Review

Overview of the potential of microRNAs and their target gene detection for cassava (Manihot esculenta) improvement

Samuel Amiteye1,2*, José M. Corral2 and Timothy F. Sharbel2

1Biotechnology and Nuclear Agriculture Research Institute, Ghana Atomic Energy Commission, P. O. Box LG 80, Legon-Accra, Ghana.
2Leibniz Institute of Plant Genetics and Crop Plant Research, International Max Planck Research School, Correnstrasse 3, 06466 Gatersleben, Germany.

Accepted 5 May, 2010

Production and utilization of cassava (Manihot esculenta) is significantly constrained by pests, diseases, poor yields and low nutritional content. Various approaches are currently being applied to mitigate these constraints. However, an aspect of plant developmental genetics little known in cassava is the role that microRNAs (miRNAs) play in gene regulation. miRNAs are 20 - 24 nucleotide long non-protein-coding RNAs that play important roles in post-transcriptional gene silencing in many organisms. Thorough understanding of the mechanisms involved in miRNAs mediated post-transcriptional gene regulation will have implications for crop production improvement. The potential of miRNAs for cassava improvement and also some data obtained on cassava miRNAs using comparative genomics analysis have been briefly discussed. 17 miRNA families and target genes in cassava that are also conserved in other plant species have been revealed. However, the ESTs representing seven of these miRNA families produced foldbacks that show more than 3 nucleotides not involved in canonical base pairing within a loop or bulge in the mature miRNA: RNA* dimer, thus were not considered miRNA secondary structures. Consistent with previous reports, majority of the target genes identified are transcription factors. Other targets appear to play roles in diverse physiological processes. Furthermore, a detailed description and insight into some of the current bioinformatic resources and approaches applicable to cassava have been discussed. Such information will further enhance the rapid discovery and analysis of more novel miRNAs in cassava towards its improvement.

Key words: Cassava, microRNAs, target genes, improvement, characterization.

INTRODUCTION

Cassava (Manihot esculenta) is one of the most important root crops in Africa. It has high adaptability to low-fertility soils and is able to withstand erratic rainfall and long drought conditions (Mathews et al., 1993). For this reason, cassava is a major food security crop and a source of cash for resource-poor farmers. Today, over half of the world’s cassava is produced in Africa. Nonetheless, it is faced with production constraints of pest, diseases and poor yields everywhere it is grown on the continent. The widespread use of cassava for food and industry places a high demand for increased production.

To achieve significant production will require the development of cultivars with higher yields, improved resistance to biotic and abiotic stresses and high nutritional content. Some significant progress is being made with various strategies in this regard. However, an aspect of plant development regulation that has not been exploited in our attempt to develop cassava varieties is the potential application of the knowledge of microRNAs (miRNAs) in gene regulation during plant growth and development. Thorough understanding of the mechanisms involved in miRNAs mediated post-transcriptional gene regulation will have implications on the strategies currently being developed to enhance the crop. Newly discovered miRNAs may play important role in cassava development.

*Corresponding author. E-mail: samamiteye@yahoo.com.
and the response to biotic and abiotic stresses. Although miRNAs have been studied for years, to date no study has been performed on cassava. Accurate detection and expression profiling of these miRNAs will enable a better understanding of their role in plant growth and development (Millar and Gubler, 2005; Rhoades and Bartel, 2004; Jagadeeswaran et al., 2009). Breakthrough in this area is likely to reveal developmental regulation and disease mechanisms related to miRNAs.

In an attempt to stimulate more research interest in cassava miRNAs, an overview of miRNAs and some current bioinformatic resources and approaches available for prediction and analysis of plant miRNAs have been discussed. Furthermore, relevant bioinformatic resources and approaches applicable to cassava as evident from data presented here have been described. Comparative genomics across vastly divergent taxa has shown that many miRNAs are highly conserved from species to species in the plant and animal kingdoms (Reinhart et al., 2002; Sunkar and Jagadeeswaran, 2008). The extensive evolutionary conservation of miRNAs provides an effective approach to their identification using comparative genomics. In plants, miRNAs have perfect to near perfect homology to their target genes and generally have only single site within their target (Subramanian et al., 2009). With the development of new algorithms based on these characteristics, computational approaches to plant miRNA prediction and characterization are on the increase. Computational methods have been shown to be efficient for the identification of putative miRNAs in a wide variety of plant species including Arabidopsis thaliana (Rhoades and Bartel, 2004; Sunkar and Zhu, 2004), Rice (Sunil and Nagaraju, 2007), Maize (Mica et al., 2006), Tomato (Pilcher et al., 2007), Irish potato (Qiang et al., 2007) and Soybean (Zhang et al., 2008).

The recent breakthrough in deciphering the cassava genome sequence (Cassava Genome Project 2009, http://www.phytozome.net/cassava) coupled with a number of available up-to-date bioinformatic approaches has now made it possible to identify and characterize miRNAs in cassava towards enhanced crop improvement.

**MICRO-RNAs: BIOGENESIS AND MECHANISM OF GENE REGULATION**

**A brief description**

miRNAs are 20 - 24 nucleotide small endogenous non-protein-coding regulatory RNA sequences that are produced by genes distinct from the genes that they regulate. Biogenesis of miRNAs involves nucleolytic processing of a precursor transcript with extensive foldback structure (Han et al., 2004; Vazquez et al., 2004; Yang et al., 2006). Evidence provided by Allen et al. (2004) and Felippes et al. (2008) show that some miRNAs evolved by inverted duplications of target gene sequences whereas others originated from random sequences that either have self-complementarity by chance or sequences that represent highly eroded inverted duplications. Since their discovery, miRNAs have become the interest of intensive research in various model organisms. Several miRNAs have been computationally and/or experimentally identified and characterized in many species. A number of studies have shown that miRNAs play key roles in regulatory functions of gene expression for most eukaryotes (Sunkar and Zhu, 2007; Jagadeeswaran et al., 2009), mainly at the posttranscriptional levels (Chapman et al., 2004; Mallory et al., 2004).

miRNA sequences have been shown to be highly conserved among plant species (Sunkar and Jagadeeswaran, 2008) and have perfect or near perfect complementarity with their specific messenger RNA (mRNA) targets. As a result, a plant miRNA guides the cleavage, degradation or translational inhibition of its target mRNA with a resultant repression of gene expression. Apart from the miRNA section, other parts of the corresponding precursor pre-miRNA sequences are less conserved (Lau et al., 2001). Initially, it was thought that the sole purpose for the miRNA gene regulatory mechanism was to defend plants against RNA viruses and transposons. However, several recent findings have implicated miRNAs in a number of biological mechanisms including leaf growth (Palatnik et al., 2003), stem growth (Mallory et al., 2004), root growth (Subramanian et al., 2009), floral organ identity, control of female gamete formation and reproductive development (Millar and Gubler, 2005; Olmedo-Monfil et al., 2010), auxin signaling (Rhoades and Bartel, 2004), biotic and abiotic stress response (Jagadeeswaran et al., 2009).

**Biogenesis pathways**

A simplified scheme of the plant miRNA biogenesis pathway is shown in Figure 1. miRNA genes are first transcribed in the nucleus by RNA polymerase II into primary transcripts referred to as pri-miRNAs, which are subsequently capped at the 5’ end and polyadenylated at the 3’ end (Lee et al., 2004). The pri-miRNA is further processed by an enzyme microprocessor complex which includes Dicer like-1 nucleases (members of an RNase III endonuclease family) and Arabidopsis hyponastic leaves (HYL1), a dsRNA-binding domain containing protein, bound to the pri-miRNA complex (Jones-Rhoades et al., 2006). The activity of the microprocessor complex culminates in pre-miRNA stem loop structures from which miRNAs are eventually processed.

In A. thaliana, all of the miRNA biogenesis steps are processed by one of four Dicer-like RNase III endonucleases (Filipowicz et al., 2008). Dicer-like-1 processes primary miRNA transcripts into an miRNA-miRNA* duplex with 2 nucleotide overhangs at the 3’ end (Reinhart et al., 2002), while the other Dicer-like enzymes are involved with another type of small RNA biogenesis and accumulation known as short interfering RNAs (siRNAs).
methyl transferase, adds methyl groups to the 3’ ends of the miRNA duplex and stabilizes it (Yu et al., 2005). The miRNA duplex is then exported into the cytoplasm with the help of HASTY (HST), a plant ortholog of a miRNA transporter enzyme, Exportin-5, which in animal systems transports pre-miRNAs to the cytoplasm (Yi et al., 2003). Based on the relative internal thermodynamic stability of the two ends of the duplex, one strand becomes the mature miRNA whereas the other complementary strand is degraded by an unknown nuclease (Khvorova et al., 2003; Schwarz et al., 2003).

**Mechanism of miRNA-mediated gene repression**

The mature strand of the miRNA is loaded into a complex of ribonucleotide proteins to form the miRNA-induced silencing complex (miRISC) to cleave its specific target mRNA or to inhibit the translation of its target transcript (Vaucheret et al., 2004; Baumberger and Baulcombe, 2005). The miRISC complex comprises mainly Argonoute (AGO1) and an RNaseH-like P-element induced wimpy testis (PIWI) domain containing protein. The AGO protein family is the most important and key component of the miRNA-RISC complex (Schwarz et al., 2003). The primary proteins in this complex are members of the Argonoute (AGO) family, each of which possesses repressive capabilities. The single stranded mature miRNA together with the PIWI domain are used as guide in the miRISC to recognize the target mRNA to which it may be perfectly or near perfectly complementary in plants unlike in animals.

In all known cases, most plant miRNAs bind to the protein-coding region of their target mRNAs induce target mRNA degradation via an RNAi-like mechanism where the AGO protein cleaves the miRNA-mRNA duplex (Figure 1), thereby repressing expression of that particular mRNA (Rhoades et al., 2002). Gene repression can also be effected by translational inhibition through deacylation of the 3’ poly (A) tail and decapping of the 5’ end in mRNAs, which leads to progressive mRNA decay and degradation (Chen, 2004; Brodersen et al., 2008). Generally, miRNA biogenesis is similar to a large extent between plants and animals, however, fundamental differences exist. A comparison of the main differences have been summarised in Table 1. These and other differences suggest that the two systems may have evolved independently, possibly as a prerequisite to suite efficiently the regulatory mechanisms of the respective developmental complexities.

**MITIGATING CASSAVA PRODUCTION AND UTILIZATION CONSTRAINTS**

Many initiatives both local and international recognize the importance of cassava in Africa. Most of these initiatives aid increased productivity through biotechnological
research for development of improved varieties, provision of improved germplasm and also offer training in new methods of cassava molecular breeding. One of such initiatives is the International Centre for Tropical Agriculture (CIAT) project “Genetic Enhancement of Cassava” (http://www.ciat.cgiar.org/africa/cassava.htm). CIAT has developed Cassava Mosaic Disease (CMD) markers and tested their suitability in marker-assisted selection for CMD resistance on several progenies. The centre has also conducted marker assisted breeding of lines to combine CMD resistance, high carotene content and good dry matter yield. These lines will further be recombined with high protein lines currently being developed.

In order to make available important biotechnology tools to complement conventional breeding programs in cassava, the International Institute of Tropical Agriculture (IITA), Nigeria has generated and characterized an 18,166 expressed sequence tag (ESTs) dataset enriched for drought responsive genes. The institute has also developed a 60-mer oligo DNA microarray comprising 35,785 ESTs as well as genetic transformation protocols for the genetic modification of model cassava genotypes.

Furthermore, the International Laboratory for Tropical Agricultural Biotechnology (ILTAB), USA with support from USAID, the Bill and Melinda Gates Foundation and in collaboration with scientists from East Africa is working to enhance virus resistance in cassava. They are employing two main strategies; ssDNA binding protein based strategy and gene silencing strategy to produce cassava disease resistant varieties. They are also examining the processing of short interfering RNAs (siRNAs) in relation to cassava geminiviruses. It is hoped that their results will have implications on the strategies being developed to control cassava geminiviruses in Africa.

Another interesting aspect of cassava being developed in collaboration with scientists from East Africa is working to enhance virus resistance in cassava. They are employing two main strategies; ssDNA binding protein based strategy and gene silencing strategy to produce cassava disease resistant varieties. They are also examining the processing of short interfering RNAs (siRNAs) in relation to cassava geminiviruses. It is hoped that their results will have implications on the strategies being developed to control cassava geminiviruses in Africa. Another interesting aspect of cassava being studied by ILTAB and an international team of cassava researchers is the development of cassava varieties with enhanced nutritional content. In this regard, the laboratory has developed transgenic cassava plants that are currently under greenhouse investigations.

As for the processing of small RNAs, there are different pathways that lead to their biogenesis in both plant and animal systems. One of these pathways is the processing of pri-miRNAs by Dicer-like enzymes in the cytoplasm of the cell. This process results in the production of mature miRNAs that can then bind to target genes and repress their expression. The presence of miRNA binding sites on target genes is essential for this process to occur.

Table 1. Comparison of miRNA biogenesis in plant and animal systems.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Plant system</th>
<th>Animal system</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-miRNAs processing</td>
<td>Processed from pri-miRNA by Dicer-like 1 in the nucleus&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Processed from pri-miRNA by Drosha protein in the nucleus&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>miRNA processing</td>
<td>Processed from pre-miRNA by Dicer-like 1 in the nucleus&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Processed from pre-miRNA by Dicer protein in the cytoplasm&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>miRNA methylation</td>
<td>miRNA duplexes are methylated at 3' ends by HEN1 enzyme&lt;sup&gt;e&lt;/sup&gt;</td>
<td>Methylation not known&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>Molecular transporters</td>
<td>miRNA is moved from nucleus to cytoplasm by the HASTY protein&lt;sup&gt;g&lt;/sup&gt;</td>
<td>pre-miRNA is moved from the nucleus to cytoplasm by Exportin-5&lt;sup&gt;h&lt;/sup&gt;</td>
</tr>
<tr>
<td>Stem-loop structure</td>
<td>Stem-loops are longer and highly variable&lt;sup&gt;i&lt;/sup&gt;</td>
<td>Stem-loops are shorter and more uniform&lt;sup&gt;j&lt;/sup&gt;</td>
</tr>
<tr>
<td>Source of miRNAs</td>
<td>Mainly from non-coding transcriptional units&lt;sup&gt;k&lt;/sup&gt;</td>
<td>Arise from introns and protein coding sequences&lt;sup&gt;l&lt;/sup&gt;</td>
</tr>
<tr>
<td>SmallRNAs in transcriptome</td>
<td>miRNAs constitute a small fraction of smallRNA populations&lt;sup&gt;m&lt;/sup&gt;</td>
<td>smallRNA populations are predominantly miRNAs&lt;sup&gt;n&lt;/sup&gt;</td>
</tr>
<tr>
<td>Presence of miRNA clusters</td>
<td>Clusters are rare&lt;sup&gt;o&lt;/sup&gt;</td>
<td>Clusters occur frequently&lt;sup&gt;p&lt;/sup&gt;</td>
</tr>
<tr>
<td>miRNA target binding</td>
<td>Have perfect or near perfect complementarity with target site&lt;sup&gt;q&lt;/sup&gt;</td>
<td>miRNA could bind imperfectly&lt;sup&gt;r&lt;/sup&gt;</td>
</tr>
<tr>
<td>miRNA binding sites</td>
<td>Predominantly bind coding regions of target genes&lt;sup&gt;s&lt;/sup&gt;</td>
<td>miRNAs bind the 3' UTRs of target genes&lt;sup&gt;t&lt;/sup&gt;</td>
</tr>
<tr>
<td>Number of miRNA binding sites</td>
<td>Generally single binding sites in target genes&lt;sup&gt;u&lt;/sup&gt;</td>
<td>Generally targets have multiple binding sites&lt;sup&gt;v&lt;/sup&gt;</td>
</tr>
<tr>
<td>Repression of gene expression</td>
<td>Generally by direct mRNA cleavage&lt;sup&gt;w&lt;/sup&gt;</td>
<td>In most cases by translational attenuation&lt;sup&gt;x&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>Filipowicz et al., (2008); <sup>b</sup>Lee et al., (2003); <sup>c</sup>Bernstein et al., (2001); <sup>d</sup>Yi et al., (2003); <sup>e</sup>Bartel (2004); <sup>f</sup>Lu et al., (2005); <sup>g</sup>Minino et al., (2006); <sup>h</sup>Millar and Waterhouse (2005); <sup>i</sup>Subramanian et al., (2009); <sup>j</sup>Sunkar and Zhu (2004); <sup>k</sup>Rajewsky (2006); <sup>l</sup>Baumberger and Baulcombe (2005).
cassava improvement. Application of induced mutation techniques in Ghana has led to the production of an improved mutant cassava. Irradiation of stem cuttings using gamma irradiation resulted in the production of “Tek bankye,” a mutant variety with high dry matter content (40%) and good poundability. BNARI is also currently using irradiation techniques in combination with molecular methods to develop disease resistant and high starch yielding cassava varieties to enhance the production and utilization of the crop.

**APPROACHES FOR CASSAVA MICRONAS AND TARGET IDENTIFICATION**

**Screening for previously known miRNA families**

miRBase (http://microrna.sanger.ac.uk/) is the main database of miRNAs from most plant and animal species as well as viruses and fungi. The latest version of miRBase released on September 14, 2009, (http://microrna.sanger.ac.uk/sequences/), consists of 10,581 miRNA and 10,883 precursor hairpin structures entries from 115 species. These miRNAs have been identified via computational and/or experimental methods. The database is constituted into three sections: the miRBase:: Registry, where novel miRNA data is uploaded prior to publication; information on miRNA sequences, nomenclature and target miRNAs can be obtained from the miRBase:: Sequences section; and, the third section miRBase:: Targets provides the prediction of the mRNA target from all published miRNAs (Griffiths-Jones et al., 2008). A total of 212 previously reported A. thaliana miRNAs and their precursor sequences were screened from miRBase:: Sequences. These miRNAs served as a reference set of miRNA sequences for the identification of putative miRNA families in cassava (Table 2). The sequences were used as query sequences for BLASTn search against the publicly available cassava ESTs in the cassava genome database (http://www.phytozome.net/cassava.php). ESTs that aligned 19 - 22 nucleotides with known A. thaliana miRNAs were selected for prediction of putative secondary stem-loop structures. The maximum number of mismatches allowed was 0 - 3 with no gaps.

Work undertaken by Smalheiser (2003) and Zhang et al. (2005) show successful identification of miRNAs by mining the repository of available ESTs. This is feasible because miRNAs are generated from long precursor hairpin structures that can be found by searching ESTs. The strategy is to look for the sequences containing conserved mature miRNAs and then check if these miRNA bearing ESTs can form stable secondary stem-loop structures (Figure 2; Table 2). Another useful miRNA database is the Plant miRNA Database (PMRD) released October 1, 2009, (http://bioinformatics.caau.edu.cn/PMRD). This first edition of the database contains 8433 miRNAs from 121 plant species, their secondary structures, target genes, expression profiles and a genome browser.

**High-throughput 454 pyrosequencing of short RNAs and transcriptomes**

Bioinformatic methods have been invaluable in the identification of many plant miRNAs, however, these approaches identify mainly conserved or evolutionarily “old” miRNAs. Besides, some non-conserved miRNAs mostly considered recently evolved “young” miRNAs have been discovered and appear to be species-specific (Allen et al., 2004; Felippes et al., 2008). Until recently, most sequence information such as ESTs or genome survey sequences (GSS) used for computational prediction of miRNAs were generated by the traditional Sanger sequencing method. Compared to conserved miRNAs, non-conserved miRNAs are often expressed at lower levels which make their detection more daunting using small-scale sequencing. The development of high-throughput 454 pyrosequencing technology has, therefore, allowed the discovery of several non-conserved or lowly expressed miRNAs through cloning and deep sequencing of small RNA and transcriptome libraries in A. thaliana (Rajagopalan et al., 2006; Fahlgren et al., 2007), wheat (Yao et al., 2007), tomato (Moxon et al., 2008).

**Prediction of miRNA secondary stem-loop structures**

Irrespective of the method used to generate sequence information, the prediction of putative pre-miRNA structures is achieved mainly using bioinformatic softwares some of which are publicly available. To predict putative pre-miRNA secondary hairpin structures for selected cassava ESTs, the programme Mfold (http://frontend.bioinfo.rpi.edu/applications/mfold/cgi-bin/rna-form1.cgi), a webserver was used. This programme is widely used for predicting the minimum fold energy (MFE) secondary structure using the algorithm originally proposed by Zuker and Stiegler, (1981). Equilibrium base-pairing probabilities of MFE structures are calculated via McCaskill’s partition function (PF) algorithm (McCaskill, 1990). Besides, RNAmFold (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi) can also be used to find potential miRNA hairpin structures from EST sequences (Hofacker, 2003). These and many other useful programmes, their applications and methodologies for miRNA identification, characterization and analysis have been recently reviewed by Unver et al. (2009).

The secondary structure output of the candidates of pre-miRNAs sequences obtained in cassava were analysed and scored for their potential to form miRNA precursors with the following widely accepted characteristics (Xie et al., 2007): (1) The RNA sequences should be able to fold into a hairpin structure, such that one arm of the hairpin contains the 20 - 22 nucleotide mature miRNA sequence. (2) The predicted secondary structure of the miRNA precursor should have lower minimal fold energy (MFE) and minimal fold energy index (MFEI) than other types of RNA (example, tRNA, rRNA). (3) The
Table 2. Characteristics of identified putative miRNAs and stem-loop structures from ESTs of cassava.

<table>
<thead>
<tr>
<th>miRNAs</th>
<th>Sequences</th>
<th>NM</th>
<th>LP</th>
<th>ARM</th>
<th>A+U%</th>
<th>AMFE</th>
<th>MFEI</th>
<th>EST ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>miR156a-f</td>
<td>UGACAGAAGAGAGAGAGCAC</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>FF535946</td>
<td></td>
</tr>
<tr>
<td>miR156g</td>
<td>UGACAGAAGAGAGAGAGCAC</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>FF535946</td>
<td></td>
</tr>
<tr>
<td>miR157a-c</td>
<td>AUGCAGAAGAGAGAGAGCAC</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>FF535480</td>
<td></td>
</tr>
<tr>
<td>miR157d</td>
<td>UGACAGAAGAGAGAGAGCAC</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>FF535480</td>
<td></td>
</tr>
<tr>
<td>miR159a</td>
<td>UUGGAUUGAAGGGAGCUCUA</td>
<td>0</td>
<td>181</td>
<td>3'</td>
<td>55.25</td>
<td>47.07</td>
<td>1.05</td>
<td>DV457265</td>
</tr>
<tr>
<td>miR159b</td>
<td>UUGGAUUGAAGGGAGCUCUA</td>
<td>1</td>
<td>89</td>
<td>3'</td>
<td>58.95</td>
<td>32.47</td>
<td>0.86</td>
<td>DB921728</td>
</tr>
<tr>
<td>miR160a-c</td>
<td>UGCCUGACUCUCUGAUUGAGCAG</td>
<td>3</td>
<td>98</td>
<td>3'</td>
<td>44.90</td>
<td>35.41</td>
<td>0.64</td>
<td>DB935104</td>
</tr>
<tr>
<td>miR162a,b</td>
<td>UCGAUAAACCUCUCAGCAG</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>FF535946</td>
<td></td>
</tr>
<tr>
<td>miR166a-g</td>
<td>UCGGACCAGGCUUCAUUCCC</td>
<td>0</td>
<td>122</td>
<td>3'</td>
<td>60.82</td>
<td>33.77</td>
<td>0.86</td>
<td>DB922056</td>
</tr>
<tr>
<td>miR166a-g</td>
<td>UCGGACCAGGCUUCAUUCCC</td>
<td>1</td>
<td>122</td>
<td>3'</td>
<td>62.92</td>
<td>33.58</td>
<td>0.91</td>
<td>DV447797</td>
</tr>
<tr>
<td>miR166a-g</td>
<td>UCGGACCAGGCUUCAUUCCC</td>
<td>1</td>
<td>171</td>
<td>3'</td>
<td>58.10</td>
<td>36.92</td>
<td>0.76</td>
<td>DV449619</td>
</tr>
<tr>
<td>miR167a,b</td>
<td>UGAAGCUGCCAGCAUGAUCU</td>
<td>0</td>
<td>177</td>
<td>3'</td>
<td>58.20</td>
<td>45.12</td>
<td>1.03</td>
<td>DV454131</td>
</tr>
<tr>
<td>miR171a</td>
<td>UGAUUGAGCCGUAAAUC</td>
<td>1</td>
<td>132</td>
<td>3'</td>
<td>59.25</td>
<td>45.15</td>
<td>1.12</td>
<td>DV454131</td>
</tr>
<tr>
<td>miR172a,b</td>
<td>AGAAUCUUGAUGAUGCUGCAU</td>
<td>0</td>
<td>73</td>
<td>3'</td>
<td>63.48</td>
<td>13.56</td>
<td>0.38</td>
<td>DV449432</td>
</tr>
<tr>
<td>miR172c,d</td>
<td>AGAAUCUUGAUGAUGCUGCAU</td>
<td>1</td>
<td>177</td>
<td>3'</td>
<td>53.67</td>
<td>36.82</td>
<td>0.81</td>
<td>CK645041</td>
</tr>
<tr>
<td>miR172e</td>
<td>AGAAUCUUGAUGAUGCUGCAU</td>
<td>1</td>
<td>73</td>
<td>5'</td>
<td>64.38</td>
<td>13.56</td>
<td>0.38</td>
<td>DV449432</td>
</tr>
<tr>
<td>miR390a,b</td>
<td>AAGCUCAGGAGAGAUA</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>DV447689</td>
<td></td>
</tr>
<tr>
<td>miR394a,b</td>
<td>UGUGCAUUCUGUCCACCU</td>
<td>0</td>
<td>151</td>
<td>5'</td>
<td>54.30</td>
<td>45.96</td>
<td>1.01</td>
<td>CK562213</td>
</tr>
<tr>
<td>miR398a</td>
<td>UGUGU UCUCAGGAGAGAUA</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>DV449231</td>
<td></td>
</tr>
<tr>
<td>miR398b,c</td>
<td>UGUGU UCUCAGGAGAGAUA</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>DV449231</td>
<td></td>
</tr>
<tr>
<td>miR399a-c</td>
<td>UGCCAAAGGAGAGAUA</td>
<td>1</td>
<td>121</td>
<td>3'</td>
<td>58.20</td>
<td>45.12</td>
<td>1.03</td>
<td>DB931132</td>
</tr>
<tr>
<td>miR399a-c</td>
<td>UGCCAAAGGAGAGAUA</td>
<td>1</td>
<td>112</td>
<td>3'</td>
<td>58.04</td>
<td>44.38</td>
<td>1.06</td>
<td>DB945966</td>
</tr>
<tr>
<td>miR399a-c</td>
<td>UGCCAAAGGAGAGAUA</td>
<td>1</td>
<td>127</td>
<td>3'</td>
<td>55.12</td>
<td>44.33</td>
<td>0.99</td>
<td>DB947384</td>
</tr>
<tr>
<td>miR399f</td>
<td>UGCCAAAGGAGAUUUGCUG</td>
<td>0</td>
<td>111</td>
<td>3'</td>
<td>52.25</td>
<td>40.81</td>
<td>0.85</td>
<td>DB932374</td>
</tr>
<tr>
<td>miR408</td>
<td>AUGCACUGCCUCUCCCUGGC</td>
<td>0</td>
<td>134</td>
<td>3'</td>
<td>50.00</td>
<td>49.03</td>
<td>0.98</td>
<td>DV444343</td>
</tr>
<tr>
<td>miR414</td>
<td>UCAUCUUCAUCAUCAUCGU</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>DV444343</td>
<td></td>
</tr>
</tbody>
</table>

NM, Nucleotide substitutions between known A. thaliana query miRNAs and the corresponding new miRNA in cassava; LP, length of pre-miRNA; ARM, miRNA location in hairpin structure; AMFE, adjusted minimum fold energy; MFEI, minimum fold energy index; EST ID, identifier of the EST from which the miRNA was derived. Italicized, underlined and bold letters show nucleotide substitutions in new cassava miRNAs. *EST could not form secondary stem-loop structures.

predicted mature miRNA should have an A+U content of 30-70%. (4) The mature miRNA sequences within the hairpin loop segment must have less than six mismatches to the opposite miRNA* sequence in the other arm. (5) Any loop or bulge in the mature miRNA:RNA* dimer should contain a maximum of 3 nucleotides not involved in canonical base pairing. Using these criteria a number of pre-miRNA sequences corresponding to identified miRNAs in cassava were obtained that are similar to hairpins previously reported for other plant species (Figure 2). However, the ESTs FF535946, FF535480, DR087131, DV447689, FF379852, DV449231 and DV444343 representing seven miRNA families produced foldback structures that show more than 3 nucleotides not involved in canonical base pairing within a loop or bulge in the mature miRNA:RNA* dimer, thus were not considered miRNA secondary structures (Table 2). This may be because the miRNA regions are located at the 5' end of the sequences. Interestingly, Song et al. (2009) could show the expression in specific citrus tissues, some miRNAs whose precursor sequences could not fold back into stable hairpins.
miR159a

```
    ga  ua  u u ----|  auu  g  c  gau  c  u  uaaau  u
    uggagcuccuu  aguccaa  gaggaucu  gc  gguag  gagcu  cugag  uauag  ccacag  ccuauc  ca  aagac  c
    aucucpaggga  uuagggu  uuuccuaa  ug  uccauu  uucga  gacuc  auac  ggguuu  ggaauag  gu  uucug  u
    ag  ug  u  uauuc^  ac-  g  u  guuc  c  -  uc----  a
```

miR160a-c

```
    gaacaca  uca--  uuca  gca  -|  cgucac  uu
    ggca  ugu  ugca  aggggg  caggc  augg  u
    cccgu  aca  acgu  ucuccuc  guccg  ugcc  c
    a-------  ucauaa  cccg  aug  a^  ucu---  cu
```

miR166a-g

```
    cucuu  a  uu  cu  g  u  ----  u-|  cuuuuuuuu  cc  aaag  g  -  gau
    ggaaguuu  uug  ggggaau  gugucg  cga  gac  cu  cu  ugauc  uu  cuc  uauaucgaaua  au  cu  \
    cccucgaa  aac  ccccuuac  cggacc  gcu  cug  ga  ga  acuag  ag  gag  auauagguau  ua  ga  c
    aauu--  c  uu  ag  g  u  uauuc  cu^  uu--------  au  ga--  g  c  aau
```

miR167a,d

```
    u  --  a  c  c  a  \-aucuagcCAAacccuu|  u
    guuggu  uugag  gguagaacuc  ac  augaucuggu  aug  ucuua  a
    caacca  aacuc  caaucucgaag  uc  uacuggacua  uac  \-agau  c
    a  cu  a  u  -  c  \------------------^  g
```

miR168a,b

```
    g  u  ua  c  u  a  ug  a  \-gcuu|  u
    uuacc  gc  gucuc  auucg  uuggcagc  ccggc  c  auu  gcuuc  gcaauag  a
    agugg  cg  cagag  uagu  aacuagcuuc  gccc  g  uaa  cgaag  uguauau  c
    -  c  gc  c  c  a  gu  g  \----^  g
```

miR171

```
    u  ca--  g  ----  c  u-|  u  a  uga
    ggagagag  ggu  uag  ugaagg  uauuggcgug  cucaauc  aa  uac  uguuuaaa  u
    ucucucuc  ccaac  auu  acucuc  auacucgcgc  gaguuag  uu  aug  acgcuaau  a
    u  uuua  a  ucuc  c  uu^  c  a  ugu
```

miR172a,b

```
    ugauc  -  g  g  u  c  \-acaacacauu  -  uu  gga-  galu
    ccau  agucac  uauu  gc  guacgc  ca  cauacgaauuc  gaag  gc  caucgc  cggu  \ggua  uacaguc  auac  cg  caucgc  gcgu  gu  guaguuccaaga  cuuc  cg  guacgc  gccc  g
    uauaca  u  a  g  c  a  \----------  u  --  ggcg^  gcau
```

Figure 2. Representative stem-loop structures of identified cassava miRNA families. Mature miRNAs are underlined, bold and shaded. miRNA precursors could be slightly longer than the sequences shown in this figure.
Computer-based procedure for prediction of miRNA targets

Several algorithms are currently available for prediction of putative miRNA-mRNA targets in plants (Hofacker, 2003; Zuker, 2003). One of the most widely employed miRNA target finding softwares is the miRU (http://bioinfo3.noble.org/miRNA/miRU.htm) (Zhang, 2005). The miRU system searches for target sites from selected databases for potential complementary target sites in miRNA-target recognition with acceptable mismatches. It is required to input a mature miRNA sequence followed by selection of the dataset for prediction of mRNA target in the organism of interest. The search could be done using default parameters otherwise, the allowable complementary mismatches between the target mRNA and the uploaded miRNA sequence can be adjusted. The output provides for each predicted miRNA target, the gene identifier, target site position, mismatch score, number of mismatches and target complementary sequence. This software was used extensively for the prediction of target genes in the identified cassava miRNA families (Table 3). In addition, targets were cross checked with BLASTn search of the Arabidopsis Information Resource (TAIR) database. Default parameters were used in these analyses.

Experimental validation of predicted miRNAs and miRNA target genes

Computationally predicted miRNAs and their targets are normally validated experimentally in order to avoid false positives (Meyers et al., 2008). This has been achieved in many cases using conventional technologies such as cloning, northern hybridization and microarray analysis. However, in plants, miRNAs are highly under-represented in the small RNA fraction rendering conventional technologies not sensitive enough to detect less abundant miRNAs (Varkonyi-Gassic et al., 2007). To circumvent these limitations, reverse transcription-polymerase chain reaction (RT-PCR) detection methods have been developed and are at present the most powerful techniques due to their comparatively higher effectiveness, specificity and sensitivity. However, because of their small size, detection of miRNAs by PCR is technically demanding. A quantitative real-time PCR (qRT-PCR) method based upon reverse transcription (RT) reaction with a miRNA specific stem-loop primer has been described (Chen et al., 2005; Tang et al., 2006, Varkonyi-Gassic et al., 2007). Following stem-loop RT, either TaqMan or SYBR Green qRT-PCR and analysis are performed using miRNA-specific forward primer and a universal reverse primer designed to bind the stem-loop RT primer sequence. Chen et al. (2005) explained that stem-loop RT primers provide better specificity and sensitivity than linear primers because of base stacking and spatial constraint of the stem-loop structure. Detection sensitivity is further enhanced by a pulsed RT reaction (Tang et al., 2006).

On the other hand, identified miRNA gene targets are commonly confirmed by 5’ RACE (Rapid Amplification of cDNA Ends) procedure. In this technique, cleaved mRNA
products of miRNA guided activity are amplified by ligation of an oligo-nucleotide adapter to the 5’ phosphate of the cleaved mRNAs followed by reverse transcription and PCR amplification with a gene specific primer (Sunkar et al., 2005). The resulting RACE products are subsequently gel purified, cloned, sequenced and analysed to ascertain the gene target of a particular miRNA.

**CASSAVA miRNA FAMILIES AND THEIR TARGET GENES**

Consistent with reported studies in other plant species, the majority of target proteins of identified cassava miRNAs are transcriptional factors, whereas others are associated with metabolism and response to environmental stress. Eight miRNA families representing about half the total number of miRNA families identified appear to target transcriptional factor encoding mRNAs. Seven other miRNA families target genes coding a range of different proteins which may play important roles in the aspects of metabolisms, development, signal transduction and stress response. Two target genes are components of the miRNA biogenesis process. Generally, plant miRNAs have single binding sites in their target genes but it was found that each of the ESTs (DV457519 and DB922056) has two binding sites for the miRNA family miR166 (Table 2). The transcription factor regulatory miRNAs identified among others include the miR156 and miR157 families. Several studies indicate that miRNA156/157 targets squamosa promoter binding protein (SBP)-like (SPL) genes (Gandikota et al., 2007). The function of these miR156/157 miRNA families is well conserved across plant species as evident from several previous reports. Other transcription factor regulatory miRNA families identified in cassava and consistent with other plant species include miR159, miR160, miR166, miR167, miR171 and miR172. Hagen and Guilfoyle (2002) showed that auxin response factors (ARF), a plant-specific family of DNA binding proteins and the target of miRNA family miR167, are involved in hormone signal transduction (Tables 2 and 3). Wu et al. (2009) also recently found that in *A. thaliana* miR172 acts downstream of miR156 to promote adult epidermal identity.

Another group of interesting miRNA families identified include the miR162 and miR168 which respectively target the miRNA biogenesis proteins Dicer-like1 and Argonaute 1 (Table 3). This indicates that the mechanism of miRNA biogenesis is self regulating. The other miRNAs could potentially play important role in plant responses to different biotic and abiotic environmental stresses. Studies into the roles of these transcription factors and other identified genes will further deepen our understanding of plant gene regulation during growth and development.

It was observed also that the identified cassava miRNAs exhibit a wide variation in the length of pre-miRNA sequences (Table 2). The pre-miRNA length varies from 73 to 391 with most between 100 and 200 nucleotides long. It has been suggested that the differences in size among the different miRNA families is indicative they have unique functions in the regulation of miRNA.

### Table 3. Potential targets of identified miRNA families in cassava.

<table>
<thead>
<tr>
<th>miRNA family</th>
<th>Target protein</th>
<th>Function of target</th>
<th>Target gene (MS, NM)</th>
<th>E-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>miR156</td>
<td>Squamosa promoter-binding-protein-like 2 (SPL2)</td>
<td>Transcription factor</td>
<td>AT5G43270 (0,0)</td>
<td>3.0e-04</td>
</tr>
<tr>
<td>miR157</td>
<td>Squamosa promoter-binding-protein-like 2 (SPL2)</td>
<td>Transcription factor</td>
<td>AT5G43270 (0,1)</td>
<td>9.0e-05</td>
</tr>
<tr>
<td>miR159</td>
<td>Myb domain protein 101 (MYB101); DNA binding</td>
<td>Transcription factor</td>
<td>AT2G32460 (2,2)</td>
<td>9.0e-05</td>
</tr>
<tr>
<td>miR160</td>
<td>Auxin Response Factor 17 (ARF17); transcription factor</td>
<td>Transcription factor</td>
<td>AT1G77850 (2,0)</td>
<td>33e-02</td>
</tr>
<tr>
<td>miR162</td>
<td>Dicer-like1 (DCL1); ATP-dependent helicase/ribonuclease III</td>
<td>miRNA biogenesis</td>
<td>AT1G01040 (2,0)</td>
<td>9.0e-05</td>
</tr>
<tr>
<td>miR166</td>
<td>DNA binding / transcription factor</td>
<td>Transcription factor</td>
<td>AT1G52150 (1.5,2)</td>
<td>9.0e-05</td>
</tr>
<tr>
<td>miR167</td>
<td>Auxin response factor 8 (ARF8)</td>
<td>Transcription factor</td>
<td>AT5G37020 (2,3)</td>
<td>3.0e-05</td>
</tr>
<tr>
<td>miR168</td>
<td>Argonaute 1; AGO1</td>
<td>miRNA biogenesis</td>
<td>AT1G48410 (3,3)</td>
<td>9.0e-05</td>
</tr>
<tr>
<td>miR171</td>
<td>Scarecrow-like transcription factor 6 (SCL6)</td>
<td>Transcription factor</td>
<td>AT4G00150 (0.5,1)</td>
<td>21e-03</td>
</tr>
<tr>
<td>miR172</td>
<td>TOE2; DNA binding / transcription factor</td>
<td>Transcription factor</td>
<td>AT5G60120 (0.5,1)</td>
<td>21e-03</td>
</tr>
<tr>
<td>miR390</td>
<td>Protein kinase family protein</td>
<td>Metabolism</td>
<td>AT5G03640 (2.5,3)</td>
<td>13e-01</td>
</tr>
<tr>
<td>miR394</td>
<td>F-box family protein</td>
<td>Signal transduction</td>
<td>AT1G27340 (1,1)</td>
<td>73e-03</td>
</tr>
<tr>
<td>miR397</td>
<td>LAC2 (laccase 2); copper ion binding / oxidoreductase</td>
<td>Metabolism</td>
<td>AT2G29130 (0.5,1)</td>
<td>50e-04</td>
</tr>
<tr>
<td>miR398</td>
<td>Heterogeneous nuclear ribonucleoprotein, putative (hnRNP)</td>
<td>Metabolism</td>
<td>AT1G08830 (3,3)</td>
<td>13e-01</td>
</tr>
<tr>
<td>miR399</td>
<td>PHO2/UBC24 (Phosphate 2); ubiquitin-protein ligase</td>
<td>Metabolism</td>
<td>AT2G33770 (2,3)</td>
<td>33e-02</td>
</tr>
<tr>
<td>miR408</td>
<td>Peptide chain release factor, putative</td>
<td>Metabolism</td>
<td>AT2G47020 (0,0)</td>
<td>9.0e-05</td>
</tr>
<tr>
<td>miR414</td>
<td>Lon protease homolog 1, mitochondrial (LON)</td>
<td>Signal transduction</td>
<td>AT1G80960 (0,0)</td>
<td>9.0e-05</td>
</tr>
</tbody>
</table>

MS, miRU Score; NM, number of mismatches.
biogenesis or gene expression. The location of the new mature miRNAs in the precursor pre-miRNAs also vary among the miRNA families. 18 miRNAs were found to be located in 3’ arm of the stem-loop hairpin structures while 7 are in 5’ arm (Table 2). These characteristics of cassava miRNAs are similar to those reported for other plant species.

Evaluation of the new miRNAs in cassava was also based on A+U content. The new miRNA precursors have A+U content ranging from 44.90 to 64.38% (Table 2). Consistent with general notion, majority of the identified cassava pre-miRNA precursors contain more A+U nucleotides than G+C (Zhang et al., 2006). It is also important to note that formation of hairpin structure is not a unique feature of miRNAs since other RNAs such as mRNA, rRNA and tRNA can also form similar structures. For this reason, uniform systems for annotating new miRNAs comprising negative minimal fold energy (MFE), adjusted minimal fold energy (AMFE) and the minimal fold energy index (MFEI) have been developed (Seffens and Digby, 1999; Bonnet et al., 2004; Zhang et al., 2006) and have become generally accepted. Zhang et al. (2006) explained that since plant miRNA precursors are highly variable in length, a comparison based only on their MFEs which are inversely proportional to length of pre-miRNA sequence (Seffens and Digby, 1999), is not appropriate hence the need to adjust the MFE values to AMFEs. AMFE is defined as the MFE of a 100-nucleotide length of sequence and is calculated as AMFE = (MFE/Sequence length) x 100 (Zhang et al., 2006). Besides, it is noteworthy that the folding energy in miRNAs is not always higher than that of other RNAs (Bonnet et al., 2004). The predictions are therefore normally based on MFEI which is considered the best criterion for prediction of miRNAs, a commonly used index for distinguishing miRNAs from other non-coding and coding RNAs. Zhang et al. (2006) indicated that most miRNA precursors identified have an MFEI greater than 0.85 which is much higher than in tRNA (0.64), rRNA (0.59), or mRNA (0.65). However, it is not out of place to find pre-miRNAs with lower MFEI's that provided the number of nucleotide substitutions in the particular miRNA compared with other species does not exceed three (Table 2). The MFEI is calculated by the equation: MFEI = AMFE/(G+C)% (Zhang et al., 2006).

CONCLUSION

Plant miRNA studies are increasingly revealing a lot of insight into a number of biological events. A better understanding of the role of miRNAs in post-transcriptional gene silencing in response to biotic and abiotic stresses will be vital in our attempt to develop superior cassava varieties with higher disease and pest resistance. As a pre-requisite, a complete catalogue of cassava miRNAs is essential to enable systematic and fundamental comparison of characteristics of miRNA species in cassava as well as in other plants. In this regard, the achievement in the identification and characterization of some conserved miRNAs in cassava may initiate further study of cassava miRNAs regulation mechanism. This first effort will also serve as a major step forward for us to learn more about the important regulatory roles of miRNAs in cassava growth and development. Besides, the fairly detailed description of currently available procedures and resources presented here for the identification and characterization of miRNAs in cassava will enable the rapid discovery of more novel miRNAs towards the improvement of the crop.

ACKNOWLEDGEMENTS

The authors thank the Institute of Plant Genetics and Crop Plant Research (IPK) for making use of their facilities for generating some of the data presented in this account. The corresponding author in particular also expresses his profound gratitude to the International Max Planck Research School (IMPRS) for his current PhD funding. We also thank the Cassava Genome Project 2009 for access to the annotated draft cassava genome sequence.

Abbreviations

ESTs, Expressed sequence tag; HYL1, hyponastic leaves; siRNAs, short interfering RNAs; miRNA, microRNA; miRISC, miRNA-induced silencing complex; AGO, argonaute; CIAT, International Centre for Tropical Agriculture; CMD, cassava mosaic disease; ILTAB, International Laboratory for Tropical Agricultural Biotechnology; BNARI, Biotechnology and Nuclear Agriculture Research Institute; IAEA, International Atomic Energy Agency; PMRD, plant miRNA database; GSS, genome survey sequences; MFE, minimum fold energy; MFEI, minimal fold energy index; AMFE, adjusted minimal fold energy; TAIR, arabidopsis information resource; PCR, polymerase chain reaction; RT-PCR, reverse transcription-PCR; qRT-PCR, quantitative real-time PCR; SBP, squamosa promoter binding protein.

REFERENCES

Seiffens W, Digby D (1999). miRNAs have greater negative folding free energies than shuffled or codon choice randomized sequences. Nucleic Acids Res. 27: 1578-1584.
microRNAs in large number of diverse plant species. BMC Plant Biol. 8: p. 37.


