Niaz Ali[2], Inayat Ur Rahman[2,3], and Mohd Munawar[1]

[1]Department of Biological Sciences, Faculty of Science, King Abdulaziz University, Jeddah-21589, **Saudi Arabia**.
[2]Department of Botany, Hazara University, Mansehra-21300, KPK, **Pakistan**
[3]William L. Brown Center, Missouri Botanical Garden, P.O. Box 299, St. Louis, MO 63166-0299, **United State of America**

***Correspondence:** muzammilshah100@outlook.com **Received 05-04-2020, Revised: 14-05-2020, Accepted: 17-05-2020 e-Published: 29-05-2020**

Lantana camara L. is a medicinal and essential oil producing plant belonging to family Verbenaceae. It is used as a medication for the treatment of various diseases and also contains important bioactive compounds i.e. steroids, flavonoids and phenylpropanoid glycosides etc. Nonetheless, very little is known about their genomic resources which might be the reason of hindering its molecular studies. In the present study we used high throughput RNA sequencing to identify SNP and SSR markers within the leaf and root transcriptomes of *L.camara*. Interestingly, we identified a total of 9,669 SNPs and 18,171 SSRs in the leaf while 9,226 SNPs and 65963 SSRs in the root transcriptomes of *L. camara.* Thus, this investigation not only provides valuable insights into molecular characterization but also these novel SSRs may be applied to genetic linkage mapping, genetic polymorphism as well as functional gene mining studies in *L. camara* and its closely related species.

Keywords: Medicinal plant, *Lantana camara*, SSRs, SNPs, RNA-Seq, Transcriptome.

INTRODUCTION

In the present era Next generation sequencing technology (NGS) has transformed, the field of life science research by increasing the potential of genetic improvement in crops as well as medicinal plants and tremendously facilitating genomic and transcriptomic studies (Grabherr et al., 2011; Martin & Wang, 2011; Mora-Ortiz et al., 2016). RNA-Seq technology has made the identification of genes, annotations, expression analysis and SSRs/SNPs mining very easy and feasible (Chabikwa et al., 2020; C. Wang et al., 2019). From the last few decades, DNA markers including random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), restriction fragment length polymorphism (RFLP), inter-simple sequence repeat (ISSR), single nucleotide polymorphism (SNP) and simple sequence repeat (SSR) have been accepted across the globe (Taheri et al., 2019; Yunsheng Wang et al., 2019). Among these markers SSRs and SNPs are known as versatile markers (Ying Wang et al., 2017). SSRs are potent markers extensively used due to its important features of high polymorphism, genome abundance, small locus size and even distribution (Thudi et al., 2011; Yunsheng Wang et al., 2019). However, SNPs are makers resulting by point mutation of genomic DNA and they are ubiquitous, nature abundant and bi-allelic (Mammadov et al., 2012; Nadeem et al., 2018).

Currently, high-throughput sequencing is becoming a rapid and feasible platform for the detection of SSRs and SNPs in various model and

non-model plants (Taheri et al., 2018; Yu et al., 2020). For instance, if there is an availability of the reference genome then thousands of SNPs can be identified through re-sequencing (Xu & Bai, 2015). However, in case of the unavailability of reference genome one can take the advantage of simplified genome parallel sequencing (RAD-tag, SLAF) and can produce millions of SNPs (Baird et al., 2008; Sun et al., 2013). Nowadays, RNA-Seq (transcriptome sequencing) has emerged as an effective and common tool for the identification of SSRs and SNPs on large scale (Jia et al., 2019). The SSRs and SNPs generated through mRNA-Seq are especially from genic regions and such markers can be used as an influential tool by molecular biologist for the identification of causative mutations (Codina-Solà et al., 2015; Djari et al., 2013). Today, the importance of these markers can be realized from the fact that these markers are extensively used in multiplex research areas i.e. ecology and evolution (De Wit & Palumbi, 2013), modern breeding (Duarte et al., 2014), association analysis (Xie et al., 2016), population biology (Gramazio et al., 2018).

Lantana camara L. belonging to family Verbenaceae is an evergreen shrub, native to the Neotropics and grown worldwide for its medicinal and ornamental value (Ghisalberti, 2000). In spite, much is known about the phytochemistry, toxicology and medicinal properties of *L. camara* (Joy et al., 2012; Khan et al., 2016) no reports are available about the genomic architecture of the plant. Recently, we have reported the complete chloroplast genome of *L.camara* (Yaradua & Shah, 2020). Here in the present study we took an advantage of high-throughput sequencing technology and sequenced leaf and root transcriptomes of *L.camara* with an aim to enrich the genetic information and detect SSRs and SNPs from the unigenes of leaf and root transcriptome of *L.camara*.

MATERIALS AND METHODS

Plant material and RNA quantification

Plants were grown in the laboratory of plant physiology King Abdulaziz University, Jeddah, Saudi Arabia under 28^0C/22^0C day and night temperature in a semi-controlled green house. Fresh leaf and root tissues were taken and immediately frozen in liquid nitrogen. RNA isolation was carried out using RNeasy Plus Mini Kit (Qiagen, Cat. No: 74134). RNA degradation and contamination was monitored on 1% agarose gel. Purity of RNA was checked on NanoPhotometer® spectrophotometer (IMPLEN, CA, USA). Quantification and integrity of RNA was assessed using the RNA Nano 6000 Assay Kit (Agilent Technologies, CA, USA).

Library preparation, Clustering, quality control and transcriptome assembly

Library preparation was performed following (Song et al., 2019). Cluster generation system (PE Cluster Kit cBot-HS Illumina) was used to perform clustering of the index coded samples following manufacturer's instructions and the library preparations were sequenced on an Illumina platform. Transcriptome assembly was performed using Trinity software with min_kmer_cov set to 2, and all other parameters with default setting (Haas et al., 2013).

SNP calling, SSR detection and primer design

SNP calling was performed using GATK3 software with standard filter method (Van der Auwera et al., 2013). MISA was used to identify SSRs and primers for each SSR were designed using Primer3 (Untergasser et al., 2012, p. 3).

RESULTS AND DISCUSSION

Detection of Single Nucleotide Polymorphism (SNPs) within leaf and root transcriptomes

SNPs are potent genetic markers having a broad range of applications in many fields including genetic diversity assessment, genetic map construction, breeding and cultivar identification (Broders et al., 2011; Chagné et al., 2008; Wu et al., 2008). Currently, NGS technology is contributing a lot in the identification of SNPs in many medicinal plants. For instance, a total of 128,921 SNPs were identified by using high-throughput sequencing technology in medicinal plant *Gastrodia elata* (Yunsheng Wang et al., 2019) and approximately 3,010,256 SNPs were identified in *platanus acerifolia* (Li et al., 2019). The importance of NGS technologies has been shown in characterizing SNPs in various plant studies.

These SNPs may then be implemented to generate various types of genotyping tools and may be used for breeding purposes (Graham et al., 2010). Due to their robustness and practical importance of SNP makers, we extended our study and identified a total of 9,669 SNPs in the leaf and 9,226 SNPs in the root transcriptomes of *L.camara* (Fig. 1, Supplementary File 1).

Identification of Simple sequence repeats (SSRs) in the leaf transcriptome

Simple sequence repeats (SSR) are 1-6 base pair repetitive nucleotide sequences present in tandem. SSR sequences are dispersed throughout the genomes of many of the known organisms ranging from viruses to eukaryotes. MISA (v1.0, default parameters; minimum number of repeats of each unit size is: 1-10; 2-6; 3-5; 4-5; 5-5; 6-5) was used for SSR detection in these unigenes. We examined a total of 72,877 sequences for the identification of SSRs in leaf transcriptome and identified 18,171 SSRs distributed in 14,829 unigenes (24.93 % of the total unigenes). Of all the SSRs containing uni genes 2,649 contained more than one SSR and 918 SSRs were present in compound formation (Table 1). Amid all these SSRs mono-nucleotide repeats were the highest (9,065), followed by di-nucleotide repeats (6,024), tri-nucleotide (2,906), tetra-nucleotide (125) and hexa-nucleotide (35). The lowest of all was penta-nucleotide repeats (16) (Fig. 2). In order to facilitate researchers to use these SSR markers, Primer3 (2.3.5 version, the default parameters) was used for primer design. More detail can be found in Supplementary File 2. These novel SSRs may be used for genetic linkage mapping, genetic polymorphism as well as functional gene mining studies in L.camara and its closely related species.

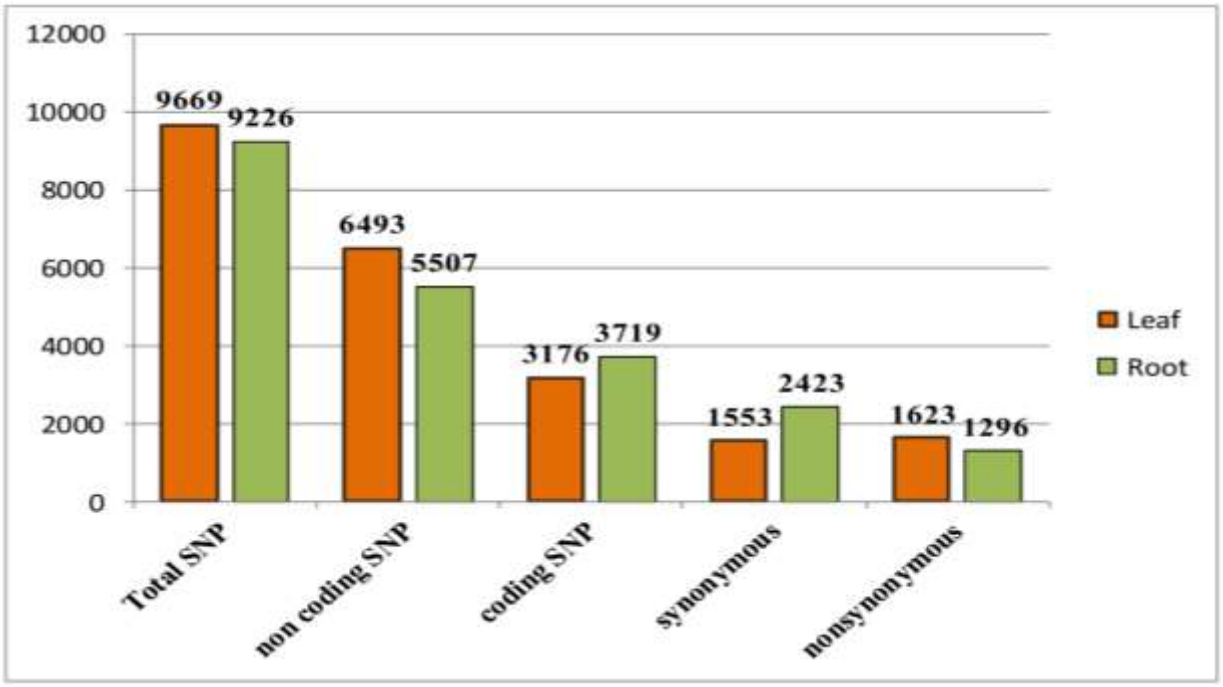

Figure1: Summary of SNPs

Table 1: SSRs markers developed in *L. camara* leaf and root transcriptomes

SSR Information	Number	
	Leaf	Root
Total number of sequences examined	72877	513985
Total size of examined sequences (bp)	83812099	336370588
Total number of identified SSRs	18171	65963
Number of SSR containing sequences	14829	52911
Number of sequences contain more than 1 SSR	2649	10146
No of SSRs present in compound formation	918	6490

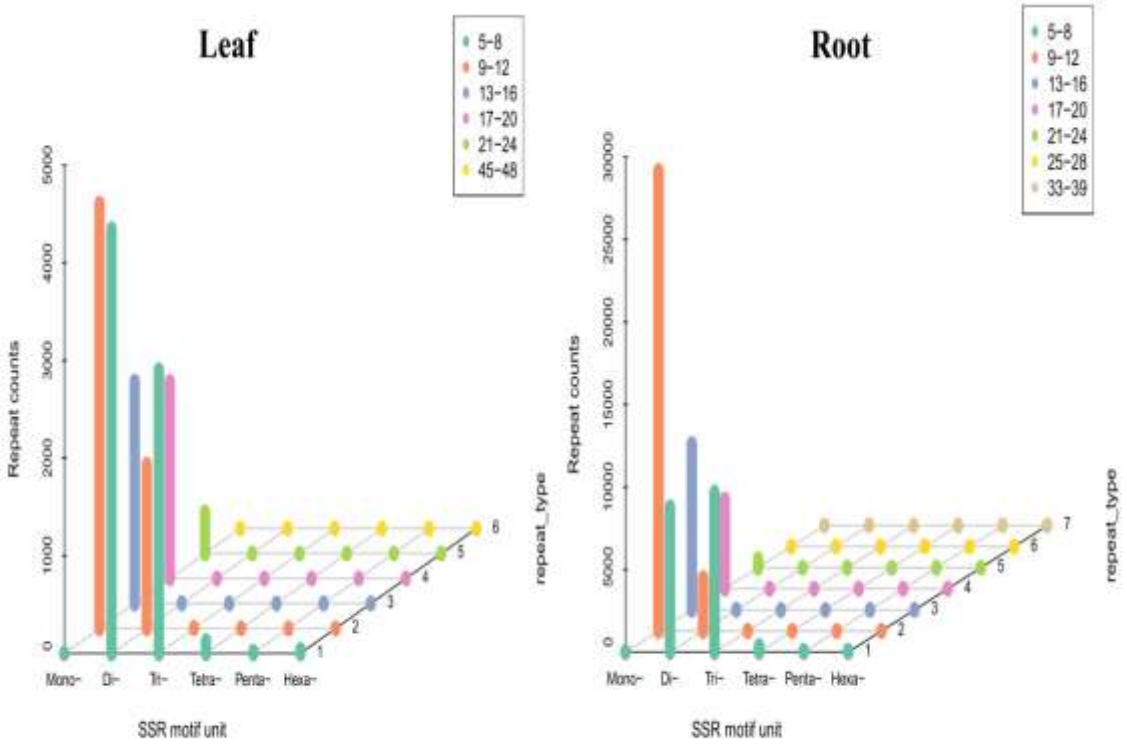

Figure 2: Distribution of SSRs motifs within leaf and root transcriptome of L.camara

Identification of Simple sequence repeats (SSRs) in the root transcriptome

For the root transcriptome SSRs analysis we examined a total of 513,985 sequences for the identification of SSRs and identified 65,963 SSRs distributed in 52,911 unigenes (10.29 % of the total unigenes). Of all the SSRs containing unigenes 2,649 contained more than one SSR and 918 SSRs were present in compound formation (Table 1). Within all these SSRs mono-nucleotide repeats were the highest (43,829), followed by di-nucleotide repeats (11961), tri-nucleotide (9,659), tetra-nucleotide (367) and hexa-nucleotide (87). The lowest of all was penta-nucleotide repeats (60) (Fig. 2). In order to facilitate future researchers in *L. camara*, primers were designed for all these SSRs using Primer3 (2.3.5 version, the default parameters).

CONCLUSION

Lantana camara L. is an important medicinal and aromatic plant. There is a lack of information in the public databases about molecular studies of *L.camara*. As a step forward, by taking an advantage of Next Generation Sequencing we identified a total of total of 9,669 and 9,226 SNPs in the leaf and root transcriptomes of *L.camara* respectively. Furthermore, we also identified 18,171 SSRs in the leaf and 65,963 SSRs in the root transcriptomes of *L.camara.* This study will serve as a base for further molecular studies as well as a tool to elevate the conservation genetics and molecular breeding of *L.camara*.

CONFLICT OF INTEREST

The authors declared that present study was performed in absence of any conflict of interest.

ACKNOWLEGEMENT

First Author is thankful to Deanship of Graduate studies at King Abdulaziz University for providing PhD Scholarship.

AUTHOR CONTRIBUTIONS

MS performed the experiments, MS collected the data, MS analyzed the data. MS wrote and drafted the manuscript, NA and IUR helped in editing of the manuscript. HFA and KRH supervised the project.

REFERENCES

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, & Johnson EA, 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PloS One, 3(10).

Broders KD, Woeste KE, SanMIGUEL PJ, Westerman RP, & Boland GJ, 2011. Discovery of single-nucleotide polymorphisms (SNPs) in the uncharacterized genome of the ascomycete Ophiognomonia clavigignenti-juglandacearum from 454 sequence data. Mole. Ecol. Res, 11(4), 693–702. https://doi.org/10.1111/j.1755-0998.2011.02998.x

Chabikwa TG, Barbier FF, Tanurdzic M, & Beveridge CA, 2020. De novo transcriptome assembly and annotation for gene discovery in avocado, macadamia and mango. Scientific Data, 7(1), 1–7.

Chagné D, Gasic K, Crowhurst RN, Han Y, Bassett HC, Bowatte DR, Lawrence TJ, Rikkerink EH, Gardiner SE, & Korban SS, 2008. Development of a set of SNP markers present in expressed genes of the apple. Genomics, 92(5), 353–358.

Codina-Solà M, Rodríguez-Santiago B, Homs A, Santoyo J, Rigau M, Aznar-Laín G, Del Campo M, Gener B, Gabau E, & Botella MP, 2015. Integrated analysis of whole-exome sequencing and transcriptome profiling in males with autism spectrum disorders. Molecular Autism, 6(1), 21.

De Wit P, & Palumbi SR, 2013. Transcriptome-wide polymorphisms of red abalone (H aliotis rufescens) reveal patterns of gene flow and local adaptation. Mole. Ecol, 22(11), 2884–2897.

Djari A, Esquerré D, Weiss B, Martins F, Meersseman C, Boussaha M, Klopp C, & Rocha D, 2013. Gene-based single nucleotide polymorphism discovery in bovine muscle using next-generation transcriptomic sequencing. BMC Genomics, 14(1), 307.

Duarte J, Rivière N, Baranger A, Aubert G, Burstin J, Cornet L, Lavaud C, Lejeune-Hénaut I, Martinant JP, & Pichon JP, 2014. Transcriptome sequencing for high throughput SNP development and genetic mapping in Pea. BMC Genomics, 15(1), 126.

Ghisalberti EL, 2000. Lantana camara L.(verbenaceae). Fitoterapia, 71(5), 467–486.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Regev A, 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotech, 29(7), 644–652. https://doi.org/10.1038/nbt.1883

Graham IA, Besser K, Blumer S, Branigan CA, Czechowski T, Elias L, Guterman I, Harvey, D, Isaac PG, & Khan AM, 2010. The genetic map of Artemisia annua L. identifies loci affecting yield of the antimalarial drug artemisinin. Science, 327(5963), 328–331.

Gramazio P, Plesa IM, Truta AM, Sestras AF, Vilanova S, Plazas M, Vicente O, Boscaiu M, Prohens J, & Sestras RE, 2018. Highly informative SSR genotyping reveals large genetic diversity and limited differentiation in European larch (Larix decidua) populations from Romania. TURK J AGRIC FOR, 42(3), 165–175.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Regev A, 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protocols, 8(8),1494–1512.

Jia Y, Bai JQ, Liu ML, Jiang ZF, Wu Y, Fang MF, & Li ZH, 2019. Transcriptome analysis of the endangered Notopterygium incisum: Cold-tolerance gene discovery and identification of EST-SSR and SNP markers. Plant Diversity, 41(1), 1–6.

Joy JM, Vamsi S, Satish C, & Nagaveni K, 2012. Lantana camara Linn: A review. Int. J Phytothera, 2(2), 66–73.

Khan M, Mahmood A, & Alkhathlan HZ, 2016. Characterization of leaves and flowers volatile constituents of Lantana camara growing in central region of Saudi Arabia.

Arab. J. Chem, 9(6), 764–774.

Li F, Wu C, Gao M, Jiao M, Qu C, Gonzalez-Uriarte A, & Luo C, 2019. Transcriptome sequencing, molecular markers, and transcription factor discovery of Platanus acerifolia in the presence of Corythucha ciliata. Scientific Data, 6(1), 1–7.

Mammadov J, Aggarwal R, Buyyarapu R, & Kumpatla S, 2012. SNP markers and their impact on plant breeding. Inter. J. Plant Genom, 2012.

Martin JA, & Wang Z, 2011. Next-generation transcriptome assembly. Nat. Rev. Gen, 12(10), 671–682.

Mora-Ortiz M, Swain MT, Vickers MJ, Hegarty MJ, Kelly R, Smith LM, & Skøt L, 2016. De-novo transcriptome assembly for gene identification, analysis, annotation, and molecular marker discovery in Onobrychis viciifolia. BMC Genomics, 17(1), 756.

Nadeem MA, Nawaz MA, Shahid MQ, Doğan Y, Comertpay G, Yıldız M, Hatipoğlu R, Ahmad F, Alsaleh A, & Labhane N, 2018. DNA molecular markers in plant breeding: Current status and recent advancements in genomic selection and genome editing. Biotechnol. Biotechnol. Equip, 32(2), 261–285.

Song Y, Chen S, Wang X, Zhang R, Tu L, Hu T, Liu X, Zhang Y, Huang L, & Gao W, 2019. A novel strategy to enhance terpenoids production using cambial meristematic cells of Tripterygium wilfordii Hook. F. Plant Methods, 15(1), 129.

Sun X, Liu D, Zhang X, Li W, Liu H, Hong W, Jiang C, Guan N, Ma C, & Zeng H, 2013. SLAF-seq: An efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing. PloS One, 8(3).

Taheri S, Abdullah TL, Rafii MY, Harikrishna JA, Werbrouck SP, Teo CH, Sahebi M, & Azizi P, 2019. De novo assembly of transcriptomes, mining, and development of novel EST-SSR markers in Curcuma alismatifolia (Zingiberaceae family) through Illumina sequencing. Sci. Rep, 9(1), 1–14.

Taheri S, Lee Abdullah T, Yusop MR, Hanafi MM, Sahebi M, Azizi P, & Shamshiri RR, 2018. Mining and development of novel SSR markers using next generation sequencing (NGS) data in plants. Mole, 23(2), 399.

Thudi M, Bohra A, Nayak SN, Varghese N, Shah TM, Penmetsa RV, Thirunavukkarasu N, Gudipati S, Gaur PM, & Kulwal PL, 2011. Novel SSR markers from BAC-end sequences, DArT arrays and a comprehensive genetic map with 1,291 marker loci for chickpea (Cicer arietinum L.). PLoS One, 6(11), e27275.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, & Rozen SG, 2012. Primer3—New capabilities and interfaces. Nucleic Acids Res, 40(15), e115. https://doi.org/10.1093/nar/gks596

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, & Thibault J, 2013. From FastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. Curr Proto Bioinformatics, 43(1), 11.10. 1-11.10. 33.

Wang C, Peng D, Zhu J, Zhao D, Shi Y, Zhang S, Ma K, Wu J, & Huang L, 2019. Transcriptome analysis of Polygonatum cyrtonema Hua: Identification of genes involved in polysaccharide biosynthesis. Plant Methods, 15(1), 65.

Wang Ying, Liu K, Bi D, Zhou S, & Shao J, 2017. Characterization of the transcriptome and EST-SSR development in Boea clarkeana, a desiccation-tolerant plant endemic to China. PeerJ, 5. https://doi.org/10.7717/peerj.3422

Wang Yunsheng, Shahid M Q, Ghouri F, Ercişli S, & Baloch FS, 2019. Development of EST-based SSR and SNP markers in Gastrodia elata (herbal medicine) by sequencing, de novo assembly and annotation of the transcriptome. 3 Biotech, 9(8), 292.

Wu SB, Wirthensohn MG, Hunt P, Gibson JP, & Sedgley M, 2008. High resolution melting analysis of almond SNPs derived from ESTs. Theo. Appl. Gene, 118(1), 1–14.

Xie J. Tian J, Du Q, Chen J, Li Y, Yang X, Li B, & Zhang D, 2016. Association genetics and transcriptome analysis reveal a gibberellin-responsive pathway involved in regulating photosynthesis. J. Expe. Bot, 67(11), 3325–3338.

Xu X, & Bai G, 2015. Whole-genome resequencing: Changing the paradigms of SNP detection, molecular mapping and gene discovery. Mol. Breed, 35(1), 33.

Yu X, Zhang M, Yu Z, Yang D, Li J, Wu G, & Li J, 2020. An SNP-Based High-Density Genetic Linkage Map for Tetraploid Potato Using Specific Length Amplified Fragment Sequencing (SLAF-Seq) Technology. Agronomy, 10(1), 114.

Yaradua, S. S., & Shah, M. (2020). The complete chloroplast genome of Lantana camara L.(Verbenaceae). Mitochondrial DNA Part B, 5(1), 918-919.