

A novel computational approach to the silencing of Sugarcane Bacilliform Guadeloupe A Virus determines potential host-derived MicroRNAs in sugarcane (*Saccharum officinarum* L.)

Fakiha Ashraf^{Equal first author, 1}, Muhammad Aleem Ashraf^{Corresp., Equal first author, 2, 3}, Xiaowen Hu⁴, Shuzhen Zhang^{Corresp. 1}

¹ Institute of Tropical Bioscience and Biotechnology, Chinese Academy of Tropical Agricultural Sciences, Haikou, Hainan, China

² Haikou Experimental Station, Chinese Academy of Tropical Agricultural Sciences, Haikou, Hainan, China

³ Department of Plant Breeding and Genetics, University College of Agriculture and Environmental Sciences, Islamia University of Bahawalpur, Bahdad-UI-Jadeed Campus, Bahawalpur, Pakistan

⁴ Zhanjiang Experimental Station, Chinese Academy of Tropical Agricultural Sciences, Zhanjiang, Guangdong, China

Corresponding Authors: Muhammad Aleem Ashraf, Shuzhen Zhang

Email address: muhammad.aleemashraf@iub.edu.pk, zhangshuzhen@itbb.org.cn

Sugarcane Bacilliform Guadeloupe A Virus (SCBGAV, genus *Badnavirus*, family Caulimoviridae) is an emerging, deleterious pathogen of sugarcane which presents a substantial barrier to producing high sugarcane earnings. Sugarcane bacilliform viruses (SCBVs) are one of the main species that infect sugarcane. During the last 30 years, significant genetic changes in SCBV strains have been observed with a high risk of disease incidence associated with crop damage. SCBV infection may lead to significant losses in biomass production in susceptible sugarcane cultivars. The circular, double-stranded (ds) DNA genome of SCBGAV (7.4 Kb) is composed of three open reading frames (ORFs) on the positive strand that replicate by a reverse transcriptase. SCBGAV can infect sugarcane in a semipersistent manner via the insect vectors sugarcane mealybug species. In the current study, we used miRNA target prediction algorithms to identify and comprehensively analyze the genome-wide sugarcane (*Saccharum officinarum* L.)-encoded microRNA (miRNA) targets against the SCBGAV. Mature miRNA target sequences were retrieved from the miRBase (miRNA database) and were further analyzed for hybridization to the SCBGAV genome. Multiple computational approaches—including miRNA-target seed pairing, multiple target positions, minimum free energy, target site accessibility, maximum complementarity, pattern recognition and minimum folding energy for attachments— were considered by all algorithms. Among them, sof-miR396 was identified as the top effective candidate, capable of targeting the vital ORF3 of the SCBGAV genome. miRanda, RNA22 and RNAhybrid algorithms predicted hybridization of sof-miR396 at common locus position 3394. The predicted sugarcane miRNAs against viral mRNA targets possess antiviral activities, leading to translational inhibition by mRNA cleavage. Interaction network of sugarcane-encoded miRNAs with SCBGAV genes, created using Circos, allow analyze new

targets. The finding of the present study acts as a first step towards the creation of SCBGAV-resistant sugarcane through the expression of the identified miRNAs.

A Novel Computational Approach to the Silencing of Sugarcane Bacilliform Guadeloupe A Virus Determines Potential Host-Derived MicroRNAs in Sugarcane (*Saccharum officinarum* L.)

Fakiha Ashraf^{1,†}, Muhammad Aleem Ashraf^{2,3,†,*}, Xiaowen Hu⁴ and Shuzhen Zhang^{1,*}

¹Institute of Tropical Bioscience and Biotechnology, Chinese Academy of Tropical Agricultural Sciences, Haikou 571101, China

²Haikou Experimental Station, Chinese Academy of Tropical Agricultural Sciences, Haikou 571101, China

³Department of Plant Breeding and Genetics, University College of Agriculture and Environmental Sciences, Islamia University of Bahawalpur, Baghdad-UI-Jadeed Campus, Bahawalpur 63100, Pakistan

⁴Zhanjiang Experimental Station, Chinese Academy of Tropical Agricultural Sciences, Zhanjiang, 524091, China

[†] These authors contributed equally to this work

* Correspondence: muhammad.aleemashraf@iub.edu.pk , zhangshuzhen@itbb.org.cn

Abstract

Sugarcane Bacilliform Guadeloupe A Virus (SCBGAV, genus *Badnavirus*, family Caulimoviridae) is an emerging, deleterious pathogen of sugarcane which presents a substantial barrier to producing high sugarcane earnings. Sugarcane bacilliform viruses (SCBVs) are one of the main species that infect sugarcane. During the last 30 years, significant genetic changes in SCBV strains have been observed with a high risk of disease incidence associated with crop damage. SCBV infection may lead to significant losses in biomass production in susceptible sugarcane cultivars. The circular, double-stranded (ds) DNA genome of SCBGAV (7.4 Kb) is composed of three open reading frames (ORFs) on the positive strand that replicate by a reverse transcriptase. SCBGAV can infect sugarcane in a semipersistent manner via the insect vectors sugarcane mealybug species. In the current study, we used miRNA target prediction algorithms to identify and comprehensively analyze the genome-wide sugarcane (*Saccharum officinarum* L.)-encoded microRNA (miRNA) targets against the SCBGAV. Mature miRNA target sequences were retrieved from the miRBase (miRNA database) and were further analyzed for hybridization to the SCBGAV genome. Multiple computational approaches—including miRNA-target seed pairing, multiple target positions, minimum free energy, target site accessibility, maximum complementarity, pattern recognition and minimum folding energy for attachments—were considered by all algorithms. Among them, sof-miR396 was identified as the top effective candidate, capable of targeting the vital ORF3 of the SCBGAV genome. miRanda, RNA22 and RNAhybrid algorithms predicted hybridization of sof-miR396 at common locus position 3394. The predicted sugarcane miRNAs against viral mRNA targets possess antiviral activities, leading to translational inhibition by mRNA cleavage. Interaction network of sugarcane-encoded

miRNAs with SCBGAV genes, created using Circos, allow analyze new targets. The finding of the present study acts as a first step towards the creation of SCBGAV-resistant sugarcane through the expression of the identified miRNAs.

Introduction

Sugarcane bacilliform virus (SCBV, genus: *Badnavirus*, family Caulimoviridae) is a circular, non-enveloped bacilliform, monopartite DNA virus that harbors about 7.5 kb double-stranded (ds) DNA with three open reading frames (ORFs) (Sun et al. 2016). SCBV was first reported in the sugarcane cultivar (B34104) in South American continent (Cuba, 1985) and disseminated into many sugarcane (*Saccharum officinarum* L.)-growing regions in the globe (AUTREY et al. 1992). SCBV is implicated in yield losses in sugarcane cultivars. SCBV infection may lead to significant losses in biomass production in susceptible sugarcane cultivars (Lockhart & Autrey 2000). SCBV infect sugarcane cultivars utilized for sugar and bioenergy production and constitute a major threat to the global exchange of sugarcane germplasm (Borah et al. 2013). Variable follicle symptoms were observed in some sugarcane varieties showing severe chlorotic mottling, stunted growth and pronounced freckle (Lockhart 1996; Viswanathan et al. 1996). SCBV has wide host range. It comprises members of plant families *Poaceae* (sugarcane, johnsongrass, rice, sorghum, *Panicum maximum*, *Rottboellia exaltata* and *Brachiaraia sp.*) and *Musaceae* (bananas) (Borah et al. 2013; Lockhart & Autrey 2000; Lockhart 1996). SCBVs are highly endemic in *S.officinarum* germplasm. These viruses can also infect commercial sugarcane (*Saccharum* interspecific hybrids), *S. barberi*, *S. robustum*, *S. spontaneum* and *S. sinensis* (Lockhart 1996). SCBV is transmitted naturally by sap-sucking insect vector mealybug species (*Saccharicoccus sacchari* and *Dysmicoccus boninsis*) through vegetative cuttings. *Agrobacterium*-mediated inoculation is used to disseminate SCBV infection experimentally (Bouhida et al. 1993; Lockhart 1996; Viswanathan et al. 1996). Previous studies have revealed that SCBV is a heterogeneous *badnavirus* species. SCBV exist in nature as a complex of viruses or distinct strains and genetic variants. Currently, four distinct SCBV species have been classified by the International Committee on Taxonomy of Viruses (ICTV). These are *Sugarcane Bacilliform Guadeloupe A Virus* (SCBGAV), *Sugarcane Bacilliform Guadeloupe D Virus* (SCBGDV) *Sugarcane bacilliform MO virus* (SCBMOV-MOR) and *Sugarcane bacilliform IM virus* (SCBIMV-QLD) originating from Guadeloupe, Morocco and Australia, respectively (Adams & Carstens 2012; Adams et al. 2016; Bouhida et al. 1993; Geijskes et al. 2002; Muller et al. 2011). SCBGAV was first identified as a sugarcane R570 hybrid variety in Guadeloupe in 2011 (Muller et al. 2011). SCBGAV was classified as a new species in the genus *badnavirus* under the SCBV-A group as assigned by ICTV(Adams et al. 2016). The disease symptoms appear in the form of chlorosis, leaf freckle and mottling. Some infected plants were also observed to exhibit no symptoms. Recently in China, SCBV-infection resulted in reduced sucrose content, juice, stalk weigh, purity and gravity in sugarcane plants(Ahmad et al. 2019). SCBGAV is a mealybug-transmitted monopartite *badnavirus* that infects sugarcane and is composed of a 7444 bp circular, dsDNA genome containing three ORFs on the positive strand that replicate by a virus-encoded reverse transcriptase (RT). The two small proteins of sizes (176-185 and 122-135 amino acids), were encoded by the ORF1 and ORF2 respectively. The ORF3 was a key component of SCBGAV genome, encoded by a polyprotein (1786-1933 amino acids) (Muller et al. 2011).

Sugarcane (*Saccharum officinarum* L.) has inherited an active immunity that is composed of microRNAs (miRNAs) to combat infection. The miRNAs are a class of 21-23 nucleotide-long, endogenous, noncoding RNAs that govern gene regulation and expression at the post-transcriptional level. In plants, this mechanism occurs through the process of mRNA cleavage or translational repression. They are synthesized after the processing of hairpin miRNA precursors using an RNase-III-like enzyme (Dicer) to control gene expression (Brodersen & Voinnet 2006). RNA silencing, through miRNAs present in the host plant, thus imparts natural immunity and resistance to the host against foreign genetic elements including plant viruses (Carbonell & Daròs 2019; Qu et al. 2007).

Viral miRNA degradation is the simplest molecular approach to combat viral infection in plants. Successful demonstration about changes in the nucleotide transcripts of mature miRNAs does not affect their biogenesis and the application of modified miRNAs to target Potyviruses in Arabidopsis was reported first time (Niu et al. 2006). Several studies have been successful in generating resistance against plant viruses using amiRNA-technology in crop plants such as tomato (Chen et al. 2019; Zhang et al. 2011) and cotton (Akmal et al. 2017; Ali et al. 2013). Most recently, amiRNA-mediated gene silencing has been demonstrated against Potexvirus and Tobamovirus successfully (Petchthai et al. 2018).

miRNA: mRNA target identification, validation and interactions are the basis for discerning the important functions of miRNAs in the greater perspective of regulatory network leading to different biotic developments. Computational algorithms for identification and analysis of miRNA targets enable the process of narrowing down possible target binding sites for experimental validation at transgenic level. *In-silico* strategies model how miRNAs target specific mRNAs and a number of algorithms are accessible, each with a diverse approach to miRNA target prediction and identification. Although, it might be valuable to have access to a variety of tools and algorithms with different capabilities, the user is challenged with a significant option in determining which algorithm/tool to apply (Peterson et al. 2014). The development of artificial microRNAs (amiRNAs)-mediated gene silencing method offers a highly precise, targeted approach that gives several advantages over RNAi-mediated gene silencing strategy. Design, construction and validation of the amiRNAs depend upon the assembly of an endogenous precursor miRNA, which is substituted with a selected miRNA nucleotide transcript complementary to the target sequence (Niu et al. 2006; Ossowski et al. 2008).

The objective of the present research work is to predict sugarcane-encoded miRNAs that have the potential to develop resistance against badnaviruses, especially the SCBGAV. The major research objective was designed to identify the potential host-derived miRNAs in the SCBGAV genome and to screen the most promising miRNAs to understand the complex host-virus interactions. The novel amiRNA silencing method has been adopted for the first time to explore as a source of creating resistance to a monopartite *badnavirus*. The predicted miRNAs may provide help for the generation of SCBGAV-resistant sugarcane plants through genetic engineering in the future.

Materials & Methods

Retrieval of Mature miRNAs of Sugarcane

The sugarcane (*Saccharum-officinarum* L) and (*Saccharum spp.*) plant miRNA sequences were accessed from the microRNA database, miRBase (<http://www.mirbase.org/cgi-bin/browse.pl/>) (Griffiths-Jones et al. 2006; Griffiths-Jones et al. 2007; Kozomara et al. 2018; Kozomara & Griffiths-Jones 2013). A total of 28 miRNAs were retrieved (Table S1).

SCBGAV Genome Retrieval and Annotation

The nucleotide sequence of SCBGAV dsDNA genome (accession number NC_038382.1) composed of 7444bp was retrieved from the NCBI database. pDRAW32 plasmid map software (v 1.1.129) was used for annotation of genome features of SCBGAV (Fig. 1).

miRNA Target Prediction in SCBGAV Genome

Four computational algorithms (Table 1) — named miRanda, RNA22, RNAhybrid and psRNATarget — were selected for the screening of mature miRNAs of sugarcane against the SCBGAV genome to identify the miRNAs target positions (Table 1). Nucleotide sequences of the SCBGAV genome and sugarcane miRNAs were recorded in FASTA format to process further using computational algorithms with defined parameters. A detailed workflow pipeline was designed to identify miRNA binding sites within the SCBGAV genome (Fig. 2).

miRanda

miRanda is a computational algorithm (John et al. 2004) used to predict and identify the potential plant genomic miRNA targets (Archak & Nagaraju 2007). It was designed to consider the following algorithm properties: sequence complementarity, RNA-RNA duplex dimerization, free energy and cross-species evolutionary conservation of molecular target sites. These properties yield output to weigh all matches and mismatch scores for gap penalties and for base pairs. It also promotes the prediction of multiple miRNA target sites including the ones with imperfect binding in the seed region within the 3'UTR of the target site thereby enhancing the specificity (Betel et al. 2008; M Witkos et al. 2011). miRanda software was downloaded from the source website (<http://cbio.mskcc.org/miRNA2003/miranda.html>), and the SCBGAV genome was assessed to determine whether there were any possible targets for sugarcane miRNAs. The miRanda algorithm was run after defining standard settings. These include: Gap Open Penalty = -9.0, Score Threshold = 130, Minimum free energy (MFE) threshold (= -15.00 kcal/mol). MFE is an important statistical parameter to screen potential targets.

RNA22

RNA22 is a novel diverse web server (<http://cm.jefferson.edu/rna22v1.0/>), designed to implement a pattern-based approach to detect potential miRNA target sites (Miranda et al. 2006). The minimum folding energy (MFE) is a key parameter to screen possible miRNA without a cross-species conservation filter (Loher & Rigoutsos 2012). Specificity and sensitivity values were selected at 61% and 63%, respectively. The value of the seed size was selected as 7 with unpaired base 1 permitted inside the seed region. There was no limit set to the maximum number of G: U wobbles in the seed region. The paired bases with minimum number and MFE were selected at 12 and -13.50 kcal/mol, respectively.

163 **Tapirhybrid**

164 Tapirhybrid is a novel web server (<http://bioinformatics.psb.ugent.be/webtools/tapir>) designed
165 for the prediction of plant microRNA targets, including microRNA target mimics, using a very
166 fast and precise algorithm (Bonnet et al. 2010). It considers many important parameters
167 including seed pairing, multiple target sites and target site accessibility. It is one of the most
168 highly recommended tools for plant microRNA prediction due to its great accuracy (Srivastava
169 et al. 2014). It was designed to run using the RNAhybrid algorithm for microRNA target
170 prediction and produces miRNA target scores (on the basis of the number of mismatches), the
171 number of GU pairs, gaps and mismatches in the seed region. It is regarded as a high precision
172 algorithm to enhance the accuracy of the predicted results. The setting for input parameters were
173 standardized as score ≤ 8 and MFE_ratio ≥ 0.5 .

174 **RNAhybrid**

175 RNAhybrid is a novel algorithm used to predict hybridization of miRNA and mRNA that is
176 based on MFE (<http://bibiserv.techfak.uni-bielefeld.de/rnahybrid>). Site complementarity, target-
177 site abundance and MFE are the unique parameters of RNAhybrid algorithm. MFE was set a
178 threshold of -20 kJ/mol . RNAhybrid can also assign a *p*-value for the miRNA: miRNA
179 interaction based on the number of binding sites within the 3' UTR sequence, which is a measure
180 of target-site abundance. The other filters remained fixed as default parameters (Krüger &
181 Rehmsmeier 2006).

182 **psRNATarget**

183 psRNATarget is a new web server (<http://plantgrn.noble.org/psRNATarget/>) that was developed
184 to predict small RNA (sRNA) in plants. It was used to analyze complementary matching
185 between target mRNA sequence and sRNA sequence based on a scoring schema (Dai et al.
186 2018). Evaluation of target-site accessibility on the basis of UPE (unpaired energy) is another
187 important feature of the psRNATarget algorithm. The standard parameter set for our analysis
188 was as follows: penalty for (extending gap = 0.5, opening gap = 2, G.U pair = 1, other
189 mismatches = 1), HSP size = 19, seed region = 2-7 nucleotides. The minimum expectation score
190 was 7.0.

191 **RNAfold Algorithm**

192 RNAfold algorithm was used to generate the stable secondary structure of pre-miRNAs from the
193 precursors of the mature miRNAs (Lorenz et al. 2011). The parameters set for secondary
194 structure prediction were: minimum free energy and partition function; avoid isolated base pairs
195 and dangling energy on both sides of helix in any case; RNA parameters; rescale energy
196 parameters at a given temperature 37°C; interactive RNA secondary structure plot; RNA
197 secondary structure plots with reliability annotation (partition function folding) and mountain
198 plot (Turner & Mathews 2009). Further, these predicted secondary structures were manually
199 curated using rules reported by (Zhang et al. 2005).

200 **Circos Algorithm**

201 A Circos plot was developed between sugarcane-encoded miRNA and SCBGAV genes by
202 applying the Circos algorithm (Krzywinski et al. 2009).

Statistical Analysis

Predicted miRNA data obtained by applying four diverse algorithms were further analyzed using R-Language (v3.1.1, software version 3.5.1) (Gandrud, 2013). The graphical representation of the predicted miRNAs was processed using in house-scripts (readxl and ggplot2 packages).

Results

miRNA Target Prediction in Genome of SCBGAV

In order to predict sugarcane miRNA targets in the SCBGAV genome, a combination of above mentioned computational algorithms were employed to magnify high accuracy of miRNA target prediction. This strategy also filters out the false-positive results. miRanda (John et al. 2004) was selected to implement various miRNA target prediction parameters. These are target site conservation and prediction of several miRNA target binding sites. Afterwards, RNA22 was widely used for pattern-recognition in the target sequence (Miranda et al. 2006). RNAhybrid is a unique algorithm which determines miRNA and mRNA hybridization using following unique parameters (Krüger & Rehmsmeier 2006). This algorithm is applied to calculate (MFE) minimum free energy and mode of target inhibition suggested as per the (Brodersen et al. 2008) conclusion. psRNATarget is the most widely used plant miRNA target prediction tool which deals with a large set of data. It was designed to predict plant sRNA targets by analyzing complementary matching between sRNA and target mRNA sequence and by evaluating target site accessibility (Dai et al. 2018). No sugarcane miRNAs were predicted in the SCBGAV genome using Tapirhybrid algorithm. Fig.3 represents all the genomic positions targeted by sugarcane miRNAs by the various algorithms applied (Table S2, S3 and File S1).

Prediction of Sugarcane-encoded miRNA Targets in the SCBGAV ORF1

There is little information available about the function of this protein. ORF1 (538-1071 nucleotides) encodes a hypothetical protein represented as P1 (177 amino acids). RNAhybrid predicted binding of sof-miR159e at locus 733, ssp-miR169 at locus 752, ssp-miR437 (a, c) at locus 978, and ssp-miR444 (a, b, c-3p) at locus 774 (Fig. 3C). In addition, psRNATarget predicted hybridization of ssp-miR166 at locus 981, ssp-miR444a at locus 775, and ssp-miR444b at position 756 (Fig. 3D)

Identification of miRNA targets in the SCBGAV ORF2

The DNA binding protein, represented as P2, is encoded by the ORF2. ORF2 has the least number of predicted targets by sugarcane miRNAs; only three miRNA of sugarcane (sof-miR167 (a, b) and ssp-miR528) were targeted at common locus 1301, indicated by RNA22 (Fig. 3B). Similarly, RNAhybrid predicted bindings of sof-miR167 (a, b) at locus 1304 and ssp-miR1432 at locus 1301 (Fig. 3C).

miRNA Target prediction in the ORF3

The polyprotein is encoded by the ORF3 which contains several functional units. These proteins are coat, movement, aspartic protease, RT, and ribonuclease H (Bouhida et al. 1993; Geijskes et al. 2002; King et al. 2011). The RT/RNase H coding region is a commonly used taxonomic marker for badnaviruses. It is considered sufficient to address the sequence diversity present within Badnavirus genomes (Bousalem et al. 2008). For the ORF3 of SCBGAV, seven different

kinds of miRNAs were predicted by miRanda: (sof-miR159 (a, b, c, d, and e), sof-miR168a, sof-miR396, ssp-miR166, ssp-miR827, ssp-miR1128 and ssp-miR1432) (Fig. 3A). RNA22 predicted hybridization of the following miRNAs: (sof-miR168 (a, b), sof-miR396, sof-miR408 (a, b, c and d) and ssp-miR444 (a, b) (Fig. 3B). Suitable miRNAs for targeting ORF3 were hypothesized by RNAhybrid to interact with the ORF3 of SCBGAV. These are sof-miR159 (a, b, d), sof-miR168a, sof-miR396, sof-miR408 (a, b, c, d and e), ssp-miR166, ssp-miR827, and ssp-miR1128 (Fig. 3C). Moreover, sof-miR159 (a, b, c, d), sof-miR167 (a, b), sof-miR396, ssp-miR528, ssp-miR444b and ssp-miR444c-3p were identified by psRNATarget (Fig. 3D).

Visualization of miRNA Target Interaction

A Circos plot was used to study the visualization of sugarcane miRNAs with SCBGAV-targets (ORFs) that display particular evidence for potential targets. We have for the first time reported sugarcane-encoded miRNAs and their targets simultaneously constructed in this manner (Fig. 4).

Sugarcane miRNAs (Predicted by a Consensus of Algorithms) for SCBGAV Silencing

Among all the predicted miRNAs for SCBGAV silencing, only four miRNAs (sof-miR167 (a, b), sof-miR396 and ssp-miR528) were predicted by all of the algorithms used (Fig. 5, Table 2 and Table S4). We have verified the efficacy of these miRNA targets against SCBGAV following suppression of RNAi-mediated badnavirus control through translational inhibition or cleavage of viral mRNA. Moreover, six miRNAs (sof-miR167 (a, b) at locus 5846 and 1310, sof-miR168a at locus 5506, sof-miR396 at locus 3394, ssp-miR444a at locus 775 and ssp-miR1128 at locus 6148) were predicted at the common locus by at least two of the algorithms used (Fig. 6). Out of 28 sugarcane miRNAs, only one miRNA (sof-miR396) was predicted by at least three algorithms used to have binding site at the same locus position 3394 (Table 3).

Prediction of Secondary Structures

Stable secondary structures of the consensus sugarcane miRNA precursor sequences was obtained using RNAfold algorithm. The precursor sequences were manually curated. Top seven stable secondary structures (based on MFEI value) of consensus sugarcane miRNAs are shown (Fig. 7). We have characterized the salient features of precursor miRNA, such as MFE (minimum free energy) to determine the secondary structure of nucleic acids. In our studies, MFE range from -56.1 to -101.7 kcal/mol, GC content (42-65%) and MFEI ranges from -0.7 to -1.17 (Table 4).

Discussion

For possible miRNA target prediction in the genome of SCBGAV, a combination of the aforementioned computational tools was used in order to filter out the false-positive results and to increase the accuracy of miRNA target prediction. miRanda was implemented to validate various parameters, from target site conservation to whole genome prediction of miRNA target genes (John et al. 2004; Riffo-Campos et al. 2016). Then, RNAhybrid and psRNATarget were used, both of which are strongly prescribed for plant miRNA target identification (Dai et al. 2018; Srivastava et al. 2014; Zhang & Verbeek 2010). RNA22 is a novel algorithm that applies an *in-silico* strategy that is highly divergent in comparison to other algorithms. Pattern-based recognition is the key feature of this algorithm (Min & Yoon 2010; Miranda et al. 2006).

The results from this study suggest that SCBGAV is susceptible to targeting by consensus sugarcane-encoded miRNAs. The genome components of SCBGAV (ORF1, ORF2 and ORF3) seemed to be principally prone sites for sugarcane-miRNA regulation. While *in vivo* demonstration requires validating functional efficacy, the degree of complementarity between target mRNA and miRNA concludes the fate of the predicted sites. A full complementarity binding between target mRNA and miRNA sequence results in endonucleolytic cleavage and disruption. Contrary to this, partial target-site complementarity characteristically down-regulates target-gene expression by suppressing translation of target mRNA.

The present study identifies suitable sugarcane-encoded miRNAs to exhibit a stronger degree of target-site complementarities within the ORF1, ORF2 and ORF3 of SCBGAV. These predicted miRNAs may be utilized to develop effective amiRNA constructs, which could be used to enhance the sugarcane immunity to SCBGAV. Pairing multiple miRNAs with a single mRNA induces effective RNA silencing (Doench & Sharp 2004). In the current study, we have predicted several miRNA targets which were associated with SCBGAV (ORF1, ORF2, and ORF3) genes at multiple loci. A deeper understanding of these vital ORFs involved in SCBGAV epidemic via miRNA-mediated control of gene expression would significantly assist in the development of molecular approaches to combat the dissemination of SCBGAV. The miRNA-target pairs ensuring MFE exceeding the threshold standards were predicted.

Our study is unique from previous studies in that it provides an *in silico* approach for comprehensive analysis of host-derived miRNAs using seven computational algorithms and advanced statistical approaches. Previous studies that have identified plant-encoded miRNA Targets in plant viruses have been limited by small numbers of miRNAs screened, small sample sizes, lack of independent validation sets, and the use of inappropriate statistical methods to present miRNA interaction data. Notably, we have validated the efficacy of the consensus seven sugarcane-encoded miRNA targets (detected two algorithms at the same locus) against SCBGAV following suppression of RNAi-mediated viral combat. These are achieved by translational inhibition or cleavage of viral miRNA.

Applying a combination of four diverse computational algorithms (miRanda, RNA22, Tapirhybrid and psRNATarget), the predicted consensus results are highly reliable and robust for assessing interaction between sugarcane miRNAs and viral genome (Oliveira et al. 2017). This miRNA-viral gene interaction was further verified by RNAhybrid (standard MFE predicted) and Circos algorithms (network visualization). False-positive miRNA-target interactions predicted by computational algorithms depend upon the mechanism of the miRNA-target recognition (Pinzón et al. 2017). In addition, MFE is one of the key factors to affect the miRNA-targets interactions for results sorting. A lower MFE value has high potential to develop miRNA-target complex (Hammell et al. 2008; Kertesz et al. 2007). We have applied a stringent cut-off -15kcal/mol for miRanda and -20kcal/mol for RNAhybrid to narrow down miRNA candidates.

In order to control false-positive prediction, union and intersection approach was applied. Union approach based on grouping of one or more algorithmic tools which results increase of number of true targets as well the number of false targets. It decreases specificity at the cost of increasing

sensitivity. Using intersection approach relies on the mixing two or more tools which improves the specificity at the cost of decreasing sensitivity (M Witkos et al. 2011). Our analysis indicated that both the strategies showed the highest performance as shown in (Fig. 5 and 6). These miRNAs were identified and considered as the most desirable sugarcane miRNAs against the genome of SCBGAV. These predicted miRNAs were identified after setting the parameters of minimum free energy, seed pairing, target site accessibility, folding energy and pattern recognition, thus integrating all aspects of miRNA target prediction. Therefore, they are the most suitable selections for gene silencing.

Similar studies have been published previously using multiple computational algorithms to identify a best fit miRNA for a desirable target to understand plant-virus interaction. First time, genome-wide identification and comprehensive analysis was carried out to predict maize (*Zea-mays*)-encoded miRNA targets against maize chlorotic mottle virus (MCMV) (Iqbal et al. 2017). We have adopted the same novel computational strategy to predict the sugarcane-encode miRNA targets in the SCBGAV genome to combat badnaviruses in sugarcane. As the amiRNAs have high specificity to the designed target gene, detrimental off-target effects can be minimized, permitting their silencing expression to be stably transmitted to future generations (Ossowski et al. 2008; Zhao et al. 2009). Furthermore, the small size of amiRNA permits for the insertion of multiple and distinct amiRNAs within a single gene expression cassette, which can then be transformed to develop transgenic plant resistant to multiple viruses simultaneously (Niu et al. 2006; Park et al. 2009; Schwab et al. 2010). We have designed future work to validate this promising amiRNA-based strategy can in fact be used to develop durable SCBGAV- resistance in transgenic sugarcane.

Conclusions

The current study details an organized computational approach for the identification of host-derived miRNAs aimed at silencing viral genes affecting the sugarcane host plant by amiRNA-mediated interference targeting different genes of the SCBGAV. This study offers an enhanced means to computationally investigate the best-candidate miRNAs against badnaviruses, prior to cloning. As our approach allows a narrow-range of match-mismatch in microRNA-mRNA attachment, it screens most of the falsely predicted attachments. These predicted miRNAs were identified after setting the parameters of MFE, seed pairing, target site accessibility, folding energy and pattern recognition, thus using key features of miRNA target prediction. Most importantly, we identified six miRNAs (sof-miR167 (a, b), sof-miR168a, sof-miR396, ssp-miR444a and ssp-miR1128) that could target the SCBGAV genome. Among them, sof-miR396 was predicted as the top-ranking effective candidate, capable of targeting the vital ORF3. To further refine our results, we also identified key interaction of miRNA networks and target genes by constructing Circos graph of consensus miRNAs and their targets. This short-listed miRNA is the best candidates to be utilized in sugarcane plant transformation for the development of SCBGAV-resistant sugarcane cultivars. These results offer an evidence of idea for the construction of innovative amiRNA-based therapeutics against emerging SCBGAV badnaviruses.

Acknowledgement

We are highly thankful to our lab colleagues for their assistance in data analysis. We thank to Dr. Zhiqiang Xia (ITBB) for assistance in constructing Circos plot. We are thankful to our senior lab colleagues Xiao-Yan Feng and Lin-Bo Shen for reading the revised manuscript as well as the editor and two referees for their valuable comments on this manuscript. We are highly thankful to a qualified native English speaking editor at the PeerJ Experts for editing proper English language.

Funding Statement

This study is supported by the National Natural Science Foundation of China (No.31771865), the Sugar Crop Research System (CARS-170301) and the “Talented Young Scientist Program” (TYSP-4th Batch 2018-2019) of China Project ID (Pakistan-18-004). MAA is also supported by “Foreign Expert Program”. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Additional Information and Declarations

Competing Interests

The authors declare that they have no competing interests.

References

- Adams MJ, and Carstens E. 2012. Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses (2012). *Archives of virology* 157:1411-1422.
- Adams MJ, Lefkowitz EJ, King AM, Harrach B, Harrison RL, Knowles NJ, Kropinski AM, Krupovic M, Kuhn JH, and Mushegian AR. 2016. Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses (2016). *Archives of virology* 161:2921-2949.
- Ahmad K, Sun S-R, Chen J-L, Huang M-T, Fu H-Y, and Gao S-J. 2019. Presence of Diverse Sugarcane Bacilliform Viruses Infecting Sugarcane in China Revealed by Pairwise Sequence Comparisons and Phylogenetic Analysis. *The plant pathology journal* 35:41.
- Akmal M, Baig MS, and Khan JA. 2017. Suppression of cotton leaf curl disease symptoms in *Gossypium hirsutum* through over expression of host-encoded miRNAs. *Journal of biotechnology* 263:21-29.
- Ali I, Amin I, Briddon RW, and Mansoor S. 2013. Artificial microRNA-mediated resistance against the monopartite begomovirus Cotton leaf curl Burewala virus. *Virology journal* 10:231.
- Archak S, and Nagaraju J. 2007. Computational prediction of rice (*Oryza sativa*) miRNA targets. *Genomics, proteomics & bioinformatics* 5:196-206.
- AUTREY L, BOOLELL S, JONES E, LOCIÇHART B, and NADIE A. 1992. Distribution of sugarcane Bacilliform Virus in various geographical regions. ISSCT. Proceedings XXI Congress. p 5-14.
- Betel D, Wilson M, Gabow A, Marks DS, and Sander C. 2008. The microRNA. org resource: targets and expression. *Nucleic Acids Research* 36:D149-D153.
- Bonnet E, He Y, Billiau K, and Van de Peer Y. 2010. TAPIR, a web server for the prediction of plant microRNA targets, including target mimics. *Bioinformatics* 26:1566-1568.
- Borah BK, Sharma S, Kant R, Johnson AA, Saigopal DVR, and Dasgupta I. 2013. Bacilliform DNA - containing plant viruses in the tropics: commonalities within a genetically diverse group. *Molecular plant pathology* 14:759-771.

- Bouhida M, Lockhart B, and Olszewski NE. 1993. An analysis of the complete sequence of a sugarcane bacilliform virus genome infectious to banana and rice. *Journal of General Virology* 74:15-22.
- Bousalem M, Douzery EJ, and Seal S. 2008. Taxonomy, molecular phylogeny and evolution of plant reverse transcribing viruses (family Caulimoviridae) inferred from full-length genome and reverse transcriptase sequences. *Archives of virology* 153:1085.
- Brodersen P, Sakvarelidze-Achard L, Bruun-Rasmussen M, Dunoyer P, Yamamoto YY, Sieburth L, and Voinnet O. 2008. Widespread translational inhibition by plant miRNAs and siRNAs. *Science* 320:1185-1190.
- Brodersen P, and Voinnet O. 2006. The diversity of RNA silencing pathways in plants. *TRENDS in Genetics* 22:268-280.
- Carbonell A, and Daròs J-A. 2019. Design, Synthesis, and Functional Analysis of Highly Specific Artificial Small RNAs with Antiviral Activity in Plants. *Antiviral Resistance in Plants*: Springer, 231-246.
- Chen L, Meng J, He XL, Zhang M, and Luan YS. 2019. Solanum lycopersicum microRNA1916 targets multiple target genes and negatively regulates the immune response in tomato. *Plant, cell & environment* 42:1393-1407.
- Dai X, Zhuang Z, and Zhao PX. 2018. psRNATarget: a plant small RNA target analysis server (2017 release). *Nucleic Acids Research* 46:W49-W54.
- Doench JG, and Sharp PA. 2004. Specificity of microRNA target selection in translational repression. *Genes & development* 18:504-511.
- Geijskes RJ, Braithwaite KS, Dale J, Harding R, and Smith GR. 2002. Sequence analysis of an Australian isolate of sugarcane bacilliform badnavirus. *Archives of virology* 147:2393-2404.
- Griffiths-Jones S, Grocock RJ, Van Dongen S, Bateman A, and Enright AJ. 2006. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Research* 34:D140-D144.
- Griffiths-Jones S, Saini HK, van Dongen S, and Enright AJ. 2007. miRBase: tools for microRNA genomics. *Nucleic Acids Research* 36:D154-D158.
- Hammell M, Long D, Zhang L, Lee A, Carmack CS, Han M, Ding Y, and Ambros V. 2008. mirWIP: microRNA target prediction based on microRNA-containing ribonucleoprotein-enriched transcripts. *Nature methods* 5:813.
- Iqbal MS, Jabbar B, Sharif MN, Ali Q, Husnain T, and Nasir IA. 2017. In silico MCMV silencing concludes potential host-derived miRNAs in maize. *Frontiers in plant science* 8:372.
- John B, Enright AJ, Aravin A, Tuschl T, Sander C, and Marks DS. 2004. Human microRNA targets. *PLoS biology* 2:e363.
- Kertesz M, Iovino N, Unnerstall U, Gaul U, and Segal E. 2007. The role of site accessibility in microRNA target recognition. *Nature genetics* 39:1278.
- King AM, Lefkowitz E, Adams MJ, and Carstens EB. 2011. *Virus taxonomy: ninth report of the International Committee on Taxonomy of Viruses*: Elsevier.
- Kozomara A, Birgaoanu M, and Griffiths-Jones S. 2018. miRBase: from microRNA sequences to function. *Nucleic Acids Research* 47:D155-D162.
- Kozomara A, and Griffiths-Jones S. 2013. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research* 42:D68-D73.
- Krüger J, and Rehmsmeier M. 2006. RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Research* 34:W451-W454.
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, and Marra MA. 2009. Circos: an information aesthetic for comparative genomics. *Genome research* 19:1639-1645.
- Lockhart B, and Autrey LJC. 2000. Sugarcane bacilliform virus. *A guide to sugarcane diseases*:266-272.

- Lockhart B, Ireyt, MJ and Comstock, JC. 1996. Sugarcane Bacilliform Virus, Sugarcane Mild Mosaic Virus and Sugarcane Yellow Leaf Syndrome. Sugarcane Germplasm Conservation and Exchange. Brisbane, Queensland, Australia: Canberra, Australia : Australian Centre for International Agricultural Research, 1996. p 108-112.
- Loher P, and Rigoutsos I. 2012. Interactive exploration of RNA22 microRNA target predictions. *Bioinformatics* 28:3322-3323.
- Lorenz R, Bernhart SH, Zu Siederdisen CH, Tafer H, Flamm C, Stadler PF, and Hofacker IL. 2011. ViennaRNA Package 2.0. *Algorithms for molecular biology* 6:26.
- M Witkos T, Koscianska E, and J Krzyzosiak W. 2011. Practical aspects of microRNA target prediction. *Current molecular medicine* 11:93-109.
- Min H, and Yoon S. 2010. Got target?: computational methods for microRNA target prediction and their extension. *Experimental & molecular medicine* 42:233.
- Miranda KC, Huynh T, Tay Y, Ang Y-S, Tam W-L, Thomson AM, Lim B, and Rigoutsos I. 2006. A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell* 126:1203-1217.
- Muller E, Dupuy V, Blondin L, Bauffe F, Dugrois J-H, Nathalie L, and Iskra-Caruana M-L. 2011. High molecular variability of sugarcane bacilliform viruses in Guadeloupe implying the existence of at least three new species. *Virus research* 160:414-419.
- Niu Q-W, Lin S-S, Reyes JL, Chen K-C, Wu H-W, Yeh S-D, and Chua N-H. 2006. Expression of artificial microRNAs in transgenic Arabidopsis thaliana confers virus resistance. *Nature biotechnology* 24:1420.
- Oliveira AC, Bovolenta LA, Nachtigall PG, Herkenhoff ME, Lemke N, and Pinhal D. 2017. Combining results from distinct MicroRNA target prediction tools enhances the performance of analyses. *Frontiers in genetics* 8:59.
- Ossowski S, Schwab R, and Weigel D. 2008. Gene silencing in plants using artificial microRNAs and other small RNAs. *The Plant Journal* 53:674-690.
- Park W, Zhai J, and Lee J-Y. 2009. Highly efficient gene silencing using perfect complementary artificial miRNA targeting AP1 or heteromeric artificial miRNA targeting AP1 and CAL genes. *Plant cell reports* 28:469-480.
- Petchthai U, Le Yee CS, and Wong S-M. 2018. Resistance to CymMV and ORSV in artificial microRNA transgenic Nicotiana benthamiana plants. *Scientific reports* 8:9958.
- Peterson SM, Thompson JA, Ufkin ML, Sathyanarayana P, Liaw L, and Congdon CB. 2014. Common features of microRNA target prediction tools. *Frontiers in genetics* 5:23.
- Pinzón N, Li B, Martinez L, Sergeeva A, Presumey J, Apparailly F, and Seitz H. 2017. microRNA target prediction programs predict many false positives. *Genome research* 27:234-245.
- Qu J, Ye J, and Fang R. 2007. Artificial microRNA-mediated virus resistance in plants. *Journal of virology* 81:6690-6699.
- Riffo-Campos Á, Riquelme I, and Brebi-Mieville P. 2016. Tools for sequence-based miRNA target prediction: what to choose? *International journal of molecular sciences* 17:1987.
- Schwab R, Ossowski S, Warthmann N, and Weigel D. 2010. Directed gene silencing with artificial microRNAs. *Plant MicroRNAs*: Springer, 71-88.
- Srivastava PK, Moturu TR, Pandey P, Baldwin IT, and Pandey SP. 2014. A comparison of performance of plant miRNA target prediction tools and the characterization of features for genome-wide target prediction. *BMC genomics* 15:348.
- Sun S-R, Damaj MB, Alabi OJ, Wu X-B, Mirkov TE, Fu H-Y, Chen R-K, and Gao S-J. 2016. Molecular characterization of two divergent variants of sugarcane bacilliform viruses infecting sugarcane in China. *European journal of plant pathology* 145:375-384.
- Turner DH, and Mathews DH. 2009. NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic Acids Research* 38:D280-D282.

509 Viswanathan R, Alexander K, and Garg I. 1996. Detection of sugarcane bacilliform virus in
510 sugarcane germplasm. *Acta virologica* 40:5-8.

511 Zhang BH, Pan XP, Wang QL, George PC, and Anderson TA. 2005. Identification and
512 characterization of new plant microRNAs using EST analysis. *Cell research* 15:336.

513 Zhang X, Li H, Zhang J, Zhang C, Gong P, Ziaf K, Xiao F, and Ye Z. 2011. Expression of
514 artificial microRNAs in tomato confers efficient and stable virus resistance in a cell-
515 autonomous manner. *Transgenic Research* 20:569-581.

516 Zhang Y, and Verbeek FJ. 2010. Comparison and integration of target prediction algorithms for
517 microRNA studies. *Journal of integrative bioinformatics* 7:169-181.

518 Zhao T, Wang W, Bai X, and Qi Y. 2009. Gene silencing by artificial microRNAs in
519 *Chlamydomonas*. *The Plant Journal* 58:157-164.

520

Figure 1

Genome annotation of sugarcane bacilliform Guadeloupe A virus

Genome organization of sugarcane bacilliform Guadeloupe A virus. Three ORFs are shown with arrows. The double stranded DNA genome of SCBGAV is composed of 7444bp.

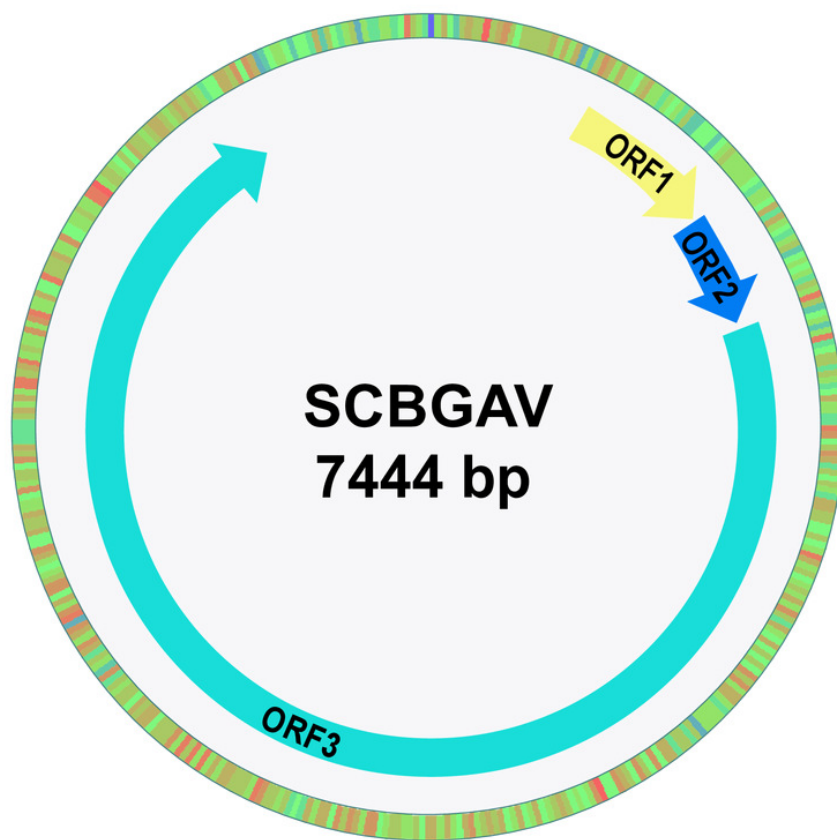


Figure 2

A schematic pipeline depicting the strategy of miRNA prediction from SCBGAV genome

A schematic representation of miRNA target prediction in SCBGAV genome pipeline.

Biological data set contains the kinds of data obtained for this study from NCBI (badnavirus genome) and miRBase (miRNAs). Computational algorithm group enlists three kinds of tools for the prediction of miRNA, secondary structure and miRNA-target interaction. R language was used to make plots and select data using in-house scripts/codes.

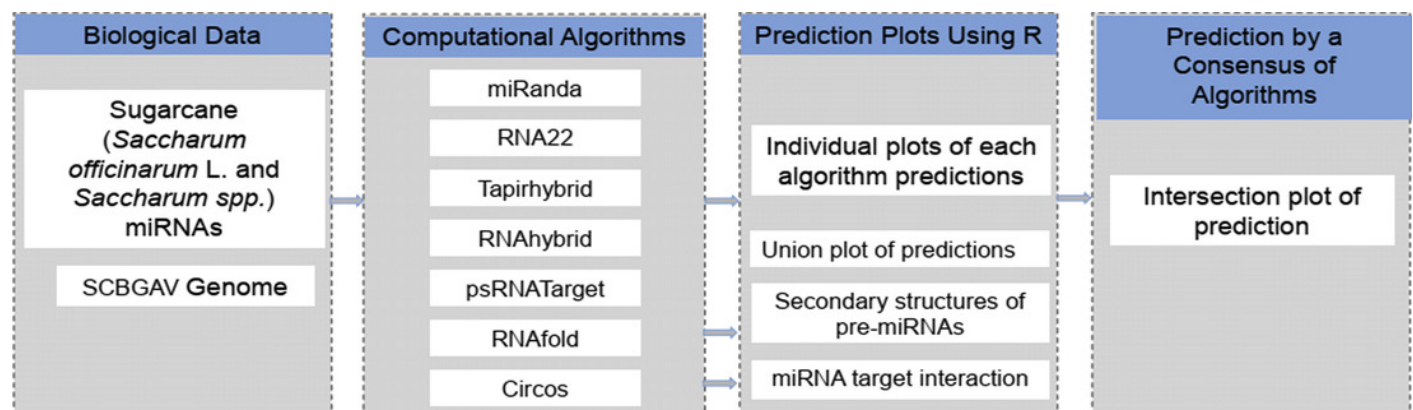


Figure 3

miRNA target prediction results of sugarcane bacilliform Guadeloupe A virus

miRNA target prediction results of sugarcane bacilliform Guadeloupe A virus. A) Indicates target prediction results from miRanda, B) target prediction by RNA22 C) target prediction by RNAhybrid and D) target prediction by psRNATarget.

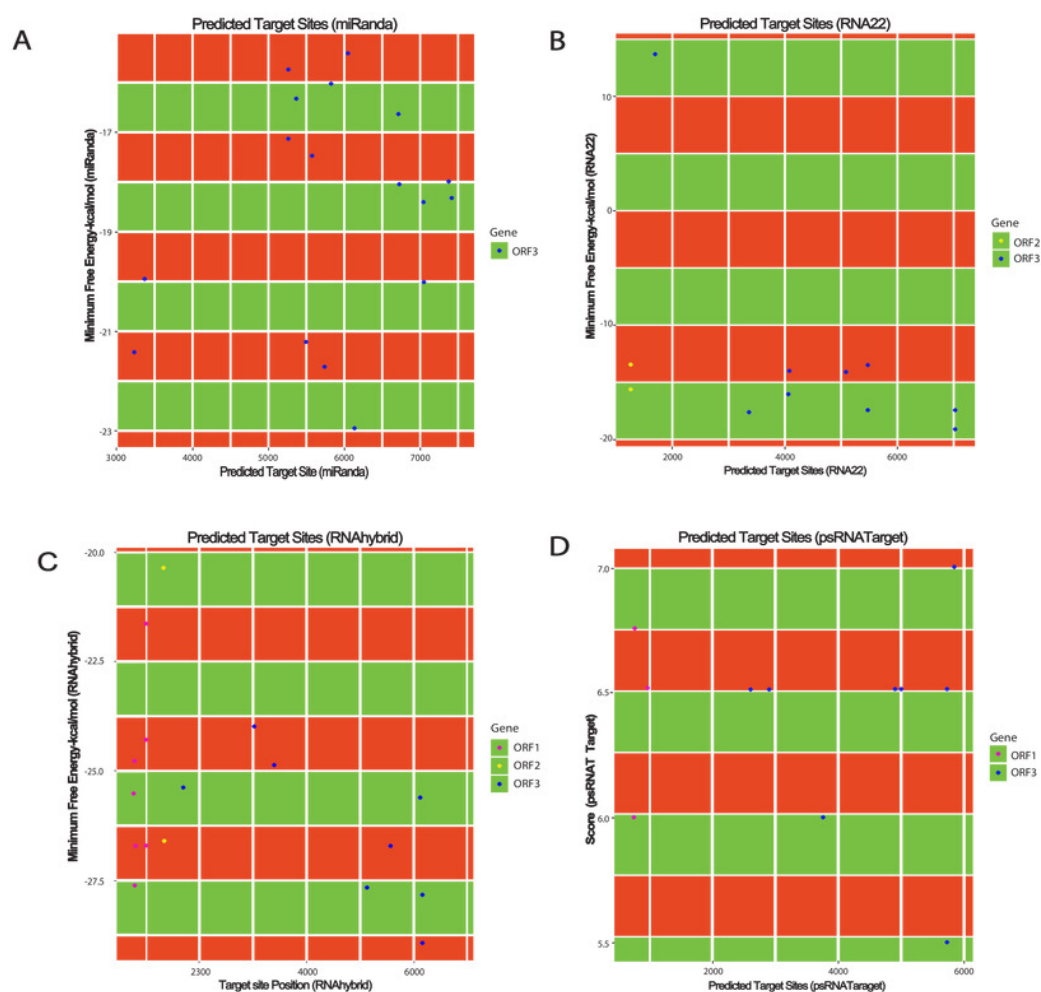


Figure 4

Circos plot showing network interaction

Circos plot showing network interaction between sugarcane-encoded miRNAs and their SCBGAV targets.

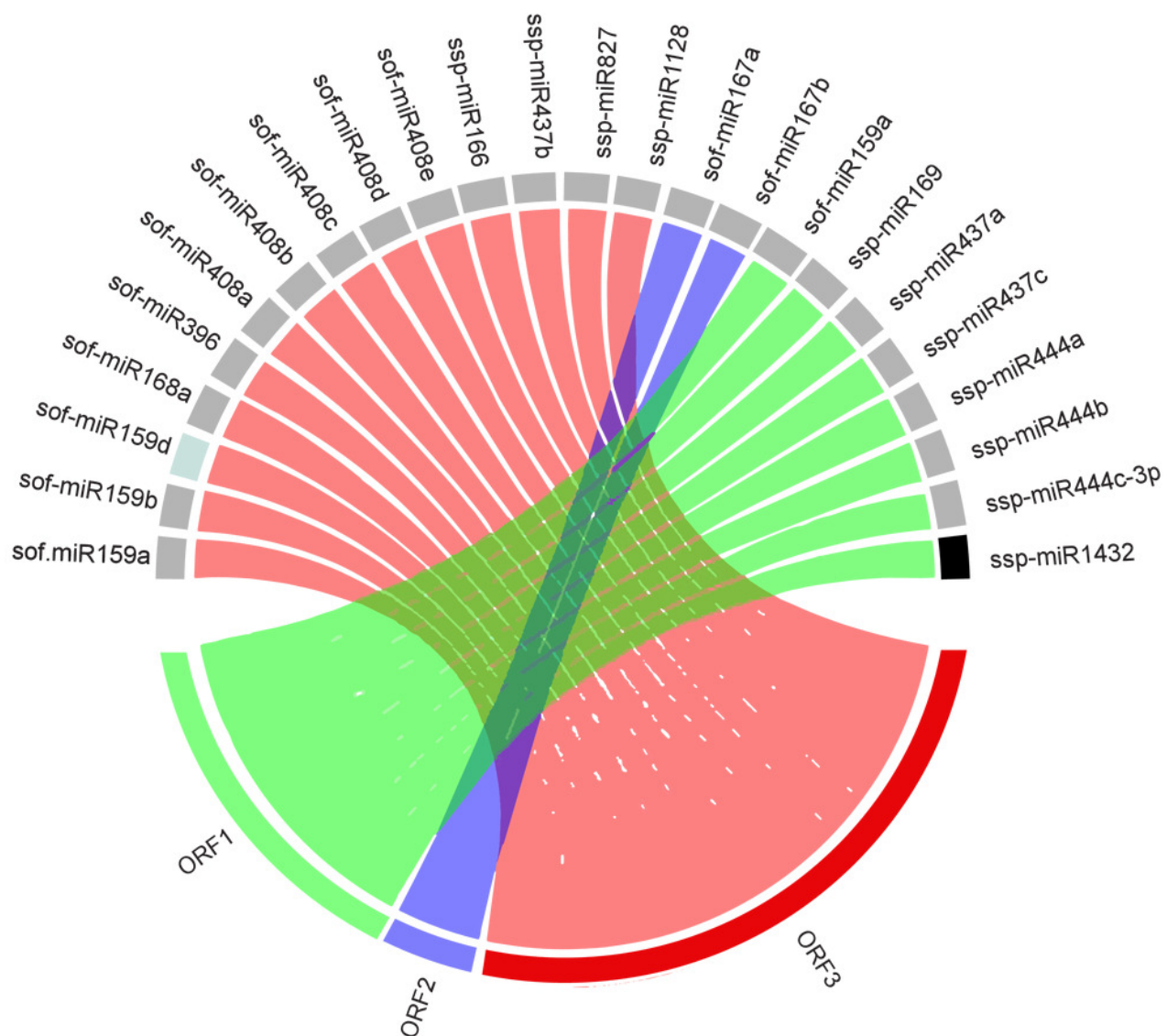


Figure 5

Venn diagram representing common sugarcane miRNAs.

Venn diagram representing common sugarcane miRNAs predicted by all algorithms.

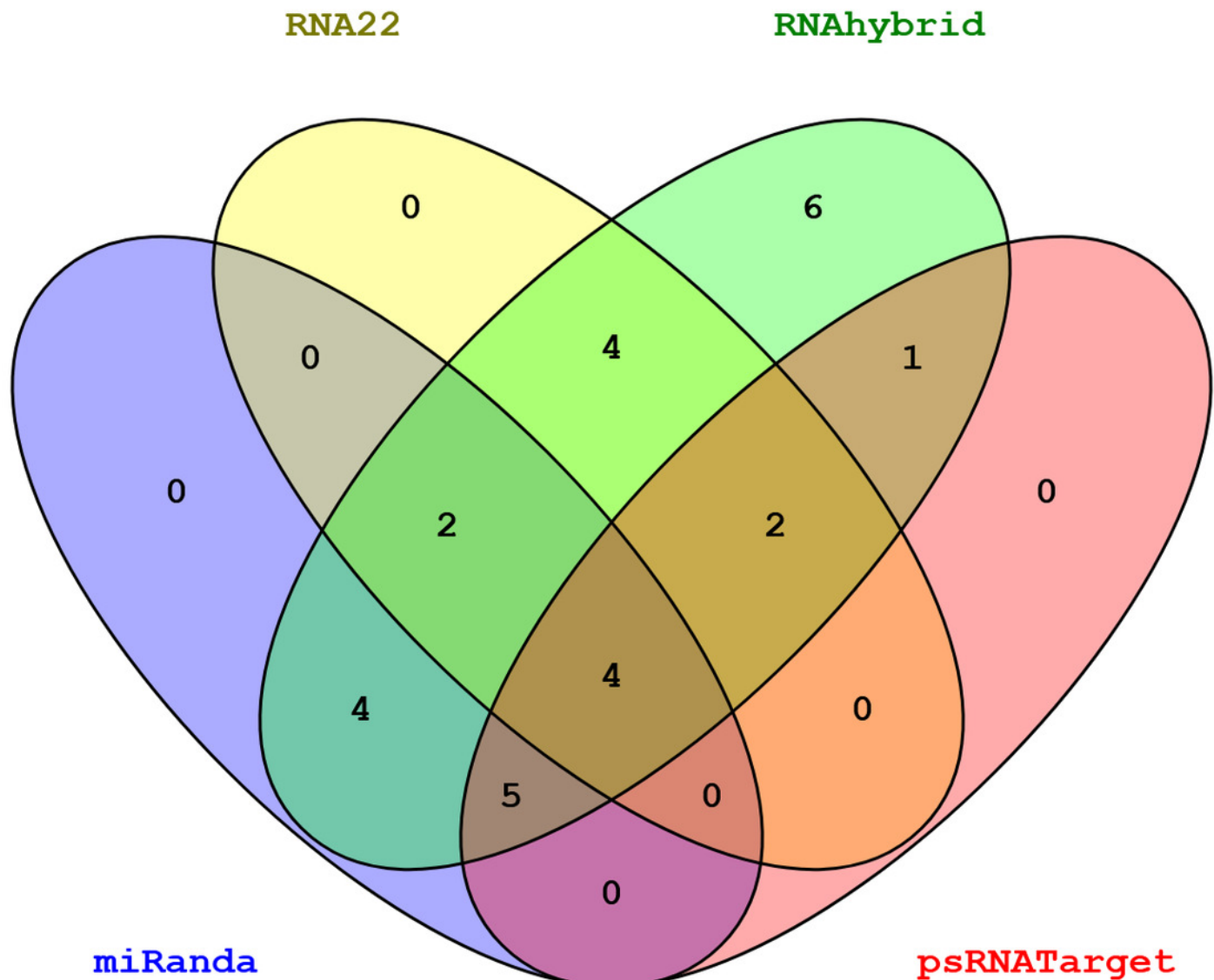


Figure 6

Intersection plot showing sugarcane miRNAs

Intersection plot showing sugarcane miRNAs predicted from at least two algorithms.

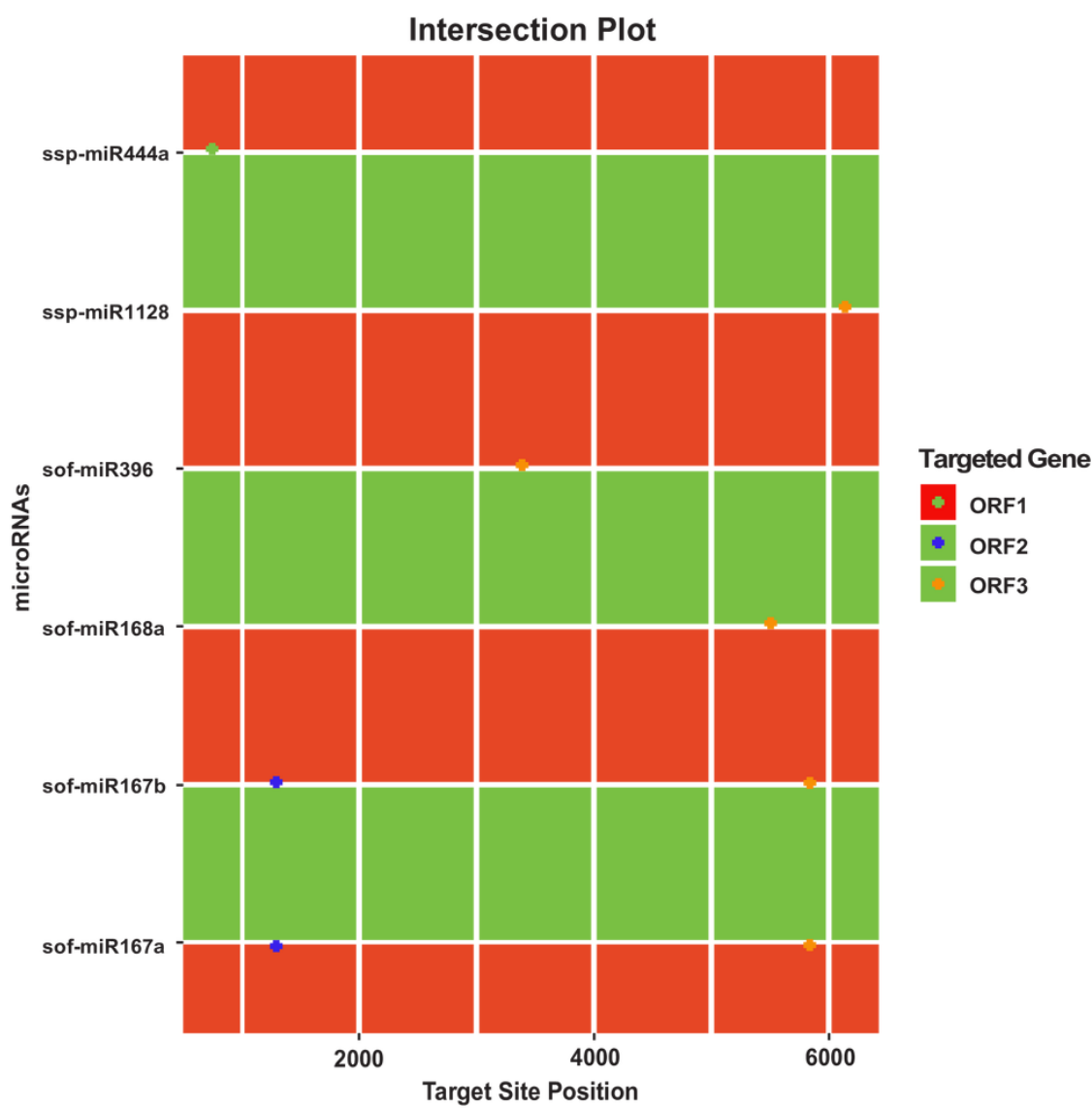


Figure 7

Stable secondary structure of top seven sugarcane miRNAs.

Stable secondary structures of top seven pre-miRNAs, precursors of the mature sugarcane miRNAs found in the study as the miRNAs detected by a consensus of algorithms. Names with minimum free energy (MFE) are represented: A) sof-miR396 (-67.40Kcal/mol) B) sof-MIR167a (-82.70Kcal/mol) C) sof-MIR167b (-86.00Kcal/mol) D) sof-MIR168a (-66.20Kcal/mol), E) sof-MIR168b (-56.10Kcal/mol), F) ssp-miR444a (-57.70Kcal/mol), G) ssp-miR1128 (-101.70).

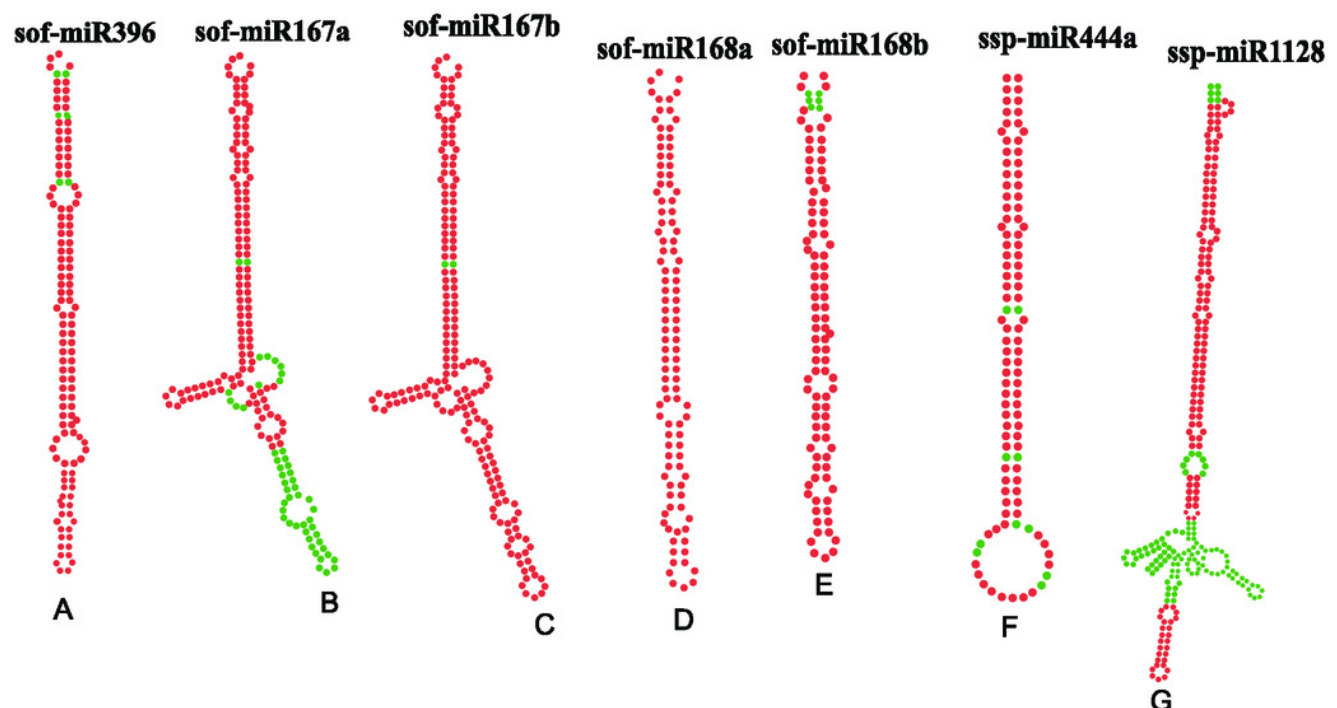


Table 1(on next page)

miRNA target prediction algorithms

Characteristic features considered by miRNA target prediction algorithms under study.

1 Table.1: Characteristic features considered by miRNA target prediction algorithms under study.

Algorithms	Main Parameters	Availability
miRanda	Seed pairing, target site accessibility, multiple target sites and translation inhibition	Web server and source code (http://cbio.mskcc.org/miRNA2003/miranda.html)
RNA22	Seed pairing, target site accessibility and multiple target sites	Only Web server https://cm.jefferson.edu/rna22/Interactive/
RNAhybrid	Seed pairing, target site accessibility, multiple target sites and translation inhibition	Web server and source code (http://bibiserv.techfak.unibielefeld.de/rnahybrid)
psRNATarget	Target site accessibility, multiple target sites and translation inhibition	Only Web server http://plantgrn.noble.org/psRNATarget/
Tapirhybrid	Seed pairing, target site accessibility and multiple target sites	Web server and source code http://bioinformatics.psb.ugent.be/webtools/tapir/
Targetfinder	Seed pairing	Only source code
Target-align	Multiple target sites	Web server and source code http://www.leonxie.com/targetAlign.php

2

Table 2(on next page)

miRNA-target pairs were selected from miRanda analysis.

miRNA-target pairs were selected from miRanda analysis. Locus position was selected by consensus analysis of at least two algorithms. MFE* and mode of target Inhibition** was determined by RNAhybrid.

- 1 Table.2: Selected miRNA-target pairs obtained after miRanda analysis. Locus position was selected by
- 2 consensus analysis of at least two algorithms.

Sugarcane microRNAs	miRNA-target pair	Locus position	MFE* (kcal/mol)
sof-miR167(a, b)	Query: 3' gucUAGUACGACCGUCGAAGu 5' : Ref: 5' aaaATCAAGTT-GCAGCCTCa 3'	5846-5865	-26.60
sof-mi396	Query: 3' guCAAGUUC-UUUCGACACCUu 5' : : Ref: 5' agGATTAGGTGATGCTGTGGAg 3'	3394-3415	-24.90
ssp-miR528	Query: 3' gaggAGACGUACGGGGAAGGu 5' Ref: 5' gcgaTCCGC-CCCCCCTTCC- 3'	7426-7444	-28.0

- 3 MFE (Minimum Free Energy)* was determined by RNAhybrid.

Table 3(on next page)

Sugarcane miRNAs and their target position.

Sugarcane miRNAs and their target positions in the SCBGAV genome identified by all four algorithms. *MFE: Minimum free energy measured in /Kcal/mol where *MFE represents minimum folding energy measured in Kcal/mol.

1 **Table3:** Sugarcane miRNAs and their target positions in the SCBGAV identified by algorithms.

miRNA Name	Position miRanda	Position RNA22	Position RNAhybrid	Position psRNATarget	MFE* miRand a	MFE** RNA22	MFE RNAhybrid	Expectation psRNATarget
sof-miR156			7104				-23.2	
sof-miR159a	5282		1659	3779	-17.18		-25.4	6
sof-miR159a(1)	6739				-18.11			
sof-miR159b	5282		1659	3779	-17.18		-25.4	6
sof-miR159b(1)	6739				-18.11			
sof-miR159c	6739		6896	3779	-16.70		-28	6
sof-miR159d	5282		1659	3779	-17.18		-25.4	6
sof-miR159d(1)	6739				-18.11			
sof-miR159e	5282		733		-15.81		-25.5	
sof-miR167a	5846	1310	1304	5846	-16.08	-15.80	-26.6	7
sof-miR167b	5846	1310	1304	5846	-16.08	-15.80	-26.6	7
sof-miR168a	5506	5506	6084		-21.24	-17.70	-25.6	
sof-miR168a(1)	7050	7050			-18.45	-17.10		
sof-miR168a(2)		5137				-14.30		
sof-miR168b	7050	7050	6937		-20.05	-19.40	-25.6	
sof-miR168b(1)		5506				-13.70		
sof-miR396	3394	3394	3394	5732	-19.99	-17.80	-24.9	5.5
sof-miR396(1)		4115				-14.20		
sof-miR408a		1735	5136			-13.70	-27.6	
sof-miR408b		1735	5136			-13.70	-27.6	
sof-miR408c		1735	5136			-13.70	-27.6	
sof-miR408d		1735	5136			-13.70	-27.6	
sof-miR408e			6152				-28.9	
ssp-miR166	3249		5572	981	-21.45		-26.7	6.5
ssp-miR166(1)	5589				-17.54			
ssp-miR166(2)	5761				-21.74			
ssp-miR169			752				-24.8	
ssp-miR437a			978				-21.7	
ssp-miR437b			6829				-21.8	
ssp-miR437c			978				-24.3	
ssp-miR528	7426	1310	7407	2897	-18.37	-13.70	-28	6.5
ssp-miR528(1)	7395				-18.05			
ssp-miR827	5378		3010		-16.40		-24	
ssp-miR444a		4103	774	775		-16.30	-27.6	6.75
ssp-miR444b		4103	774	756		-16.30	-27.6	6.0
ssp-miR444b(1)				2615				6.5
ssp-miR444b(2)				5009				6.5
ssp-miR444b(3)				5731				6.5
ssp-miR444c-3p			774	4940			-26.7	6.5
ssp-miR1128	6141		6148		-22.96		-27.8	
ssp-miR1432	6070		1301		-15.48		-20.4	

2 *MFE: Minimum free energy measured in /Kcal/mol where *MFE represents minimum folding
3 energy measured in Kcal/mol.

Table 4(on next page)

features of secondary structures of precursor miRNAs

Predicted characterized features of secondary structures of precursor miRNAs.

¹MFE represents minimum free energy. ²AMFE is an adjusted minimum free energy calculated by MFE/length of precursor*100. MFEI defines as minimum free energy index which was calculated by AMFE/ (G+C) %.

1 Table.4: Predicted characterized features of secondary structures of precursor miRNAs.

miRNA ID	miRNA Sequence	Length miRNA	Length precursor	MFE ¹ (Kcal/mol)	AMFE ²	MFEI ³	(G+C)%
sof-miR396	UUCCACAGCUUUCUUGAACUG	21	134	-67.40	-50.29	-1.17	42.85
sof-miR167a	UGAAGCUGCCAGCAUGAUCUG	21	188	-82.70	-43.98	-0.83	52.38
sof-miR167b	UGAAGCUGCCAGCAUGAUCUG	21	188	-86.00	-45.74	-0.87	52.38
sof-miR168a	UCGCUUGGUGCAGAUCCGGGAC	21	104	-66.20	-63.65	-1.02	61.90
sof-miR168b	UCGCUUGGGCAGAUCCGGGAC	20	103	-56.10	-54.46	-0.83	65.00
ssp-miR444a	UGCAGUUGUUGCCUCAAGCUU	21	105	-57.70	-54.95	-1.14	47.61
ssp-miR1128	UACUACUCCCUCCGUCCCAA	21	275	-101.70	-36.98	-0.70	52.38

2 ¹MFE represents minimum free energy. ²AMFE is an adjusted minimum free energy calculated by MFE/length of

3 precursor*100. MFEI defines as minimum free energy index which was calculated by AMFE/ (G+C) %.