Genome-Wide Mining, Characterization and Development of miRNA-SSRs in *Arabidopsis thaliana*

Anuj Kumar[1,2], Aditi Chauhan[1], Mansi Sharma[2], Sai Kumar Kompelli[3], Vijay Gahlaut[4], Johny Ijaq[2], Krishna Pal Singh[1], MNV Prasad Gajula[5*], Prashanth Suravajhala[3,6,7], Harindra Singh Balyan[3], and Pushpendra Kumar Gupta[3]

1. Advance Centre for Computational and Applied Biotechnology, Uttarakhand Council for Biotechnology (UCB), Dehradun-248007, India
2. Bioclues.org, Kukatpally, Hyderabad 500072, Telangana, India
3. ICMR-ICPO (Institute of Cytology and Preventive Oncology), NOIDA, I-7, Sec-39, Uttar Pradesh
4.Molecular Biology Laboratory, Department of Genetics & Plant Breeding, Ch. Charan Singh University, Meerut-250004, India
4. Institute of Biotechnology, PJTSAU, Rajendra Nagar, Hyderabad-500030, India
5. Bioinformatics Organization, 28 Pope st, Hudson, MA 01749, USA
6. Department of Biotechnology and Bioinformatics, Birla Institute of Scientific Research, Statue Circle, 302001, RJ, India

***Corresponding authors**: email2gajula@gmail.com*

Abstract

Simple Sequence Repeats (SSRs), also known as microsatellites are short tandem repeats of DNA sequences that are 1-6 bp long. In plants, SSRs serve as a source of important class of molecular markers because of their hypervariabile and co-dominant nature, making them useful both for the genetic studies and marker-assisted breeding. The SSRs are widespread throughout the genome of an organism, so that a large number of SSR datasets are available, most of them from either protein-coding regions or untranslated regions. It is only recently, that their occurrence within microRNAs (miRNA) genes has received attention. As is widely known, miRNA themselves are a class of non-coding RNAs (ncRNAs) with varying length of 19-22 nucleotides (nts), which play an important role in regulating gene expression in plants under different biotic and abiotic stresses. In this communication, we describe the results of a study, where miRNA-SSRs in full length pre-miRNA sequences of *Arabidopsis thaliana* were mined. The sequences were retrieved by annotations available at EnsemblPlants using BatchPrimer3 server with miRNA-SSR flanking primers found to be well distributed. Our analysis shows that miRNA-SSRs are relatively rare in protein-coding regions but abundant in non-coding region. All the observed 147 di-, tri-, tetra-, penta- and hexanucleotide SSRs were located in non-coding regions of all the 5 chromosomes of *A. thaliana*. While we confirm that miRNA-SSRs were commonly spread across the full length pre-miRNAs, we

envisage that such studies would allow us to identify newly discovered markers for breeding studies.

***Keywords**:* MicroRNA, miRNA-SSRs, Genome-wide identification studies, noncoding RNAs, gene expression

1. Introduction

MicroRNAs (miRNA) represent a class of non-coding RNA (ncRNA) with varying length of 19-22 nucleotides (nts) (Bartel, 2004). These miRNAs are endogenous in origin, and are found to play a major role in regulating the gene expressions in plants, fungi and animals, with bulk of the sequences linked to transcription factors (Bartel and Bartel, 2003). The miRNA are involved regulation of genes implicated in different processes including the following: (i) response to different biotic and abiotic stresses (Khraiwesh et al. 2012; Kompelli et al. 2015); (ii) different development and protein degradation processes (Eldem et al., 2012), (iii) pathogen invasion, signal transduction etc. (Jones-Rhoades et al. 2006; Jung et al. 2009).

Simple Sequence Repeats (SSRs), also known as microsatellites are short tandem repeats of DNA sequences that are 1-6 bp long (Gupta et al. 1996; Chen et al. 2009). The SSRs are found both in prokaryotic and eukaryotic genomes (Toth et al. 2000; Katti et al. 2001). SSRs are co-dominant, and multi-allelic by nature and due to constant variation in the number of tandem repeats; they are known to be, robust, highly polymorphic (Brandstrom et al. 2008, Heesacker et al. 2008), locus-specific and co-dominant, thus becoming the markers of choice. (Gupta et al. 1996; Ni et al. 2002; Lightfoot and Iqbal, 2013; Senan et al. 2014; Wang et al. 2015). Previous reports show that SSRs are selectively neutral and are randomly distributed in the eukaryotic genome (Schlotterer, 2000; Schlotterer, 2004). Although many of them are found in protein coding (Madsen et al., 2000), non-coding (Riley and Krieger, 2009a, 2009b) or untranslated regions (Mondal and Ganie, 2015) of plant genome, mainstream SSRs are regularly found in non-coding regions and relatively rare in protein coding regions (Madsen et al. 2008). Furthermore, with SSRs known to have numerous applications, application of SSRs in construction of genetic maps has led to significant interest (Gupta et al. 1996; Li et al. 2002; Usdin, 2008). While SSRs aid in chromatin organization (Cuadrado and Schwarzacher, 1998), available evidence show that SSRs located in promoter regions may affect the level of gene expression (Young et al. 2000). It has been reported that they are widely considered as a hot spots for recombination (Jeffreys et al. 1998; Templeton et al. 2000).

. Recently, SSRs have been reported in pre-miRNA sequences in some plant species. For instance, Chen et al. 2010 carried out a comprehensive analysis for the prediction of SSRs in 8,619 premiRNA sequences from 87 species, including Arthropoda, Nematoda, Platyhelminthes, Urochordata, Vertebrata, Mycetozoa, Protistate, Viridiplantae, and Viruses. In another studies, salt responsive (trait specific) miRNA-SSRs were reported in rice genome (Ganie and Mondal, 2015; Mondal and Ganie) linking them to phenotype and expression of genes. Furthermore, studies on role of transcriptional profiling of SSR specific long noncoding RNAs (lncRNAs) are studied in Banana and sugarcane which supports the hypothesis there is a major role of SSRs in non-coding genome in both small and larger noncoding elements (Cardoso-Silva et al. 2014; Yang et al. 2015). However, no study has so far been conducted to study SSRs in Pre-miRNA full length transcripts of *A. thaliana*, which is a model plant system with a small genome that was the first higher plant genome to be fully sequenced (The Arabidopsis Genome Initiative, 2000). Because of enormous utilities of miRNA as well as SSRs, there is a need for development of markers associated with miRNA, so that markers may be developed for traits influenced by miRNAs. Keeping this in view of the prospective development markers from the noncoding regions, we discovered miRNA-SSRs in full length genomic sequences of pre-miRNAs of *A. thaliana*.

2. Methodology

2.1. Computational identification and discovery of miRNA-SSRs in *A. thaliana* genome

A total of 325 pre-miRNAs of *A. thaliana* were downloaded from miRBase 21.0 (http://www.mirbase.org/) (Kozomara et al. 2014) and full length genomic transcripts representing pre-miRNA were extracted in FASTA format using BioMart-Ensembl genomes (Kasprzyk, 2011) available in EnsemblPlants (Bolser et al.2015) (***see Supplementary Table* 1**); among 325 pre-RNAs, only 169 pre-miRNA sequences were found **(*see Supplementary Table* 2)** whose full length genomic sequences are available in EnsemblPlants. After downloading all full length premiRNA genomic transcripts from EnsemblPlants, manual annotation was done to confirm the transcripts (>1000bp + premiRNA) for the discovery of SSRs belonging to miRNA genes (i.e., promoter, 5' UTR, primRNA, or 3' UTR but not pre- or mature miRNA). The search for miRNA-SSRs and the designing of primers flanking miRNA-SSRs was carried out in full length premiRNA transcripts from all 5 chromosomes using BatchPrime3 v1.0 (You et al. 2008) with default parameters. A flow chart showing the pipeline used in this study is presented in Figure 1.

2.2. Computational Prediction of SSRs-containing miRNAs

As earlier documented, plant miRNAs predominantly target different families of transcription factors (TFs) (Llave et al. 2002; Chen, 2004; Brodersen et al. 2008; Gupta et al. 2015; Gahlaut et al. 2016). However, subsequent studies suggested that miRNAs also target plant functional protein encoding genes, which control various physiological processes, such as root growth and development, stress responses, signal transduction, leaf morphogenesis, plant defenses, and biogenesis of sRNA (Brousse et al. 2014). Unlike in animals, miRNAs in plants identify their target mRNAs through perfect or near-perfect complementarity and initiate cleavage.

The putative target sites of SSRs-containing miRNAs were predicted by aligning the miRNA sequences either perfectly or near-perfectly binding to complementary sites on their target mRNA sequences by using homology search-based psRNATarget server (Dai and Zhao, 2011). Transcripts of SSRs-containing miRNAs were used as a query against updated version of *A.thaliana* transcripts available on The Arabidopsis Information Resource (TAIR) (https://www.arabidopsis.org/). Following parameters embedded in psRNATarget algorithm were used: maximum expectation: 2.0, length for complementarity scoring (hspsize): 20, target accessibility-allowed maximum energy to unpair the target site (UPE): 25.0, flanking length around target site for target accessibility analysis: 17 bp in upstream and 13 bp in downstream, Range of central mismatch leading to translation inhibition: 9–11nt.

2.3. Prediction of genes adjacent to identified miRNA-SSRs and analysis of enriched gene ontologies (GO)

Genes adjacent to identified novel miRNA-SSRs were manually predicted using the TAIR 9 browser embedded in windows based integrated genome browser (IGB) (Nicol et al. 2009). The criteria for manual curation was based on location of SSRs and nearby gene located on 5' untranslated region (5′ UTR) and 3' untranslated region (3′ UTR) sites on a particular chromosome of *A. thaliana* genome. Further predicted adjacent transcripts were retrieved from the EnsemblPlants (Bolser et al., 2015) in FASTA format. Arabidopsis adjacent transcripts were used as input for Gene ontology analysis using agriGO (Du et al. 2010) and REVIGO (Supek et al. 2011) server.

3. Result and Discussion

3.1. Dinucleotide repeats were found to outnumber other repeats

In the present study, 147 miRNA-SSRs were discovered among 169 pre-miRNA genomic transcripts of *A. thaliana* genome (**Table. 1**). We found that dinucleotide SSR repeats (48/147) outnumbered the other repeats; primers designed for 45 of these dinucleotide repeats while no primers were designed for the remaining three SSRs including $(AC)_7$ associated with miR164b, $(AT)_7$ associated with miR165b and and $(TA)_{10}$ associated with miR832A. Ten (10) different classes of dinucleotide SSR repeats were found in all premiRNA transcripts of *A.thaliana* and the largest count of dinucleotide repeat was TA. **(Fig.2).** While trinucleotide miRNA-SSR repeats were found to be less than dinucleotide repeats, only one of 38 repeats was found with no SSR flanking primer (TTC with miR837a and SSR length - 12). Nevertheless, there were 37 SSR flanking primers found to be associated with them. Within 15 different classes of trinucleotide miRNA-SSRs repeats, TTC and CTT with same number of counts formed the highest count of trinucleotide repeats (**Fig.2**)

The tetranucleotide miRNA-SSRs (46) were found to be more than trinucleotide repeats but less than dinucleotide repeats. Primers flanking two SSRs *viz.* $(TTTA)_n$, and $(TTAT)_n$ for miR164c and miR394a, respectively could not be designed $(TTTA)_n$ repeats was most abundant among the tetranucleotide repeats in discovered miRNA-SSRs. (**Fig. 2**). The pentanucleotide SSRs in pre-miRNA transcripts of *A. thaliana* were least frequent. Out of the 12 of the 147 miRNA-SSRs, were pentanucleotide repeats. Primers flanking to 11 miRNA-SSRs were designed and no primers could be designed for, $(TTGTT)_3$ associated with miR777a. Only eight classes of pentanucleotide SSR repeats were found in all pre-miRNA transcripts of *A. thaliana* and TTTTA was found as topmost count of pentanucleotide SSRs (**Fig. 2**).The hexanucleotide miRNA-SSRs were least common and these belonged to $(GTTTGA)_n$, $(GGGAGG)_n$, $(ACAAAT)_n$, and $(CGTTTC)_n$ classes to be associated with flanking primers and remarkably distributed across all 5 chromosome in *A.thaliana* genome (**Fig. 3**). The chromosomes 1 and 5 have maximum miRNA-SSRs, while chromosome 3 has minimum number of miRNA-SSRs (**Fig. 3**).

3.2. Conservation of SSR loci spanning flanking regions

The miRNA-SSR polymorphism will provide trait-related molecular markers at the specific chromosomal loci, which in turn would depend on the number of indels in the flanking regions. Whether or not they are dinucleotide repeats or compound repeats is dependent not only on variances at the each repeat unit of the sequences, but also on how they are arranged

or distributed across the genome. As we observed such repeats, it would be interesting to examine their locus specific polymorphism to allow their physically mapping. It would be interesting to see if they can serve as unknown tagged sites which in turn would depend on the presence of a particular sequence tagged region or sequence tagged sites (STS). These STS' in principle can be used as potential markers.

3.3. SSRs-containing miRNAs targeted diverse set of TFs

On the basis of the biogenesis of miRNAs in plants, a homology search-based method was used to predict the targets for SSRs-containing miRNA in *A. thaliana* using psRNATarget. The SSR-containing miRNAs were used as queries to predict potential mRNA targets in the Arabidopsis genome annotation (TAIR10). This search revealed that 90 SSR-containing miRNAs identified 698 target genes, with each SSR-containing miRNA predicting more than one gene (Table S1). Most of the SSR-containing miRNAs targeted a number of TFs families including WRKY, MADS, MYB, NAC, bHLH, AP2/EREBP, ARF etc., which play an important role in different metabolic and regulatory processes such as stress response, transcriptional regulation, signal transduction, growth, development, nutrient uptake, nutrient transport and nutrient assimilation (**Table 2**). The values of UPE for targeted gene ranged from 3.238 to 24.941.

Targeted TFs could be utilized for developing next generation microsatellites, Transcription Factor Gene-Derived Microsatellite (TFGM) Markers which have potential in marker-assisted genetic improvement and genotyping applications through marker assisted selection (MAS) breeding program to develop the drought/heat responsive and nutrient efficient cultivars for cereal crops (Gupta and Prasad, 2009; Kujur et al. 2013, 2014; Liu et al. 2015). However in plants, (TFGM) markers have only been reported in chickpea and *Medicago truncatula* to date (Kujur et al. 2013; Liu et al. 2015).

3.4. Prediction of genes adjacent to identified miRNA-SSRs and GO analysis

In order to predict the genes adjacent to SSR containing miRNAs, representing 5′ UTR and 3′ UTR sites TAIR 9 was manually curated. Based on length and chromosomal location, a diverse set of adjacent genes were predicted both in n5′ UTR and 3′ UTR regions (**Table. 2**). Predicted adjacent transcripts revealed that SSR containing miRNAs are associated with different genes in network form, which play a pivotal role in gene regulation. However effect of miR-SSR on adjacent genes and vice- versa need to be studied in detail.

To evaluate the biological significance of the adjacent genes to SSR containing miRNAs in Arabidopsis it is important to have the gene ontology (GO) descriptions i.e., detailed annotations of gene function, biological process it is involved, and cellular location of the gene product. The potential functions were predicted by searching against GO database using agriGO and REVIGO server. Predicted adjacent transcripts were subjected to singular enrichment analysis (SEA) embedded in agriGO to identify enriched GOs. SEA designed to identify enriched GO terms in a list form of microarray probe sets or gene identifiers available in database. Finding different enriched GO terms corresponds to finding enriched biological facts, and term enrichment level was judged by comparing query list to a background population from which the query list is derived. In this study the background query list comprised of 27,416 protein coding genes from the updated TAIR (https://www.arabidopsis.org/index.jsp). **Fig. 4** wholly reflects the categorization of adjacent genes based on biological process, cellular component and molecular function. Adjacent genes were divided into 14 GO categories. Among the adjacent gene transcripts, GOs associated with response to stimulus, cellular biosynthetic process, nitrogen compound metabolic process, nucleobase, nucleoside, cellular macromolecule metabolic process, protein metabolic process, transport activity, RNA metabolic process, gene regulation and binding (**Fig.5**).

In order to reduce the number of GO terms, enriched GO categories with false discovery rates (FDR) < 0.05 from AgriGO analysis were submitted to the REVIGO (REduce and Visualize GO) server. Using the Uniprot (http://www.uniprot.org/) as background and the default semantic similarity measure (Simrel), this analysis clearly showed that biological processes associated with metabolism, localization, nitrogen regulation, regulation of transcription were significantly overrepresented among the adjacent genes to SSR containing miRNAs in Arabidopsis (**Fig.6**).

3.5. Taking an analogy with long non-coding RNAs

If we may consider an analogy of this keeping in view of their larger non-coding peers, viz. lncRNAs, we might expect SSRs to be mapped to the lncRNAs as well. What remains a challenge is to see if the miRNAs/lncRNAs have a coding potential of transcripts in noncoding RNA as these are associated with “unknown transcripts” which eventually are unmapped. Can the SSR-miRNAs that code for non-coding elements prove to be real candidates for understanding gene expression in plants underlying to various traits as discussed above? If it were the case, with breakthrough in genome technology in the form of

clustered regulatory interspaced short palindromic repeats/CRISPR-associated protein 9(CRISPR-Cas9) technology (Sander and Joung, 2014; Jain, 2015), it would be interesting to explore probable CRISPR loci that play a role into regulatory roles of these ncRNAs esp. the smaller miRNAs (Yi et al. 2015).

4. Conclusion

In the present study, we discovered total 147 miRNA-SSRs from 169 pre-miRNAs representing full length genomic transcripts of *A. thaliana*. Our result shows that all the di-, tri-, tetra-, penta and hexanucleotide SSRs were located in non-coding repertoire of all the 5 chromosomes of *A.thaliana* (**Fig. 3**). While dinucleotide miRNA-SSRs were found to be higher, hexanucleotide miRNA-SSRs were found to be lowest repeats in the pre-miRNA transcripts. It was observed that miRNA-SSRs flanking primers were larger in number for discovered miRNA-SSRs. We firmly consider these candidates could be extended for experimentation for allelic variation. It is important to know that these miRNA-SSRs serve as a source of highly informative molecular markers and aids as a reference for marker assisted breeding in plants. We hope this first report on genome-wide identification and characterization of miRNA-SSRs in *A. thaliana* could serve as a reference for identifying more sequences from non-coding repertoire of the genomes.

Acknowledgments

AK would like to give his sincere thanks to Mr. Deepak Kumar, Secretary, IT, ST & BT Government of Uttarakhand for encouragement, suggestions and timely help. PKG was awarded a National Academy of Sciences India (NASI) Senior Scientist Platinum Jubilee Fellowship, and INSA Senior Scientist positions during the tenure of which this study was conducted; VG was awarded a Junior Research Fellowship under the same program, and was later awarded the position of SRF/ RA under a DBT project.

Authors Contributions

AK, AC, SKK, and VG performed the data analysis; KPS and MNVPG manually crosschecked the annotation. KPS assisted AK and AC for preparing the first draft. PS, HSB and PKG conceived, supervised, edited, and finalized the manuscript.

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Bartel B, Bartel DP (2003) MicroRNAs: At the root of plant development?. Plant Physiol 132:709-717

Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116:281-297

Bolser DM, Kerhornou A, Walts B, Kersey P (2015) Triticeae resources in Ensembl Plants. Plant Cell Physiol 56:e3

Brandstrom M, Bagshaw AT, Gemmell NJ, Ellegren H (2008) The relationship between microsatellite polymorphism and recombination hot spots in the human genome. Mol Biol Evol 25: 2579-87

Brodersen P, Sakvarelidze-Achard L, Bruun-Rasmussen M, Dunoyer P, Yamamoto YY, Sieburth L, Voinnet O (2008) Widespread translational inhibition by plant miRNAs and siRNAs. Science 320:1185–1190

Brousse C, Liu Q, Beauclair L, Deremetz A, Axtell MJ, Bouché N (2014) A non-canonicial microRNA target site. Nucleic Acids Res 42: 5270-5279

Cardoso-Silva CB, Costa EA, Mancini MC, Balsalobre TW, Canesin LE, Pinto LR, Carneiro MS, Garcia AA, de Souza AP, Vicentini R (2014) De novo assembly and transcriptome analysis of contrasting sugarcane varieties. PLoS One 9:e88462

Chen M, Tan Z, Zeng G, Peng J (2010) Comprehensive analysis of simple sequence repeats in Pre-miRNA. MolBiolEvol 27:2227-2232

Chen X (2004) A microRNA as a translational repressor of APETALA2 in Arabidopsis flower development. Science 303:2022–2025

Chen M, Tan Z, Jiang J, Li M, Chen H, Shen G, Yu R (2009) Similar distribution of simple sequence repeats in diverse completed Human Immunodeficiency Virus Type 1 genomes. FEBS Lett 583:2959-2963

Cuadrado A, Schwarzacher T (1998) The chromosomal organization of simple sequence repeats in wheat and rye genomes. Chromosoma 107:587-594

Dai X, Zhao PX (2011) psRNATarget: a plant small RNA target analysis server. Nucleic Acids Res 39: 155-159

Du Z, Zhou X, Ling Y, Zhang Z, Su Z (2010)agriGO: a GO analysis toolkit for theagricultural community. Nucleic Acids Res. 38(W): 64–70

Eldem V, Okay S, Ünver T (2013). Plant microRNAs: new players in functional genomics. Turk J Agric For 37:1-21

Gahlaut V, Jaiswal V, Kumar A, Gupta PK (2016) Transcription factors involved in drought tolerance and their possible role in developing drought tolerant cultivars with emphasis on wheat (Triticum aestivum L.). Theor Appl Genet 129: 2019-2042

Ganie SA Mondal TK (2015) Genome-wide development of novel miRNA-based microsatellite markers of rice (Oryza sativa) for genotyping applications. Mol Breeding 35:51

Gupta PK (2015) MicroRNAs and target mimics for crop improvement. Curr Sci 108: 1624-1633

Gupta PK, Balyan HS, Sharma PC, Ramesh B (1996) Microsatellites in plants: A new class of molecular markers. Curr Sci 70:45-53

Gupta S, Prasad M (2009) Development and characterization of genic SSR markers in Medicago truncatula and their transferability in leguminous and non-leguminous species. Genome. 52: 761–771

Heesacker A, Kishore VK, Gao W, Tang S, Kolkman JM, Gingle A, Matvienko M, Kozik A, Michelmore RM, Lai Z, Rieseberg LH, Knapp SJ (2008) SSRs and INDELs mined from the sunflower EST database: abundance, polymorphisms, and cross-taxa utility. Theor Appl Genet 117:1021-1029

Jain M (2015) Function genomics of abiotic stress tolerance in plants: a CRISPR approach. Front Plant Sci 6:375

Jeffreys AJ, Murray J, Neumann R (1998) High-resolution mapping of crossovers in human sperm defines a minisatellite associated recombination hotspot. Mol Cell 2: 267-273

Jones-Rhoades MW, Bartel DP, Bartel, B (2006) MicroRNAs and their regulatory roles in plants. Annu Rev Plant Biol 57: 19–53

Jung JH, Seo PJ, Park CM (2009) MicroRNA biogenesis and function in higher plants. Plant Biotechnol Rep 3: 111–126

Kasprzyk A (2011) BioMart: driving a paradigm change in biological data management. Database (Oxford) 13:2011:bar049

Katti MV, Ranjekar PK, Gupta VS (2001). Differential distribution of simple sequence repeats in eukaryotic genome sequences. MolBiolEvol 18: 1161-1167

Khraiwesh B, Zhu JK, Zhu J (2012) Role of miRNAs and siRNAs in biotic and abiotic stress responses in plants. Biochem Biophys Acta 1819:137-148

Kompelli SK, Kompelli VSP, Enjala C, Suravajhala P (2015) Genome-wide identification of miRNAs in pigeonpea (Cajanus cajan L.) Aust J Crop Sci 9:215-222

Kozomara A, Griffiths-Jones S (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res 42(D):68-73

Kujur A, Bajaj D, Saxena M, Tripathi S, Upadhyaya HD, Gowda CL, Singh S, Tyagi A, Jain M, Parida S (2014) An efficient and cost-effective approach for genic microsatellite marker-based large-scale trait association mapping: Identification of candidate genes for seed weight in chickpea. Mol Breed 34: 241–265

Kujur A, Bajaj D, Saxena MS, Tripathi S, Upadhaya HD, Gowada CL, Singh S, Jain M, Tyagi AK, Parida SK (2013) Functionally relevant microsatellite markers from chickpea transcription factor genes efficient genotyping applications and trait association mapping. DNA Res 20: 355-374

Li YC, Korol AB, Fahima T, Beiles A, Nevo E (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. MolEcol 11:2453-2465

Lightfoot DA, Iqbal MJ (2013) Molecular mapping and breeding with microsatellite markers. Methods MolBiol 1006: 297-317

Liu W, Jia X, Liu Z, Zhang Z, Wang Y, Liu Z, Xie W (2015) Development and Characterization of Transcription Factor Gene-Derived Microsatellite(TFGM) Markers in Medicagotruncatula and Their Transferability in Leguminous and Non Leguminous Species. Molecules 20:8759-8771

Llave C, Xie Z, Kasschau KD, Carrington JC (2002) Cleavage of scarecrow-like mRNA targets directed by a class of Arabidopsis miRNAs. Science 297: 2053–2056

Madsen BE, Villesen P, Wiuf C (2008) Short tandem repeats in human exons: a target for disease mutations. BMC Genomics 9:410

Mondal TK, Ganie SA (2014) Identification and characterization of salt responsive miRNA-SSR markers in rice (Oryza sativa). Gene 535:204–209

Ni J, Colowit PM, Mackill DJ (2002) Evaluation of genetic diversity in rice subspecies using microsatellite markers. Crop Sci 42: 601–607

Nicol JW, Helt GA, Blanchard SG Jr, Raja A, Loraine AE (2009) The Integrated Genome Browser (IGB): free software for distribution and exploration of genome-scale datasets. Bioinformatics 25:2730-1

Riley DE, Krieger JN (2009a) Embryonic nervous system genes predominate in searches for dinucleotide simple sequence repeats flanked by conserved sequences. Gene 429:74-79

Riley DE, Krieger JN (2009b) UTR dinucleotide simple sequence repeat evolution exhibits recurring patterns including regulatory sequence motif replacements. Gene 429:80-86

Sander JD, Joung JK (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol 32:347-55

Schlotterer C (2000) Evolutionary dynamics of microsatellite DNA. Chromosoma 109: 5844-5849

Schlötterer C (2004) The evolution of molecular markers-Just a matter of fashion? Nat Rev Genet 5: 63–69

Senan S, Kizhakayil D, Sasikumar B, Sheeja TE (2014) Methods for Development of Microsatellite Markers: An Overview. Not Sci Biol 6:1-13

Supek F, Bosnjak M, Skunca N, Smuc T (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. PloS One 6:e21800

Templeton AR, Clark AG, Weiss KM, Nickerson DA, Boerwinkle E, Sing CF (2000) Recombinational and mutational hot spots within the human lipoprotein lipase gene. Am J Hum Genet 66:69-83

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408:796–815

Toth G, Gaspari Z, Jurka J (2000) Microsatellites in different eukaryotic genomes: survey and analysis. Genome Res 10: 967-981

Usdin K (2008) The biological effects of simple tandem repeats: lessons from the repeat expansion diseases. Genome Res 18:1011-1019

Wang Y, Yang C, Jin Q, Zhou D, Wang S, Yu Y, Yang L (2015) Genome-wide distribution comparative and composition analysis of the SSRs in Poaceae. BMC Genetics 16:18

Yang QS, Gao J, He WD, Dou TX, Ding LJ, Wu JH, Li CY, Peng XX, Zhang S, Yi GJ (2015) Comparative transcriptomics analysis reveals difference of key gene expression between banana and plantain in response to cold stress. BMC Genomics 16:446

Yi X, Zhang Z, Ling Y, Xu W, Su Z (2015) PNRD: a plant non-coding RNA database. Nucleic Acids Res 43:D982-989

You FM, Huo N, Gu YQ, Luo MC, Ma Y, Hane D, Lazo GR, Dvorak J, Anderson OD (2008) BatchPrimer3: a high throughput web application for PCR and sequencing primer design. BMC Bioinformatics 9:253

Young ET, Sloan JS, van Riper K (2000) Trinucleotide repeats are clustered in regulatory genes in Saccharomyces cerevisiae. Genetics 154:1053-1068

Legend

Figure1. Pipeline used for discovery of miRNA-SSRs in *A. thaliana.*

Figure 2 **.**Incidence and number of di, tri, tetra, and pentanucleotide miRNA-SSRs.

Figure 3. Chromosomal locations of discovered miRNA-SSRs in *A.thaliana* geneome.

Figure 4. GO classifications of adjacent genes to SSR containing miRNAs.

Figure 5. **GO analysis of adjacent genes to SSR containing miRNAs:** box reflects the GO term number, the p-value in parenthesis, and GO term. The first pair of numerals shows the number of adjacent genes in the input list associated with that GO term and the number of genes in the input list. The second pair of numerals represents the number of genes associated

with the particular GO term in the TAIR database and the total number of Arabidopsis genes with GO annotations in the TAIR database. The box colours indicates levels of statistical significance with yellow = 0.05; orange = e-05 and red = e-09.

Figure 6. **GO analysis of adjacent genes to SSR containing miRNAs using REVIGO:** The scatter plot represents the cluster representatives (terms remaining after reducing redundancy) in a two-dimensional space derived by applying multi-dimensional scaling to a matrix of GO terms semantic similarities. Bubble color indicates the p-value for the false discovery rates derived from the AgriGO analysis. The circle size represents the frequency of the GO term in the uniprot database (more general terms are represented by larger size bubbles).

Fig 1.

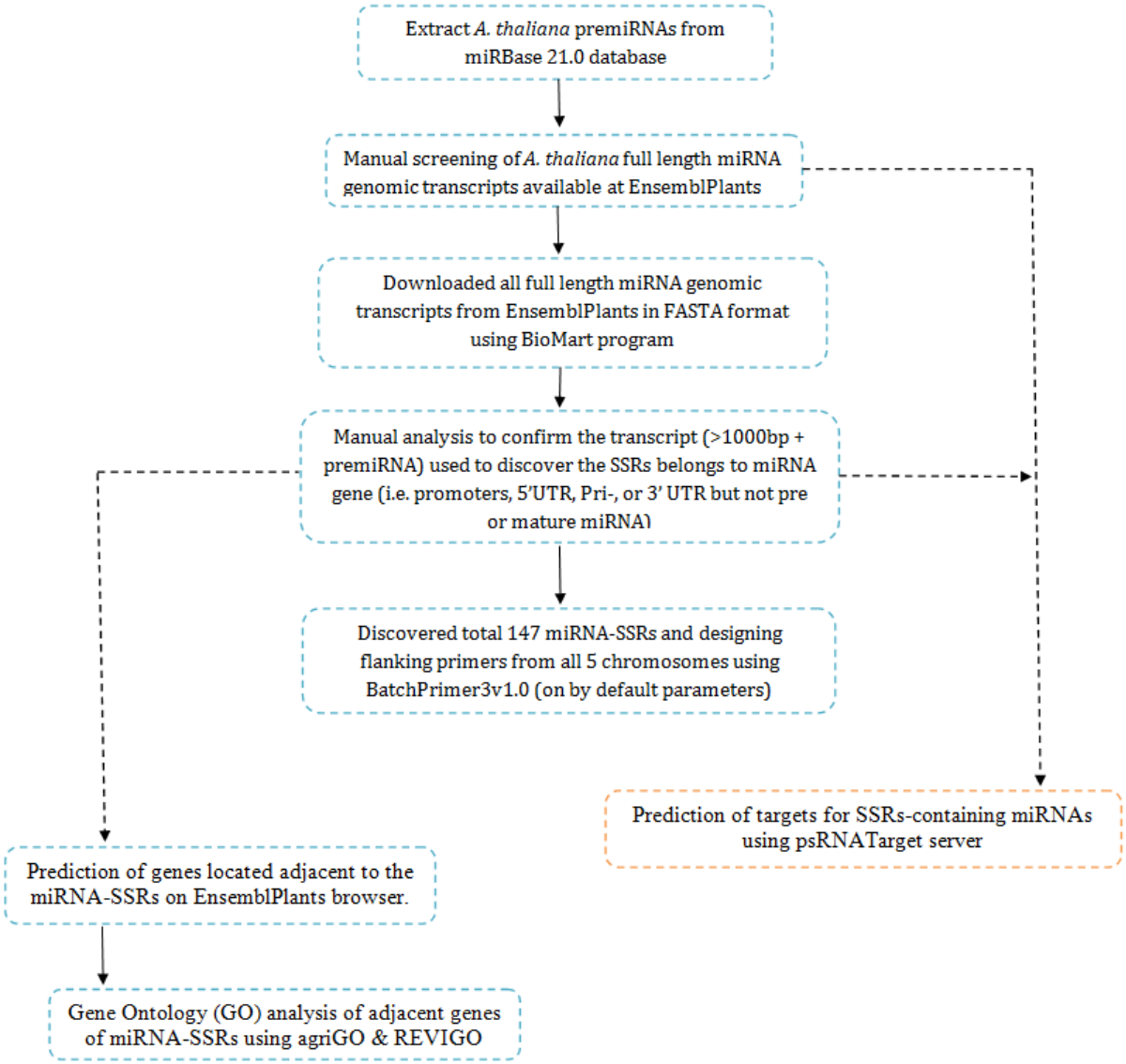


Fig 2.

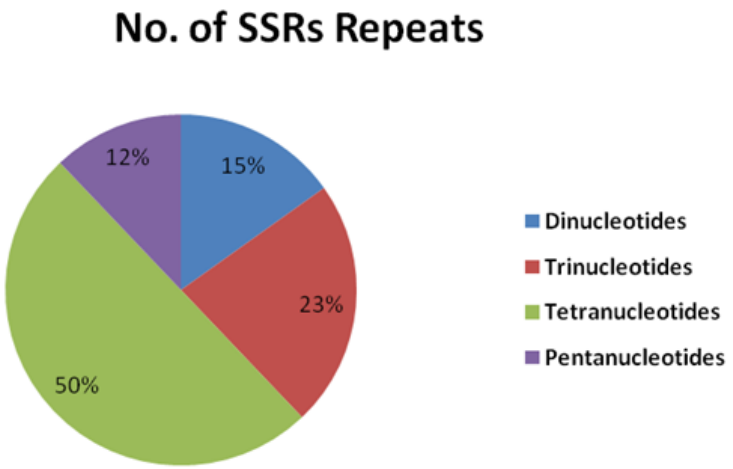

No. of SSRs Repeats
12%
15%
23%
50%
Dinucleotides
Trinucleotides
Tetranucleotides
Pentanucleotides

Fig 3.

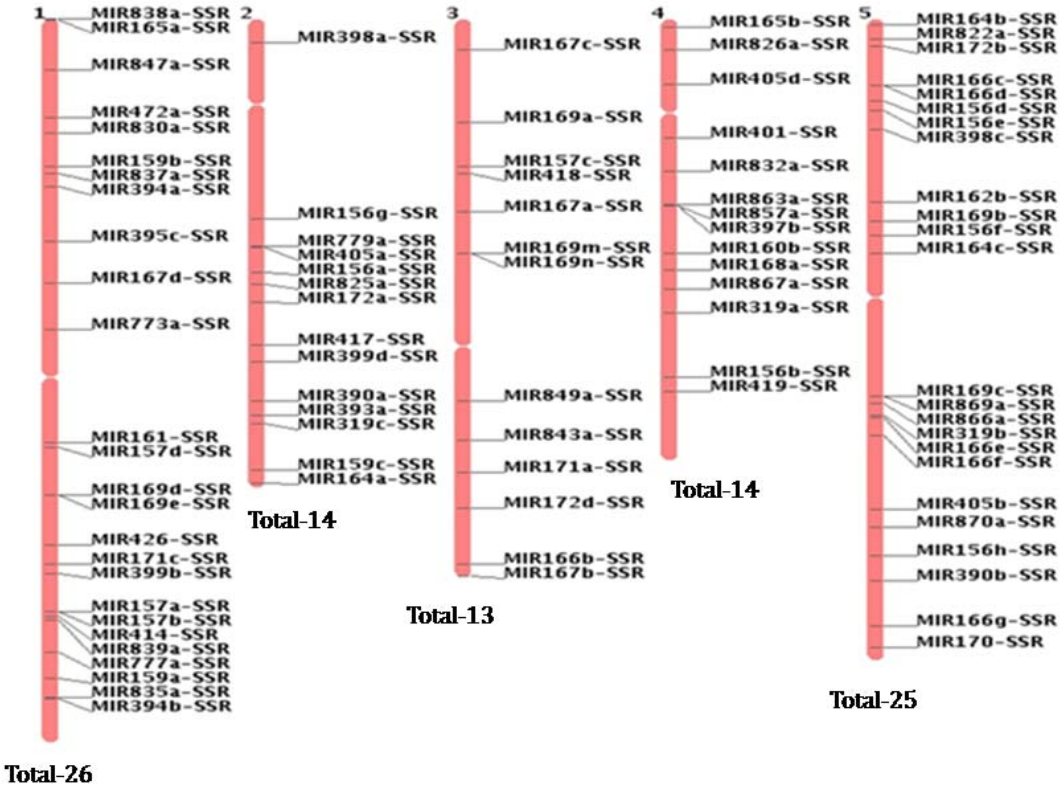

1
MIR838a-SSR
MIR165a-SSR
MIR847a-SSR
MIR472a-SSR
MIR830a-SSR
MIR159b-SSR
MIR837a-SSR
MIR394a-SSR
MIR395c-SSR
MIR167d-SSR
MIR773a-SSR
MIR161-SSR
MIR157d-SSR
MIR169d-SSR
MIR169e-SSR
MIR426-SSR
MIR171c-SSR
MIR399b-SSR
MIR157a-SSR
MIR157b-SSR
MIR414-SSR
MIR839a-SSR
MIR777a-SSR
MIR159a-SSR
MIR835a-SSR
MIR394b-SSR
Total-26
2
MIR398a-SSR
MIR156g-SSR
MIR779a-SSR
MIR405a-SSR
MIR156a-SSR
MIR825a-SSR
MIR172a-SSR
MIR417-SSR
MIR399d-SSR
MIR390a-SSR
MIR393a-SSR
MIR319c-SSR
MIR159c-SSR
MIR164a-SSR
Total-14
3
MIR167c-SSR
MIR169a-SSR
MIR157c-SSR
MIR418-SSR
MIR167a-SSR
MIR169m-SSR
MIR169n-SSR
MIR849a-SSR
MIR843a-SSR
MIR171a-SSR
MIR172d-SSR
MIR166b-SSR
MIR167b-SSR
Total-13
4
MIR165b-SSR
MIR826a-SSR
MIR405d-SSR
MIR401-SSR
MIR832a-SSR
MIR863a-SSR
MIR857a-SSR
MIR397b-SSR
MIR160b-SSR
MIR168a-SSR
MIR867a-SSR
MIR319a-SSR
MIR156b-SSR
MIR419-SSR
Total-14
5
MIR164b-SSR
MIR822a-SSR
MIR172b-SSR
MIR166c-SSR
MIR166d-SSR
MIR156d-SSR
MIR156e-SSR
MIR398c-SSR
MIR162b-SSR
MIR169b-SSR
MIR156f-SSR
MIR164c-SSR
MIR169c-SSR
MIR869a-SSR
MIR866a-SSR
MIR319b-SSR
MIR166e-SSR
MIR166f-SSR
MIR405b-SSR
MIR870a-SSR
MIR156h-SSR
MIR390b-SSR
MIR166g-SSR
MIR170-SSR
Total-25

Fig 4.

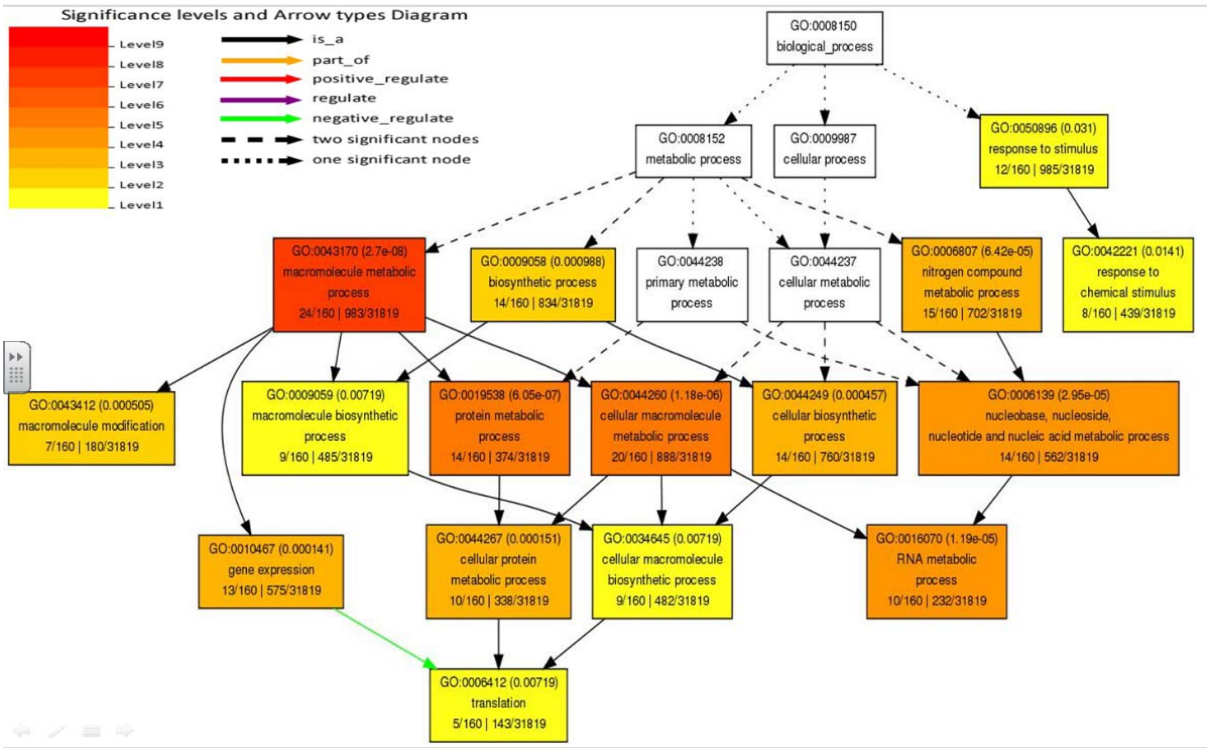

Significance levels and Arrow types Diagram
Level9
Level8
Level7
Level6
Level5
Level4
Level3
Level2
Level1
is_a
part_of
positive_regulate
regulate
negative_regulate
two significant nodes
one significant node
GO:0008150 biological_process
GO:0008152 metabolic process
GO:0009987 cellular process
GO:0050896 (0.031) response to stimulus 12/160 | 985/31819
GO:0043170 (2.7e-08) macromolecule metabolic process 24/160 | 983/31819
GO:0009058 (0.000988) biosynthetic process 14/160 | 834/31819
GO:0044238 primary metabolic process
GO:0044237 cellular metabolic process
GO:0006807 (6.42e-05) nitrogen compound metabolic process 15/160 | 702/31819
GO:0042221 (0.0141) response to chemical stimulus 8/160 | 439/31819
GO:0043412 (0.000505) macromolecule modification 7/160 | 180/31819
GO:0009059 (0.00719) macromolecule biosynthetic process 9/160 | 485/31819
GO:0019538 (6.05e-07) protein metabolic process 14/160 | 374/31819
GO:0044260 (1.18e-06) cellular macromolecule metabolic process 20/160 | 888/31819
GO:0044249 (0.000457) cellular biosynthetic process 14/160 | 760/31819
GO:0006139 (2.95e-05) nucleobase, nucleoside, nucleotide and nucleic acid metabolic process 14/160 | 562/31819
GO:0010467 (0.000141) gene expression 13/160 | 575/31819
GO:0044267 (0.000151) cellular protein metabolic process 10/160 | 338/31819
GO:0034645 (0.00719) cellular macromolecule biosynthetic process 9/160 | 482/31819
GO:0016070 (1.19e-05) RNA metabolic process 10/160 | 232/31819
GO:0006412 (0.00719) translation 5/160 | 143/31819

Fig 5.

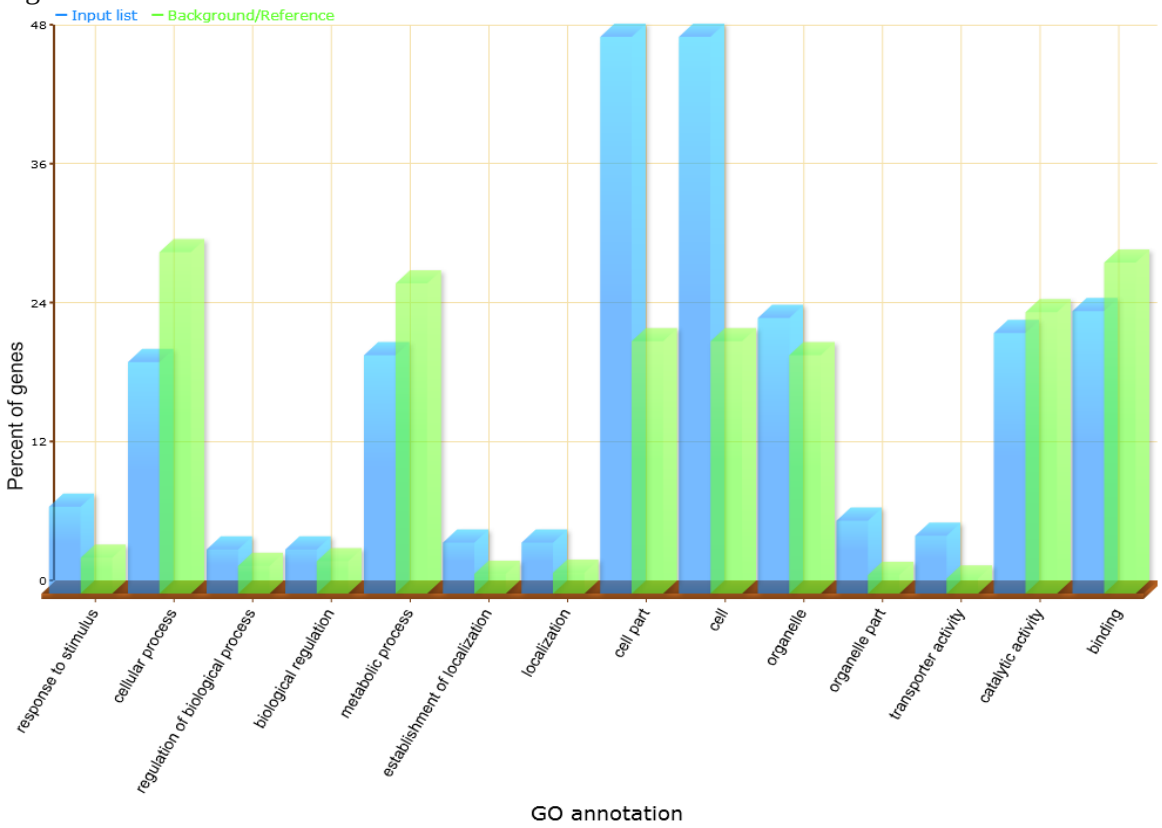
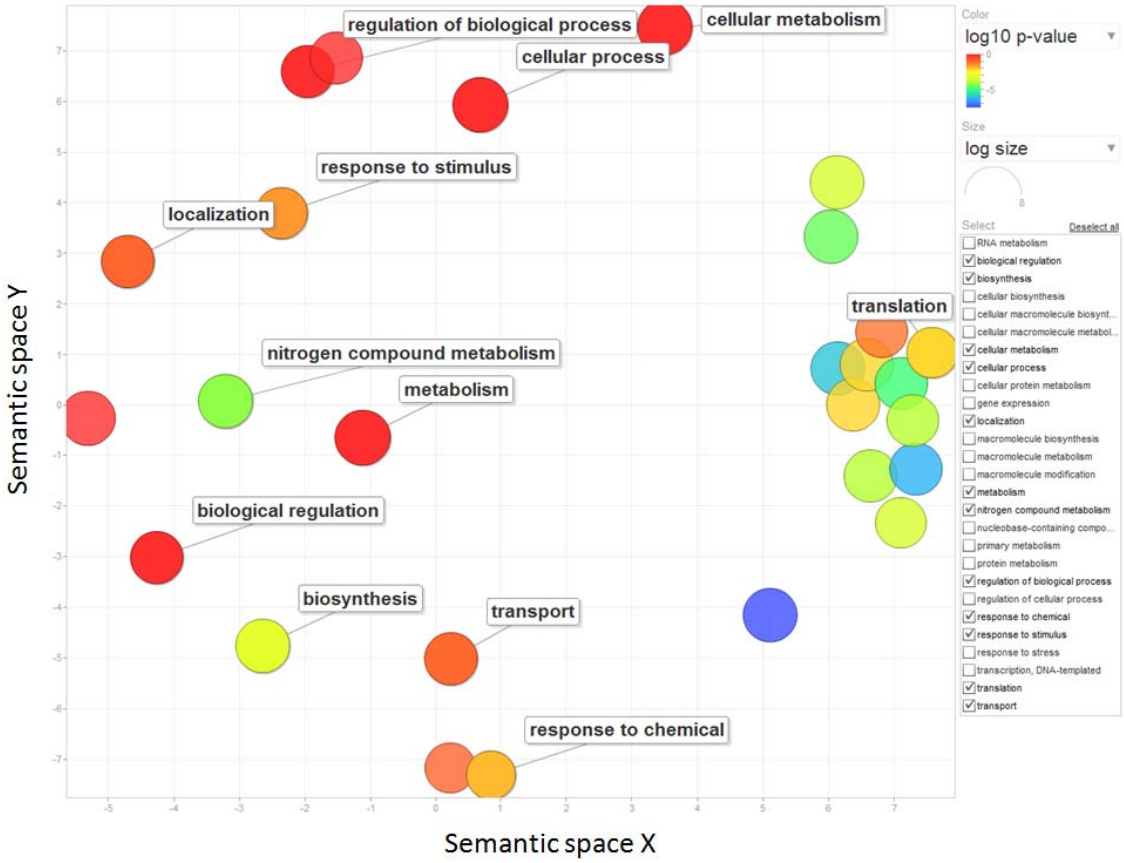

Input list
Background/Reference
48
36
24
12
0
Percent of genes
response to stimulus
cellular process
regulation of biological process
biological regulation
metabolic process
establishment of localization
localization
cell part
cell
organelle
organelle part
transporter activity
catalytic activity
binding
GO annotation

Fig 6.


regulation of biological process
cellular metabolism
cellular process
response to stimulus
localization
translation
nitrogen compound metabolism
metabolism
biological regulation
biosynthesis
transport
response to chemical
Semantic space Y
Semantic space X
Color
log10 p-value
Size
log size
Select
Deselect all
RNA metabolism
biological regulation
biosynthesis
cellular biosynthesis
cellular macromolecule biosynt...
cellular macromolecule metabol...
cellular metabolism
cellular process
cellular protein metabolism
gene expression
localization
macromolecule biosynthesis
macromolecule metabolism
macromolecule modification
metabolism
nitrogen compound metabolism
nucleobase-containing compo...
primary metabolism
protein metabolism
regulation of biological process
regulation of cellular process
response to chemical
response to stimulus
response to stress
transcription, DNA-templated
translation
transport

Tables:

Table 1. Designed flanking primers for discovered miRNA-SSRs using BatchPrimer3 server.

MIRNA	Motif	Primer Sequence		*T*m (^{0}C)	Product Size
		Forward	Reverse		
MIR156a	$(TCTT)_3$	ACAAGAGCCATAAAGAAAGGT	AGGGTTTTTGTCTAAATCGAG	55	154
MIR156b	$(CT)_{11}$	CACCTCTCTTTCTGTCAGTTG	ACACATCACTAGCAAAAGTGC	55	140
MIR156d	$(TTC)_6$	TCTCCATCATCTCTGTTTCAC	GGCTGCTTTACTTCTCTCTCT	55	154
	$(GAGT)_3$	TGTTGGATTCATTCTTCATTC	AAGGAGATAAACTCAGAATTGC	55	176
MIR156e	$(TC)_7$	TTGAAGCTATGTGTGCTTACTC	ACTTTGATCCGTTTGATGATA	55	153
MIR156f	$(CT)_9$	GAAGCTATGTGTGCTCACTCT	GTAAAACCAAAAGAATGGATG	54	139
	$(AT)_8$	ACTCTCTTCTTGTCTCCTGCT	AACCATACACAGAGACGTTTG	55	153
MIR156g	$(GAT)_4$	AAACGTAGCTAGTGGTCAGTG	AAACGTAGCTAGTGGTCAGTG	55	152
	$(TATC)_3$	ACTCTCTTCTTGTCTCCTGCT	AACCATACACAGAGACGTTTG	55	153
	$(CATG)_3$	ACTCTCTTCTTGTCTCCTGCT	AACCATACACAGAGACGTTTG	55	153
MIR156h	$(TTC)_5$	TCATCCTCTTCGCTATAAATG	AGGTTGTGCTCTCTTTCTTCT	55	144
MIR157a	$(TCA)_4$	TTCGTTGCTCTCTATGTGTCT	TTCGTTGCTCTCTATGTGTCT	55	161
	$(AAGT)_3$	GTTCGTAGTCTTCTCAAATCG	AACCATCAAACCTTATGGAAT	55	170
MIR157b	$(TCA)_4$	TAGCTGTCCTCTATGCGTCTA	TCAAGAACTTGATTAAACACCA	55	187
MIR157c	$(TA)_8$	TATTTTCCCTTTGTCACTTCA	GTCAACAACCAACATCACTCT	55	150
	$(AT)_{22}$	TTTGGTAACCTGATCTCCATA	CCAAACTATCAAACCAAACTG	54	137
MIR157d	$(TA)_{13}$	TATGCTTCTGTCATCACCTTT	ACTTTTCTCACACCAAAACAA	55	156
	$(AAAG)_3$	GATGCTATGCAAAACAGACAC	GGTGATGACAGAAGCATAGAG	55	151
MIR159a	$(TC)_6$	ATTTCTTCCAAAACATGACG	CAAAAACACCAAAAGAGGTAA	55	158
MIR159b	$(TC)_7$	TAATGGCTTCACTCTTCTTTG	CTACTCAAGATCCATCATCCA	55	153
	$(AAT)_4$	GACAATAAGATTTACTGCCAAA	AAAGAGCATCAACCCTAGTCT	54	141
MIR159c	$(CAT)_4$	ATAATCGTCCCAAGGAGTAGA	AAACTATGGAAAGAGGGAGAA	55	141
	$(AAAT)_4$	CACCCTAACCGTATCTCTCTC	TCTACTCCTTGGGACGATTAT	55	190
MIR160b	$(TA)_6$	CCAATCATATTTAAGGGTTCC	TTGGTCATGCTTGACTACTCT	55	150
MIR161	$(AGA)_4$	CTTTGTTTGAGATTGCATCAT	TGACTACCAGTCTACCACTATGT	55	158
MIR162b	$(TTTA)_3$	GTTTGTTCATCAACCGATTT	TCGATTCTTGCTTTTGTAAAC	55	153
	$(TTGT)_3$	GATTCGATAAAGTCTTCTCAGC	TGATCTGTTACCCAAAACAAT	55	173

MIR164a	$(CA)_9$	TTGCCTTACGTAAAACACACT	TGAGAACTTTGGTTATGGAAA	38	137
	$(AC)_7$	No SSR flanking primer found			
MIR164b	$(TA)_7$	GGAATCACGTTTTCAAATATC	AAGTGCGAGTGTTGTTTATGT	54	149
	$(TC)_{10}$	ATCATACCCCCAAGGTAACTA	ATTCTCTCCGACCACATAACT	55	153
	$(TG)_6$	AGTTATGTGGTCGGAGAGAAT	TCATCCATATCATCACACTCA	55	165
	$(ACAT)_3$	ATCATACCCCCAAGGTAACTA	ATTCTCTCCGACCACATAACT	55	153
	$(TACG)_3$	GAGGAAGTAGATACCCTCTGC	GATCAAGATGCGTGATCATA	54	135
MIR164c	$(TTTA)_4$	No-SSR flanking primer found			
MIR165a	$(AT)_7$	ACTATGAAACCTTCACGCATA	CCTCATCATAACACCATCATC	54	154
	$(CT)_7$	CCTCATCATAACACCATCATC	TAATATCCTCGATCCAGACAA	55	157
MIR165b	$(AT)_7$	No-SSR flanking primer found			
	$(TC)_7$	ACGACTTATTTCAGCCTTCTT	GCAGCTCAATCTTATGTGAGT	55	155
	$(TC)_{14}$	TTTGGCTCTCTCTCCACTTAC	GGCTAAGATCAAGGAAGAGAA	56	146
	$(AAG)_5$	TTCTCTTCCTTGATCTTAGCC	AGAAAAATCCCTCTTTAAATCC	55	159
MIR166c	$(TC)_6$	CACGTCACAATTCACATCTTA	TTAAGTCTCGTCGTTGTCTTC	54	161
	$(TCT)_4$	GTGGTCATTTGTGCCTCTAT	CCACGTTATCAAGAAGAGAAA	55	150
	$(CTTT)_3$	CACTCGAATTAATTTGGAAGA	GGTCGCAAGATAGAACAAATA	55	150
MIR166d	$(CTT)_7$	AATATTCGCCTCACACATAGA	TCAATCTACCGATCTTCTTCA	55	141
MIR166e	$(TC)_8$	CCCTCTCTTCTTTTCATCATT	CTCAAAAGGAAAAGCTTCACT	55	152
MIR166f	$(GAT)_5$	GTCTTTCTGAGCCAAAAGTTC	CTTGAAGATTGAGAAGCAAAA	56	146
MIR166g	$(CTT)_5$	TAGGGCTTAGATCTTTTGTCC	AACCCTAAATCGCTTCACTAT	55	162
MIR167a	$(AAAG)_4$	CCAAAAACCAAGAATAAGAAGA	CCAAAAACCAAGAATAAGAAGA	55	162
MIR167b	$(GA)_{11}$	TGGAGTCAAACTAAGAATGGA	TATATCTCCACCACCTGTGAC	55	173
	$(CT)_7$	TCACAGGTGGTGGAGATATAG	TTAAAGAAGCCTGAAACAGTG	55	150
MIR167c	$(AG)_7$	AGCATGATCTTGTCTTCCTCT	TCTCCTTCATGCTACAATCAT	55	158
	$(AGA)_7$	GAGAGAGACTAGGTCATGCTG	TTCATGAGATCCTCTTTCTGA	54	129
MIR167d	$(TG)_7$	AACAAGGATCTGTGTAACGTG	GAAAAATGCTCAGCTTGATAA	55	152
	$(GT)_7$	ATGTATGTGGTGTGTGTGTCA	GAGGGATCGTAAAAGTTAAGG	55	157
MIR168a	$(CTT)_4$	AGTGTGAAAGCGAAAATCTCT	TAATGGGGAAATGAGGTTTAT	56	157
	$(AATA)_3$	CACGTGCTTCTCAAAAAGATA	GTCTCTTTTCACCCGAGAGT	55	186

MIR169b	$(TTCA)_3$	TGAACATATTTCTGGCAAGTT	CTCATACGGTCGATGTTAATC	55	134
MIR169c	$(TTTA)_3$	TTGAGATGCTAAAGTAGAGCAA	CGAAGTTGAATTTTGACATTG	55	178
	$(TTAT)_4$	GGCTCAACATGTAGGAAAGTA	GATTGGAGCAAACTAAACTCTT	55	167
MIR169d	$(CGAT)_3$	TAATACCGAAAACCCAAAACT	CCACCTGTCGTACTTTTCTTA	55	162
MIR169e	$(ATG)_5$	TCATCATGAGTTAGGGTTTTG	TCATCATGAGTTAGGGTTTTG	55	140
	$(ATC)_4$	AAAGATTCCTCCCTTCTTTTT	GCTGCAAGTACAAGTGTTGA	55	160
MIR169m	$(AT)_{14}$	AGATGGACATGACAAGAAAAA	ATCCATGTTCTTCCACAATC	55	165
MIR169n	$(TA)_6$	AAACACGTCTAAAGTTGCATT	GTCGGTTCATTCACTAAATTG	55	144
	$(AT)_{14}$	AGATGGACATGACAAGAAAAA	AGATGGACATGACAAGAAAAA	55	165
MIR170	$(CTT)_4$	GTGCATTGAGAGTAGCAGAGT	GGACTCTCTCGGAAACATAGT	55	157
MIR171a	$(AG)_6$	TTGAGGTTTTGTAAAAAGCAG	ATAAATTTTGAGGGAATCTCG	55	139
	$(AGAA)_4$	GCAGAGAAAGAGAGAGAGAGG	ATCGATGAAGATGCTTTGTAA	55	142
MIR171c	$(TCAC)_3$	GCCCAATGTTATAAAGGGTAG	GACACCTTCAATTTCGTGATA	56	172
MIR172a	$(TC)_{11}$	ACAGTCACATCTCTTACTGTGC	TTGGAAGCCATATATTAACCA	55	118
	$(CT)_7$	TGATTCACTCTCCACAAAGTT	ACCTACCTGAAGAAGATCTGG	55	142
	$(GTTTGA)_5$	TGAAGGTACGAGTTTCTAGTGTC	CGGAAATTAGTCTTCCATTTT	55	182
MIR172b	$(TTC)_4$	TCTTATGACGTAAAAGGACCA	TTCGATCTCTATTTTCTTGGA	55	171
MIR172d	$(CT)_9$	GTATCTTCGATTACGATGTGC	GGAAGAGATTTAGGGTGAAGA	55	155
	$(TA)_6$	TCAGAAATCCAGATCCTCATA	ATCATTCATCATCGTTTTGTC	55	163
MIR319a	$(CT)_6$	ATCTACCATCCCTTTTCTACG	AGAGATGGGAAAAGAAGATGA	55	144
	$(ATAC)_3$	ATCTACCATCCCTTTTCTACG	AGAGATGGGAAAAGAAGATGA	55	144
	$(ATAC)_3$	GTTCCAAACGCTCTATCTCTT	CGAAAAACCATGATTTAGAAG	55	154
MIR319b	$(AATG)_3$	CCAAAATTCAAACTAGACTCG	TAGTGGATCAAGCATGTTTTT	54	157
	$(AATG)_3$	TCCACTCATGGAGTAATATGTG	CTTCAGTCCAAGCATAGAGAA	55	146
MIR319c	$(AAT)_5$	TCTTCGGTTATGACGACTATG	AATAAATCAGGGAGGAAAATG	55	148
	$(ATA)_4$	TCTTCGGTTATGACGACTATG	AATAAATCAGGGAGGAAAATG	55	148
MIR390a	$(ATTA)_3$	GTCGGGTAAGTTTCATCTGTA	GTCGGGTAAGTTTCATCTGTA	54	144
MIR390b	$(TA)_7$	TGTAATATGGGGACACTTAGC	CATCCATAGGTATGCATCTTC	54.	164
	$(TA)_{14}$	GCTATTTCCGAAAACTTTTGT	CAAACTACCAAGTAAGCATGAA	55	155
	$(AAG)_4$	CAACCTTGTATCTCAAGCCTA	AAATCCAATGAAGAAGAAAGC	55	162

MIR393a	$(AAAT)_4$	CGTCTGGTTTACTAGCTCCAT	GATCGTGTTCCTCTTGATTTT	56	149
	$(TTAT)_5$	No SSR-flanking primers found!			
MIR394b	$(TC)_7$	TGCCTCTTTCTCAATCTCATA	CGAATGTAACATCGAGAGGTA	55	149
MIR395c	$(TTTGG)_4$	TTTGTTTACACCCAAACCTAA	AATGCGAGTGACAGTCATTAT	55	133
MIR397b	$(TTTTA)_3$	ATGAAGAAAACACCCAAAAAG	TCTCCACAATAGTCACGCTAC	56	148
MIR398a	$(TCT)_4$	CCAAAACCAACTAAAACTGAA	GCTTTGGAATAAACAGAGGAG	55	134
MIR398c	$(CTT)_4$	GTACGAGTATCCGTAGAGCAG	AAACTCGAACCAGAACAAACT	55	151
	$(TGTTG)_3$	ATCAGTTTCGCAGTACACAAT	CACAACAAATGATGAAAGGAT	55	159
MIR399b	$(CATG)_3$	AAAAATGACATGGTGTACTCA	TTCAGAGAGGGTTGTTTGATA	53	146
MIR399d	$(TTG)_4$	AACACAATCGTCTTTCATCAC	TGGTTCTTTCTTTCTTTCCTC	55	138
MIR399d	$(TTCT)_3$	TCATACGGTTCTCGAAGAATA	GCAACTCAAAATTTGTGAAAC	55	146
	$(GAAA)_3$	GATTCTTTCTTTCTTCTGTTGG	TAAGGAATGGTTGATGACACT	55	147
MIR401	$(TA)_{11}$	CCAACATTCAAGATCCTTCTA	CAAGTTCCCCTTTGTTTACTC	55	151
MIR405a	$(AACCC)_3$	TTGTTACTAGGGGTGTCAAAA	CCCATCAAATGAAATGAGTTA	55	144
MIR405b	$(GTTGG)_3$	CCCATCAAATGAAATGAGTTA	TTAAGTTCATTCCTGTGGGTA	55	157
MIR405d	$(ATTA)_3$	GATTTTCCCGTCTAAAAATGT	GATGGGTTGAGTTGTTAAATG	55	168
	$(GTTGG)_3$	GGGTCTAACCCATAACTCATT	GCAACATTCTCCTTTTCTTTT	55	168
MIR414	$(CA)_6$	AGTCACACAACCTTTGACATC	AGAGGGCAGATAGAGTTGAAG	55	151
	$(AT)_6$	AGTCACACAACCTTTGACATC	AGAGGGCAGATAGAGTTGAAG	55	151
	$(TTC)_4$	TAATGTTTATCTCCGACTCCA	GCATCCTTAGACCAGTCTTTA	55	145
	$(ATC)_4$	TATTAGATGGTGGTGAGGATG	GATGACGATGATGATGAAGAT	55	134
	$(TCA)_6$	GCTTGAAGTCGAAGATAAAGA	TTGCTTCTCAACTCAAATCTC	54	157
MIR417	$(AAAT)_3$	AGGTTGTACTTATGTGGTGGA	AGATAATGTAGGTGGGAGATACA	55	147
MIR418	$(CAAA)_5$	AGGTGTCAGGTTCTACACAAA	CCAATACATGTGTTAGGATTTTT	55	150
	$(TTTTA)_3$	AAATACCCCAAAAAGAGACAC	AAATACCCCAAAAAGAGACAC	55	146
MIR419	$(TTGC)_3$	GCTGAGGATGTTGTTATTACG	GGTTCATGACTTGTTTTCTTG	55	158
MIR426	$(TAAA)_4$	GTGGACCAAAAGACATACAAT	TGGTGTTGTTTCTTTCCTCTA	54	200
	$(GGGAGG)_3$	TGCAATGGATCAGTTAGAATAG	ATCGTCATGTGGACAAGTATT	55	151
MIR472a	$(TGTA)_3$	AAGGGGAGTCATATTCTCATC	CAAACACCAAAACCTTACAAA	55	200

	$(AAT)_4$	TGTCTAAGAGAGTTTTTAGCAAG	GTTATTGGGCTTTTATTGGAT	53	292
MIR773a	$(TTAT)_3$	CTGGTACATTCATAGTTGTTGC	CAAAACTCTACTCCGTGTTTG	55	151
	$(TTGTT)_3$	No-SSR flanking primer found			
MIR779a	$(TGTTT)_3$	GTTAGCTGAGCAACCATACTT	CTCATTAAGCACAATGCTTTC	54	150
MIR822a	$(TA)_{20}$	GTTTCAGAAAGGGAAAACATT	CGAAATCGAGTTTGTTAATTC	55	202
MIR825a	$(CTAT)_3$	ACAGGTCAATGGTGTTAGAAA	AACTGCACAAAGTCTACAAGC	55	139
MIR826a	$(TGCA)_3$	TTATTATTTGGAGCCATCAAC	GTCTGTTTCTGTGTGATTCGT	55	167
	$(ACAAAT)_3$	CCCTAAAGTATGGGTTCACTT	GCACATGCACATGTACAATAA	55	140
MIR830a	$(TTTTG)_3$	TGACACTTGTTAAAAACTCAGC	TAGCGAGACTCTGGTGAAATA	55	150
	$(TA)_{10}$	No SSR-flanking primers found!			
MIR832a	$(TTTG)_3$	GCGTTGAGTTTAAATTTTCCT	TATTTTCCTCTTCCATTCCTC	55	149
	$(CGTTTC)_3$	AAAAATCGTTTCTCATTTCC	CCTCATCCTTCTAACATTGTG	53	146
MIR835a	$(TTG)_4$	TTATCTAAATCCGTCGTCGT	AAAATTTTCGATCCTGGTG	55	152
	$(TA)_9$	TCTACAGAGGATGGAAAGTCA	ACGAACAAGAAACTGATGAAA	55	157
MIR837a	$(TTC)_4$	No SSR flanking primer found			
	$(TAAA)_3$	TGGAAAAACATGAGGACTTTA	AACATGAAAGAAACAGATCCA	55	210
MIR838a	$(TA)_7$	ATGTTACTCGCTGTTCAACTC	TCAAGGCTTCAAGAATCTACA	55	152
MIR839a	$(CTCA)_3$	CAACTTCTCGGTTGATGTTTA	ATGCTACTCTTTCTGCTCACA	55	165
MIR843a	$(AGA)_4$	ATTAAACCAGCAGTGAAACAA	TGAAGAAGCTAAAGGTTGGAT	55	153
MIR847a	$(TCT)_7$	GACTCGAAGGTTGAAGAAAGT	TATGGTGACGGATTTACAAAG	55	151
MIR849a	$(TTTA)_3$	AGCTTTTCTTCTGGGTTATGT	TGGTCTAGTAGTTGTCCAATCA	55	165
MIR857a	$(TTTTA)_3$	ATGAAGAAAACACCCAAAAAG	TCTCCACAATAGTCACGCTAC	56	148
MIR863a	$(TATT)_3$	GGGGAAAACTCTTTCTTATGT	CTCTCAATCGCATTGGTATAA	54	213
MIR866a	$(ATC)_4$	TTTTCTCTTTCGACTCCTCTT	TCAAGGGTGTGAATCATTTAG	55	155
	$(ATTA)_3$	AACATCAAACCAACTTTCTGA	TCAATTGTCTTTTCGAATCTC	55	166
MIR867a	$(AAG)_4$	CAAAACTGATTTAAAGTTTGTGG	TGTCTATTGGGCTTACAAGAA	56	152
	$(GAA)_4$	AAAAGAAGAAGAAGAACGATG	TGATATTGGGCATTTGTCTAT	55	127
MIR869a	$(AT)_{10}$	TAACAGTATTCGTGGGAAAAA	CTTATCCAACAACTACCACCA	55	149
	$(AT)_6$	TGGTGGTAGTTGTTGGATAAG	AGGAGTTTTCTCAAGAAGGTG	55	153
MIR870a	$(TCT)_4$	AAACAATCGATCAACATCATC	CAAAAATTTCAAATCCCATC	55	154

$(AGA)_4$	TTCGTAAAGAAACATTTGGTC	TGTTGCAAATGTTAGGAGTCT	55	152

Table 2. Genes located adjacent to the miRNA-SSRs.

miRNA	Accession Number	Chr no	5' UTR genes	Gene Description	3' UTR genes	Gene Description
MIR838a	AT1G01046	Chr 1	AT1G01040	Encodes a Dicer homolog.	AT1G01070	Nodulin MtN21-like transporter family protein
MIR165a	AT1G01183	Chr 1	AT1G01180	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	AT1G01210	DNA-directed RNA polymerase
MIR847a	AT1G07051	Chr 1	AT1G07050	CCT motif family protein	AT1G07060	Unknown protein
MIR472a	AT1G12294	Chr 1	AT1G12290	Disease resistance protein (CC-NBS-LRR class) family	AT1G12300	Tetratricopeptide repeat (TPR)-like superfamily protein
MIR830a	AT1G14071	Chr 1	AT1G14060	GCK domain-containing protein	AT1G14090	Pseudogene
MIR159b	AT1G18075	Chr 1	AT1G18070	Translation elongation factor EF1A/initiation factor IF2gamma family protein	AT1G18080	Encodes the Arabidopsis thaliana homolog of the tobacco WD-40 repeat ArcA gene
MIR837a	AT1G18879	Chr 1	AT1G18871	Unknown protein; LOCATED IN: endomembrane system	AT1G18880	NITRATE TRANSPORTER
MIR394a	AT1G20375	Chr 1	AT1G20370	Pseudouridine synthase family protein	AT1G20380	Prolyl oligopeptidase family protein
MIR395c	AT1G26985	Chr 1	AT1G26976	Unknown protein; FUNCTIONS IN: molecular_function unknown	AT1G26990	Transposable element gene
MIR167d	AT1G31173	Chr 1	AT1G31166	Transposable element gene	AT1G31175	Unknown protein
MIR773a	AT1G35501	Chr 1	AT1G35500	Unknown protein	AT1G35510	O-fucosyltransferase family protein
MIR161	AT1G48267	Chr 1	AT1G48260	Encodes a member of the SNF1-related kinase (SnRK) gene family	AT1G48270	Unknown protein

MIR157d	AT1G48742	Chr 1	AT1G48740	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	AT1G48745	Unknown protein
MIR169d	AT1G53683	Chr 1	AT1G53660	Nucleotide/sugar transporter family protein	AT1G53687	MICRORNA169E
MIR169e	AT1G53687	Chr 1	AT1G53683	Encodes a microRNA that targets several HAP2 family members	AT1G53690	Protein of unknown function that is homologous to At5g41010
MIR426	AT1G60025	Chr 1	AT1G60020	Transposable element gene	AT1G60050	Nodulin MtN21-like transporter family protein
MIR171c	AT1G62035	Chr 1	AT1G62030	Cysteine/Histidine-rich C1 domain family protein	AT1G62045	BEST Arabidopsis thaliana protein match is: ankyrin repeat family protein (TAIR:AT1G11740.1)
MIR399b	AT1G63005	Chr 1	AT1G62981	Protein of unknown function (DUF1191)	AT1G63010	Major Facilitator Superfamily with SPX (SYG1/Pho81/XPR1) domain-containing protein
MIR157a	AT1G66783	Chr 1	AT1G66780	MATE efflux family protein	AT1G66790	Unknown protein
MIR157b	AT1G66795	Chr 1	AT1G66790	Unknown protein	AT1G66800	Unknown protein
MIR414	AT1G67195	Chr 1	AT1G67190	F-box/RNI-like superfamily protein	AT1G67200	Pseudogene
MIR839a	AT1G67481	Chr 1	AT1G67480	Galactose oxidase/kelch repeat superfamily protein	AT1G67510	Leucine-rich repeat protein kinase family protein
MIR777a	AT1G70645	Chr 1	AT1G70640	Octicosapeptide/Phox/Bem1p (PB1) domain-containing	AT1G70650	Ran BP2/NZF zinc finger-like superfamily protein
MIR159a	AT1G73687	Chr 1	AT1G73680	Encodes an alpha dioxygenase	AT1G73690	CYCLIN-DEPENDENT KINASE D1
MIR835a	AT1G76062	Chr 1	AT1G76050	Pseudouridine synthase family protein	AT1G76065	LYR family of Fe/S cluster biogenesis protein
MIR394b	AT1G76135	Chr 1	AT1G76120	Pseudouridine synthase family protein	AT1G76140	Prolyl oligopeptidase family protein
MIR398a	AT2G03445	Chr 2	AT2G03430	Ankyrin repeat family protein	AT2G03460	Galactose oxidase/kelch repeat superfamily protein
MIR156g	AT2G19425	Chr 2	AT2G19420	Unknown protein	AT2G19415	Hydroxyproline-rich glycoprotein family protein

MIR779a	AT2G22496	Chr 2	AT2G22482	Unknown protein	AT2G22510	Polynucleotidyl transferase
MIR405a	AT2G22668	Chr 2	N/A	N/A	N/A	N/A
MIR156a	AT2G25095	Chr 2	AT2G25090	Encodes a member of the SNF1-related kinase (SnRK) gene family	AT2G25100	Polynucleotidyl transferase
MIR825a	AT2G26211	Chr 2	AT2G26210	Ankyrin repeat family protein	AT2G26215	Transposable_element_gene
MIR172a	AT2G28056	Chr 2	AT2G28053	Transposable element gene	AT2G28060	5'-AMP-activated protein kinase beta-2 subunit protein
MIR417	AT2G32273	Chr 2	AT2G32240	Unknown protein	AT2G32275	Expressed protein
MIR399d	AT2G34202	Chr 2	AT2G34200	RING/FYVE/PHD zinc finger superfamily protein	AT2G34210	Transcription elongation factor Spt5
MIR390a	AT2G38325	Chr 2	AT2G38304	Unknown protein	AT2G38330	MATE efflux family protein
MIR393a	AT2G39885	Chr 2	AT2G39870	Unknown protein	AT2G39900	Encodes a member of the Arabidopsis LIM proteins
MIR319c	AT2G40805	Chr 2	AT2G40802	Unknown protein	AT2G40815	Calcium-dependent lipid-binding (CaLB domain) family protein
MIR159c	AT2G46255	Chr 2	AT2G46250	Myosin heavy chain-related	AT2G46260	Encodes a member of the Arabidopsis LIM proteins
MIR164a	AT2G47585	Chr 2	AT2G47570	Ribosomal protein L18e/L15 superfamily protein	AT2G47610	Ribosomal protein L7Ae/L30e/S12e/Gadd45 family protein
MIR167c	AT3G04765	Chr 3	AT3G04760	Pentatricopeptide repeat (PPR-like) superfamily protein	AT3G04780	Thioredoxin-like protein
MIR169a	AT3G13405	Chr 3	AT3G13403	Encodes a defensin-like (DEFL) family protein.	AT3G13410	Unknown protein
MIR157c	AT3G18217	Chr 3	AT3G18215	Protein of unknown function, DUF599	AT3G18220	LIPID PHOSPHATE PHOSPHATASE 4
MIR418	AT3G18895	Chr 3	AT3G18890	NAD(P)-binding Rossmann-fold superfamily protein	AT3G18900	FUNCTIONS IN: molecular_function unknown

MIR167a	AT3G22886	Chr 3	AT3G22870	F-box and associated interaction domains-containing protein	AT3G22910	ATPase E1-E2 type family protein / haloacid dehalogenase-like hydrolase family protein
MIR169m	AT3G26818	Chr 3	AT3G26816	Encodes a microRNA that targets several HAP2 family members	AT3G26819	MICRORNA169N
MIR169n	AT3G26819	Chr 3	AT3G26818	Encodes a microRNA that targets several HAP2 family members	AT3G26820	Esterase/lipase/thioesterase family protein
MIR849a	AT3G44444	Chr 3	AT3G44440	unknown protein	AT3G44450	unknown protein
MIR843a	AT3G48057	Chr 3	AT3G48050	'SHUTTLE' IN CHINESE, SUO	AT3G48058	pseudogene of Rac-like GTP-binding protein
MIR171a	AT3G51375	Chr 3	AT3G51370	Protein phosphatase 2C family protein	AT3G51390	DHHC-type zinc finger family protein
MIR172d	AT3G55512	Chr 3	AT3G55490	GINS complex protein	AT3G55520	FKBP-like peptidyl-prolyl cis-trans isomerase family protein
MIR166b	AT3G61897	Chr 3	AT3G61870	unknown protein	AT3G61898	unknown protein
MIR167b	AT3G63375	Chr 3	AT3G63360	Encodes a defensin-like (DEFL) family protein.	AT3G63380	ATPase E1-E2 type family protein / haloacid dehalogenase-like hydrolase family protein
MIR165b	AT4G00885	Chr 4	AT4G00880	SAUR-like auxin-responsive protein family	AT4G00890	Encodes a putative glycosyl hydrolase family 10 protein (xylanase).
MIR826a	AT4G03039	Chr 4	AT4G03030	Galactose oxidase/kelch repeat superfamily protein	AT4G03038	Unknown gene
MIR405d	AT4G05508	Chr 4	N/A	N/A	N/A	N/A
MIR401	AT4G08116	Chr 4	N/A	N/A	N/A	N/A
MIR832a	AT4G10345	Chr 4	AT4G10330	Glycine-rich protein	AT4G10360	TRAM
MIR863a	AT4G13494	Chr 4	AT4G13495	Unknown gene	AT4G13500	Unknown protein
MIR857a	AT4G13554	Chr 4	AT4G13550	Triglyceride lipases	AT4G13555	MICRORNA397B

MIR397b	AT4G13555	Chr 4	AT4G13554	Encodes a microRNA that targets a Laccase family member	AT4G13575	unknown protein
MIR160b	AT4G17788	Chr 4	AT4G17780	F-box and associated interaction domains-containing protein	AT4G17790	SNARE associated Golgi protein family
MIR168a	AT4G19395	Chr 4	AT4G19390	Uncharacterised protein family (UPF0114)	AT4G19400	Profilin family protein
MIR867a	AT4G21362	Chr 4	AT4G21360	Transposable element gene	AT4G21363	transposable element gene
MIR319a	AT4G23713	Chr 4	AT4G23690	Encodes a homodimeric all-beta dirigent protein in the superfamily of calycins	AT4G23720	Protein of unknown function (DUF1191)
MIR156b	AT4G30972	Chr 4	AT4G30970	Unknown protein	AT4G30975	Unknown gene
MIR419	AT4G32445	Chr 4	AT4G32440	Plant Tudor-like RNA-binding protein	AT4G32450	Pentatricopeptide repeat (PPR) superfamily protein
MIR164b	AT5G01747	Chr 5	AT5G01740	Nuclear transport factor 2 (NTF2) family protein	AT5G01750	Protein of unknown function (DUF567)
MIR822a	AT5G03552	Chr 5	AT5G03550	TRAF-like family protein	AT5G03555	NUCLEOBASE CATION SYMPORTER 1
MIR172b	AT5G04275	Chr 5	AT5G04270	DHHC-type zinc finger family protein	AT5G04280	ATRZ-1C
MIR166c	AT5G08712	Chr 5	AT5G08710	Regulator of Chr condensation (RCC1) family protein	AT5G08720	CONTAINS InterPro DOMAIN/s: Streptomyces cyclase/dehydrase (InterPro:IPR005031)
MIR166d	AT5G08717	Chr 5	AT5G08710	Regulator of Chr condensation (RCC1) family protein	AT5G08720	CONTAINS InterPro DOMAIN/s: Streptomyces cyclase/dehydrase (InterPro:IPR005031)
MIR156d	AT5G10945	Chr 5	AT5G10946	Unknown protein	AT5G10950	Tudor/PWWP/MBT superfamily protein
MIR156e	AT5G11977	Chr 5	AT5G11970	Protein of unknown function (DUF3511)	AT5G11980	Conserved oligomeric Golgi complex component-related / COG complex component-related
MIR398c	AT5G14565	Chr 5	AT5G14560	Unknown protein	AT5G14580	polyribonucleotide nucleotidyltransferase

MIR162b	AT5G23065	Chr 5	AT5G23035	Encodes a defensin-like (DEFL) family protein.	AT5G23070	Thymidine kinase
MIR169b	AT5G24825	Chr 5	AT5G24820	Eukaryotic aspartyl protease family protein	AT5G24830	Tetratricopeptide repeat (TPR)-like superfamily protein
MIR156f	AT5G26147	Chr 5	AT5G26140	LONELY GUY 9 (LOG9)	AT5G26146	Potential natural antisense gene
MIR164c	AT5G27807	Chr 5	AT5G27800	Class II aminoacyl-tRNA and biotin synthetases superfamily protein	AT5G27810	MADS-box transcription factor family protein
MIR169c	AT5G39635	Chr 5	AT5G39630	Vesicle transport v-SNARE family protein	AT5G39640	Putative endonuclease or glycosyl hydrolase
MIR869a	AT5G39693	Chr 5	AT5G39670	Calcium-binding EF-hand family protein	AT5G39730	AIG2-like (avirulence induced gene) family protein
MIR866a	AT5G40384	Chr 5	AT5G40382	Cytochrome c oxidase subunit Vc family protein	AT5G40400	Pentatricopeptide repeat (PPR) superfamily protein
MIR319b	AT5G41663	Chr 5	AT5G41660	Unknown protein	AT5G41670	6-phosphogluconate dehydrogenase family protein
MIR166e	AT5G41905	Chr 5	AT5G41900	alpha/beta-Hydrolases superfamily protein	AT5G41908	Unknown protein
MIR166f	AT5G43603	Chr 5	AT5G43590	Acyl transferase/acyl hydrolase/lysophospholipase superfamily protein	AT5G43620	Pre-mRNA cleavage complex II
MIR405b	AT5G50717	Chr 5	N/A	N/A	N/A	N/A
MIR870a	AT5G52797	Chr 5	AT5G52790	FUNCTIONS IN: molecular_function unknown	AT5G52780	Protein of unknown function (DUF3464)
MIR156h	AT5G55835	Chr 5	AT5G55830	Concanavalin A-like lectin protein kinase family protein	AT5G55840	Pentatricopeptide repeat (PPR) superfamily protein
MIR390b	AT5G58465	Chr 5	AT5G58450	Tetratricopeptide repeat (TPR)-like superfamily protein	AT5G58480	O-Glycosyl hydrolases family 17 protein

MIR166g	AT5G63715	Chr 5	AT5G63710	Leucine-rich repeat protein kinase family protein	AT5G63720	KOKOPELLI, KPL
MIR170	AT5G66045	Chr 5	AT5G66010	RNA-binding (RRM/RBD/RNP motifs) family protein	AT5G66050	Wound-responsive family protein

534