Dynamic Biochemistry, Process Biotechnology and Molecular Biology ©2012 Global Science Books

Computational Identification of Conserved MicroRNA and their Targets in *Coffea canephora* by EST Analysis

Shibin Mohanan Nellikunnumal • Arun Chandrashekar*

Plant Cell Biotechnology Department, Central Food Technological Research Institute (A constituent Laboratory of CSIR, India), Mysore, Karnataka, 570020, India

Corresponding author: * **achandrashekar85@gmail.com**

ABSTRACT

Comparative genomic tools have been successfully used to predict new miRNAs in different plant and animal species using expressed sequence tag (EST) and genome survey sequence (GSS) analysis. In the present study we identified conserved microRNAs in *Coffea canephora* identified by EST analysis using a new modified comparative genomics method. Use of this method eliminates false positives to a greater extent. Conserved microRNA belonging to 12 families was identified. All identified miRNA were used to search their potential target genes from the SGN (Solanaceace Genome Network) EST database and a total of 42 potential targets were identified for miRNA families from *C. canephora*. Most of the miRNA targets were transcription factors which appeared to be involved in plant growth, development and stress responses according to the mRNA target information provided by NCBI. The newly identified microRNA can help in better understanding of growth and development *Coffea*. The improved method can be used for more accurate prediction of microRNA from other plant species with sufficient EST database.

Keywords: *Coffea canephora*, computational identification, EST analysis, microrna
Abbreviations: EST, expressed sequence tags; **GSS**, Genome survey sequences; **mRNA**, messenger RNA; **nt**, nucleotide; **srcDNA**, small RNA cDNA

INTRODUCTION

MicroRNAs (miRNA) are extensive class of tiny RNA molecules of ~19-22nt length that regulate the expression if target genes by means of complementary base pair interactions. In plants, miRNAs are processed from transcripts that fold into stable hairpin and the maturation of miRNA is divided into a two step process that involves drosha and a dicer like enzyme (DCL1) to form a miRNA: miRNA* duplex. Finally the single stranded mature miRNA is incorporated into a complex known as RNA-induced silencing complex (RISC) that bind to the target mRNA (Bartel 2004). Plant miRNAs generally interact with their targets through perfect or near perfect complementarity and regulate the gene expression by direct mRNA degradation (Llave *et al*. 2002; Rhoades *et al*. 2002; Bartel 2009; Meng *et al*. 2011; Thakur *et al*. 2011) or repressed at translational level.

Several studies have shown that miRNAs are involved in the regulation of various plant physiological processes such as meristem cell identity (Mallory *et al*. 2004a, 2004b; Guo *et al*. 2005; Zhang *et al*. 2006b; Nonogaki 2010; Wang *et al*. 2011), leaf organ morphogenesis, rosette leaf expansion and curvature (Palatnik *et al*. 2003, 2007), floral differentiation and development (Chen 2004; Mallory *et al*. 2004a), stress responses (Kasschau *et al*. 2003; Llave 2004; Sunkar and Zhu 2004; Fujii *et al*. 2005; Sunkar *et al*. 2006; Buhtz *et al*. 2010; Gao *et al*. 2010; Kong and Yang 2010; Martin *et al*. 2010; Xu 2010) embryonic, vegetative, floral organ boundary formation (Mallory *et al*. 2004a; Peaucelle and Laufs 2006; Nodine and Bartel 2010) and root development (Meng *et al*. 2010).

MiRNAs are highly conserved in plant kingdom (Zhang *et al*. 2006a), although there are rare evidences of conservation between plants and animals (Arteaga-Vázquez *et al*. 2006). Recent studies have shown that miRNA genes are evolutionarily conserved in all plant species (Allen *et al*. 2004; Zhang *et al*. 2006a, 2006b; Han *et al*. 2010; Taylor *et al*. 2010; Meng *et al*. 2011).

Identification of miRNAs in various species have been done using different methods like genetic screening: the first two miRNA *lin-4* and *let-7* were identified by this method (Lee *et al*. 1993); direct cloning: through small RNA library construction (Aravin and Tuschl 2005; Sunkar *et al*. 2005); computational approach: based on whole genome sequences (Bonnet *et al*. 2004).

Although large number of miRNA has been identified in recent years from different organisms, relatively fewer numbers of miRNAs have been identified from plants. Computational identification of miRNA is successful not only for those plant species for which full genomic and large EST database is available but also for those with incomplete genomic information. A large number of miRNA are identified computational methods and then validated experimentally, in contrast only few are directly cloned. The computational identification of miRNA usually results in large number of false positive and requires a careful screening of the potential candidate sequences.

Coffee is a one of the most important beverage crop consisting of 103 species genera. Commercial coffee production mostly relies on *Coffea arabica* (Arabica coffee) and *Coffea canephora* (Robusta coffee). *C. arabica* is cultivated extensively and represent 70% of the market share and the remaining 30% consists mainly of *C. canephora* and other species like *C. liberica* and *C. dewevrei and C. racemosa* produced only to satisfy local consumption. It is also a great source of secondary metabolites like polyphenols, caffeine, theobromine and chlorogenic acids which have medicinal properties.

Recently the molecular biology of coffee plants has been the most kinetic research fields of coffee research. With the recent progress in functional genomics research, based on large-scale EST generation (Mueller *et al*. 2005), cloning and analysis of genes has provided a critical significance on elucidating the molecular mechanism of growth,

Received: 1 August, 2011. Accepted: 7 February, 2012.

Original Research Paper

development, differentiation, metabolism, quality, yield, and stress resistance, as well as genetic manipulation via biotechnological approaches (Ashihara *et al*. 2008; Kochko *et al*. 2010). Even though *C. arabica* is economically more predominant due to the allotetraploid nature, diploid *C. canephora* was selected in the study as the EST database is freely available and moreover the whole genome sequencing of the *C. canephora* is being carried out and is expected to be released in the near future.

Hundreds of miRNAs have been identified from different plant species yet there is no report of miRNA from any of the coffee species. In the present study, all the identified miRNA from different plant species were used to identify its conserved homologs in coffee using a new modified procedure for more accurate prediction of potential miRNAs.

MATERIALS AND METHODS

Mature miRNA reference sequences

All mature miRNA sequences previously identified in 13 plant species like *Arabidopsis thaliana, Glycine max*, *Medicago truncatula*, *Oryza sativa*, *Physcomitrella patens*, *Populus trichocarpa*, *Saccharum officinarum*, *Sorghum bicolour* and *Zea mays* were downloaded from mirBASE miRNA sequence database (Kozomara and Griffiths-Jones 2011) release 9.0 October 2006 and 10.0 November 2008 (Griffiths-Jones *et al*. 2006, 2008). Due to high degree of conservation of miRNAs in plants, all the repeat sequences were removed from the different plant species in order to avoid repeat search of redundant and overlapping miRNAs, remaining unique set of sequences were defined as reference set.

Coffee expressed sequence tags (EST)

A total of 30000 *C. canephora* EST sequences were obtained from Solanaceae genomics network database (SGN) (Mueller *et al*. 2005) (version II).

Availability of software

Comparative software BLAST-2.2.15 was downloaded from the NCBI Genebank. MFold 4.0 web server (Markham and Zuker 2008) was used online to analyze secondary structure of RNAs. Another version of the same program UNA Fold 4.5 was downloaded from and run on Ubuntu 9.10 In order to improve the efficiency of base pairing, identification of miRNA and their targets algorithms such as MiRNAassist (Zhang *et al*. 2006b; Xie *et al*. 2007) and mirEval (Ritchie *et al*. 2008) was used.

Prediction of miRNA

In order to predict potential miRNA from *C. canephora*, the reference set was subjected to BLASTX 2.2.15 against Coffee EST database to find coffee miRNA homologs using following parameters: Score value for 100; alignment for 1000; except value for 1000. Low complexity was chosen as the sequence filter; the number of descriptions and alignments was raised to 1,000. The default word-match size between the query and database sequences was seven. All BLASTn results were saved. If the matched sequence was shorter than the queried miRNA sequence, the aligned and non-aligned parts were manually inspected and compared to determine the number of matching nucleotides. The EST sequences were considered miRNA candidates only if they fit the following criteria: (1) The candidate sequence should be at least 18 nt in length (2) Only 0-3 nt mismatch was allowed with the previously known mature plant miRNAs. The selected set of miRNA candidates were further characterized by following steps (1) the entire EST sequence was selected to predict the secondary structures and to screen for miRNA precursor sequences; (2) the selected ESTs were further compared with each other to eliminate redundancies; (3) the candidate EST should be non- protein coding sequences and these sequences were subjected to evaluation for miRNA prediction properties using mirEval software (Ritchie *et al*. 2008) and miRNAassist.

Prediction of secondary structure

Prediction of secondary structure of precursor sequences was done using Zunker folding algorithm MFold (Zunker 2003), which is publically available and UNAFold (Markham and Zuker 2008). Following parameters were used in predicting the secondary structures: (1) linear RNA sequence; (2) folding temperatures were fixed at 37°C with ionic conditions of 1M NaCl with no divalent ions; (3) percent sub optimality number of 5; (4) maximum interior/bulge loop size of 30; (5) the grid lines in energy dot plot turned on. All other parameters were set with default values.

Following criteria was used to designate a RNA sequence as a potential miRNA precursors or pre-miRNA: (1) Identified pre-miRNA within the competent sequence should fold into a typical stem loop structure with ~22 nt mature miRNA in one arm; (2) miRNA precursors with secondary structures must have higher negative minimal free energies (MFEs) and minimal free energy index (MFEI) when compared to that of other types of RNAs. The adjusted MFE (AMFE) is defined as MFE of 100 nucleotide length of sequence AMFE = (MFE/sequence) 100. Minimal free energy index is calculated as MFEI = AMFE/(G+C) %; (3) Potential miRNAs should have 30-70% A+U content; (4) predicted mature miRNAs had no more than six mismatches with the miRNA* sequence in the other arm; (5) A maximum size of 3-5 nt for a bulge in the miRNA sequence; and (6) no large loops or break in were allowed in the miRNA sequences. These criteria significantly reduced false positives and required that the predicted miRNAs fit the criteria proposed by Ambros *et al*. (2003).

Prediction of miRNA targets in *C. canephora*

Previous studies have shown that plant miRNAs bind to the protein coding region of their mRNA targets with a perfect or near perfect sequence complementarity and degrade the target mRNA. The perfect or near perfect complementarity of miRNA with its target mRNA has greatly simplified the identification of the miRNA targets in plants. All the mature miRNA identified in coffee were used to search against the coffee EST database (SGN, Cornell) using a web server miRU (Dai and Zhao 2011). Not more than 4 mismatches were allowed between miRNAs and its potential targets.

Construction of small RNA library and RT-PCR for specific microRNA

Small RNA was isolated using mirVANA miRNA isolation kit (Ambion, TX, USA) following manufacturer's instruction. Small RNA library was prepared by A-tailing small RNA using poly A-tailing kit (Ambion, TX, USA). Small RNA cDNA (srcDNA) library was prepared by using ~2 µg of tailed small RNA and 1 µg of RTQ primer mixed in a 26 µl reaction volume, followed by an incubation at 65°C for 10 min and annealing at 4°C for 20 min. Reverse transcription was carried out with 200 U of enhanced Avian Reverse-transcriptase (eAMV) (Sigma-Aldrich, USA), 1 µl dNTPs mix (10 mM each) and 8 µl 5X buffer in a final reaction of 40 µl at 50°C for 60 min. Finally, the reverse transcriptase was inactivated by incubation at 70°C for 15 min and 1.5U of RNaseH (Ambion, TX USA) was added to remove the small RNAs.

All srcDNA samples were diluted to the same concentration of 25 ng/µl (total about 500 µl). A small RNA-specific primer and a universal reverse primer, RTQ-UNIr, were used for specific amplification of small RNAs (Ro *et al*. 2006).

PCR was performed in a final reaction volume of 25 µl containing 1 µl (25 ng) of the synthesized srcDNA and 5 µM of the primers. A 3-steps PCR protocol (95°C for 10 min; then 40 cycles of 95°C for 15 sec; 50°C for 30 sec and 60°C for 30 sec) was used. The annealing temperature was adjusted according to the Tm of each of the small RNAs and checked on 2% agarose (Sigma-Aldrich, USA) (Ro *et al*. 2006). A touchdown PCR using double sense and antisense primers served as negative control.

Northern blot analysis for small RNA

Small RNA (100 ng) isolated using mirVANA miRNA isolation kit (Ambion, TX, USA) was run on 15% denaturing PAGE containing

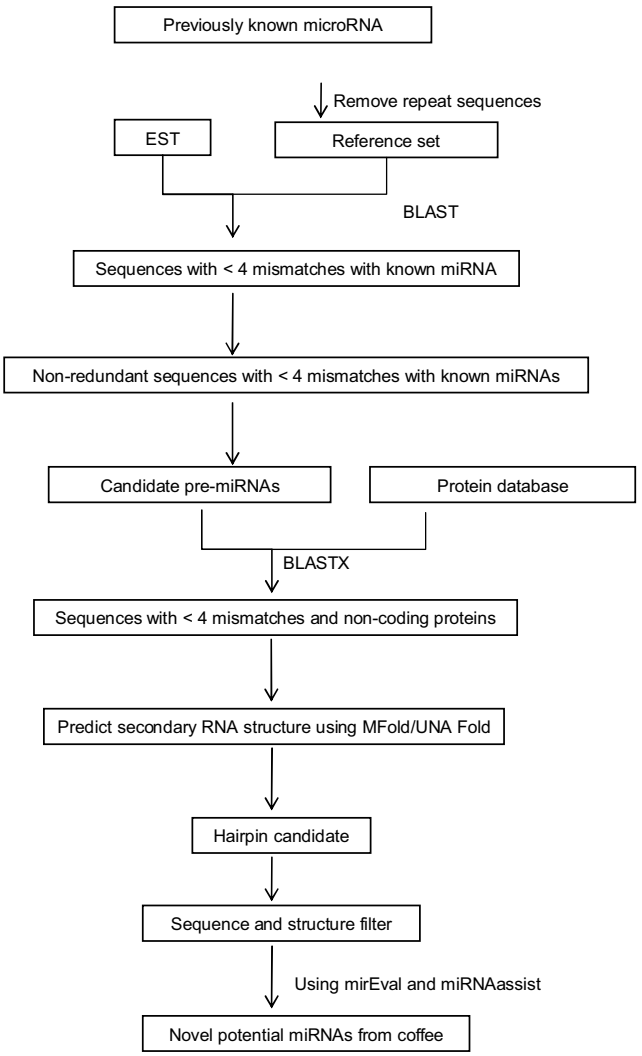

Fig. 1 Procedure for potential miRNAs from *C. canephora* by identifying homologs of previously known plant microRNAs. The prediction was performed using the publically available EST database. An additional step involving analysis using miEVAL and mirassist reduced the chances of false positives.

7M urea was run for 5 h at 5 V/cm^2. Small RNA was transferred on to a nylon membrane using turbo transfer method using transfer buffer provided in Northern Max kit (Ambion, TX USA) or electroblotted using semi-dry transfer apparatus (Bio-Rad, Munich, Germany) using 0.5X TBE. Pre-hybridization and hybridization was performed using Northern MAX kit following manufacturer's instruction (Ambion). Biotin labelled synthetic oligos complementary to the predicted miRNAs was used as probes for hybridization.

RESULTS

To predict miRNA from coffee a defined set of sequence and structural properties of known miRNA was used to screen candidate miRNA from the coffee EST database. **Fig. 1** summarizes the modified search and filtering procedure of identifying potential miRNA in *C. canephora.*

A total of 318 EST sequences were identified by BLAST search using 1220 miRNAs from plant species deposited at miRNA database. After eliminating the repeat sequences and protein coding ESTs, 78 sequences met the preliminary criteria i.e. mature miRNAs were 19~23 nt in length with 0-4 mismatches with known miRNAs. These non protein coding homologs were predicted as potential miRNA candidates and their structures predicted using MFold 4.0. After this filtering 18 conserved miRNAs belonging to 12 miRNA families were identified from coffee EST database (**Table 1**). These potential miRNA candidates were evaluated for all the criteria mentioned in materials and methods.

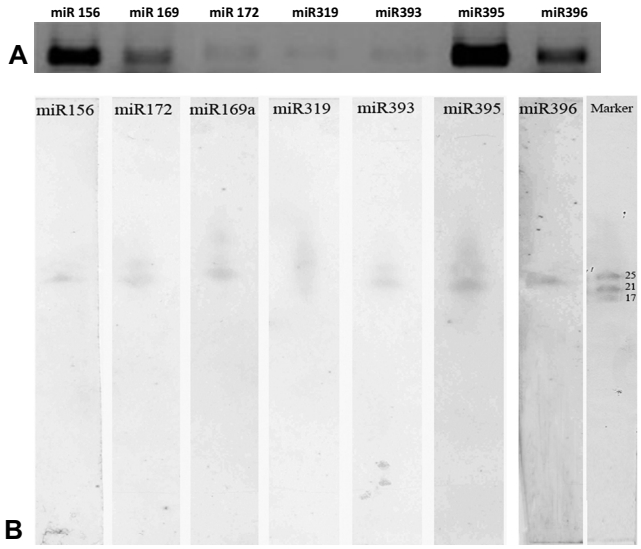

Fig. 2 (**A**) Expression of newly identified microRNAs from *C. canephora* leaves. The microRNAs expressed at different levels under our experimental conditions. (**B**) Northern Blot analysis of PCR positive microRNAs using complementary synthetic oligos as probes.

In order to detect the expression of these miRNAs, RT-PCR was performed in order to amplify individual miRNAs using a sense primer specific to each miRNA family and universal antisense primer By this approach the expression of 7 miRNA families out of 12 was detected, namely Mir156, mir169, mir172, mir319, mir393, mir395, mir396 (**Fig. 2A**).

The northern blot analysis of the PCR positive microRNA using complementary oligos further confirmed the presence of these conserved miRNA in *C. canephora* (**Fig. 2B**).

To identify the potential regulatory targets, the EST database was searched for complementary mRNA sequences having less than 4 mismatches to the miRNAs with no gaps, G: U and other non-canonical pairs were not allowed and were considered as mismatches. Plant miRNAs identifies their target mRNA by perfect or near perfect complementary base pair interactions. Using the newly identified miRNAs, 42 target genes were identified from the mRNA database of *C. canephora* (**Table 2**). The miRNAs were found to target transcription factors and genes involved in various aspects of metabolism, signal transduction, stress responses, growth and development.

DISCUSSION

The evolutionarily conserved nature of plant miRNAs has simplified the prediction of microRNA from the plant species by EST and GSS analysis. This strategy has been used to predict a lot of conserved microRNAs in divergent plant species including Arabidopsis (Zhang *et al*. 2005), *Gossypium hirsutum* (Qiu et al. 2007), tomato (Yin *et al*. 2008), peach (Gao *et al*. 2012), and wormwood (Pani *et al*. 2011). All previously known miRNA were used search for the homologs in *C. canephora*. By computational analysis a total of 18 microRNAs belonging to 12 families was identified from coffee for the first time. They are miR156a/b, miR172, miR393, miR395, miR414, miR834, miR838, miR854 and miR829. All the identified miRNAs is considered valid as they fulfil almost all the criteria of empirical formula for microRNA formation and biogenesis (Ambros *et al*. 2003). miR414 and mir854 are rare microRNAs reported from few species. These are speculated to be involved in speciation (Gao *et al*. 2012). For many highly conserved miRNAs found in different plant species, such as miR162 and miR164 were not identified in this study. This may be due the stringent method followed and the there may be mutations in evolution of miRNA biogenesis similar condition was reported in Asiatic cotton (Wang *et al*. 2012).

Table 1 Identified microRNA from *Coffea canephora* by EST analysis. The predicted miRNAs were named in accordance to mirBASE. Mature sequences were designated as miR with prefix 'cca' for *Coffea canephora*.

miRNAs	Sequence	LM	EST No	Length of precursor	Location	(A+U)%	MFE	MFEI
Cca-miR 156	GUGCUCUCUCUCUUCUGUCA	20	SGN-U350056	160	5'	50	-94.49	0.99
Cca-miR169a	UGCUCAGUCCUCUUCUGUCA	20	SGN-U356265	197	3'	45	-79.7	0.97
Cca-miR172	UGCAGCAUUCUUCAGAUUCU	20	SGN-U352029	65	3'	50	-215	1
Cca-miR172a	UCCAGCAUCGUCAAGAUUCA	20	SGN-U361682	71	3'	55	-103.24	0.81
Cca-miR319	GGCGCUAUCCUACCUGAGCUU	21	SGN-U351786	56	3'	43	-102.14	0.58
Cca-miR393	ACCAUUGUGAUCUCUUUGGA	20	SGN-U362786	54	3'	55	-86.78	0.64
Cca-miR395	AGUUCACCCAAGCAGUUCAG	20	SGN-355421	98	5'	40	-156.07	0.78
Cca-miR396	CGGUUCAAUAAAGCUGUGGGA	21	SGN-U351910	139	5'	52	-175.9	1.23
Cca-miR414	UAACGGUGAUGAUGAGGAUGA	21	SGN-U352699	99	5'	57.1	-133.84	1.15
cca-miR414	UGAUGAUGAUGAUGAUGAUGA	21	SGN-U355224	205	5'	58	-176.7	0.68
Cca-miR414	UGAUGAUAAUGAUGAGGAUGA	21	SGN-U358642	439	5'	61.9	-107.81	0.73
Cca-miR414	UGAUAGUGAUGAUGAAGAUGA	21	SGN-U361343	189	5'	59	-133.56	0.88
Cca-miR854	CUCCUCCUGCCUCUCCUCGUC	21	SGN-U347346	214	5'	37	-146.54	0.74
Cca-miR827	CUUUGUUGAGGGUUAUCUGA	21	SGN-U361418	83	3'	57.6	-104.96	0.87
Cca-miR829	UAUCAGUUGGCAUCAGAGCU	20	SGN-U353787	103	3'	55	-157	0.80
Cca-miR838	GUGAAAGAGGUUGAAGAAAG	21	SGN-U356064	124	3'	57	-161.96	0.91
Cca-miR854	CUUUUUCUCUCUAUCUUCAUU	21	SGN-U354509	52	3'	61	-99.87	0.68
Cca-miR854b	CUCUUCUUCCUUUUUCUCAUC	21	SGN-U352507	257	3'	61.7	-104.26	0.74

Table 2 List of potential targets for predicted miRNAs in *C. canephora*.

miRNA	Target protein	Target gene function	Target gene	Conserved
miR156a	Squamosa promoter binding protein (SBP)	Transcription factor	SGN-U352279	Yes
	Serine/Threonine protein phosphatase 2A (PP2A) 55 kDa regulatory subunit B	Metabolism	SGN-U361686	No
miR156h	Squamosa promoter-binding protein-like 4 (SPL4)	Transcription factor	SGN-U352752	yes
miR157a	Squamosa promoter-binding protein-like 4 (SPL4)	Transcription factor	SGN-U352279	Yes
	far-red impaired responsive protein	Metabolism	SGN-U359652	No
miR157d	Squamosa promoter binding protein SBP	Transcription factor	SGN-U352279	Yes
miR159	TCP transcription factor	Transcription factor	SGN-U348885	Yes
miR163	Short chain alcohol degenerase	Metabolism	SGN-U349256	No
miR167	Auxin responsive factor (ARF6/ARF8)	Transcription factor	SGN-U355688	Yes
	Gibberellin 20-oxidase	Metabolism	SGN-U350908	No
	Basic helix-loop-helix (bHLH) family protein	Transcription factor	SGN-U356572	No
miR169a	Zinc finger (GATA type) family protein	Transcription factor	SGN-U347940	Yes
	GATA-binding transcription factor-like protein	Transcription factor	SGN-U347940	No
miR172	NAD-dependent sorbitol dehydrogenase	Metabolism	SGN-U351178	No
miR319	TCP family transcription factor	Transcription factor	SGN-U354024	yes
miR391	Trypsin and protease inhibitor protein similar to LeMir (Miraculin homolog)	Metabolism	SGN-U360818	
miR393a	Transport inhibitor protein (TIR1)	Transcription factor	SGN-U351616	Yes
miR394	F-Box protein	Metabolism	SGN-U360601	Yes
miR395	Gylcosyl hydrolase family 1 protein	Metabolism	SGN-U348455	No
miR396a	Cysteine proteinase (RD21A)	Metabolism	SGN-U352616	Yes
miR396b	Transcription activator GRL1	Transcription factor	SGN-U355115	Yes
miR414a	DNAJ heat shock protein similar to PGR	metabolism	SGN-U350203	No
miR414b	Importin beta-2 subunit	Metabolism	SGN-U350775	No
miR414c	pfkB-type carbohydrate kinase family protein	Metabolism	SGN-U352445	No
miR414d	Transcription factor IIA large subunit	Transcription factor	SGN-U353395	No
miR414f	RNA polymerase I specific transcription factor RRN3	Transcription factor	SGN-U357053	No
miR414g	Surfeit locus protein 2 family	Metabolism	SGN-U359530	No
miR414h	Ribosomal protein S6 family protein	Metabolism	SGN-U359530	No
miR447c	Lanthionine synthase C-like protein	Metabolism	SGN-U352847	No
miR771	Auxilin-related	Metabolism	SGN-U350660	No
miR773	Protein phosphatase 2C	Metabolism	SGN-U349417	No
miR827	ATP-dependent Clp protease ATP-binding subunit (ClpD)	Metabolism	SGN-U356843	No
miR834	Heavy-metal associated domain contain protein/copper chaperone (CCH) related	Transcription factor	SGN-U349064	No
miR834	Late embryogenesis abundant 3 family protein (LEA3)	Metabolism	SGN-U353803	No
miR835-3p	Amine oxidase family protein similar to polyamine oxidase	Metabolism	SGN-U351784	No
miR838	Thioredoxin H-type protein	Metabolism	SGN-U362502	No
miR842	Hypersensitive-induced response protein (band 7 family protein)	Metabolism	SGN-U347327	No
miR863	SET domain containing protein (ASHH1)	Transcription factor	SGN-U362144	No
miR849	DEAD/DEAH box helicase	Metabolism	SGN-U348522	No
miR854a	Haloacid dehalogenasehydrolase family protein/cytochrome B5	Metabolism	SGN-U347346	No
miR854b	Platid specific 30S ribosomal protein 3 (PSRP-3)	Metabolism	SGN-U348964	No
miR855	Cinnamoyl - CoA reductase protein family/Proliferation cell nuclear antigen	Metabolism	SGN-U348803	No

When compared to the number of miRNA reported in other plant species and on the basis of the principle of prediction of one mature microRNA for every 10,000 EST sequence, at least 50 mature should have been identified (Zhang *et al*. 2006a, 2006b). The smaller microRNA family size in coffee can be attributed by the fewer EST sequences available for *C. canephora* i.e. 55,694 when compared to that of other plant species. However, using a similar method, 21 microRNAs were identified from 293,350 lettuce sequences and 30 from *Gossypium hirsutum* (Qiu *et al*. 2007; Han

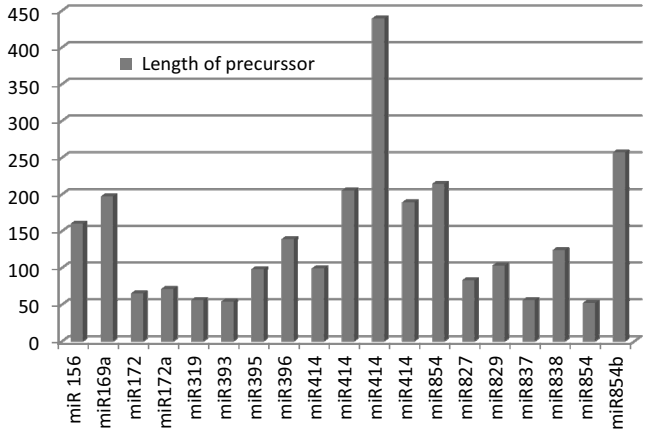

Fig. 3 Graph showing length of precursors in miRNA families, the length varies from 46 to 447 nt.

et al. 2010). A comparatively lower frequency of predictions may also be due to the additional stringent filtering of the primary candidate sequences.

The analysis of various characteristics of conserved microRNA identified agrees to all the criteria laid out by previous workers for a potential miRNA even though the miRNA precursors vary significantly (**Figs. 3, 4**). The presence of variable length of pre-miRNA within the same family suggests that the members of the same family can have different expression pattern such as in spatiotemporal or tissue specific manner (Zhang *et al*. 2008). Even though the minimal free energy (MFE) varied, the minimal free energy indexes (MFEI) were below 2.00 (**Table 2**) suggesting that these miRNAs identified in *C. canephora* are likely to be true miRNAs.

The RT-PCR and northern blotting confirmed the expression of these miRNAs in *C. canephora*. The miRNA showed a showed varied expression level which may be indicative to their specific roles to individual microRNA families. However, expression of mir414, mi854, mir827, mir829, mir837, mir838 and mir854 were not detected which may be due to low expression levels or due to expression in response to specific stimuli.

MiRNAs in plants are known to target transcription factors. In the present study 12 microRNA families were identified which target different transcription factors like miR156/157 targeting Squamosa promoter binding protein (SBP/SPL4) (Nodine and Bartel 2010; Nonogaki 2010), miR167 targeting auxin responsive factor (ARF6) (Ru *et al*. 2006; Bartel 2009; Meng *et al*. 2009), TCP family transcription factor and TIR1 targeted by 319 and 393 respectively (Ori *et al*. 2007; Schommer *et al*. 2008; Addo-Quaye *et al*. 2009) (**Table 2**). It was also observed that in *C. canephora* in addition to the function of regulating transcription factors, the same microRNAs also regulate divergent aspects of development by targeting protein involved in stress response, organ development and so on, for example, miR167 may also target gibberellins 20-oxidase and bHLH protein in addition to the role of regulating auxin responsive factors. In plants like *Arabidopsis*, rice and maize more than a dozen Squamosa promoter binding proteins are reported by miRNAs 156/157 (Rhoades *et al*. 2002; Zhang *et al*. 2010).

MiRNAassist and miRU web server (Zhang 2005) were used to analyze microRNA and their targets in *C. canephora* with that in *Arabidopsis*. It was found that many potential target genes were conserved within *Arabidopsis* and other plant species indicating that miRNA-mediated gene regulation mechanism is evolutionarily conserved among plant species. Certain miRNAs are specific to a small group of related plant species (Gao *et al*. 2012). High throughput sequencing methods should be employed to identify these miRNAs.

CONCLUSIONS

To conclude, using a publicly available *C. canephora* EST database, a total of 18 conserved miRNAs belonging to 12 families were identified using an improved EST analysis method. RT-PCR and northern blotting was performed to validate the expression of these microRNAs even though the varied expression evident from the gel has to be further investigated. Most of the potential microRNA precursors exhibited diversity in length of precursors. Finally, a total of 42 potential targets were identified for the predicted miRNAs and they were involved in regulation of growth and development. The presence of these conserved microRNAs in coffee brooks further investigation into the role of microRNA in *C. canephora* growth and development. The method described can be used for more stringent and accurate prediction of microRNA from EST and GSS databases for plants.

ACKNOWLEDGEMENTS

NSM is recipient of CSIR Research fellowship. The project funding from Department of Biotechnology, Govt. of India is gratefully acknowledged. Authors are grateful to Dr. Z. M. Yang, Nanjing Agricultural University, China for providing MiRNAassist program and Mr. Kalidiswarane, Informedia technologies, Mysore for his technical help during the course of the present study.

REFERENCES

Addo-Quaye C, Snyder JA, Park YB, Li YF, Sunkar R, Axtell MJ (2009) Sliced microRNA targets and precise loop-first processing of MIR319 hairpins revealed by analysis of the *Physcomitrella patens* degradome. *RNA* **15**, 2112-2121

Allen E, Xie Z, Gustafson AM, Sung GH, Spatafora JW, Carrington JC (2004) Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana. Nature Genetics* **36**, 1282-1290

Ambros V, Bartel B, Bartel DP, Burge CB, Carrington JC, Chen X, Dreyfuss G, Eddy SR, Griffiths-Jones S, Marshall M, Matzke M, Ruvkun G, Tuschl T (2003) A uniform system for microRNA annotation. *RNA* **9**, 277-279

Aravin A, Tuschl T (2005) Identification and characterization of small RNAs involved in RNA silencing. *FEBS Letters* **579**, 5830-5840

Arteaga-Vázquez M, Caballero-Pérez J, Vielle-Calzada JP (2006) A family of microRNAs present in plants and animals. *Plant Cell* **18**, 3355-3369

Ashihara H, Sano H, Crozier A (2008) Caffeine and related purine alkaloids: Biosynthesis, catabolism, function and genetic engineering. *Phytochemistry* **69**, 841-856

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

Bartel DP (2009) MicroRNAs: Target recognition and regulatory functions. *Cell* **136**, 215-233

Bonnet E, Wuyts J, Rouze P, Van de Peer Y (2004) Detection of 91 potential conserved plant microRNAs in *Arabidopsis thaliana* and *Oryza sativa* identifies important target genes. *Proceedings of the National Academy of Sciences USA* **101**, 11511-11115

Buhtz A, Pieritz J, Springer F, Kehr J (2010) Phloem small RNAs, nutrient stress responses, and systemic mobility. *BMC Plant Biology* **10**, 64

Chen X (2004) A microRNA as a translational repressor of APETALA 2 in *Arabidopsis* flower development. *Science* **303**, 2022-2025

Dai X, Zhao PX (2011) psRNATarget: A plant small RNA target analysis server. *Nucleic Acids Research* **39**, W155-159

Fujii H, Chiou TJ, Lin SI, Aung K, Zhu JK (2005) A miRNA involved in phosphate-starvation response in *Arabidopsis*. *Current Biology* **15**, 2038-2043

Gao P, Bai X, Yang L, Lv D, Pan X, Li Y, Cai H, Ji W, Chen Q, Zhu Y (2010) osa-MIR393: A salinity- and alkaline stress-related microRNA gene. *Molecular Biology Reporter* **38**, 237-242

Gao Z, Luo X, Shi T, Cai B, Zhang Z, Cheng Z, Zhuang W (2012) Identification and validation of potential conserved microRNAs and their targets in peach (*Prunus persica*). *Molecules and Cells* **34**, 239-249

Griffiths-Jones S, Grocock RJ, Dongen van S, Bateman A, Enright AJ (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Research (Database Issue)* **34**, D140-D144

Griffiths-Jones S, Saini HK, Dongen van S, Enright AJ (2008) miRBase: Tools for microRNA genomics. *Nucleic Acids Research* **(Database Issue) 36**, D154-D158

Guo HS, Xie Q, Fei JF, Chua NH (2005) MicroRNA directs mRNA cleavage of the transcription factor NAC1 to downregulate auxin signal for *Arabidopsis* lateral root development. *Plant Cell* **17**, 1376-1386

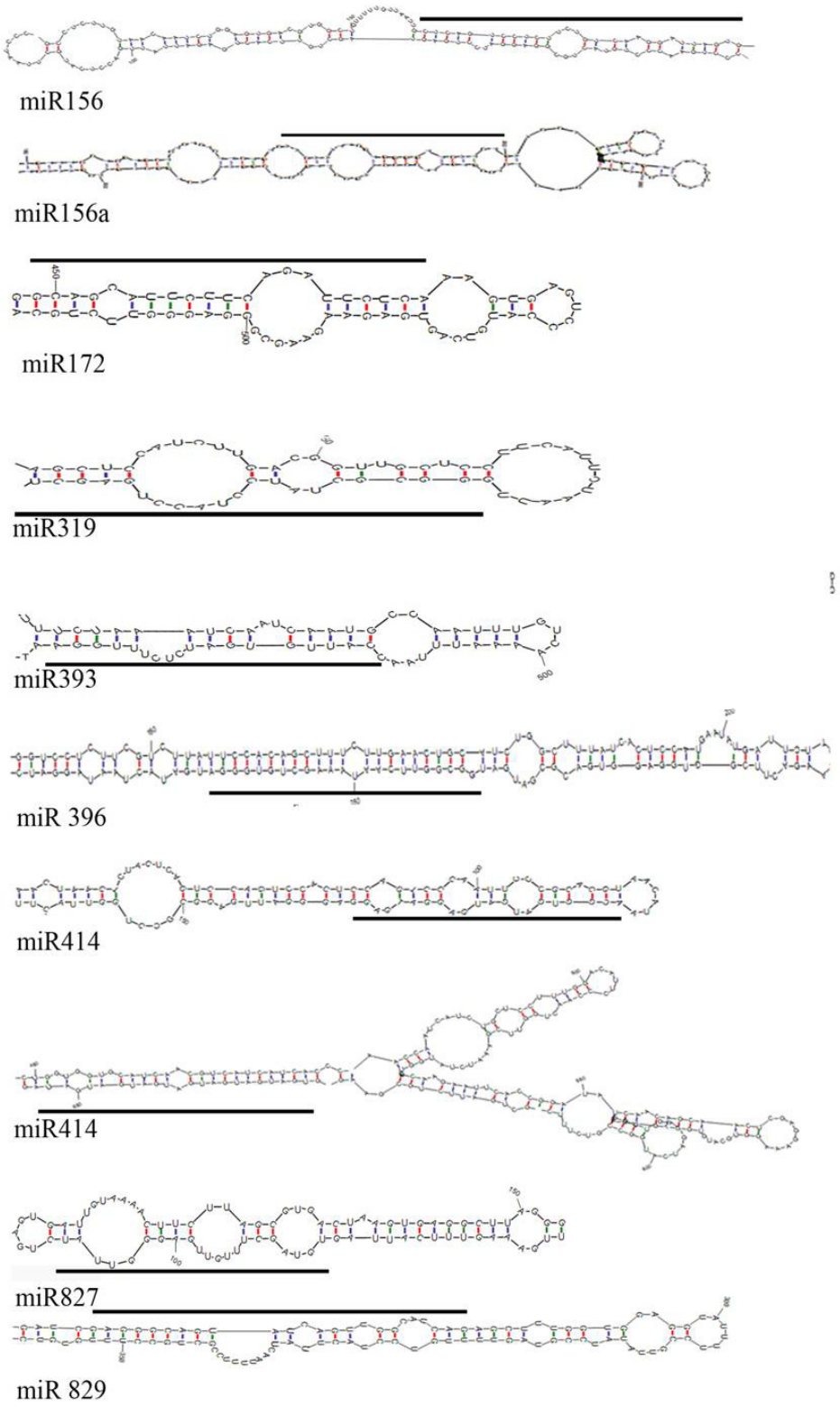

Fig. 4 Mature and precursor sequences and the predicted stem loop structures of newly cloned miRNAs from *C. canephora*.

Han Y, Zhu B, Luan F, Zhu H, Shao Y, Chen A, Lu C, Luo Y (2010) Conserved microRNA and their target identification in lettuce (*Lactuca*) by EST analysis. *Gene* **463**, 1-7

Kasschau KD, Xie Z, Allen E, Llave C, Chapman EJ, Krizan KA, Carrington JC (2003) P1/HC-Pro, a viral suppressor of RNA silencing interferes with *Arabidopsis* development and miRNA function. *Developmental Cell* **4**, 205-217

Kochko AD, Akaffou S, Andrade AC, Campa C, Crouzillat D, Guyot R, Hamon P, Ming R, Mueller LA, Poncet V, Tranchant-Dubreuil C, Hamon S (2010) Advances in *Coffea* genomics. In: Jean-Claude K, Michel D (Eds) *Advances in Botanical Research*, Academic Press, London, pp 23-63

Kong WW, Yang ZM (2010) Identification of iron-deficiency responsive microRNA genes and cis-elements in *Arabidopsis*. *Plant Physiology and Biochemistry* **48**, 153-159

Kozomara A, Griffiths-Jones S (2011) miRBase: Integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Research* **39**, D152-D157

Lee RC, Feinbaum RL, Ambros V (1993) The *Caenorhabditis elegans* heterochronic gene *lin-4* encodes for small RNAs with antisense complementarity to *lin-14*. *Cell* **75**, 843-854

Llave C (2004) MicroRNAs: More than a role in plant development? *Molecular Plant Pathology* **5**, 361-366

Llave C, Kasschau KD, Rector MA, Carrington JC (2002) Endogenous and silencing-associated small RNAs in plants. *Plant Cell* **14**, 1605-1619

Mallory AC, Dugas DV, Bartel DP, Bartel B (2004a) MicroRNA regulation of NAC-domain targets is required for proper formation and separation of adjacent embryonic, vegetative, and floral organs. *Current Biology* **14**, 1035-1046

Mallory AC, Reinhart BJ, Jones-Rhoades MW, Tang G, Zamore PD, Barton MK, Bartel DP (2004b) MicroRNA control of PHABULOSA in leaf

development: Importance of pairing to the microRNA 5' region. *EMBO Journal* **23**, 3356-3364

Markham NR, Zuker M (2008) UNAFold: Software for nucleic acid folding and hybridization. *Methods Molecular Biology* **453**, 3-31

Martin RC, Liu PP, Goloviznina NA, Nonogaki H (2010) microRNA, seeds, and Darwin? Diverse function of miRNA in seed biology and plant responses to stress. *Journal of Experimental Botany* **61**, 2229-2234

Meng Y, Gou L, Chen D, Mao C, Jin Y, Wu P, Chen M (2011) PmiRKB: A plant microRNA knowledge base. *Nucleic Acids Research* **39**, D181-187

Meng Y, Huang F, Shi Q, Cao J, Chen D, Zhang J, Ni J, Wu P, Chen M (2009) Genome-wide survey of rice microRNAs and microRNA-target pairs in the root of a novel auxin-resistant mutant. *Planta* **230**, 883-898

Meng Y, Ma X, Chen D, Wu P, Chen M (2010) MicroRNA-mediated signaling involved in plant root development. *Biochemical and Biophysical Research Communications* **393**, 345-349

Mueller LA, Solow TH, Taylor N, Skwarecki B, Buels R, Binns J, Lin C, Wright MH, Ahrens R, Wang Y, Herbst EV, Keyder ER, Menda N, Zamir D, Tanksley SD (2005) The SOL Genomics Network: A comparative resource for Solanaceae biology and beyond. *Plant Physiology* **138**, 1310-1317

Nodine MD, Bartel DP (2010) MicroRNAs prevent precocious gene expression and enable pattern formation during plant embryogenesis. *Genes and Development* **24**, 2678-2692

Nonogaki H (2010) MicroRNA gene regulation cascades during early stages of plant development. *Plant and Cell Physiology* **51**, 1840-1846

Ori N, Cohen AR, Etzioni A, Brand A, Yanai O, Shleizer S, Menda N, Amsellem Z, Efroni I, Pekker I, Alvarez JP, Blum E, Zamir D, Eshed Y (2007) Regulation of LANCEOLATE by miR319 is required for compound-leaf development in tomato. *Nature Genetics* **39**, 787-791

Palatnik JF, Allen E, Wu X, Schommer C, Schwab R, Carrington JC, Weigel D (2003) Control of leaf morphogenesis by microRNAs. *Nature* **425**, 257-263

Palatnik JF, Wollmann H, Schommer C, Schwab R, Boisbouvier J, Rodriguez R, Warthmann N, Allen E, Dezulian T, Huson D, Carrington JC, Weigel D (2007) Sequence and expression differences underlie functional specialization of *Arabidopsis* microRNAs miR159 and miR319. *Developmental Cell* **13**, 115-125

Pani A, Mahapatra RK, Behera N, Naik PK (2011) Computational identification of sweet wormwood (*Artemisia annua*) microRNA and their mRNA targets. *Genomics*, *Proteomics and Bioinformatics* **9**, 200-210

Peaucelle A, Laufs P (2006) The UN peace-keeping force in the plants: A microRNA, miR164, stabilizes the boundary domain in the meristem. *Medicine Sciences (Paris)* **22**, 796-798

Qiu CX, Xie FI, Zhu YY, Guo K, Huang SQ, Nie L, Yang ZM (2007) Computational identification of microRNAs and their targets in *Gossypium hirsutum* expressed sequence tags. *Gene* **395**, 49-61

Rhoades MW, Reinhart BJ, Lim LP, Burge CB, Bartel B, Bartel DP (2002) Prediction of plant microRNA targets. *Cell* **110**, 513-520

Ritchie W, Theodule FX, Gautheret D (2008) Mireval: A web tool for simple microRNA prediction in genome sequences. *Bioinformatics* **24**, 1394-1396

Ro S, Park C, Jin J, Sanders KM, Yan W (2006) A PCR-based method for detection and quantification of small RNAs. *Biochemical and Biophysical Research Communications* **351**, 756-763

Ru P, Xu L, Ma H, Huang H (2006) Plant fertility defects induced by the enhanced expression of microRNA167. *Cell Research* **16**, 457-465

Schommer C, Palatnik JF, Aggarwal P, Chételat A, Cubas P, Farmer EE, Nath U, Weigel D (2008) Control of jasmonate biosynthesis and senescence by miR319 targets. *PLoS Biology* **6**, e230

Sunkar R, Girke T, Jain PK, Zhu JK (2005) Cloning and characterization of microRNAs from rice. *Plant Cell* **17**, 1397-1411

Sunkar R, Kapoor A, Zhu JK (2006) Posttranscriptional induction of two Cu/Zn superoxide dismutase genes in *Arabidopsis* is mediated by down-regulation of miR398 and important for oxidative stress tolerence. *Plant Cell* **18**, 2051-2065

Sunkar R, Zhu JK (2004) Novel and stress-regulated microRNAs and other small RNAs in *Arabidopsis*. *Plant Cell* **16**, 2001-2019

Taylor PF, Xie FL, Freistaedter A, Burklew CE, Zhang B (2010) Identification and characterization of microRNAs and their target genes in tobacco (*Nicotiana tabacum*). *Planta* **232**, 1289-1308

Thakur V, Wanchana S, Xu M, Bruskiewich R, Quick WP, Mosig A, Zhu XG (2011) Characterization of statistical features for plant microRNA prediction. *BMC Genomics* **12**, 108

Wang L, Gu X, Xu D, Wang W, Wang H, Zeng M, Chang Z, Huang H, Cui X (2011) miR396-targeted AtGRF transcription factors are required for coordination of cell division and differentiation during leaf development in *Arabidopsis*. *Journal of Experimental Botany* **62**, 761-773

Wang M, Wang Q, Wang B (2012) Identification and characterization of microRNAs in Asiatic cotton (*Gossypium arboreum* L.). *PLoS One* **7**, e33696

Xie FL, Huang SQ, Guo K, Xiang AL, Zhu YY, Nie L, Yang ZM (2007) Computational identification of novel microRNAs and targets in *Brassica napus*. *FEBS Letters* **581**, 1464-1474

Xu ZH (2010) Advances in plant microRNA and stresses response. *Yi Chuan* **32**, 1018-1030

Yin Z, Li C, Han X, Shen F (2008) Identification of conserved microRNAs and their target genes in tomato (*Lycopersicon esculentum*). *Gene* **414**, 60-66

Zhang B, Pan X, Cannon CH, Cobb GP, Anderson TA (2006a) Conservation and divergence of plant microRNA genes. *Plant Journal* **46**, 243-259

Zhang B, Pan XP, Anderson TA (2006b) Identification of 188 conserved maize microRNAs and their targets. *FEBS Letters* **580**, 3753-3762

Zhang J, Zeng R, Chen J, Liu X, Liao Q (2008) Identification of conserved microRNAs and their targets from *Solanum lycopersicum* Mill. *Gene* **423**, 1-7

Zhang Y (2005) miRU: An automated plant miRNA target prediction server. *Nucleic Acids Research* **33**, W701-W704

Zhang Z, Yu J, Li D, Liu F, Zhou X, Wang T, Ling Y, Su Z (2010) PMRD: Plant microRNA database. *Nucleic Acids Research* **38**, D806-813

Zunker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* **31**, 3406-3415