

Identification of Novel MicroRNAs and their Target Prediction in *Stevia rebaudiana*

Vibha Mandhan and Kashmir Singh*

Department of Biotechnology, Plant molecular biology lab, Punjab University, BMS Block-I, Sector 25, Chandigarh 160014, India

Abstract

Stevia rebaudiana Bertoni, due to its nutritional importance and its emerging use as natural sweetener, is gaining attention worldwide. To better understand the gene regulatory network of this plant, we have initiated high-throughput small RNAs sequencing to discover novel microRNAs (miRNAs). miRNAs are a class of short endogenous non-coding small RNA molecules of about 18-22 nucleotides in length whose main function is to down regulate gene expression in different manners like translational repression, mRNA cleavage and epigenetic modification. We constructed *S. rebaudiana* sRNA library and after sequencing it using illumina genome analyzer II, a total of 30,472,534 reads representing 2,509,190 distinct sequences were obtained. From these reads, twelve 12 novel miRNAs were predicted whose precursors were potentially generated from stevia EST and nucleotide sequences. All novel sequences have not been described earlier in stevia or other plant species. Putative target genes were predicted for most novel miRNAs which include mainly mRNA encoding enzymes regulating essential plant metabolic and signaling pathways. Our result has increased the number of miRNAs in stevia, which should be useful for further investigation into the biological functions and evolution of miRNAs in stevia and other plant species.

Keywords: *Stevia rebaudiana*; Small RNA sequencing; miRNAs; Novel, Target prediction

Introduction

miRNAs are a class of endogenous, small, noncoding, ~22 nucleotides single-stranded RNAs that act as posttranscriptional regulators in eukaryotes [1]. They have reported to be located mostly within noncoding regions of genomes, and usually transcribed from RNA polymerase II promoters [2,3]. The generation of mature miRNA is a complicated enzyme-catalyzed process, from the initial transcript pri-miRNA to the precursor (pre-miRNA) with a characteristic hairpin structure, then a miRNA duplex (miRNA:miRNA*) [4]. In the end, the mature miRNAs function within large complexes to negatively regulate specific target mRNAs. Perfect complementarity between miRNA: mRNA strand generally results in cleavage, such as in plants, whereas imperfect base-pairing leads to translational repression [4,5]. MiRNA genes represent about 1%-2% of the known eukaryotic genomes and constitute an important class of fine-tuning regulators that are involved in several physiological or disease-associated cellular processes [6]. Plant miRNAs target a large number of genes with functions in a range of development processes, including meristem cell identity [7-9], leaf organ morphogenesis and polarity [8,10,11] and floral differentiation and development [12,13]. miRNAs are also reported to be involved in plant responses to biotic and environmental stresses [14-18]. Increasing evidences showed that miRNA repertoire of plant or animal species includes a set of conserved (ancient, abundant) and non-conserved or novel (species specific, recently evolved) miRNAs [16]. However, the biggest challenge in plants is to identify novel miRNAs and to understand their mode of action, role in various metabolic processes. The availability of high throughput next generation sequencing (NGS) technologies such as 454 and Illumina have further revolutionized sRNA discovery. Whereas, transcribed sequences such as ESTs has led to the identification of only conserved (abundant) miRNAs or miRNAs previously identified in other species, on the other hand, NGS provides high throughput tools to make new discoveries of additional species specific or lowly expressed miRNAs in plants irrespective of whether their genome is sequenced or not, e.g. *Arabidopsis* [17-19], *Oryza sativa* [20], *Populus trichocarpa* [21], *Triticum aestivum* [22], *Brachypodium distachyon* [23], *Vitis vinifera* [24], *Arachis hypogaea* [25], *Citrus trifoliata* [26], *Carthamus tinctorius* [27] and many more. NGS strategy for discovery of miRNAs may be successful not only for these plant

species of full genomic and sufficient EST database available, but as well as for those with incomplete genomic information but with sufficient EST sequences available that is the case of our model plant *Stevia rebaudiana*. The latest miRBase release (v20, June 2013) contains 24,521 microRNA loci from 206 species, processed to produce 30,424 mature microRNA products. The rate of deposition of novel microRNAs and the number of researchers involved in their discovery continue to increase, driven largely by sRNA deep sequencing experiments. *Stevia rebaudiana* Bertoni (family Asteraceae) is a perennial herb which accumulates up to 30% (w/w leaf dry weight) diterpenoid steviol glycosides (SGs) [28,29]. SGs are glucosylated derivatives of diterpenoid alcohol steviol. Stevioside and rebaudioside A are the major SGs found in stevia. Sweetness indices of these and other related compounds ranges between 30 to 300 times higher than that of sucrose [30] and are used as non-calorific sweeteners in many countries of the world. In addition to being a sweetener, stevia has been suggested to exert beneficial effects on human health and is considered to have anti-hyperglycemic, anti-hypersensitive, anti-oxidant, anti-tumor/anti-inflammatory, immunomodulatory, antidiarrheal, antimicrobial and anti-rotavirus activities [28]. Stevioside has been reported to lower the postprandial blood glucose level of Type II diabetes patients and blood pressure in mildly hypertensive patients [31,32]. SGs being important compounds of stevia, imparting medicinal properties to the plant, so the molecular studies done on this plant were mainly focused on understanding the biosynthesis and regulation of the genes involved in SG biosynthesis. A control over the regulation of these early and late genes can help to manipulate the diterpenoid contents [33]. Further, the unraveling of sRNA guided circuitry in stevia has further enhanced the value of its

*Corresponding author: SDr Kashmir Singh, Assistant Professor, Department of Biotechnology, Plant molecular biology lab, Punjab University, BMS Block-I, Sector 25, Chandigarh 160014, India, Tel: +91 172 2534085; E-mail: kashmirbio@pu.ac.in, kashmir123@gmail.com

Received April 23, 2015; Accepted May 08, 2015; Published May 11, 2015

Citation: Mandhan V, Singh K (2015) Identification of Novel MicroRNAs and their Target Prediction in *Stevia rebaudiana*. Transcriptomics 2: 107. doi:10.4172/2329-8936.1000107

Copyright: © 2015 Mandhan V, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

gene and EST information and improved our ability to devise strategies to enhance certain essential features of stevia that are less amenable to functional genomics analysis leading to its enhanced nutritive value. The identification of miRNA and their targets is important not only to help us learn more about the roles of miRNAs in stevia development and physiology but also to provide a framework for further designing RNAi based experiments for regulation of gene expression in this species. In our previous study, we generated a sRNA library of *S. rebaudiana* and identified 100 miRNAs belonging to 34 highly conserved families along with 12 novel miRNAs [34]. Genome of *S. rebaudiana* has not been sequenced, thus, it was not possible to perform an extensive study to discover stevia miRNAs using computational analysis with this limited number of available sequences. In recent time, nucleotide sequence information on *Stevia* has increased with the availability of 29,874 cDNA sequences at NCBI database. Thus we used our sRNA library to search new sequences and this led to identification of 12 putative novel miRNAs along with their precursors. A scan of the *S. rebaudiana* EST and nucleotide databases with novel miRNA sequences revealed their putative targets.

Results

Identification of novel stevia miRNAs

In this study, we predicted 12 potential novel miRNAs based on bioinformatics analysis using the stevia EST (5646) and nt (29874) sequences available in NCBI database. Using 36 cycled single end sequencing by genome analyzer II, a total of 30,472,534 sequences were obtained from smRNAome of *S. rebaudiana*. After removing the low quality and adaptor sequences, 17,295,850 sequences were obtained. Among these sequences 15,327,722 sequences ranged from 16-28 nt in length. Then, we removed the sequencing tags that can be annotated as non- coding RNAs including tRNA or rRNAs and then filtered the known miRNA sequences from further analysis. A total of 6,075,552 redundant sequences survived the above filtration steps and

627,678 unique sequences were analyzed for novel miRNA prediction. These unique sRNA sequences were mapped to EST and nt sequences by searching all *S. rebaudiana* sRNAs using the BLAST standalone program. The generated Blast output file showed the following characteristics as given in Table 1. Out of the total 70,081,886 hits to the database (length: 3,637,171) containing 5,646 EST and 29,874 nucleotide sequences (Total: 35,520), sequences better than e-value 10 i.e. 3,186,710 sequences were taken into further consideration (Table 1). Out of these sequences, sequences showing exact matches i.e. 2401 sequences were filtered for further secondary structure prediction. A distinguishing feature of miRNAs is the ability of their pre miRNA sequences to adopt canonical stem loop hairpin structure. These above unique sRNA sequences were then mapped to EST and nt sequences by searching all *S. rebaudiana* sRNAs and predicting the secondary structure of sequences surrounding them (± 150 bp) using RNA fold annotation tool provided in the UEA sRNA tool kit. We identified additional 12 novel miRNA sequences that satisfied the secondary structure and other criteria (Table 2). The hairpin structures of all the novel miRNAs are given in supplementary file S1. A sRNA is considered as a potential miRNA candidate only if it meets the following strict criteria as defined by Zhang et al. [35]; 1) the sequence could fold into an appropriate stem loop hairpin secondary structure; 2) the sRNA sits in one arm of hairpin structure; 3) no more than 6 mismatches between mature miRNA sequence and its opposite miRNA' sequence in the secondary structure; 4) no loop or break in the miRNA or miRNA' sequences; 5) predicted secondary structure has negative minimal folding free energy; 6) minimum A+U% (30-70%) as mature miRNAs contain more A+U nucleotides than G+C and 7) usually begins with 5' U which is one of the characteristic feature of miRNAs.

Prediction of novel miRNA targets

miRNA target prediction in plants is easier due to high and significant complementarities of miRNA- mRNA targets and targets of most plant miRNAs possesses a single perfect or near perfect complementary site in coding region [36]. Assuming this to be generally the case, the *Stevia* transcript library (EST and nucleotide databases) was searched for complementarity with the sequences of identified novel miRNAs using the two target prediction web server's psRNA target and TAPIR. Almost all the targets predicted through psRNA target server were further authenticated by TAPIR which showed the same targets. Further, TAPIR predicted additional putative targets. A deep insight into miRNA targets helps us in understanding the range of sRNA expression, regulation and their functional importance. Further, multiple targets were predicted for only 8 out of 12 novel miRNAs were predicted using stevia transcript library. The predicted targets are mainly mRNA

BLAST Characteristics	Number
No. of Sequences	35,520
No. of Hits to DB	70,081,886
No. of extensions	25,687,690
No. of successful extensions	25,687,690
No. of sequences better than 10.0	3,186,710
No. of HSP's gapped	25,681,316
No. of HSP's successfully gapped	25,681,237
Length of database	3,637,171

Table 1: Characteristics of BLAST output file.

miRNA	Sequence	Length	Clones	EST/nt accession no.	Strand	A+U content	Mismatch	miRNA start	miRNA end	EST/nt start	EST/nt end	e-value	Precursor
Stv_13	UCAACCGUCGCGGACUGGAUUC	22	2	BG523957	Plus	45.4%	0	1	22	444	465	2e-06	YES
Stv_14	UGAAAAAUGAUCCUGAAUUU	20	2	BG523167	Plus	75%	0	1	20	577	596	2e-05	YES
Stv_15	AGAACUCGGAAGACCUAGAA	20	1	BG524766	Plus	55%	0	1	20	94	113	2e-05	YES
Stv_16	UUAGGACUCAGAUUGAUGAUU	22	1	BG525073	Plus	63.6%	0	1	22	297	318	2e-06	YES
Stv_17	UCUGUAUCUGUGAUGACUUCU	21	1	BG522152	Plus	61.9%	0	1	21	166	186	7e-06	YES
Stv_18	AACACGAUGACAUUCUGGC	19	1	BG526040	Plus	52.6%	0	1	19	126	144	8e-05	YES
Stv_19	UUACAUAAAAGCAUCCAUUGUA	21	1	AY215486	Plus	71.4%	0	1	21	784	804	7e-06	YES
Stv_20	GGAUUUCUCGCGCUCACGCCAUC	21	3	AB457301	Plus	42.8%	0	1	21	309	329	7e-06	YES
Stv_21	AUUAAACGAACACAGAUCUCGUC	23	1	BG526368	Plus	56.5%	0	1	23	110	132	5e-07	YES
Stv_22	CUGGAAACGAUUCUCGACA	20	1	BG524830	Plus	55%	0	1	20	273	292	2e-05	YES
Stv_23	AUCAUCGGAACUUUCGUCC	19	2	BG525875	Plus	52.6%	0	1	19	468	486	8e-05	YES
Stv_24	CCGAUGGACUACGAGCCUUGA	22	2	AY215182	Plus	45.4%	0	1	22	163	184	2e-06	YES

Table 2: Summary of novel miRNA candidates identified in *S. rebaudiana*: Clones: No. of times the miRNA repeats itself in library; EST/nt accession number: Accession no. of EST or gene showing complementarity with miRNA; Strand: miRNA complementary or reverse complementary to EST or nt; miRNA start- miRNA end: stretch of miRNA length or nt showing complementarity with stevia ESTs or nt; EST/nt start-EST/nt end: Stretch of EST or gene showing complementarity with miRNA; Mismatch: No. of mismatched nt in the complementarity region between miRNA and EST or gene of stevia; Precursor: stem loop secondary structure of respective miRNA predicted (YES) or not (NO) using Vienna RNA package; A+U content: Percentage of sum of adenine and uracil in the miRNA sequence.

Novel miRNA	Target accession	Score	Target annotation	miRNA: target alignment	Target inhibition	UPE
Stv_13	NIL					
Stv_14	BG525503	3.0	PSRP1 [<i>Citrus sinensis</i>]	miRNA 20 UUUUAGUCCUAGUAAAAAGU 1 ::: ::::: ::::: ::::: ::::: Target 130 AAAAUCCAGGUUCAUUUUUUUA 149	Translation	8.853
	BG526779	3.0	3-hydroxyisobutyryl-CoA hydrolase-like [<i>Glycine max</i>]	miRNA 20 UUUUAGUCCUAGUAAAAAGU 1 ::: ::::: ::::: ::::: ::::: Target 638 AAAUUCAUUUUUUUUUUUUA 657	Cleavage	10.598
	GANE01006171	2.5	6-phosphofructokinase 4, [<i>Vitis vinifera</i>]	miRNA 20 UUUUAGUCCUAGUAAAAAGU 1 ::: ::::: ::::: ::::: ::::: Target 279 AAAUUUAGGAUCAUUUUUACA 298	Cleavage	13.548
	GANE01015186	3.0	Pyridoxal phosphatase [<i>R. communis</i>]	miRNA 20 UUUUAGUCCUAGUAAAAAGU 1 ::: ::::: ::::: ::::: ::::: Target 264 AAAUUUUAUGAUUUUUUUUCU 283	Cleavage	11.191
Stv_15	NIL					
Stv_16	BG522304	3.5	Elongation factor Ts [<i>Morus notabilis</i>]	miRNA 20 AGUAGUUAGACUCAGGGAUU 1 ::: ::::: ::::: ::::: ::::: Target 439 UCCUCAAUUUGAUUCUCUAG 458	Cleavage	15.91
	BG526206	4.0	Magnesium-chelatase subunit chlI family protein [<i>P. trichocarpa</i>]	miRNA 20 AGUAGUUAGACUCAGGGAUU 1 ::: ::::: ::::: ::::: ::::: Target 270 UCCUCAAUUCUGUGUCUCAA 289	Translation	12.482
	BG526183	4.0	Zinc finger A20 and AN1 domain-containing [<i>Triticum urartu</i>]	miRNA 21 UAGUAGUUAGACUCAGGGAUU 1 ::: ::::: . ::::: ::::: ::::: Target 171 AUCAUCAACUCAAGUCCUGA 191	Translation	16.581
Stv_17	BG524980	3.0	Histone H3.3 [<i>A. thaliana</i>]	miRNA 21 UCUUCAGUAGUGUCUAUGUCU 1 ::: ::: : ::::: ::::: ::::: Target159 AGAAGCC-UCACAGAUACAGA 178	Cleavage	19.369
	BG524941	4.0	hypoxanthine-guanine phosphoribosyltransferase, putative [<i>R. communis</i>]	miRNA 20 CUUCAGUAGUGUCUAUGUCU 1 ::: : ::::: ::::: ::::: Target300 GAAGAUUUGUGGAUACAGG 319	Cleavage	18.089
Stv_18	BG524022	3	hypersensitive-induced response protein [<i>Carica papaya</i>]	miRNA 19 CGGUCUUACAGUAGCACAA1 oo. . target 405 GUUUGAAUGUCCUGUGUU424	Cleavage	-
	BG524910	3	hypersensitive-induced response protein 1- [<i>Oryza brachyantha</i>]	miRNA 19 CGGUCUUACAGUAGCACAA1 oo. . target 608 GUUUGAAUGUCCUGUGUU627	Cleavage	-

encoding enzymes regulating essential plant metabolic and signaling pathways which include structural proteins, elongation factors, protein kinases, heat shock proteins, dehydrogenases, transferases, synthases etc. Using stevia transcript library as reference library Stv_14 targets PSRP1 which is a ribosome-binding factor, 3-hydroxyisobutyryl-CoA hydrolase, 6-phosphofructokinase and pyridoxal phosphatase involved in various physiological processes (Table 3). Stv_16 targets elongation factor Ts involved in translation, magnesium-chelatase chII family protein, and zinc finger A20 and AN1 domain-containing stress-associated protein 8 involved in various stress responses (Table 3). Further, stv_17 targets Histone H3 involved in DNA structure and hypoxanthine-guanine phosphoribosyl transferase and stv_18 targets hypersensitive induced protein which is responsible for protein histidine kinase binding in certain stress responses (Table 3). Further, stv_19 predicted targets include lipid transfer protein and certain hypothetical protein of unknown function. Stv_21 is predicted to target receptor protein like kinase and GATA transcription factor and Stv_22 targets some interesting genes i.e. cholinephosphate cytidyltransferase and Elongation factor P involved in translational machinery. Lastly, Stv_24 is predicted to target, alpha/beta-hydrolases super family protein, unnamed protein and a 50S ribosomal protein involved in ribosome biogenesis (Table 3). Each of these novel miRNAs were found to bind to some other genes but they bound those genes outside their respective ORF (Open reading frame) regions, so they were not considered as their respective targets. Further, mode of action of all the novel miRNAs on targets was found to be both through translation and cleavage but predominantly through cleavage as is the case seen in other plants [37,38]. No targets were predicted for stv_13, stv_15, stv_20 and stv_23 using *S. rebaudiana* transcript library as reference. So, targets of only 8 out of 12 novel miRNAs were predicted speculating that rest of the miRNA may be typically present in stevia and needs further experimental work to locate their gene targets. The above described information regarding predicted targets could be utilized efficiently to assess the regulatory roles of these novel miRNAs in stevia. Further if any of these putative novel miRNA is found to have control over the regulation of the early and late genes in SGs biosynthetic pathway as seen in case of conserved miRNA 414 [34] can help to manipulate the diterpenoid contents.

Discussion

Computational approaches are successful in identifying novel or species specific miRNAs in many plants, but they require knowledge of the complete genome sequence, which is unavailable for most plant species. This has resulted into a sort of knowledge skew where most of the miRNAs have been reported only for those species whose genomic sequences are available or homologous sequences are known. Only 16% of total reported miRNAs in miRBase are from species whose genome is not sequenced. Also, the majority of these 16% miRNAs has been identified using homology search and exhibit a very few species specific miRNAs [39]. However, large genomic fragmented data in the form of GSSs, high-throughput genomics sequences (HTGSSs), nonredundant nucleotides (NRs), as well as ESTs, are available for several plant species and can be used for identification of novel miRNAs. Thus, barring a few model organisms, there is almost negligible miRNA information for most of the species. In our case, although the *S. rebaudiana* full genome sequence is not available, the large number of stevia expressed sequence tags and nucleotide sequences are an excellent source for precursor identification. Further, the development of NGS technology has greatly improved the capacity to identify low abundance or tissue-specific miRNAs, and has enhanced the discovery of several conserved, non-conserved or lowly expressed miRNAs through cloning and deep

sequencing of sRNA and transcriptome libraries in *A. thaliana* [40,41], *Triticum aestivum* (wheat; [42], *Solanum lycopersicum* (tomato; [43]), *Oryza sativa* (rice; [44]), and *Manihot esculenta* (Cassava; [45]). But still a large number of plant species are unexplored and stevia was one of them until 2012 when in our pioneer study we identified 34 conserved miRNA families and 12 novel miRNA families using deep sequencing [34]. A vast survey of miRNAs in stevia will provide useful information to elucidate their physiological functions, gene regulation, biogenesis and evolutionary roles in plants. The objective of this work was, thus, to identify and catalogue more novel miRNAs in *S. rebaudiana*. The identification of miRNA and, subsequently, their targets will lay the foundation for unraveling the complex miRNA-mediated regulatory networks controlling development and other physiological processes. In our study, we predicted 12 new novel miRNAs and their targets whose precursors were potentially generated from stevia EST and nt sequences. All novel sequences have not been earlier described in other plant species. According to the criterion mentioned by Zhang et al [35] to be considered as a potential miRNA candidate, all of the predicted 12 novel miRNAs i.e. Stv_12 to Stv_24 folded into perfect stem loop hairpin secondary structure in which the miRNA sequence was placed on one arm of the hairpin structure. Further, satisfied the other criteria with no more than 6 mismatches and no break or loop in the miRNA and miRNA* sequences (Table 2 and Supplementary Figure S1). Further, all the predicted novel miRNAs possessed A+U% within the range of 42.1-75% with Stv_14 showing the highest A+U% i.e. 75% [43]. Lastly, the criteria which says that miRNAs usually begins with 5' U was satisfied as 5 of the 12 novel miRNAs i.e. Stv_13, Stv_14, Stv_16, Stv_17 and Stv_19 have 5' U as the starting nucleotide (Table 2). The predicted novel miRNAs exhibited much lower expression levels, consistent with the notion that non-conserved miRNAs are often expressed at lower levels than conserved miRNAs [46]. Similarly, in our case, except Stv_20 with 3 clones all the rest 11 novel miRNAs were sequenced only once or twice (Table 2). The low abundance of novel miRNAs might suggest a specific role for these miRNAs under various growth conditions, in specific tissues, or during specific developmental stages. This suggests miRNAs identified in stevia might represent only a meager portion of novel miRNAs due to the fact that sRNA library was constructed from young plant leaves under normal conditions. In general, novel miRNAs represents either lineage-specific or species specific miRNAs are expressed at low levels. Whether these low abundance miRNAs are expressed at higher levels in other tissues and organs, or whether they are regulated by environmental stresses, remain to be further investigated. Furthermore, based on BLASTn searches and hairpin structure prediction, we found potential precursors for all the 12 novel miRNAs. Lacking of genomic information might be the main reason that leads to the identification of only 12 novel miRNAs. Their mature sequences were aligned with all the miRNAs registered in the miRNA database miRBase (<http://www.mirbase.org/>) [47] and PMRD (<http://bioinformatics.cau.edu.cn/PMRD/>) [48], CSRDB (<http://sundarlab.ucdavis.edu/smrnas/>) and non-redundant sequences in Genbank. There was no homolog matching with them, which implied the 12 miRNAs found in our data are novel and might be stevia-specific miRNAs and could play more species-specific roles. The high degree of sequence complementarity between plant miRNAs and their target mRNAs has facilitated the bioinformatic prediction of miRNA targets. Plant miRNAs have been predicted or confirmed to regulate a wider variety of developmental and physiological processes than animal miRNAs [49]. With these increasing evidence, it is concluded that miRNAs regulatory impact on plants is more pervasive than was previously suspected. miRNA target prediction in plants is easier due to high and significant complementarities of miRNA- mRNA targets and targets of most plant miRNAs possesses a single perfect or near perfect

complementary site in coding region [36]. Assuming this to be generally the case, the *Stevia* transcript library (EST and nucleotide databases) was searched for complementarity with the sequences of identified novel miRNAs using the two target prediction web server's psRNATarget (<http://bioinfo3.noble.org/psRNATarget/>) and TAPIR (<http://bioinformatics.psb.ugent.be/webtools/tapir>). Almost all the targets predicted through psRNA target server were further authenticated by TAPIR which showed the same targets. Further, TAPIR and psRNA target program individually predicted additional putative targets. A deep insight into miRNA targets helps us in understanding the range of sRNA expression, regulation and their functional importance. To analyze probable sRNA targets is significant in plants because the complimentary sites of potential sRNA can exist anywhere along the target mRNA rather than at 3'UTR in case of animals. As usually the targets are conserved between plant species but to know specific targets in our model plant *S. rebaudiana* we predicted putative target genes for 8 out of 12 *stevia* specific or novel miRNAs taking *stevia* transcript library as reference library. Similarly, mode of action of miRNAs on targets was found to be predominantly through cleavage which is usually the case seen in plants [37,38]. Stv_14 targets 3-hydroxyisobutyryl-CoA hydrolase which participates in 3 metabolic pathways: valine, leucine and isoleucine degradation, beta-alanine metabolism, and propanoate metabolism and ribosome binding factor PRSP1 involved in translational machinery. Stv_16 targets Zinc finger A20 and AN1 domain-containing stress-associated protein involved in environmental stress responses induced by cold, dehydration and salt stresses. Further, Stv_17 targets an interesting gene i.e. Histone H3 which is one of the five main histone proteins involved in the structure of chromatin in eukaryotic cells [49]. Stv_18 and Stv_19 are predicted to target certain stress proteins i.e. TMV resistance protein and hypersensitive induced protein respectively proving the indispensable roles of miRNAs in various stress responses. Further as miRNAs regulation is seen at both transcriptional and translational level, stv_21 is found to target GATA transcription factors which are a family of transcription factors characterized by their ability to bind to the DNA sequence "GATA" [50] and Stv_22 on the other hand targets Elongation factor P, an important component of translation machinery. Lastly, Stv_24 is also predicted to regulate certain important proteins involved in translation i.e. 50S ribosomal protein for the biogenesis of the ribosomes involved in protein synthesis. Further, mode of action of all the novel miRNAs on targets was found to be both through translation and cleavage but predominantly through cleavage as in reported in all other plants [1,6]. No targets were predicted for stv_10, stv_11, stv_12 and stv_23 using *S. rebaudiana* transcript library as reference library. Target of only 8 out of 12 novel miRNAs were predicted speculating that rest of the miRNA may be typically present in *stevia* and needs further experimental work to locate their gene targets. The above described information regarding predicted targets could be utilized efficiently to assess the regulatory roles of these novel miRNAs in *stevia*. Further if any of these putative novel miRNA is found to have control over the regulation of the early and late genes in SGs biosynthetic pathway can help to manipulate the diterpenoid contents. As an evolutionary important plant genus, *stevia* might contain a certain amount of novel genes, including miRNA genes. With the availability of more genomic sequence information on *stevia* genus, more miRNAs will be identified in this species. The cloning and identification of these genes and figuring out their regulation relationships would be very helpful for exploiting new genes and regulatory pathways and their evolution in plant. The identification of miRNA and their targets is important not only to help us learn more about the roles of miRNAs in *stevia* development and physiology but also to provide a framework for further designing RNAi based experiments for regulation of gene expression in this species.

Materials and Methods

Plant Material, RNA isolation, sRNA library preparation and sequencing

Plants of *S. rebaudiana* were grown and maintained in Punjab University, Chandigarh, India. Young leaves were harvested by snap freezing method and stored at -80 °C till further use. Total RNA enriched with sRNA fraction was isolated using protocol developed by Ghawana et al. [51] combined with miRNA easy spin kit (Qiagen, Germany) as mentioned in Mandhan et al (2012) [34]. RNA samples were sent to the Microarray and Genomic Core Facility, Huntsman cancer institute, University of Utah, Salt Lake City, USA, for preparation of the sRNA library and sequencing using the Genome Analyzer II (Illumina, USA) [34].

Bioinformatics analysis

Processing of raw sequences: To analyze the sRNA data, online version of the UEA sRNA toolkit (<http://srna-tools.cmp.uea.ac.uk/>) [52] was used. Toolkit provides a package of various tools for the analysis of high-throughput sRNA data including the sequence preprocessing tool which aids in conversion of FASTQ to FASTA format, removal of adaptor sequences, extracting sequences of defined size range and their abundance calculation. Further, filter tool was used to remove t/r-RNAs (non-coding) sequences from further analysis by comparing the t/rRNA sequences available at Rfam database, genomic tRNA database and EMBL release 95. Finally, sRNAs in the size range of 19nt to 24nt were extracted which constitutes the ideal miRNA size range.

Prediction of novel miRNAs: *Stevia* 5548 ESTs and 29874 nt sequences were downloaded from the NCBI database and conserved miRNA sequences were removed from the above processing file using Filter tool available in the UEA sRNA toolkit and the rest were used to perform BLASTN searches against *stevia* sequences in order to obtain precursors for identifying potential novel miRNAs. The selected EST and nt sequences which showed significant complementarity with the sRNA in the dataset were then folded into a secondary structure using RNA fold annotation tool provided in the UEA sRNA tool kit which uses Vienna RNA package to fold miRNA precursors using the minimum free energy algorithm [53] and yields a single optimal structure. The sRNA was considered to be a novel miRNA if a perfect stem loop structure was formed with sRNA sequence at one arm of the stem as well as keeping other criterion given by Meyers et al. [54] under consideration.

Target gene prediction: The potential targets of novel miRNAs were predicted using the psRNATarget program [55] and secondly checked with new web server called TAPIR [56], designed for the prediction of plant miRNA targets. psRNA target program was executed with parameters at their default settings i.e. score or maximum expectation at 3, length of complementary scoring at 20, target accessibility at 25, flanking length around target 17 bp upstream and 13 bp downstream and keeping range of central mismatch leading to translational inhibition in between 9-11. Newly identified novel miRNA sequences were used as custom miRNA sequences and *S. rebaudiana* transcript library (EST and nucleotide databases) was used as custom plant database. All predicted target genes were evaluated under the psRNA target server by scoring, and the criteria used were as follows: each G:U wobble pairing was assigned 0.5 points, each indel was assigned 2.0 points, and all other non-canonical Watson-Crick pairings were assigned 1.0 point each. The total score for an alignment was calculated based on 20 nt. When the query exceeded 20 nt length, scores for all possible consecutive 20 nt subsequences were computed, and the

minimum score was considered the total score for the query-subject alignment. Because targets complementary to the miRNA 5' end appear to be critical, mismatches other than G: U wobbles at positions 2-7 at the 5' end were further penalized by 0.5 points in the final score [57]. Sequences were considered to be miRNA targets if the total score was less than 3.0 points. For further validation of the above miRNA targets TAPIR was used keeping the score cutoff value (default 4) and the free energy ratio cutoff value (default 0.7) using FASTA search engine. TAPIR indicates several parameter values (names, free energy ratio, start position of the duplex on the mRNA, seed and non-seed mismatches, gaps and G-U pairs) including a full representation of the miRNA-mRNA duplex with an alignment string. Sensitivity and specificity tests for the TAPIR fast method was done with psRNA target server using various parameters. All the results showed that TAPIR with the FASTA search engine (score cutoff 4) has a higher rate of true positive, while keeping the false positives to values that are similar to those of psRNA target [58]. Once potential target mRNA sequences were obtained, BLAST was performed using these target sequence and the NCBI database as reference to predict functions of potential targets.

Competing Interests

The authors declare no conflict of interests.

Acknowledgments

Authors are thankful to Dr. Brian Dalley, Director, Microarray and Genomic Core Facility, Huntsman cancer institute, University of Utah, Salt Lake City-UT, USA for providing services for library preparation and GAI sequencing. Vibha Mandhan is thankful to University Grants commission, India for providing research fellowship.

References

- Unver T, Namuth-Covert DM, Budak H (2009) Review of current methodological approaches for characterizing MicroRNAs in plants. *International Journal of Plant Genomics* Article ID 262463.
- Megraw M, Hatzigeorgiou AG (2010) MicroRNA promoter analysis. *Methods Mol Biol* 592: 149-161.
- Lee Y, Kim M, Han J, Yeom KH, Lee S, et al. (2004) MicroRNA genes are transcribed by RNA polymerase II. *EMBO J* 23: 4051-4060.
- Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and Function. *Cell* 116: 281-297.
- Carrington JC, Ambros V (2003) Role of microRNAs in plant and animal development. *Science* 301: 336-338.
- Barbato C, Arisi I, Frizzo ME, Brandi R, Sacco L Da, et al. (2009) Computational challenges in miRNA target predictions: to be or not to be a true target. *Journal of Biomedicine and Biotechnology* Article ID 803069.
- Laufs P, Peaucelle A, Morin H, Traas J (2004) MicroRNA regulation of the CUC genes is required for boundary size control in Arabidopsis meristems. *Development* 131: 4311-4322.
- Mallory AC, Dugas DV, Bartel DP, Bartel B (2004) MicroRNA regulation of NAC-domain targets is required for proper formation and separation of adjacent embryonic, vegetative and floral organs. *Curr. Biol* 14: 1035-1076.
- Guo HS, Xie Q, Fei JF, Chua NH (2005) MicroRNA directs mRNA cleavage of the transcription factor NAC1 to downregulate auxin signals for arabidopsis lateral root development. *Plant Cell* 17: 1376-1386.
- Emery JF, Floyd SK, Alvarez J, Eshed Y, Hawker NP, et al. (2003) Radial patterning of Arabidopsis shoots by class III HD-ZIP and KANADI genes. *Curr Biol* 13: 1768-1774.
- Juarez MT, Kui JS, Thomas J, Heller BA Timmermans MC (2004) microRNA-mediated repression of rolled leaf specifies maize leaf polarity. *Nature* 428: 84-88.
- Aukerman MJ, Sakai H (2003) Regulation of flowering time and floral organ identity by a microRNA and its APETALA2-like target genes. *Plant Cell* 15: 2730-2741.
- Chen X (2004) A microRNA as a translational repressor of APETALA2 in Arabidopsis flower development. *Science* 303: 2022-2025.
- Kasschau KD, Xie Z, Allen E, Llave C, Chapman EJ, et al. (2003) P1/HC-Pro, a viral suppressor of RNA silencing, interferes with Arabidopsis development and miRNA function. *Developmental Cell* 4: 205-217.
- Llave C (2004) MicroRNAs: more than a role in plant development. *Mol Plant Pathol* 5: 361-366.
- Sunkar R, Zhu JK (2004) Novel and stress-regulated microRNAs and other small RNAs from Arabidopsis. *Plant Cell* 16: 2001-2019.
- Fujii H, Chiou TJ, Lin SI, Aung K, Zhu JK (2005) A miRNA involved in phosphate-starvation response in Arabidopsis. *Curr Biol* 15: 2038-2043.
- Sunkar R, Kapoor A, Zhu JK (2006) Posttranscriptional induction of two Cu/Zn superoxide dismutase genes in Arabidopsis is mediated by down regulation of miR398 and important for oxidative stress tolerance. *Plant Cell* 18: 2051-2065.
- Rajagopalan R, Vaucheret H, Trejo J Bartel DP (2006) A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. *Genes Dev* 20: 3407-3425.
- Sunkar R, Zhou X, Zheng Y, Zhang W Zhu JK (2008) Identification of novel and candidate miRNAs in rice by high throughput sequencing. *BMC Plant Biol* 8: 25.
- Klevebring D, Street NR, Fahlgren N, Kasschau KD, Carrington JC, et al. (2009) A Genome-wide profiling of populous small RNAs. *BMC Genomics* 10: 620.
- Wei B, Cai T, Zhang R, Li A, Huo N, et al. (2009) Novel microRNAs uncovered by deep sequencing of small RNA transcriptomes in bread wheat (*Triticum aestivum* L.) and *Brachypodium distachyon* (L.) *Beauv Funct Integr Genomics* 9: 499-511.
- Zhang J, Xu Y, Huan Q, Chong K (2009) Deep sequencing of *Brachypodium* small RNAs at the global genome level identifies microRNAs involved in cold stress response. *BMC Genomics* 10: 449.
- Pantaleo V, Szittyta G, Moxon S, Miozzi L, Moulton V, et al. (2010) Identification of grapevine microRNAs and their targets using high-throughput sequencing and degradome analysis. *Plant J* 62: 960-976.
- Zhao CZ, Xia H, Frazier TP, Yao YY, Bi YP, et al. (2010) Deep sequencing identifies novel and conserved microRNAs in peanuts (*Arachis hypogaea* L.). *BMC Plant Biol* 5: 10.
- Song C, Wang C, Zhang C, Korir NK, Yu H, et al. (2010) Deep sequencing discovery of novel and conserved microRNAs in trifoliolate orange (*Citrus trifoliata*). *BMC Genomics* 11: 431.
- Li H, Dong Y, Sun Y, Zhu E, Yang J, et al. (2011) Investigation of the microRNAs in safflower seed, leaf, and petal by high-throughput sequencing. *Planta* 233: 611-619.
- Madan S, Ahmad S, Singh GN, Kohli K, Kumar Y, et al. (2010) *Stevia rebaudiana* (Bert.) Bertonii - a review. *Indian J Nat Prod Resour* 1: 267-286.
- Cardello HM, Da Silva MA, Damasio MH (1999) Measurement of the relative sweetness of stevia extract, aspartame and cyclamate/saccharin blend as compared to sucrose at different concentrations. *Plant Foods Hum Nutr* 54: 119-130.
- Crammer B, Ikan R (1987) Progress in chemistry and properties of the rebaudiosides. In *Development in Sweeteners*. Elsevier 45-64 (1stedn) Grenby TH, New York.
- Hsieh MH, Chan P, Sue YM, Liu JC, Liang TH, Huang TY, et al. (2003) Efficacy and tolerability of oral stevioside in patients with mild essential hypertension: a two-year, randomized, placebocontrolled study. *Clin Ther* 25: 2797-2808.
- Gregersen S, Jeppesen PB, Holst JJ, Hermansen K (2004) Antihyperglycemic effects of stevioside in type 2 diabetic subjects. *Metabolism* 53: 73-76.
- Kumar H, Kaul K, Gupta SB, Kaul VK, Kumar S (2011) A comprehensive analysis of fifteen genes of steviol glycosides biosynthesis pathway in *Stevia rebaudiana* (Bertonii). *Gene* 492: 276-284.
- Mandhan V, Kaur J, Singh K (2012) smRNAome profiling to identify conserved and novel microRNAs in *Stevia rebaudiana* Bertonii. *BMC Plant Biol* 12: 197.
- Zhang BH, Pan XP, Wang QL, Cobb GP, Anderson TA (2005) Identification and characterization of new plant microRNAs using EST analysis. *Cell Res* 15: 336-60.
- Adai A, Johnson C, Mlotshwa S, Archer-Evans S, Manocha V, et al. (2005) Computational prediction of miRNAs in Arabidopsis thaliana. *Genome Res* 15: 78-91.
- Llave C, Xie Z, Kasschau KD, Carrington JC (2002) Cleavage of Scarecrow

- like mRNA targets directed by a class of Arabidopsis miRNA. *Science* 297: 2053-2056.
38. Jones- Rhoades MW, Bartel DP (2004) Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Mol Cell* 14: 787-799.
39. Jha A, Shankar R (2013) miReader: Discovering Novel miRNAs in Species without Sequenced Genome. *PLoS One* 8: e66857.
40. Rajagopalan R, Vaucheret H, Trejo J, Bartel DP (2006) A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. *Genes & Dev* 20: 3407-3425.
41. Fahlgren N, Howell MD, Kasschau KD, Chapman EJ, Sullivan CM, et al. (2007) High-throughput sequencing of *Arabidopsis* microRNAs: Evidence for frequent birth and death of MIRNA genes. *PLoS ONE* 2: e219.
42. Yao Y, Guo G, Ni Z, Sunkar R, Du J, et al. (2007) Cloning and characterization of microRNAs from wheat (*Triticum aestivum* L.). *Genome Biol* 8: R96.
43. Moxon S, Jing R, Szittyá G, Schwach F, Pilcher RLR, et al. (2008) Deep sequencing of tomato short RNAs identifies microRNAs targeting genes involved in fruit ripening. *Genome Res* 18: 1602-1609.
44. Sunkar R, Girke T, Jain PK, Zhu JK (2005) Cloning and characterization of microRNAs from rice. *Plant Cell* 17: 1397-1411.
45. Amiteye S, Corral JM, Sharbel TF (2011) Overview of the potential of microRNAs and their target gene detection for cassava (*Manihot esculenta*) improvement. *African Journal of Biotechnology* 10: 2562-2573.
46. Liu N, Yang J, Guo S, Xu Y, Zhang M (2013) Genome-wide identification and comparative analysis of conserved and novel microRNAs in grafted watermelon by high-throughput sequencing. *PLoS One* 8: e57359.
47. Kozomara A, Griffiths Jones (2011) S.mIRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res* 39: D152-157.
48. Zhang Z, Yu J, Li D, Zhang Z, Liu F, et al. (2010) PMRD: plant microRNA database. *Nucleic Acids Res* 38: D806-813.
49. Dugas DV, Bartel B (2004) microRNA regulation of gene expression in plants. *Current Opinion in Plant Biology* 7: 512-520.
50. Bhasin M, Reinherz EL, Reche PA (2006) Recognition and classification of histones using support vector machine. *J Comput Biol* 13: 102-112.
51. Ko LJ, Engel JD (1993) DNA-binding specificities of the GATA transcription factor family. *Mol Cell Biol*. 13: 4011-4022.
52. Ghawana S, Paul A, Kumar H, Kumar A, Singh H, et al. (2011) An RNA isolation system for plant tissues rich in secondary metabolites. *BMC Res Notes* 4: 85.
53. Moxon S, Schwach F, Dalmay T, Maclean D, Studholme DJ, et al. (2008) A toolkit for analysing large-scale plant small RNA datasets. *Bioinformatics* 24: 2252-2253.
54. Zuker M, Stiegler P (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res* 9: 133-148.
55. Meyers BC, Axtell MJ, Bartel B, Bartel DP, Baulcombe D, et al. (2008) Criteria for annotation of plant microRNAs. *Plant Cell* 20: 3186-3190.
56. psRNATarget: A Plant Small RNA Target Analysis Server.
57. Zhang Y (2005) mi RU: An automated plant miRNA target prediction server. *Nucleic Acids Res* 33: W701-704.
58. Bonnet E, He Y, Billiau K, Van de Peer Y (2010) TAPIR, a web server for the prediction of plant microRNA targets, including target mimics. *Bioinformatics* 26: 1566-1568.