DOI : 10.5958/0974-0112.2019.00020.3

Indian J. Hort. 76(1), March 2019: 141-147

Identification of SSR and miRNA from transcriptome of tuberose

Jayanthi Madhavan*, Pawan Jayaswal, Kanchan BM Singh, Prasoon Kumar Thakur*** and Uma Rao**

Division of Nematology, ICAR-Indian Agricultural Research Institute, New Delhi 110012

ABSTRACT

For identification of SSRs, the tuberose transcripts were searched with perl script MISA and a total of 1,656 SSRs in 1,280 transcripts with frequency of one SSR per 9.33 kb of the sequence were identified. In 7876 EST sequence of tuberose, we have identified 87 copies of miRNA. The SSR markers identified in this study can be used for genetic diversity analysis of tuberose and can be used to select tuberose varieties with required traits. The data generated in this study will help in better understanding of the possible role of the transcribed microsatellites and miRNA. Further the data will provide deeper knowledge of genetic mechanisms, evolutionary relationships between tuberose and other plants.

Keywords: *Polianthes tuberosa*, microsatellites, simple sequence repeats (SSR), microRNA.

INTRODUCTION

Tuberose (*Polianthes tuberosa*) is an ornamental perennial herbaceous plant. It is cultivated in many countries including India, Hawaii, Taiwan, Mexico, China, Kenya and many tropical and subtropical areas across the world. Appealing beauty of its flower and the unique fragrance makes it a commercially popular crop. Tuberose flowers are widely been used as loose and cut flowers in the floriculture market for decoration purposes and in making attractive garlands. Essential oils from tuberose are widely used in perfume industry due to rich aromatic fragrance with ability to induce positive thoughts (Guenther *et al.,* 8).

Since the genomic information on tuberose is not available till date, the genetic diversity of this species is unclear. In this study, a high-throughput, RNA sequencing (RNA-seq) of the *Polianthes tuberosa* flower transcriptome was carried out to generate a database that will be useful for further functional analyses. It is a well-known fact that transcriptome sequencing will provide new insights such as new gene discovery, gene functions, differential expressions and molecular marker development. Illumina paired end sequencing technique has provided a novel method of transcriptome analysis and has been widely used in research on model and non-model plants.

Microsatellites, which are also known as simple sequence repeats (SSRs), are short tandem repetitive DNA sequences with a repeat length of 1–6 base pairs, and are often genotype-specific and are therefore used to characterize genotypic diversity (Grover and Sharma, 7). The distribution of SSR motifs in the transcriptome sequences was studied and validated. These reported SSRs can be used in further diversity analysis. Also increasing evidence shows that miRNAs play important roles in developmental processes and gene regulation upon biotic and abiotic stresses. miRNAs were investigated in detail in model plants but with the development of second generation sequencing techniques it has become feasible to carry out small RNA studies in crop species. Though many important crops have been studied and both species-specific and conserved miRNAs have been discovered in, there has not been any report for *P. tuberosa*.

MATERIALS AND METHODS

Total RNA was extracted from fully opened flowers of tuberose cultivar Shringar using 596 Nucleospin RNA isolation kit (Macherey-Nagel GmbH & Co. KG, Duren, Germany) according to the manufacturer's instructions. Extracted RNA was assessed for quality and quantity using an Agilent 2100 Bioanalyzer (Agilent Technologies). RNA with an RNA integrity number (RIN) of 8.0 was used for mRNA purification. mRNA was purified from 1 mg of intact total RNA using oligodT beads (Illumina® TruSeq® RNA Sample Preparation Kit v2). The purified mRNA was fragmented at elevated temperature (90°C) in the presence of divalent cations and reverse transcribed with Superscript II Reverse Transcriptase (Invitrogen Life Technologies) by priming with random hexamers. Second strand

*Corresponding author's E-mail: jayman21@gmail.com

**National Research Centre for Plant Biotechnology, Pusa, New Delhi 110012

***Institute of Molecular Genetics of the ASCR, Vídeňská 1083, 142 20 Prague 4, Czech Republic

cDNA was synthesized in the presence of DNA polymerase I and RNaseH. The cDNA was cleaned using Agencourt Ampure XP SPRI beads (Beckman-Coulter). Illumina adapters were ligated to the cDNA molecules after end repair and the addition of an 'A' base followed by SPRI clean-up. The resultant cDNA library was amplified using PCR for the enrichment of adapter-ligated fragments, quantified using a Nanodrop spectrophotometer (Thermo Scientific) and validated for quality with a Bioanalyzer (Agilent Technologies). It was then sequenced on the Illumina Hiseq 2000 platform at SciGenom Next-Gen sequencing facility, Cochin, India.

The low quality reads (Phred score <30) were removed and sequencing statistics was generated with the help of NGSQC Toolkit version v2.3.3 (Patel and Jain, 15).This high quality paired end filtered reads (15.9 gb) has been used for *de-novo* assembly using Velvet (v.1.2.08) and Oases (v.0.2.08) pipeline. Velveth assembly was done with various k-mer range (71- 83) and optimal assembly was attained at k-mer 83. Oases tool was used to identify non-overlapping isoforms/splice variants at minimum transcript length 100. Since our initial target was to identify unique genes. Thus, transcripts were subjected for clustering using CD-HIT-EST 90% similarity. ORF Predictor web server (http://bioinformatics.ysu.edu/tools/OrfPredictor.html) was used to predict proteins from all the non-redundant transcripts (≥ 100bp) using the default cut-off value of 1e–5, and 7,876 proteins were predicted which were considered for the annotation. The raw sequence data generated has been deposited in the SRA database (http://www.ncbi.nlm.nih.gov/bioproject/321962) for public access (BioSample accession ID: SAMN05006898).

Transcriptome data of tuberose was used for the identification of SSR. Repeat elements were identified using RepeatMasker v.4.0.5 (Smit *et al.,* 17). Repeat Masker was installed with RMBlast v 2.2.28 and Tandem Repeats Finder 4.04 (Benson, 3). The transcripts were queried with Repbase v.20140131 using default parameters. The perl script program MISA (MIcroSAtellite; http://pgrc.ipk-gatersleben.de/misa) was used for identification of short sequence repeats (SSRs) with default parameters. Stringent detection criteria were implemented to get perfect SSR motifs. Threshold search criteria in MISA script was considered to be appropriate, if repeats of mono-nucleotide were present 12 times, di-nucleotides repeat more than 8 times and tri-, tetra-, penta- and hexa-nucleotide repeats 5 times.

In order to validate microsatellite loci, a total of 20 primer pairs (Table 1) were designed using the Batch Primer3 software http://probes.pw.usda.gov/cgi-bin/batchprimer3/batchprimer3.cgi based on the unigene sequences derived from the combined non-redundant unigene database. The parameters were designed for identification of di-, tri-, tetra-, penta and hexanucleotide motifs. Primer design parameters were selected as follows: length range 18 to 21 nucleotides with 20 as optimum; PCR product size range 100 to 300 bp; optimum annealing temperature 60°C; and GC content 40-60% with 50% as optimum. Genomic DNA of different tuberose varieties was extracted using CTAB method. For PCR amplification a final reaction volume of 25µL containing 1X PCR buffer,2.5 mM dNTPs, 0.6 unit Taq polymerase, 4 µM each primer and 50 ng DNA template was used for PCR amplification. The PCR reaction conditions were the following: 95°C for 4 min, followed by 35 cycles at 94 °C for 1min, 55-58°C for 30s, and 72 °C for 1 min, with final step at 72 °C for 10min. The PCR products were then electrophoresed on 3.0% Agarose gels.

The miRNA genes were identified using BLASTN search (e >1*10^{-5}, top hits) of sequences present in the 21st release of miRNA database (35828 mature miRNA from 223 species, http://www.mirbase.org/), allowing no more than three mismatches (Griffiths-Jones *et al.*, 6) against the tuberose assembled transcriptome.

RESULTS AND DISCUSSION

Tuberose is a member of Asparagaceae family and has attractive long spiked fragrant flowers which are grown commercially in tropical and subtropical regions of the world. According to Anonymous, 1 tuberose cultivation is spread across 7.95 lakh hectares with 27.71 million tonnes production of loose and 1560 lakhs of cut flowers. Genetic diversity of tuberose is ambiguous due to lack of molecular markers. SSRs are variable number of DNA repeat sequences of 2-6 base pairs, typically codominant. They are extensively used as molecular marker in genetic mapping/pedigree and molecular breeding in crops. Development of SSR markers by traditional methods is expensive and time consuming. In this context, using high throughput sequencing technology to develop SSR markers is highly efficient, convenient and low in cost. The transcriptome data (Jayanthi *et al.,* 9) was used to identify repeat elements present in tuberose transcripts since, there is no information on genome-wide repeats available. These elements are known to influence expression of genes by various mechanisms such as, silencing by hetero chromatisation, up-regulation by providing alternative promoters and novel expression patterns through

Table 1. SSR primers designed using primer batch 3 for tuberose.

S. No.	Seq ID	Tm	GC%	Seq	Prod Size	Motif	SSR (Length)
1	TR28F	59.34	50	TCCAAGAGATGACGGACAGA	161	CT	CTCTCTCTCTCTCTCT (16)
	TR28R	57.07	40	GCAAAGGAAAGACAAGGAAA			
2	TR40F	59.79	45	AATGGAATGGAACGGAACAG	183	GAATG	GAATGGAATGGAATGGAATGGAATG (25)
	TR40R	59.96	50	ATTACATTCCCCTGCACTCG			
3	TR55F	60.73	52.6	CCTTTGGCTCATTCCAAGC	120	CA	CACACACACACACACA (16)
	TR55R	60.00	55	AGCGATCTTCTCGCTCTCTG			
4	TR133F	60.39	55	GGGGCAGACAGACAGAGAAA	113	AG	AGAGAGAGAGAGAG (14)
	TR133R	58.40	60	CTCTCTCACTCCCTCCCTGT			
5	TR133.2F	59.38	55	GAGACAGACGGGGAGAGAAA	194	AGAC	AGACAGACAGACAGACAGACAGAC (24)
	TR133.2R	59.98	60	GGCTAGGGCCTTCTCTCTGT			
6	TR137F	59.59	52.38	GGAAAGGACTGGAGTGGACTT	165	GAATG	GAATGGAATGGAATGGAATGGAATG (25)
	TR137R	60.45	40	TTGCCGTTCCATTCAATTCT			
7	TR141F	59.85	45	CAAGAAAACGGGAAAGCAAG	138	GA	GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA (38)
	TR141R	60.01	55	CTTAGCCTCCCAAAGTGCTG			
8	TR142F	59.85	45	CAAGAAAACGGGAAAGCAAG	129	GA	GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA (38)
	TR142R	60.67	52.63	CCAAAGTGCTGGGATGACA			
9	TR164F	59.60	55	AGCTTCTGGCTGTTGAGGAC	167	CGG	CGGCGGCGGCGGCGG (15)
	TR164R	60.22	55	ACCGACGAGGAGATGATGAG			
10	TR164F	60.68	60	CCGTCTCTCCTCATCACCAC	181	CGA	CGACGACGACGACGA (15)
	TR164R	60.39	55	CCTCCTCCCTTCCTATTCCA			
11	TR_240F	59.95	40	TCCATTCCATTCCATTCCAT	177	ATTCC	ATTCCATTCCATTCCATTCCATTCC (25)
	TR_240R	59.81	50	GAATCAACCTGAGCGGAAAG			
12	TR_268F	59.89	43.48	GCCTTTCTTTTCCTTTGTCTCTC	198	GGA	GGAGGAGGAGGAGGAGGA (18)
	TR_268R	59.87	50	CCTGTTGAAGTGGTTGAGCA			
13	TR_180F	60.02	55	CGACACAATACACCCAGTCG	220	AT	ATATATATATATATAT (16)
	TR_180R	60.03	50	TCAGATCCGACCGAAATAGG			
14	TR_306F	62.94	52.63	AACGATGAGCGACGACGAA	201	CCA	CCACCACCACCACCACCACCACCA (24)
	TR_306R	61.93	60	TAGCCGTGGGCGTAGTAGGT			
15	TR_324F	62.58	52.63	TGGAACTGGGATCGGGATT	206	AAC	AACAACAACAACAACAACAACAAC (24)
	TR_324R	60.93	52.38	TAGTGGGTGATGAGGACGATG			
16	TR_324.2F	60.96	55	CATCGTCCTCATCACCCACT	226	ATC	ATCATCATCATCATCATC (18)
	TR_324.2R	60.86	52.63	CCGGAGCTTTATTGCTTGC			
17	TR_349F	60.33	57.89	GGGCAAAGAGAGAGGTGGA	198	GAC	GACGACGACGACGAC (15)
	TR_349R	61.71	52.38	CCTGTTGTTGGTGTTCCTTCC			
18	TR_366F	61.45	57.89	CTGAACGCTGCCGTGTATG	218	AC	ACACACACACACACACACACACACACACACACACACACAC (42)
	TR_366R	61.27	47.62	AATGAAGGAAGGGAGGAAGGA			
19	TR_380F	63.03	57.14	GCTCATCTCCACACCGATCAC	153	CCG	CCGCCGCCGCCGCCG (15)
	TR_380R	60.31	57.89	GCTGCTTCACACCCACAAC			
20	TR_398F	65.53	57.89	TGCTTGCCTGCTTGCCTGT	234	TTCC	TTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCC (40)
	TR_398R	65.20	60	AAGTCAGCCGGTGTGGTGGT			

alternative splicing and polyadenylation (Chow *et al.,* 4; Lev Maor *et al.,*12). The Repeat Masker program (Smit *et al.,* 17) was used to identify different repetitive elements among the transcripts. Approximately 4.04% of the total transcript was found to be encoded by different repetitive elements (Table 2). Simple and satellites repeats regions encoded the maximum amount (approximately 72%) of transcripts from repetitive elements (Fig. 1). Whereas, Interspersed repeats are present in the least amount, it comprises 5% of the total repeat element. A total of 51 retro elements were found among transcripts, with 15 long interspersed repeat elements (LINEs), 30 short interspersed repeat elements (SINEs) was found. Among retroelements, the number of long terminal repeats (LTRs) was lower (6) than non-LTR elements. In addition, 9% low complexity repeats and 14% small RNAs were identified. The transcript markers are important resource for determining functional genetic variation. For identification of SSRs, the tuberose transcripts were searched with perl script MISA. We identified a total of 1,656 SSRs in 1,280 transcripts of tuberose with frequency of one SSR per 9.33 kb of the sequence. MISA results showed that 241 sequences containing more than one SSR. In addition, we observed Maximal 100 bases interrupting two SSRs in a compound microsatellite and 190 SSRs present in compound formation. The di-nucleotide SSRs represented the largest fraction (37.3%, 618) of SSRs identified followed by tri-nucleotide (29.8%, 494) and mono-nucleotide (27.23%, 451) SSRs. Although only a small fraction of tetra- (33) and penta-(55) copies of SSRs were identified in tuberose transcripts. The distribution of top 20 SSR has been shown in Fig. 2. A collection of randomly selected primer pairs were used for PCR amplification of SSR markers and validated experimentally. Amplification of the desired products was successful among all the six varieties used. A representative of some SSR primer amplification is shown among different varieties of tuberose (Fig. 3).

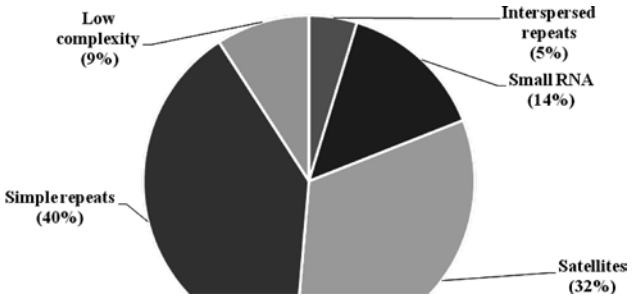

Fig. 1. Distribution of Repeative elements in tuberose transcriptome.

Table 2. Distribution of different types of repeat elements in the tuberose transcriptome data.

Repeat category	Number of elements	Length of occupied	Percentage sequence
SINEs:	30	5982 bp	0.10%
ALUs	30	5982 bp	0.10%
MIRs	0	0 bp	0%
LINEs:	15	3628 bp	0.06%
LINE1	12	3459 bp	0.06%
LINE2	2	95 bp	0%
L3/CR1	0	0 bp	0%
LTR elements:	6	1567 bp	0.03%
ERVL	1	461 bp	0.01%
ERVL-MaLRs	0	0 bp	0%
ERV_classI	2	377 bp	0.01%
ERV_classII	1	621 bp	0.01%
DNA elements:	5	292 bp	0%
hAT-Charlie	1	53 bp	0%
TcMar-Tigger	3	183 bp	0%
Unclassified:	1	139 bp	0%
Total interspersed repeats:		11608 bp	0.19%
Small RNA:	111	36198 bp	0.58%
Satellites:	109	81745 bp	1.31%
Simple repeats:	2169	99480 bp	1.60%
Low complexity:	348	23330 bp	0.37%

MicroRNAs (miRNAs) are single-stranded RNAs of 20–24 nucleotides (nt) in length that are generated from endogenous hairpin-shaped transcripts (Kim, 11). miRNAs play important roles in plant post-transcriptional gene regulation by targeting mRNAs for cleavage or repressing translation. miRNAs are involved in plant development, signal transduction, protein degradation, response to environmental stress and pathogen invasion, and regulate their own biogenesis. miRNAs regulate the expression of many important genes; a majority of these genes are transcriptional factors (Bartel, 2). It has been estimated that organisms probably contain about 1% miRNA genes of the total protein-coding genes (Lim *et al.,* 14). Recently, bioinformatics studies indicated that the proportion of genes in the genome under miRNA regulation may be much larger than previously thought. For example, Lewis *et al.,* (13) proposed

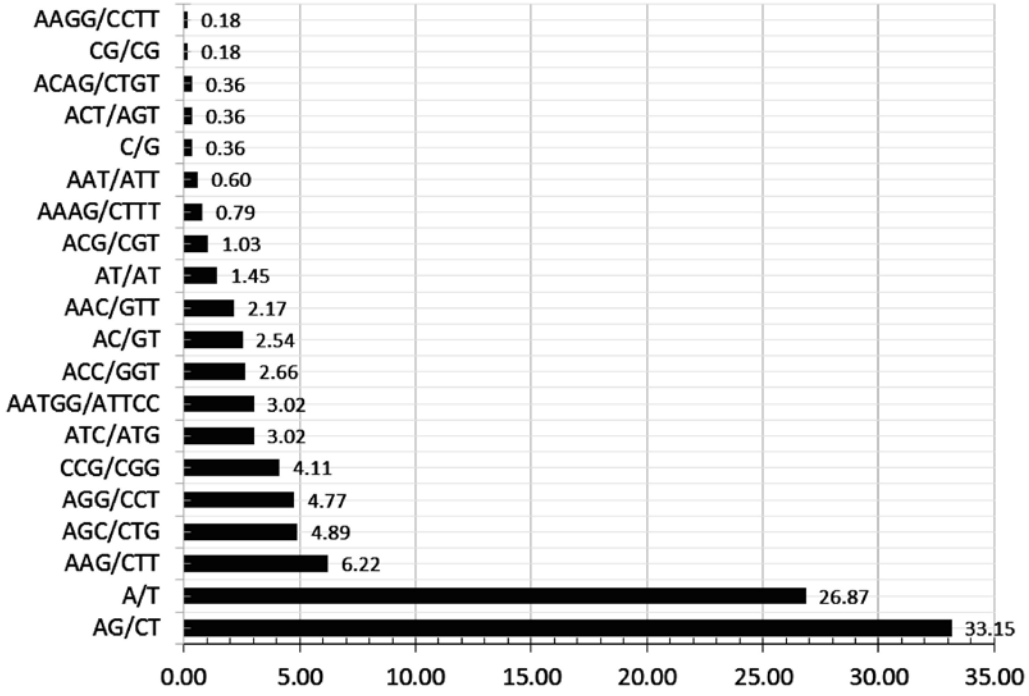

Fig. 2. Distribution of top 20 different categories of SSRs maker.

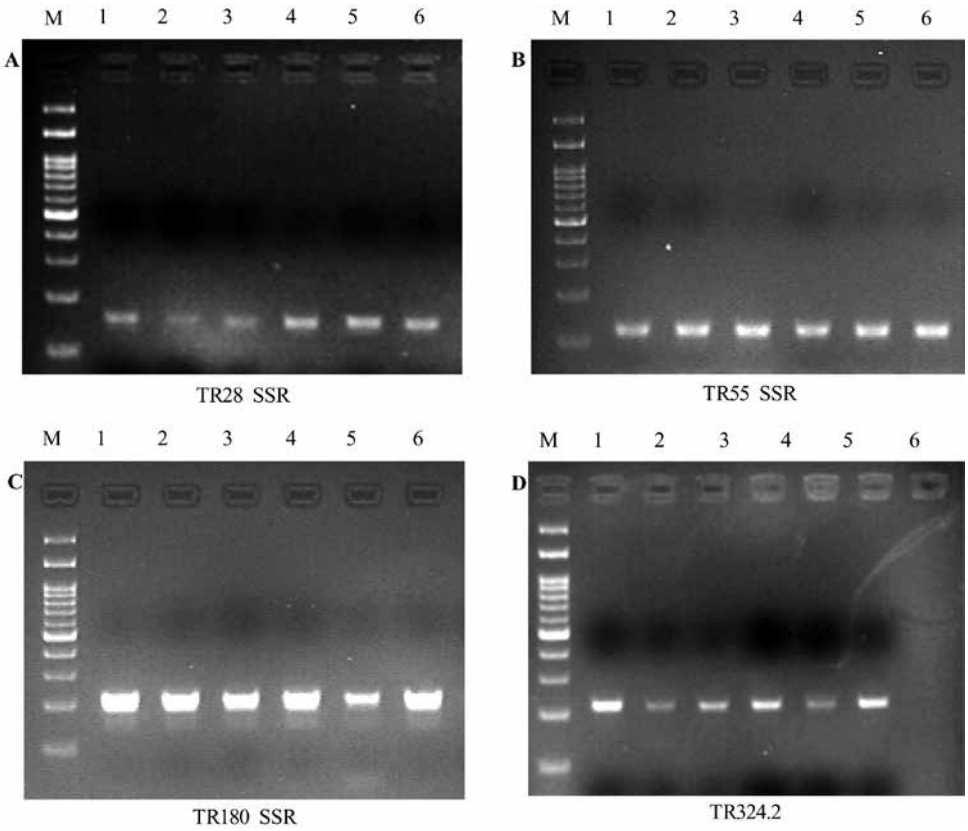

Fig. 3. SSR primer amplification among different varieties (1. Phule rajani 2.Calcutta double 3.Bidhan ujwal 4.Calcutta single 5. Hyderbad double 6. Prajwal) of tuberose.

that about 30% of all human genes may be regulated by miRNAs. However, approximately 872 miRNAs have been identified in 71 plant species (Griffiths Jones, 5). This is far very less than the proposed number. Therefore, a total of 87 miRNA genes were identified using miRNA database (miRBase, http://www.mirbase.org/) out of which seventy one copies were from different families of miRNA (Table 3). In all predicted miRNA, 13 miRNA families were having more than one copy of miRNA. miR-3533 shared five copies number of miRNA out of 87 which has highest among all predicted miRNA.

In our unigene sequences we found 1656 motif sequence types far more than that of *Hevea brasiliensis* (Salgado *et al.,* 16) and *Oenanthe javanica* (Jiang *et al.*, 10). MicroRNAs (miRNAs) have been identified as key molecules in regulatory networks. The fine-tuning role of miRNAs in addition to the regulatory role of transcription factors has shown that molecular events during development are tightly regulated. In addition, several miRNAs play crucial roles in the response to abiotic stress induced by drought, salinity, low temperatures, and metals such as aluminium. Interestingly, several miRNAs have overlapping roles with regard to development, stress responses, and nutrient homeostasis. Moreover, in response to the same abiotic stresses, different expression patterns for some conserved miRNA families among different plant species revealed different metabolic adjustments. The use of deep sequencing technologies for the characterization of miRNA frequency and the identification of new miRNAs adds complexity to regulatory networks in plants. Therefore, we identified 87 miRNA genes using miRNA database (miRBase, http://www.mirbase.org/) out of which seventy one copies were from different families of miRNA. In all predicted miRNA, 13 miRNA families were having more than one copy of miRNA. miR-3533 shared five copies number of miRNA out of 87 which has highest among all predicted miRNA. The data generated during this study will help in the better understanding of expression patterns and their relation to function and regulation, possible role of transcribed microsatellites and micro RNAs and also the genetic mechanisms, evolutionary relationships between tuberose and other plants.

Table 3. Distribution of miRNA in tuberose transcriptome.

S. No.	miRNAData	Count	S. No.	miRNAData	Count
1	miR-3533	5	37	miR5019	1
2	miR1134	2	38	miR5141	1
3	miR164b	2	39	miR5282	1
4	miR-3800	2	40	miR5303i	1
5	miR395o	2	41	miR5503	1
6	miR-4459	2	42	miR5523	1
7	miR5538	2	43	miR6113	1
8	miR-619-5p	2	44	miR-6577-5p	1
9	miR6482	2	45	miR-6767-3p	1
10	miR-6958-5p	2	46	miR-6795-5p	1
11	miR-8109	2	47	miR-6866-5p	1
12	miR-9340	2	48	miR-6940-5p	1
13	miR-K12-6-5p	2	49	miR-7131c-3p	1
14	miR1127b-3p	1	50	miR-71a-1-3p	1
15	miR-1538	1	51	miR-7389-5p	1
16	miR168b-3p	1	52	miR-7408-3p	1
17	miR171c	1	53	miR-74-5p	1
18	miR172f	1	54	miR7494b	1
19	miR-1795	1	55	miR7502	1
20	miR-2005-5p	1	56	miR7531	1
21	miR-207	1	57	miR7540b	1
22	miR-219-5p	1	58	miR773b-5p	1
23	miR-2502	1	59	miR-7847-3p	1
24	miR-2889	1	60	miR-7880n-5p	1
25	miR-2d-5p	1	61	miR8044-5p	1
26	miR-309a-3p	1	62	miR-8112	1
27	miR-3314	1	63	miR-8218a-5p	1
28	miR-3585-5p	1	64	miR-8282c-5p	1
29	miR-3618-5p	1	65	miR-877-3p	1
30	miR393a	1	66	miR-8861	1
31	miR414	1	67	miR9666b-3p	1
32	miR-4644	1	68	miR9728	1
33	miR-4778-5p	1	69	miR-H3	1
34	miR-4852	1	70	miR-I5-3p	1
35	miR-4897	1	71	miR-rR1-5	1
36	miR-4968-3p	1			

ACKNOWLEDGMENTS

The authors gratefully acknowledge Director and Joint Director, ICAR-Indian Agricultural Research Institute, New Delhi for providing the research facilities.

REFERENCES

1. Anonymous 2013. Horticultural statistics at a glance. National Horticultural Board, Gurugram, Haryana.
2. Bartel, D.P. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**: 281-297.
3. Benson, G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**: 573-80.
4. Chow, J.C., Ciaudo, C., Fazzari, M.J., Mise, N., Servant, N., *et al.* 2010. LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation. *Cell*, **141**: 956-69.
5. Griffiths-Jones, S. 2004. The microRNA registry. *Nucleic Acids Res.* **32**: D109-D111.
6. Griffiths-Jones, S., Saini, H.K., Dongen, S.V., Enright, A.J. 2008. miRBase: tools for microRNA genomics. *Nucleic Acids Res.* **36**: D154-D158.
7. Grover, A. and Sharma, P.C. 2014. Development and use of molecular markers: past and present. *Crit. Rev. Biotech.* **36**: 290-302.
8. Guenther, E., Gilbertson, G., Koenig, R.T. 1975. Essential Oils and Related Products. *AnalChem.* **47**: 139-57.
9. Jayanthi M., Jayaswal P., Singh K.B.M., Rao U. 2018. Identification of putative flowering genes and transcription factors from flower *de novo* transcriptome dataset of tuberose (*Polianthes tuberosa* L.). *Data in Brief*, **20**: 2027-35.
10. Jiang, Q., Wang, F., Tan, H.W., Li, M.Y., Xu, Z.S., *et al.* 2015. *De novo* transcriptome assembly, gene annotation, marker development, and miRNA potential target genes validation under abiotic stresses in *Oenanthe javanica*. *Mol. Genet. Genomics*, **290**: 671-83.
11. Kim, V.N. 2005. MicroRNA biogenesis: coordinated cropping and dicing. *Nat. Rev. Mol. Cell Biol.* **6**: 376-85.
12. Lev-Maor, G., Sorek, R., Shomron, N., Ast, G. 2003. The birth of an alternatively spliced exon: 3'splice-site selection in Alu exons. *Science*, **300**: 1288-91.
13. Lewis, B.P., Burge, C.B., Bartel, D.P. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, **120**: 15-20.
14. Lim, L.P., Glasner, M.E., Yekta, S., Burge, C.B., Bartel, D.P. 2003a. Vertebrate microRNA genes. *Science*, **299**: 1540.
15. Patel R.K., Jain M. 2012. NGS QC Toolkit: A toolkit for quality control of next generation sequencing data. *PLoS ONE*, **7**: e30619.
16. Salgado, L.R., Koop, D.M., Pinheiro, D.G., Rivallan, R., Le Guen, V., *et al.* 2014. *De novo* transcriptome analysis of *Hevea brasiliensis* tissues by RNA-seq and screening for molecular markers. *BMC Genomics*, **15**: 236.
17. Smit, A.F.A., Hubley, R., Green, P. 1996-2010. *RepeatMasker Open-3.0.* <http://www.repeatmasker.org>.

Received : May, 2018; Revised : February, 2019;
Accepted : February, 2019