Plant Omics Journal POJ

POJ 8(6):544-548 (2015) ISSN:1836-3644

Genome-wide *in silico* screening of genetic variability of microRNA genes in seven plant species

Fekonja Simon, Zorc Minja, Kunej Tanja*

Department of Animal Science, Biotechnical Faculty, University of Ljubljana, Groblje 3, 1230 Domzale, Slovenia

***Corresponding author: tanja.kunej@bf.uni-lj.si**

Abstract

MicroRNAs (miRNAs) are non-coding RNAs, about 21 nucleotides in length, with a significant role in post-transcriptional regulation of gene expression. MicroRNA genes play an important role in regulatory functions, affect disease susceptibility and are associated with almost all biological and metabolic processes. It has been shown that miRNA polymorphisms are linked with important plant traits for agricultural production. Genetic variability of plant miRNA genes has been systematically studied only in rice. In this study we used the upgraded online tool miRNA SNiPer version 5.0 (http://integromics-time.com/miRNA-SNiPer/) which now enables analysis in 22 genomes, including seven plant species. Genome-wide *in silico* screening of the seven plant genomes revealed that miRNA genes are most polymorphic in tomato (100%; 70/70) and mouse-ear cress followed by sorghum, maize, rice, grape and false purple brome (2.6%; 8/310). In average 13.8% of miRNA genes have variations within seed region in seven plant species, the highest number in mouse-ear cress, (43.4%; 141/325). Several miRNA genes had long regions of consecutive single nucleotide polymorphisms (SNPs), the longest being *ath-MIR5998a* with 11 consecutive SNPs. Mouse-ear cress miRNA genes also contained a large number of insertions and deletions (indels), multiple nucleotide polymorphisms (MNPs) and substitutions. Our study presents a baseline for further functional research and associational studies to explore the importance of genetic variability of microRNA genes and lead to a new miRNA-based biotechnology, for improving plant yield, quality and tolerance to environmental biotic and abiotic stresses.

Keywords: *In silico;* microRNA (miRNA); multiple nucleotide polymorphisms (MNP); plants; seed region; single nucleotide polymorphism (SNP).
Abbreviations: DNP_double nucleotide polymorphism; HMDD_Human MicroRNA Disease Database; miRNA_microRNA; MNP_ multiple nucleotide polymorphism; SNP_single nucleotide polymorphism; QTL_quantitative trait locus.

Introduction

MicroRNAs (miRNAs) are non-coding RNAs, with an important role in post-transcriptional regulation of gene expression. MicroRNA are about 21 nucleotides in length and contain three distinct regions; the seed region, the mature region and the pre-miRNA region (Zorc et al., 2012). The seed region is the key binding location for translational suppression, resides within the mature miRNA sequence and is located 2-7 or 2-8 consecutive nucleotides from the 5'-end of the miRNA (Hibio et al., 2012). Abundance of miRNA genes, their expression patterns in different tissues or different stages of development in addition to their evolutionary conservation, suggest they play an important role in regulatory functions (Reinhart et al., 2002). An important factor in miRNA target recognition is associated with complementarity with the miRNA seed region that can be affected by polymorphisms. MicroRNA polymorphisms, especially in the seed and mature regions can effect phenotypic variability and disease development (Zorc et al., 2012). MicroRNAs play vital part in almost all biological and metabolic processes, of which many are directly or indirectly connected to important plant traits for agricultural production, such as plant responses to environmental stress and regulation of plant development and plant architecture (Zhou and Luo, 2013). For example, genome wide screening for miRNA polymorphisms in rice revealed that SNP in *osa-miR156* perturbs normal miR156-SPL14 interactions and leads to producing improved plant architecture (Liu et al., 2013). Furthermore, a GG/AA polymorphism in the loop structure of miRNA gene *osa-miR2923* in the japonica and indica rice varieties was reported to be associated with grain length and length-width ratio. Significant association between seed length and the GA/AA polymorphism was found in a population consisting of 72 landraces (Wang et al., 2013). Another study reported that a naturally occurring polymorphism within the mouse-ear cress miRNA gene *ath-MIR164a* influences leaf shape and shoot architecture, with the effects being modified by additional loci in the genome. The predictability of the miRNA:miRNA* duplex is altered by a single base pair substitution in the miRNA complementary sequence. This causes a reduction in the miRNA accumulation, probably because it interferes with precursor processing (Todesco et al., 2012). Bioinformatics tool miRNA SNiPer version 4.0 enables genome-wide screening for genetic variability of 15 animal species (Zorc et al., 2015). Genetic variability of miRNA genes in animal species has been extensively reported in our previous publications (Zorc et al., 2012; Zorc et al., Zorc et al., 2015a; Zorc et al., 2015b). In this study we used the upgraded tool version 5.0, which enables identification of miRNA genetic variability in 22 genomes, including seven plant species. The main objective of this study was to use the upgraded tool to gather information about miRNA genetic variability and to display polymorphisms and their distribution in miRNA genes of seven plant species. Additionally, miRNA genes

with high number of polymorphisms were highlighted. Further research and work in this field could contribute to the development of new potential biomarkers for the plant species, improving potential future applications, such as genome mapping, marker assisted breeding, quantitative trait locus (QTL) analysis and genome association analysis.

Results

The upgraded miRNA SNiPer tool version 5.0 (http://integromics-time.com/miRNA-SNiPer/) enables analysis of genetic variations residing within 22 genomes, including seven plant species: maize (*Zea mays*), grape (*Vitis vinifera*), sorghum (*Sorghum bicolor*), mouse-ear cress (*Arabidopsis thaliana*), tomato (*Solanum lycopersicum*), rice (*Oryza sativa*), and purple false brome (*Brachypodium distachyon*). The seven plant species were selected because their genome assemblies matched and they were therefore included in the miRNA SNiPer tool 5.0. Plant genomes were screened for polymorphisms within miRNA genes and a total of 1803 miRNA genes were utilized in the analysis. The miRNA SNiPer tool displays polymorphisms within pre-mature, mature and seed miRNA regions of the plant miRNAs (Fig. 1).

Based on current versions of genomic databases 49.4% (890/1803) of the miRNA genes in seven plant species are polymorphic. Out of all polymorphic plant miRNA genes, 50.1% (446/890) had polymorphisms within the pre-miRNA regions, 21.9% (195/890) within the mature region and 28% (249/890) within the seed region (Tables 1 and 2). The highest number of polymorphic miRNA genes is found in tomato, mouse-ear cress, sorghum and maize, followed by rice, grape and false purple brome, respectively (Table 1, Fig. 2). The number of polymorphic miRNA genes ranges from 2.6% (8/310) in false purple brome to 100% (70/70) in tomato. Mouse-ear cress miRNA genes are also significantly polymorphic, with 97.8% (318/325) polymorphic miRNA genes. The most polymorphic seed regions are present in mouse-ear cress, in which 43.4% (141/325) of the seed regions are polymorphic. Ath-MIR5998a is an example of a mouse-ear cress miRNA gene with 11 consecutive polymorphisms, including SNPs, indels, DNPs, MNPs, and multiallelic substitutions spanning complete seed and four nucleotides of the mature region (Fig. 1). Similarly, tomato and maize also have very polymorphic seed regions, 30% (21/70) and 28% (44/157) seed regions comprised polymorphisms, respectively. In average 13.8% of seed regions were polymorphic in the seven plant species.

Among collected polymorphisms most were SNPs, followed by indels, and MNPs. Indels were present in maize, rice, tomato and mouse-ear cress. Out of all the indels, in tomato, 37% (70/187), in mouse-ear cress 27% (160/585) and in rice 10% (7/71) are longer than 4 bp. All 38 indels in maize are shorter than 4 bp. In tomato and mouse-ear cress 294 and 28 multiple nucleotide length polymorphisms (MNLPs) were present. Those polymorphisms greatly vary in length, from exchange of one with two nucleotides in *ath-MIR8178* ENSVATH10422654 (GA/C), to the exchange of one with 149 nucleotides in *ath-MIR5651* ENSVATH10412482 (Supplementary Table 1). Similarly, in tomato the length of alleles ranged from exchange of one with two nucleotides in *sly-MIR6026* (vcZWN8OH), to a multiallelic substitution of four alleles with 4/29/30/29 bps in length in *sly-MIR399* (vcZ153U1I) (Supplementary Table 2).

Discussion

Out of 1803 miRNA genes in seven plant species, 49.4% (890/1803) were polymorphic and among them 28% (249/890) had polymorphisms within seed regions. The longest region with consecutive SNPs (11) resides in the mouse-ear cress miRNA gene *ath-MIR5998a*. This miRNA is an interesting candidate for further analysis, because consecutive polymorphisms comprise the complete seed region and part of the mature region. Mouse-ear cress miRNA genes also contained 585 indels, of which 160 are longer than 4 bp. In mouse-ear cress and tomato MNLPs were present. This type of polymorphisms has been previously described in cattle genome and has shown to effect promoter activity. A functional study detected differences in gene expression activity between two MNLP alleles and bioinformatics analysis predicted gain/loss of regulatory binding sites (Jiang et al., 2007). Bioinformatics tools are required to advance and develop rapidly in order to keep up with the fast growing amount of genome information. The current version of the miRNA SNiPer tool combines 15 animal genomes which were already available in SNiPer 4.0, with seven plant genomes added in the version 5.0 and enables genome-wide analysis of miRNA polymorphisms. After searching scientific literature we came to the conclusion that this is the first bioinformatics tool that enables identification of genetic variations within plant miRNA genes. With new releases of the source databases miRBase and Ensembl, the miRNA SNiPer tool can also be upgraded, by adding additional genomes to the tool. The high variability in miRNA SNPs between species is more likely a consequence of differences in the intensity of genome variability research, the continued development of resource databases and ongoing sequencing projects than because of biological reasons. Plant miRNA genes have not been systematically sequenced and SNPs have not yet been validated, which could be a source of errors in the databases.
The high level of polymorphic miRNAs in several plants and the long consecutive SNPs in certain miRNAs are potentially highly interesting for further studies, where the information could be experimentally validated. Functional research and associational studies could also be performed in light of the displayed information for further analysis of different plant species. With the use of such studies and experimental approaches, a new miRNA-based biotechnology could be developed, for improving plant yield, quality and tolerance to environmental biotic and abiotic stresses (Sun, 2012).

It has been previously reported that polymorphic miRNA play a key factor in plant trait variability (Todesco et al., 2012; Wang et al., 2013; Zhou and Luo, 2013). Polymorphisms within miRNA can alter the binding region and change gene regulation, therefore the information about the genetic variability of the seven screened plant genomes serves as a foundation for the research community to experimentally evaluate and research the effects of identified miRNA genetic variations on plant traits. A better understanding of these polymorphisms is vital for advancing in the agricultural, botanical and other related research fields.

Materials and Methods

Our previously developed bioinformatics tool miRNA SNiPer version 5.0 (Zorc et al., 2015b) was used for the

Table 1. Statistics of data obtained using miRNA SNiPer 5.0 in seven plant species.

Statistics	Maize	Grape	Sorghum	Tomato	Rice	False purple brome	Mouse-ear cress
Total number of known miRNA	157	159	205	70	577	310	325
miRNA genes without variations	53	134	55	0	362	302	7
Polymorphic miRNA genes	104	25	150	70	215	8	318
Total number of SNPs within all miRNA genes	1070	122	547	1827	340	95	7433
Number of miRNA genes with SNPs within seed region	44	7	12	21	23	1	141
Number of miRNA with SNPs in mature region (without seed)	32	2	16	20	37	2	86
Number of miRNA with SNPs in pre-mature region (without mature region)	28	16	122	29	155	5	91
Total number of Indels	38	0	0	187	71	0	585
Total number of long Indels (longer than 4 bp)	0	0	0	70	7	0	160

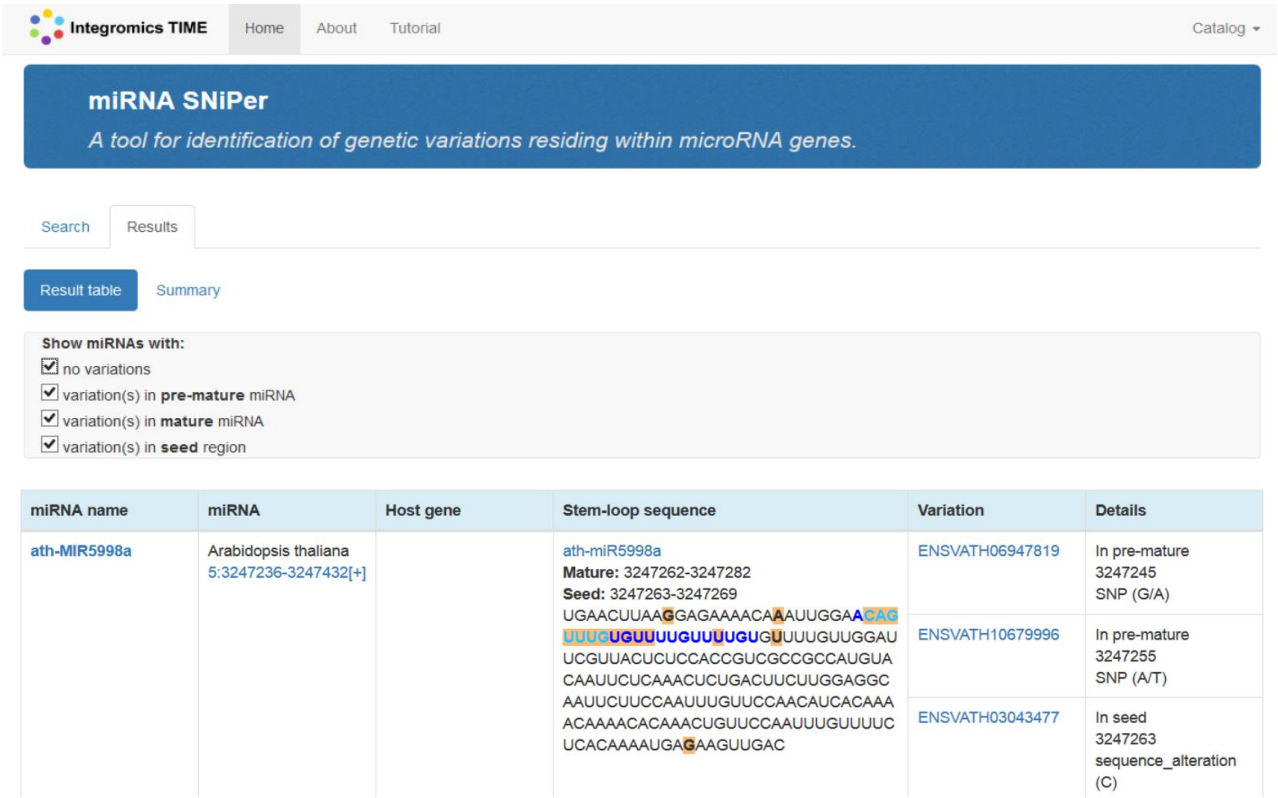

Fig 1. Genetic variability of the *ath-MIR5998a* gene displayed using the miRNA SNiPer tool 5.0.

collection of miRNA polymorphisms. The current version of the tool integrates six source databases: Ensembl Variation database release 78, miRBase 21, TargetScan 6.2, mirTarBase 4.5, Human MicroRNA Disease Database (HMDD) and SNPchiMp v.3. The latest matching database assemblies of MiRBase 21 and Ensembl Variation database, release 78, provide miRNA SNiPer with information regarding miRNA gene location and genetic variability. Data about genome variants such as SNPs and indels are included in the Ensembl Variation database. TargetScan 6.2 was used to obtain the locations of seed miRNA regions. MicroRNA SNPs refer to single nucleotide variations in the miRNA gene sequences. Alteration of two nucleotides compared to the reference sequence is regarded as a double nucleotide polymorphism (DNP), while alterations of three or more consecutive nucleotides are multiple nucleotide polymorphisms (MNPs) (Rosenfeld et al., 2010).

Table 2. Most polymorphic miRNA genes in seven plant species.

MicroRNA name	Number of polymorphisms within: pre-mature region	mature region	seed region	Information regarding consecutive polymorphisms
Maize				
zma-MIR395g	10	2	2	DNP in seed region
zma-MIR156i	6	4	1	-
zma-MIR167h (3p)	7	4	1	3 consecutive SNPs in mature region
Grape				
vvi-MIR172b	2	3	0	-
vvi-MIR3625 (3p)	6	2	1	DNP in pre-mature region
vvi-MIR169n	3	1	1	DNP in pre-mature region
Sorghum Bicolor				
sbi-MIR6228	4	2	3	DNP (one SNP is in the seed region)
sbi-MIR5389	6	2	0	DNP in mature region
sbi-miR6225 (3p)	12	0	2	-
Tomato				
sly-MIR9477 (3p)	33	2	2	DNP in seed, DNP in mature and 5x DNP in pre-mature region and an Indel region
sly-MIR9470 (3p)	60	3	3	3 consecutive SNPs in seed region, 6x DNP in pre-mature region and an Indel region
sly-MIR9479	32	3	2	DNP in mature region, 4x DNP in pre-mature region and an Indel region
Rice				
osa-MIR5544	22	1	4	4 consecutive SNPs in seed region, 4x DNP in pre-mature region
osa-MIR2931	39	6	2	2x DNP in mature region, 4 consecutive SNP's and 10x DNP
osa-MIR1871	29	6	2	2x DNP in mature region, 4 consecutive SNPs and 3 consecutive SNP's
False purple brome				
bdi-MIR7757 (3p.1)	13	0	1	-
bdi-MIR167c	3	0	0	DNP in pre-mature region
bdi-MIR7767	3	0	0	DNP in pre-mature region
Mouse-ear cress				
ath-MIR5998a	4	5	7	11 consecutive SNPs (7 in seed region, 4 in mature region)
ath-MIR8169	18	10	3	DNP in seed region, 7 consecutive SNPs in mature region
ath-MIR829 (5p)	43	11	2	8 consecutive SNPs in mature region
ath-MIR405a	59	6	5	3 consecutive SNPs in seed region, 4 consecutive SNPs (2 in seed, 2 in mature), 7 consecutive in pre-mature region

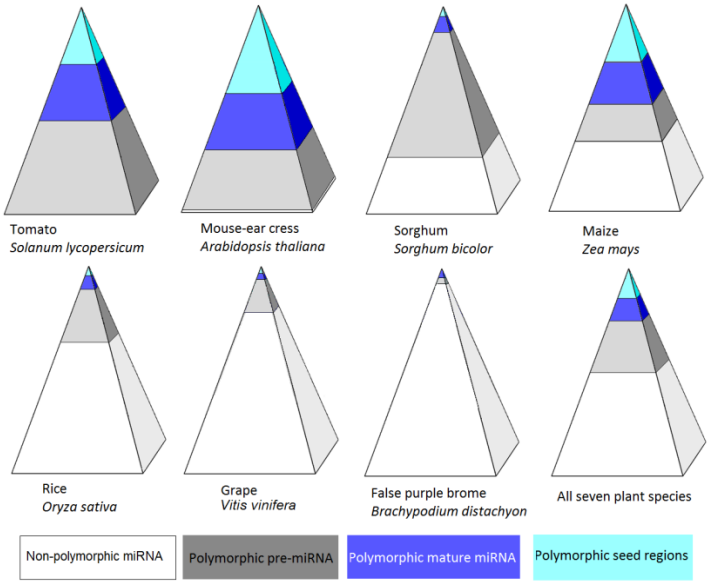

Fig 2. Number of polymorphic pre-miRNA, mature miRNA and seed regions in seven plant species included in the miRNA SNiPer tool 5.0.

Conclusions

Our bioinformatics study presents genetic variability information that can serve as a baseline for further functional research and associational studies to explore the importance of genetic variability of miRNA genes and lead to a new miRNA-based biotechnology, for improving plant yield, quality and tolerance to environmental biotic and abiotic stresses.

Acknowledgements

We thank the anonymous reviewers for helpful comments to the manuscript.

Funding

This work was supported by the Slovenian Research Agency (ARRS) through the Research programme Comparative genomics and genome biodiversity [grant number P4-0220].

References

Hibio N, Hino K, Shimizu E, Nagata Y, Ui-Tei K (2012) Stability of miRNA 5'terminal and seed regions is correlated with experimentally observed miRNA-mediated silencing efficacy. Sci Rep. 2:996.

Jiang Z, Wang Z, Kunej T, Williams GA, Michal JJ, Wu XL, Magnuson NS (2007) A novel type of sequence variation: multiple-nucleotide length polymorphisms discovered in the bovine genome. Genetics. 176: 403-407.

Liu Q, Wang H, Zhu L, Hu H, Sun Y (2013) Genome-wide identification and analysis of miRNA-related single nucleotide polymorphisms (SNPs) in rice. Rice (N Y) 6: 10

Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP (2002) MicroRNAs in plants. Genes Dev. 16: 1616-1626.

Rosenfeld JA, Malhotra AK, Lencz T (2010) Novel multi-nucleotide polymorphisms in the human genome characterized by whole genome and exome sequencing. Nucleic Acids Res. 38: 6102-6111.

Sun G (2012) MicroRNAs and their diverse functions in plants. Plant Mol Biol. 80: 17-36.

Todesco M, Balasubramanian S, Cao J, Ott F, Sureshkumar S, Schneeberger K, Meyer RC, Altmann T, Weigel D (2012) Natural variation in biogenesis efficiency of individual Arabidopsis thaliana microRNAs. Curr Biol. 22: 166-170.

Wang C, Ye J, Tang W, Liu Z, Zhu C, Wang M, Wan J (2013) Loop nucleotide polymorphism in a putative miRNA precursor associated with seed length in rice (Oryza sativa L.). Int J Biol Sci. 9: 578-586.

Zhou M, Luo H (2013) MicroRNA-mediated gene regulation: potential applications for plant genetic engineering. Plant Mol Biol. 83: 59-75.

Zorc M, Skok DJ, Godnic I, Calin GA, Horvat S, Jiang Z, Dovc P, Kunej T (2012) Catalog of microRNA seed polymorphisms in vertebrates. PLoS One. 7: e30737

Zorc M, Obsteter J, Dovc P, Kunej T (2015a) Genetic variability of microRNA genes in 15 animal species. *J Genomics 3: 51-56*.

Zorc M, Omejec S, Tercic D, Holcman A, Dovc P, Kunej T (2015b). Catalog of genetic variants within mature microRNA seed regions in chicken. Poult Sci. 94(9):2037-40.