CHAPTER IV

COMPUTATIONAL IDENTIFICATION OF miRNAs AND THEIR TARGETS

COMPUTATIONAL IDENTIFICATION OF miRNAs AND THEIR TARGETS

4. 1 Review of literature

Genomes of higher eukaryotes encode not only proteins but also diverse non-coding RNAs, particularly small (20 to 30 nt) regulatory RNAs (Zhu et al. 2008). Small RNAs include miRNAs, siRNAs, and piwi-interacting RNAs (piRNAs) (Chapman and Carrington 2007, Matranga and Zamore 2007). The miRNAs are single stranded RNAs that are generated from endogenous hairpin shaped long precursor transcripts (Kim 2005, Zhou et al. 2007). Precursor RNAs (pre-miRNAs) are usually ~60-80 nt in animals and more variable in plants (Murchison and Hannon 2004, Kim 2005). Mature miRNAs always present in one arm of the hairpin precursor without having large internal loops or bulges (Krol et al. 2004). They bind to the complementary sites on target mRNAs and repress post transcriptional gene expression in both animals and plants (Singh and Nagaraju 2008, Zhang et al. 2006a). In higher eukaryotes, miRNAs are one of the largest gene families (>200 members per species) accounting for ~ 1% of the genome which have a key role in diverse regulatory pathways (Kim 2005).

4.1.1 Biogenesis of miRNAs and their functional roles

Since the discovery of first miRNA in *lin-14* gene in *C. elegans* (Lee et al. 1993), it has been extensively studied in both animals and plants for discovering new miRNAs and their biogenesis pathways with functional roles. However, first miRNA in plant was reported only in the year 2002 in *Arabidopsis,* much later than its animal counterpart (Reinhart et al. 2002). It is now known that miRNA genes reside in regions of the genome as distinct transcriptional units as well as in clusters of polycistronic units carrying the information of several miRNAs (Bartel 2004, Chapman and Carrington 2007). Studies suggest that approximately half of known miRNAs reside in non-protein coding RNAs (intron and exon) or within the intron of protein coding genes (Chapman and Carrington 2007). In animals, pre-miRNAs are processed in the nucleus into shorter hairpin RNAs of 65 nt by the microprocessor complex containing RNaseIII enzyme Drosha and its cofactor DGCR8/Pasha, a dsRNA binding protein (Zhu 2008). Precursor miRNA is then exported to the cytoplasm,

where it is further processed by another RNase III enzyme, Dicer, to release a 22 nt miRNA/miRNA duplex (Matranga and Zamore 2007, Han et al. 2006). Dicer function also requires a dsRNA-binding protein, TRBP, as a cofactor. The miRNA is loaded into the effector complex, known as RISC, to direct complementary or partially complementary mRNAs for cleavage or translational repression (Zhu et al. 2008). In plants, the two-step processing of pre-miRNAs into mature miRNAs occurs entirely in the nucleus and is carried out by a single RNase III enzyme, DCL1 (Dicer-like 1) (Jones-Rhoades et al. 2006). In addition to DCL1, genetic analysis revealed that HYL1, a dsRNA binding protein, and SE, a C2H2-type zinc finger, are also required for processing pre-miRNAs and for accumulation of mature miRNAs (Zhu 2008). However, whether DCL1 alone is active in processing pre-miRNAs into miRNAs and how HYL1 and SE may function in the processing steps are not known. Recent studies suggested that DCL1 is inaccurate in catalyzing the release of miRNAs from pre-miRNAs and HYL1 or SE could improve the accuracy resulting in 80% of the *in vitro* processing products being miRNAs (Dong et al. 2008). The importance of DCL1, HYL1, and SE in plant growth and development was evident early on the severe and pleitropic plant phenotypes exhibited by their loss-of-function mutant alleles (Zhu 2008, Park et al. 2002).

It has been well-demonstrated that miRNAs play critical roles in diverse biological processes such as development, cellular differentiation, cell-cycle control, apoptosis and oncogenesis. As in other organisms, in plants miRNAs play crucial roles in various stages of development and maintenance of organ identity such as leaf morphology and polarity, roots formation, transition from juvenile to adult vegetative phase and vegetative to flowering phase, flowering time, floral organ identity and reproduction (Wang et al. 2007, Mallory and Vaucheret 2006, Sunkar et al. 2007). They are also found to be involved in response to pathogen invasion (Zhang et al. 2006a), hormone signaling (Eckardt 2005, Guo et al. 2005), environmental stress such as cold, salinity, drought stress (Jones-Rhoades and Bartel 2004, Sunkar and Zhu 2004, Lu and Huang 2008) and promotion of anti-viral defence (Lu et al. 2008a).

4.1.2 Strategies for miRNA identification and characterization

Identification of miRNAs and their functional classification has become one of the most active research fields in biology in the recent years. The miRNA identification largely relies on two main reverse genetics strategies: 1) computational and (2) experimental approaches. A third identification approach, forward genetics, is rarely used in miRNA discovery (Unver et al. 2009).

Identification of miRNAs using computational tools is one of the most widely used methods, contributing considerably to the prediction of new miRNAs in both animal and plant systems (Zhang et al. 2006b). This is largely due to low cost, high efficiency, fast and comprehensive methodology of bioinformatics (Griffiths-Jones et al. 2006). The main theory behind this approach is finding homologous sequences of known miRNAs both within a single genome and across genomes of related organisms (Jones-Rhoades and Bartel 2004). Sequence and structure homologies are used for computer-based predictions of miRNAs (Jones-Rhoades and Bartel 2004). Therefore, prior DNA or RNA sequence information of the organism selected for miRNA identification is required in order to run the softwares. On the other hand, the computationally predicted miRNAs should also be confirmed via experimental methods. In recent years, a number of programs and bioinformatics tools have been developed and used successfully for the identification and analysis of miRNAs and their targets (Rajewsky and Socci 2004, Zhang et al. 2009b). Computational identification of miRNAs were reported by homology search in ESTs (Zhang et al. 2005, Guo et al. 2007, Nasaruddin et al. 2007, Jin et al. 2008), GSS (Zhang et al. 2006c, Qiu et al. 2007, Lu and Yang 2010) as well as in whole genome (Zhang et al. 2009a) from diverse plant species. Importantly, large portion of the identified miRNAs were reported from diverse crop plants species (Table 4.1).

Identification and characterization of miRNAs through cloning and sequencing of small RNA libraries represents an experimental approach. Next generation massive sequencing techniques such as pyrosequencing and Illumina are also applied to identify new miRNAs in plants (Moxon et al. 2008, Szittya et al. 2008). Such approaches for miRNA identification also have some limitations. First, most of the miRNAs are tissue and time specific, and generally their expression level is low. In

Table 4.1: Identification of miRNAs in important crop plants

	Number of miRNAs			
Crop plants	**Total**	**Exp.**	**Computational**	**References**
Medicago truncatula	395	356	39	Zhang et al. 2005, Dezulian et al. 2005, Jagadeeswaran et al. 2009, Zhou et al. 2008, Szittya et al. 2008, Lelandais-Briere et al. 2009
Vigna unguiculata	50	-	50	Wang et al. 2009, Lu and Yang (2010)
Glycine max	166	80	86	Zhang et al. 2006b, Zhang et al. 2005, Subramanian et al. 2008, Wang et al. 2009, Zhang et al. 2008a
Phaseolus sp.	12	8	4	Zhang et al. 2005, Zhang et al. 2006b, Arenas-Huertero et al. 2009, Sunkar and Jagadeeswaran 2008, Jagadeeswaran et al. 2009
Oryza sativa	2641	305	2336	Reinhart et al. 2002, Rhoades et al. 2002, Jones-Rhoades and Bartel 2004, Sunkar et al. 2005, Guddeti et al. 2005, Jones-Rhoades et al. 2006, Archak et al. 2007, Morin et al. 2008, Lu et al. 2008b, Sunkar et al. 2008, Zhang et al. 2006b, Xue et al. 2009, Sunkar et al. 2005, Lindow et al. 2007, Zhu et al. 2008, Lacombe et al. 2008, Wu et al. 2009, Huang et al. 2009, Johnson et al. 2009, Jian et al. 2010

(Table 4.1 continued....)

Brassica sp.	82	70	12	Xie et al. 2007, Wang et al. 2007, Pant et al. 2009, Kutter et al. 2007, He et al. 2008, Hsieh et al. 2009
Zea mays	269	49	220	Zhang et al. 2005, Zhang et al. 2006b, Dezulian et al. 2005, Juarez et al. 2004, Maher et al. 2004, Zhang et al. 2006c Zhang et al. 2009, Johnson et al. 2009
Allium cepa	7	-	7	Zhang et al. 2005, Zhang et al. 2006b
Arachis hypogaea	36	-	36	Zhao et al. 2010
Glycine max	229	201	28	Subramanian et al. 2008, Zhang et al. 2005, Zhang et al. 2008a, Dezulian et al. 2005
Hordeum vulgare	17	-	17	Zhang et al. 2005, Zhang et al. 2006b, Sunkar and Jagadeeswaran 2008
Saccharum officinarum	32	-	32	Sunkar and Jagadeeswaran 2008, Dezulian et al. 2005
Solanum lycopersicum	37	21	16	Zhang et al. 2005, Zhang et al. 2006b, Sunkar and Jagadeeswaran 2008, Zhang et al. 2008b, Moxon et al. 2008
Solanum tuberosum	62	-	62	Zhang et al. 2005, Zhang et al. 2006b, Sunkar and Jagadeeswaran 2008, Zhang et al. 2009b
Sorghum bicolor	150	-	150	Zhang et al. 2005, Zhang et al. 2006b, Sunkar and Jagadeeswaran 2008, Dezulian et al. 2005, Maher et al. 2004, Bedell et al. 2005, Paterson et al. 2009
Triticum aestivum	85	71	14	Zhang et al. 2005, Zhang et al. 2006b, Yao et al. 2007, Wei et al. 2009

(Table 4.1 continued....)

Ricinus communis	87	63	24	Griffiths-Jones et al. 2003, Dezulian et al. 2005, Zeng et al. 2010
Gossypium sp.	89	19	70	Khan Barozai et al. 2008, Qiu et al. 2007, Zhang et al. 2007, Pang et al. 2009, Kwak et al. 2009
Malus x domestica	5	-	5	Zhang et al. 2005, Zhang et al. 2006b
Vitis vinifera	144	-	144	Zhang et al. 2005, Jaillon et al. 2007, Hsieh et al. 2009
Citrus sp.	126	-	126	Zhang et al. 2005, Zhang et al. 2006b, Sunkar and Jagadeeswaran 2008, Wu et al. 2009, Song et al. 2009, Song et al. 2010

addition, they mostly express in response to specific environmental stimuli. They also coexist with their cleaved and degraded target mRNAs, hence cloning small RNAs (miRNA and siRNA) is difficult (Unver et al. 2009). Forward genetics approach is time consuming, expensive, and less efficient; hence it is rarely used in plant miRNA identification. There is only one example using a forward genetics experimental approach to identify miRNA in plants. Baker et al. (2005) identified a miRNA loss of function allele by a transposon insertion upstream of the predicted *MIR164c* stem-loop. The miRNA mutant resulted in a flower phenotype with extra petals.

4.1.3 Identification of miRNA targets

Target identification of miRNAs is important to know their functional roles. Specific miRNA targets in plant genomes and transcriptomes have been identified with computational methods and subsequently validated experimentally. Predicting miRNA targets in plants is much easier due to high and significant complementarities to miRNA-mRNA targets (Moxon et al. 2008). High degree of complementarity between plant miRNAs and their target mRNAs has allowed the prediction of targets using computational algorithms and tools (Jones-Rhoades and Bartel 2004). Ability for plant miRNA to target mRNA with perfect sequence complementary matches was first shown with miR171 (Llave et al. 2002). It was shown that miR171 has perfect antisense complementarity with three Scarecrow-like (SCL) transcription factors in the *Arabidopsis* genome. Predicting conserved miRNA targets in different organisms has revealed that homologous mRNAs are targeted by conserved miRNAs within a miRNA family, yet allowing more gaps and more mis-matches between an individual miRNA and its target (Laufs et al. 2004). Successful prediction of miRNA targets have been well demonstrated in several plant species including *Arabidopsis* (Rhoades et al. 2002) and rice (Li et al. 2010). Most of the computationally identified targets in different studies were reported as transcription factors which play important roles in plants growth and developments (Zhang et al. 2006a, Han et al. 2010, Lu and Yang 2010). As with computational approaches, experimental approaches have been utilized widely to predict plant miRNA-mRNA target sites. Genome-wide expression profiling to search for miRNA targets can be applied on expression arrays. In one example, array data showed that five transcripts encoding *TCP* genes were down-

regulated via overexpression of miR319a (miR-JAW) in *Arabidopsis*. Those five TCP transcription factor mRNAs show up to five mismatches, or four mismatches when G: U wobble counts 0.5 mismatch (Palatnik et al. 2003). Additionally Schwab et al. (2005) overexpressed four different miRNAs in each *Arabidposis* plant and examined each expression profile to experimentally establish parameters for target cleavage guided by plant miRNAs. However, they found no new target mRNAs other than previously identified by computational approaches. At present, the most powerful method to confirm miRNA-mRNA targets is the 5′ RACE procedure. The 5′ RACE technique has been used successfully to identify miRNA targets in many plants (Sunkar et al. 2005, Palatnik et al. 2003, Mallory et al. 2005). Cleaved mRNA products in plants have two diagnostic properties. One is that the 5′ phosphate of a cleaved mRNA product can be ligated to an RNA adaptor with T4 RNA ligase. Second, in general, the precise target cleavage position is that mRNA target nt pair with the tenth nucleotide of miRNA (Sunkar et al. 2005, Kasschau et al. 2003). Cleaved mRNA products by miRNA guided activity can be amplified with ligation of an oligo-nucleotide adaptor to the 5′ end, followed by reverse transcription and PCR amplification with a gene specific primer (Sunkar et al. 2005).

MiRNA studies in plants have already explained a number of biological events in response to abiotic stresses. Improved understanding of molecular mechanisms of miRNA in plants will lead to the development of novel and more precise techniques that will help in better understanding of some post-transcriptional gene silencing in response to stresses. Accumulating knowledge on the roles of plant miRNA's in molecular biology is leading to the development of more efficient and reliable tools for their characterization.

4.2 Material and methods

Discoveries of miRNAs in model and crop plants as well as their growing evidence of involvement in a variety of functional roles have produced a great deal of excitement in agricultural biotechnologly. Despite the limited genome resources of tea, published ESTs and full length nucleotide sequences have provided the scope to get more genetic information. In this study, new conserved miRNAs were mined in local tea sequence database for the purpose of understanding their roles in regulating growth and development, metabolism and other physiological processes.

4.2.1 Collection of reference miRNAs, full-length nucleotides and EST sequences

A total of 9275 plant miRNAs (on 20 September, 2010) and their fold back sequences were obtained from miRBase (Griffiths-Jones et al. 2008). Homolog miRNAs were eliminated and the rest were defined as reference for searching homolog tea miRNAs. Tea nucleotide and EST sequences (14819 as on 8 August, 2010) including those generated in present investigation were downloaded from NCBI's nucleotide and dbEST database (http://www. ncbi.nlm.nih.gov/). All redundant and poor quality sequences were eliminated and created a local nucleotide database.

4.2.2 Prediction of miRNAs and their precursors

The procedure for searching conserved miRNA homologues in tea is summarized in Fig. 4.1. Since, only mature miRNAs were conserved in plants, the reference sequences were used as a query in homology search against the reference nucleotide database at e-value threshold less than 0.01 through BLAST+ 2.2.22 program (Altschul et al. 1997). Target sequences with no more than four mismatches were considered and predicted the secondary structure using Mfold v 3.2 (Zuker 2003). Precursor sequences were searched in 50 nt upstream or downstream from the location of mature miRNAs with an increment of 10 nt. While selecting a RNA sequence as a candidate miRNA precursor, following criteria were used according to Zhang et al. (2006b) with minor modifications as: i) a RNA sequence can fold into an appropriate stem-loop hairpin secondary structure, ii) a mature miRNA sequence site in one arm of the hairpin structure, iii) miRNAs had less than seven mismatches

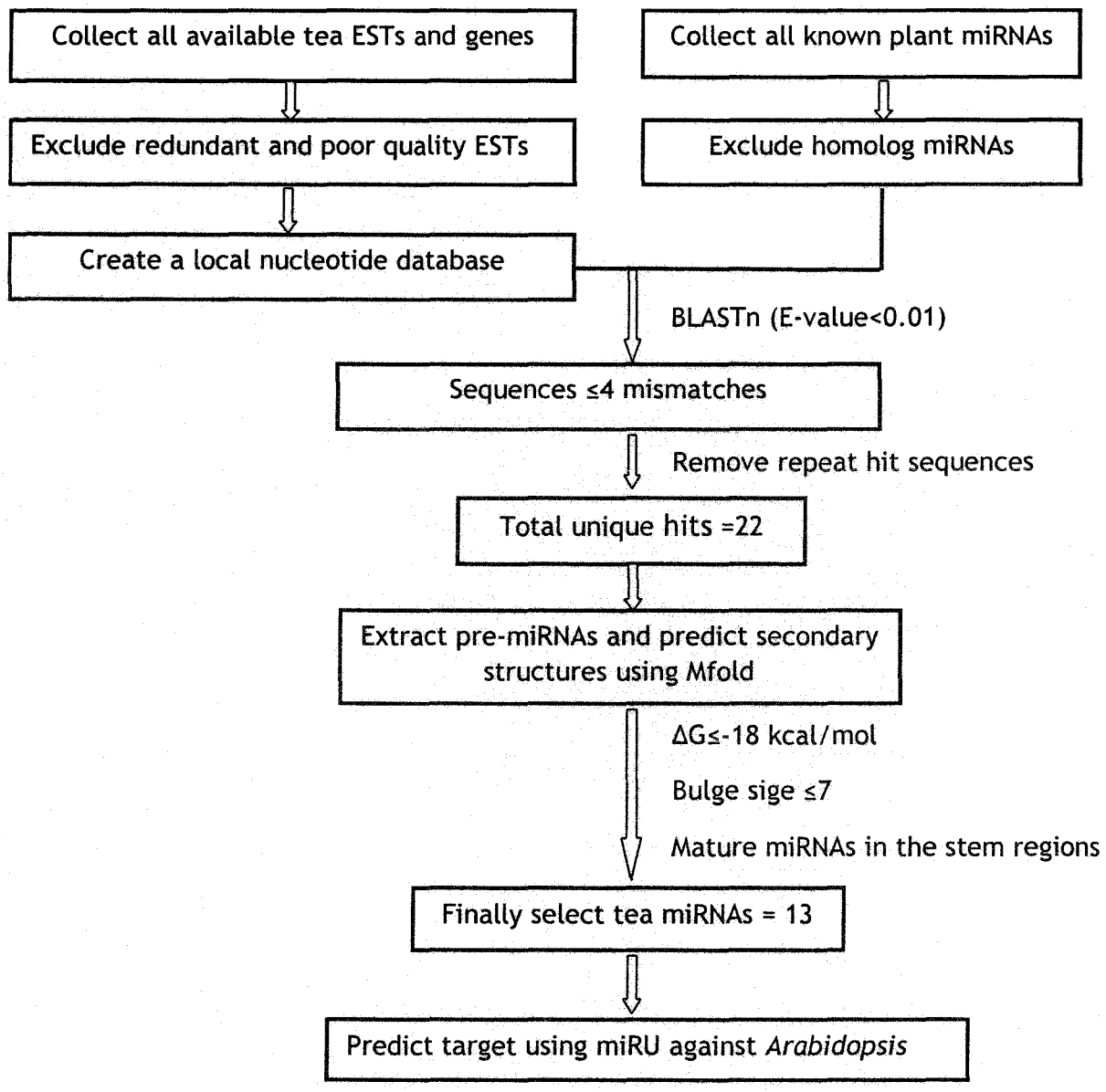


Fig. 4.1: Different steps involved in the identification process of conserved miRNAs and their targets.

with opposite miRNA sequence in the other arm, iv) no loop or break in miRNA sequences, v) predicted secondary structures had higher negative energy MFEs (≤-18 kcal/mol), and iv) 40-70% A + U contents.

4.2.3 Identification of potential miRNA targets

As for tea, since only few gene sequences are available, *Arabidopsis* was used as a reference system for finding targets of the candidate miRNAs. The predicted tea miRNAs were used as query against the *Arabidopsis thaliana* DFCI gene index (AGI) release 13 using miRU (http://bioinfo3.noble.org/psRNATarget/) following the criteria as i) maximum expectation value 3 ii) multiplicity of target sites 2 iii) range of central mismatch for translational inhibition 9 to 11 nt iv) maximum mismatches at the complementary site ≤4 nt without any gaps.

4.2.4 Evaluation of conserved nature in miRNAs and their phylogenetic relationships

Due to the conserved nature of miRNAs, orthologue discovery can be done through bioinformatics analysis. Here, tea miRNA conservation was analysed with their orthologues. A homology search of candidate tea miRNAs was done against all plant miRNAs using BLAST+ 2.2.22 (Altschul et al. 1997) allowing maximum of 3 mismatches and e-value <0.001. The corresponding precursor sequences of homolog small RNA's were identified and collected. The collected sequences of diverse plant species were aligned with the tea miRNA homolog using Clustal W (Thompson et al. 1994).

A query of tea small RNAs against known miRNA families (miRBase, release 15) allowed to identify 3 previously reported large families. The precursor sequences of three known family members were selected along with respective precursor sequences of tea minas. Then, the maximum likelihood trees were constructed for each family based on Tamura-Nei model (Tamura and Nei 1993) with default bootstrap values using MEGA 4.0 (Tamura et al. 2007) to illustrate the evolutionary relationships among the members of the family.

4.2.5 Nomenclature of miRNAs

The predicted tea miRNAs were named in accordance with miRBase (Griffiths-Jones et al. 2006). Mature sequences were designated 'miR', and the precursor hairpins were labeled as 'mir' with prefix '*csi*' for *C. sinensis*, '*cja*' for *C. japonica* and '*cas*' for *Camellia assamica*. In the cases where distinct precursor sequences have identical miRNAs with different mismatch pattern, they were named as *csi-mir-1-a* and *csi-mir-1-b*.

4.3 Results

Sequence and structure homologies are the main theory behind the computer-based approach for miRNA prediction. At present, four kinds of databases namely genome, GSS, EST and nucleotide are mainly used for miRNA mining. Considering the unavailability of genome and genomic survey sequences of *Camellia*, both dbEST and nucleotide databases was mined for miRNA identification.

4.3.1 Identification of miRNAs and their precursors

A total of 14819 tea sequences containing 2023 full length nt and 12796 ESTs were obtained from available individual database in GenBank. Out of these, 22 sequences had less than five mismatches with previously known plant miRNAs. After carefully evaluating the hairpin structures using the defined criteria, 13 miRNAs were finally identified. Details of tea miRNAs such as source sequences, location in the source sequences, length of precursor sequences and their minimum folding free energies and A + U content were given in Table 4.2. A total of 9 miRNAs were predicted from ESTs whereas 4 were from full length nucleotide sequences. Five of them were located in the direct strand and rests were in the indirect strand. Newly identified pre-miRNAs have minimum folding free energies (mfe) ranging from -186.83 to -18.5 kcal/mol, with an average of about -72.69 kcal/mol and A+U content ranging from 45.24 to 69.83% with an average of 53.79% (Table 4.2, Appendix-4.1). Length of tea miRNA precursors ranges from 65 to 663 nt with an average of 248 nt and mature sequences ranges from 20 to 25 nt. The newly predicted two tea miRNA (cja-miR2910, csi-miRf10132-akr) sequences were perfectly (100%) matched with the corresponding homologues of *populus* and rice, whereas the remaining 11 mature tea miRNA sequences differ by 1 to 4 nt from their homologues. All mature sequences of tea miRNAs were found in the stem portion of the hairpin structures (Fig. 4.2) containing less than 7 mismatches in the other arm without break or loop inside the mature sequences. It was found that tea miRNA (csi-miR408) has been conserved with diverse plant species (Fig. 4.3, Appendix-D 4.2) from monocotyledonous plants such as rice, maize to dicotyledonous plants such as populous.

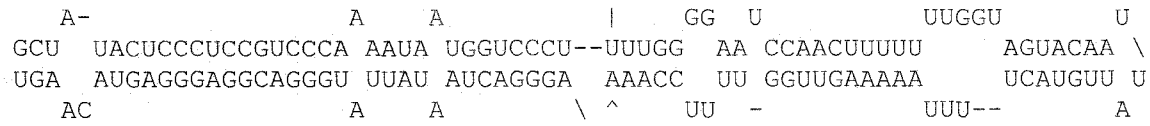
Table 4.2: Predicted miRNAs of *Camellia sp.*

New miRNAs	NS*	Gene ID	Strand	SP*	EP*	NM (nt)	Mature miRNAs	e-value	PL*	A+U (%)	MFE
csi-miR 408	EST	206583693	3′	137	117	3	CUGCACUGCCUCUUCCCUGAG	0.001	336	45.24	-20.10
csi-miR1171	EST	171355265	5′	286	308	1	UGGAGUGGAGUGAAGUGGAGUGG	3E-04	181	56.98	-5.97
csi-miR414a	EST	206583641	3′	757	637	3	UCUUCCUCAUCAUCAUCUUCU	0.001	663	57.32	-3.18
csi-miR414d	EST	284026209	3′	186	166	1	UCAUCGUCAUCGUCAUCAUCU	0.004	193	61.14	-7.72
csi-miR414f	EST	212378632	5′	122	142	1	UCAUCAUCAUCAUCAUCUUCA	6E-05	68	57.35	-18.50
cas-miR1122	FL	214011104	5′	214	237	4	UACUCCCUCCGUCCCAAAAUAAUG	6E-05	294	69.83	-1.23
csi-miR414g	EST	51453040	3′	474	454	3	CCUUCCUCAUCAUCAUCGUCC	0.001	70	45.71	-25.20
csi-miRf10132-akr	EST	51453383	3′	58	34	0	GCGAGCUUCUCGAAGAUGUCGUUGA	9E-08	200	49.00	-69.50
cja-miR2910	FL	1777723	5′	1262	1282	0	UAGUUGGUGGAGCGAUUUGUC	1E-05	301	49.83	-91.00
csi-miR2914	FL	34787361	5′	345	367	1	UAUGGUGGUGACGGGUGACGGAG	5E-06	65	49.23	-20.90
cas-miRf10185-akr	EST	221071827	3′	232	212	3	GAAAGGGGAAAACAUUGUAGC	0.004	139	48.92	-51.10
cas-miR11590-akr	EST	212379609	3′	113	94	3	UUUUGGUGUGCCUUCAACCU	0.003	75	53.33	-23.80
csi-miR414h	EST	295345415	3′	79	58	4	UCAUCCUCAUCAUCGUCAGAA	0.004	644	55.36	-86.83

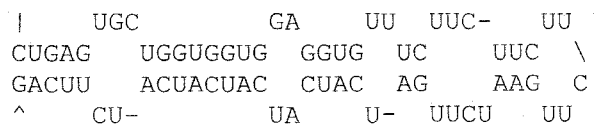
*NS = nucleotide source, FL = full-length, SP = start point, EP = end point, ME = No match, PL = pre-miRNA length

Fig. 4.2: Predicted hairpin secondary structures of candidate pre-miRNAs. The miRNAs were highlighted (red color) in stem portion

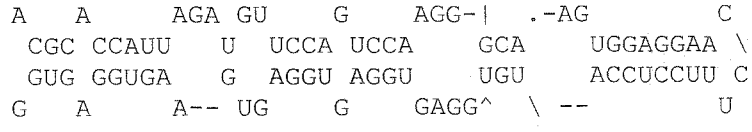
cas-miR1122



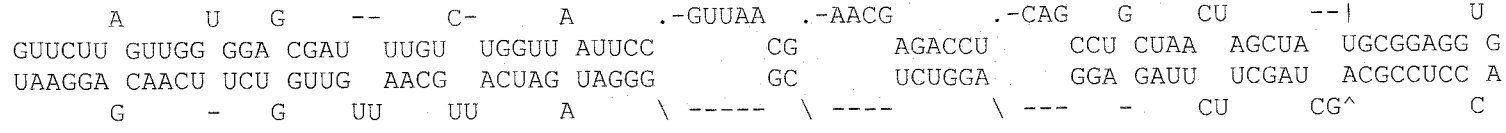
csi-miR414f



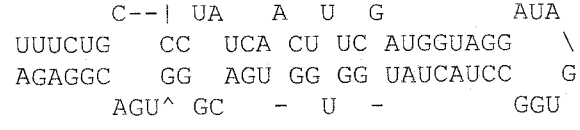
csi-miR1171



cja-miR2910

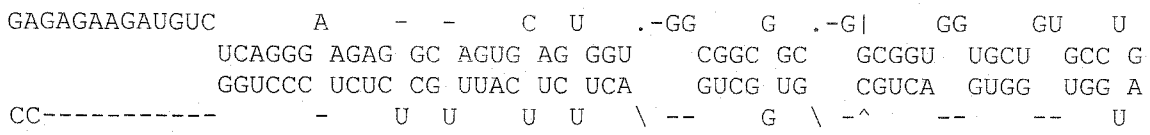


csi-miR2914

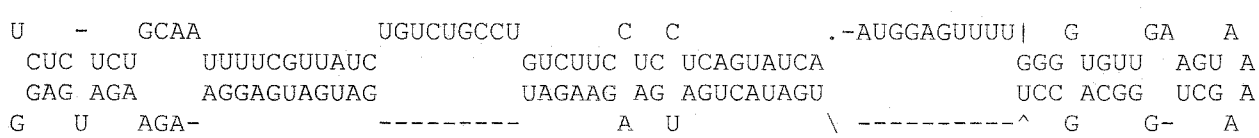


(Fig. 4.2 continued...)

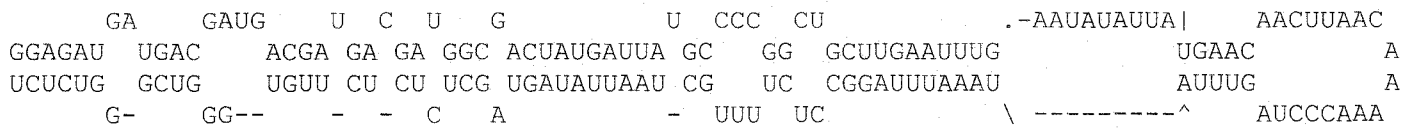
csi-miR 408



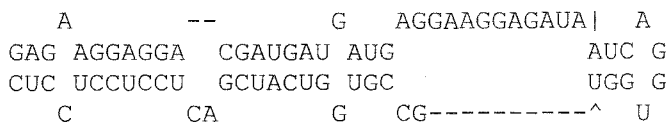
csi-miR414a



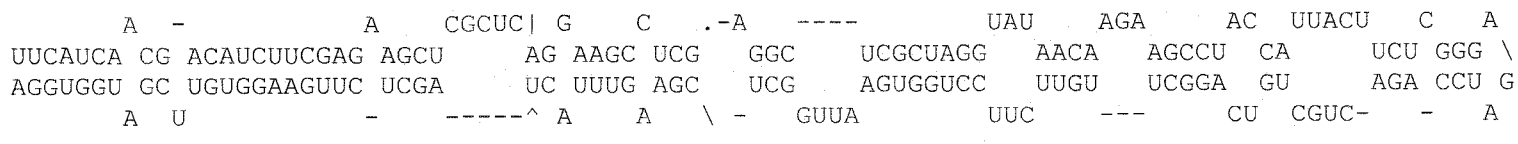
csi-miR414d



csi-miR414g



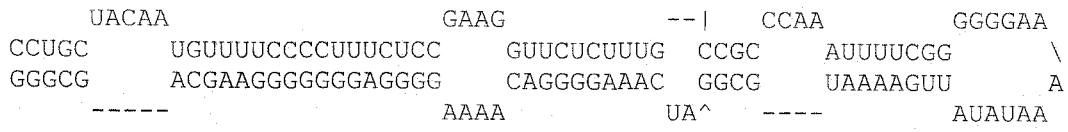
csi-miRf10132-akr



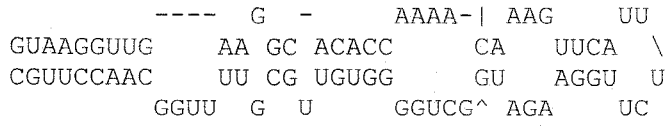
.

(Fig. 4.2 continued...)

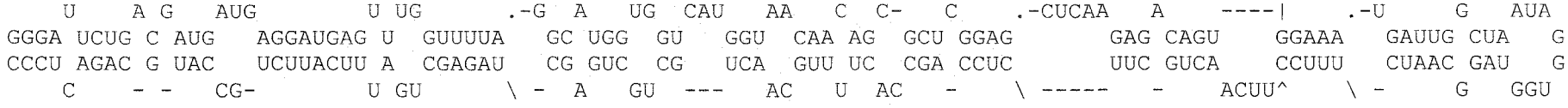
cas-miRf10185-akr



cas-miR11590-akr



csi-miR414h



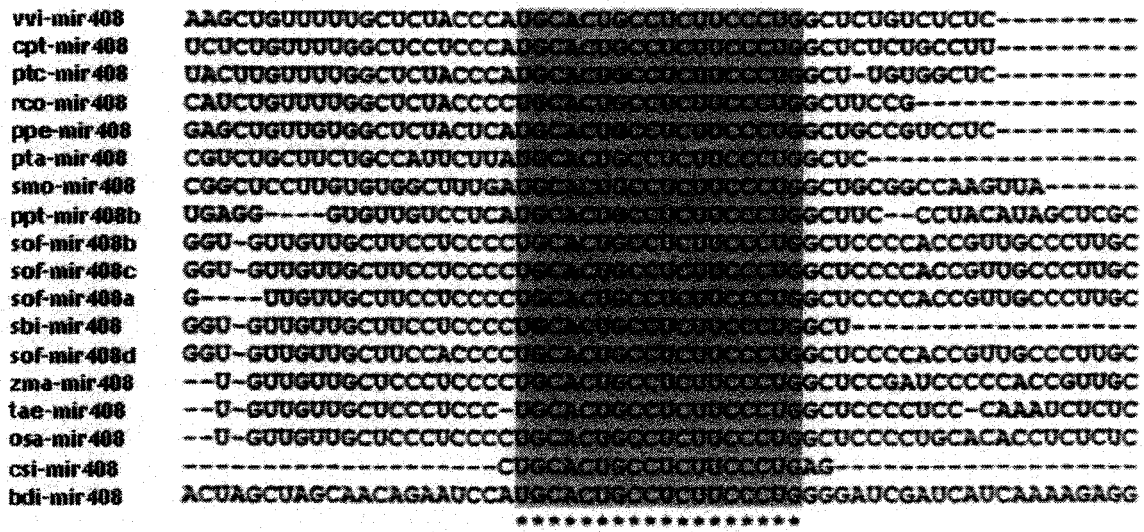


Fig. 4.3: Conservation of tea miRNA (csi-miR408) in diverse plant species. Alignment of pre-miRNAs of different plant species (Appendix-D 4.2) with homolog tea miRNA (csi-miR408) demonstrate the conserve nature of mature miRNAs (highlighted).

4.3.2 Phylogenetic analysis

Newly identified tea miRNAs belong to 9 miRNA families including three known independent large miRNA families (mir 408, mir414 and mir1122). There are one tea miRNA namely csi-miR 408 and cas-miR1122 in each family of mir-408 and mir-1122, respectively. However, five members of family mir408 were found in tea (csi-miR414a, csi-miR414d, csi-miR414f, csi-miR414g, and csi-miR414h). Comparison of tea miRNA precursor sequences with other members in the same family showed that most members could be found to have a high degree of sequence similarity with others. Phylogenetic trees among the members of each family illustrated the evolutionary relationships of tea miRNAs (Fig. 4.4, Appendix-D 4.3-4.5).

4.3.3 Identification of potential miRNA targets

A total of 37 potential targets were identified for the 7 predicted miRNA families which include 11 miRNAs based on their perfect or nearly perfect complementarity with their target sequences in *Arabidopsis* (Table 4.3, Fig. 4.5). For all the miRNAs, single binding site was found in the targets without any gaps in the complementary region and expectation value ranges from 0 to 3. These potential miRNA targets were belonged to a number of gene families that involved in different biological functions such as regulation of cell cycle, metal ion transportation, starch metabolic processes etc. There were 8% of genes encoding transcription factors, 30% of genes encoding different enzymes and 14% of genes encoding transporters as well as 48% of genes encoding various proteins of physiological and metabolic processes (Table 4.3). The miRNA family 'miR414' showed the highest 30 numbers of independent target genes followed by 'miR408' family with 2 numbers of target genes. Rest miRNA families were with single target genes in *Arabidopsis* (Table 4.3). The 'miR1171' and 'miR1122' miRNA family members did not bind to any target sequences within the defined filtration criteria.

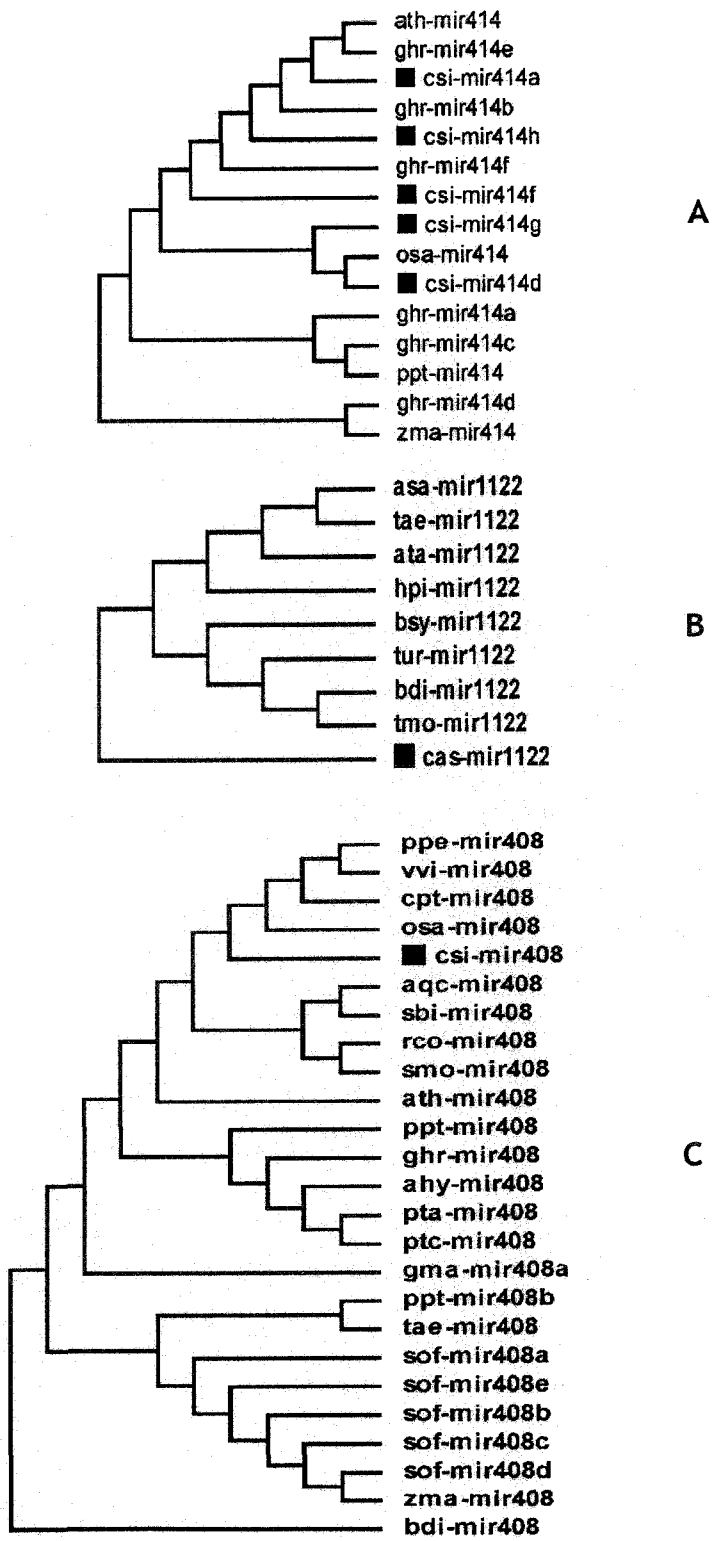


Fig. 4.4. Phylogenetic relationship among the members of each miRNA family, namely: A) miRNA414, B) miRNA1122 and C) miRNA408.

Table 4.3: Potential targets of the candidate miRNA families

Families	TS*	Targeted proteins	Targets involved in	E-value	Gene IDs
miR408	1	Cyclin-dependent protein kinase	regulation of cell cycle	3.0	TC284764
	1	Copper ion binding protein	metal ion transport	3.0	TC300377
miR414	1	RAN GTPase activating protein 2	cytokinesis	1.0	TC283936
	1	50S ribosomal protein L21	translation process	1.0	TC290468
	1	26S proteasome AAA-ATPase subunit	proteasomal protein catabolic process	1.0	TC304323
	1	SEC12p-like transporter	ER to golgi vesicle mediated transport	1.5	NP225634
	1	Nucleotide binding protein	nucleotide binding	2.0	TC280880
	1	MYB transcription factor	regulation of circadian rythm	2.0	TC283178
	1	Aldose 1-epimerase	carbohydrate metabolic process	0.5	TC285263
	1	Phosphatase 2C-like protein	protein amino acid dephosphorylation	0.5	TC285483
	1	Phosphatidylinositol phosphatase	inositol or phosphatidylinositol activity	0.5	TC312518
	1	Starch branching enzyme class II	starch metabolic process	1.0	AA586097
	1	Reproductive meristem protein 1	regulation of transcription	1.0	TC284340
	1	Calcium ion binding protein	calcium ion binding	1.0	TC299015
	1	Emb protein	RNA processing	1.0	TC294192
	1	Zinc ion binding protein	regulation of transcription	1.0	TC299076
	1	Calmodulin-4	calcium ion binding	1.5	TC294389
	1	Translation initiation factor 3 subunit 8	translation initiation	0	TC280751
	1	SMC3 protein	chromosome segregation process	0	TC281006

(Table 4.3 continued...)

	1	MLO-like protein 3	cell death	0	TC293173
	1	Metalloendopeptidase	proteolysis	0	TC299050
	1	Ubiquitin conjugating enzyme	catabolic process	0.5	CA781750
	1	Zinc finger protein	regulation of transcription	0.5	NP030706
	1	Plastid protein	protein targetting to chloroplast	0.5	TC290688
	1	Ubiquitin thiolesterase	catabolic process	1.5,1.5	TC280710, TC298096
				1.5	TC305774
	1	Methionyl-tRNA synthetase	methionyl-tRNA aminoacylation	1.5	TC282196
	1	Metal ion binding protein	metal ion binding	1.5	TC305801
	1	Sfc4 protein	xylem or phloem pattern formation	2.0	TC280894
	1	Transcription factor	regulation of transcription	2.0, 2.0	TC293875, TC294793
	1	Synaptosomal-associated protein 25	vesicle mediated transport	2.5	TC293303
	1	ATP binding	protein amino acid phosphorylation	2.5	TC299397
	1	ADP-ribosylation factor-like protein	intracellular protein transport	3.0	TC289959
miRf10132	1	Histone H2B like protein	nucleosome assembly	1.5, 1.5	TC297551, TC313314
				2.5, 1.5	TC313977, TC294144
miR2910	1	Extracellular matrix structural constituent	matrix organisation	0	TC310823,
miR2914	1	Glutamate semialdehyde dehydrogenase	glutamate metabolism	2.0	TC287905
miRf10185	1	Carboxylic ester hydrolase	hydrolase activity	3.0, 3.0	TC298946, TC308821
miR11590	1	FRIGIDA protein	regulation of flower development	2.0	TC309547

*TS=targeted site

	ACUUCUACUACUACUACUACU		csi-miR414f
	::::::::::::::::::::		
646	UGAAGAUGAUGAUGAUGAUGA	666	eukaryotic initiation factor 8
	UCUACUACUGCUACUGCUACU		csi-miR414d
	:::::::::::::::.:::::		
312	AGAUGAUGACGAUGAUGAUGA	332	phosphatase 2C-like protein
	UCUUCUACUACUACUCCUUCU		csi-miR414a
	:::.:::::::::::::::.::		
287	AGAGGAUGAUGAUGAGGAGGA	307	50S ribosomal protein L21
	CUGCUACUACUACUCCUUCC		csi-miR414g
	::.::::::::::::::::::		
1435	GAUGAUGAUGAUGAGGAAGA	1454	ubiquitin thiolesterase
	AGUUGCUGUAGAAGCUCUUCGAGCG		csi-miR10132
	:.::::::.::::::::::::.::		
356	UUAACGAUAUCUUCGAGAAGCUUGC	380	histone H2B like protein
	UAAGCUGCUACUACUCCUACU		csi-miR414h
	.::::::.:::::::: :::::		
815	GUUCGAUGAUGAUGAAGAUGA	836	ATP binding protein
	GAGUCCCUUCUCCGUCACGUC		csi-miR408
	:::: : .:::::::::::.::		
1239	CUCAUGAGAGAGGCAGUGUAG	1259	cyclin dependant protein kinase
	GAUGUUACAAAAGGGGAAAG		cas-miR10185
	.:.:::::::::::. ::::		
1328	UUGCAAUGUUUUCCUGUUUC	1347	carboxylic ester hydrolase
	CUGUUUAGCGAGGUGGUUGAU		cja-miR2910
	::::::::::::::::::::::		
452	GACAAAUCGCUCCACCAACUA	472	extracellular matrix structural constituent
	UCCAACUUCCGUGUGGUUUU		cas-miR11590
	:: ::::::: :::::::::		
2124	AGUUUGAAGGAACACCAAAA	2143	FRIGIDA protein
	GCAGUGGGCAGUGGUGGUAU		csi-miR2914
	.::::::::::::.:.::::.		
900	UGUCACCCGUCAUCGCCAUG	919	Glutamate semialdehyde dehydrogenase

Fig. 4.5: Predicted miRNA targets (black in colour, numbers indicate the position of nt in the target sequence) and their complementary sites with miRNAs (red in colour).

4.4 Discussion

With the availability of sequence resources in public databases, computer based miRNA identification methods has been focused more and more in the recent years due to its advantages of low cost and high efficiency. The number and sorts of miRNAs predicted in tea supported the fact that software-based approach is feasible and effective (Khan-Barozai et al. 2008, Jones-Rhoades and Bartel 2004, Singh and Nagraju 2008).

The identified tea miRNAs belong to 9 families where 'miR414' family has 5 members and the rest have single member in each. This familial distribution of miRNAs was also observed in *Arabidopsis*, rice and maize (Bonnet et al. 2004). This may be an indicative of dominant nature of 'miR414' family in miRNA-mediated gene regulation in tea. The identified miRNAs were found diverse in nature such as location of mature miRNA sequences and length of precursor sequences. Average length of precursor sequence was 248 nt; however a majority of them (62%) have 65-200 nt. This finding is similar to other plants where the length of precursors varied in contrast to consistent miRNA length of animal miRNAs (70-80 nt) (Bartel 2004, Ambros 2004). In tea miRNAs, diversity was also observed within the members of same family which was also reported in maize (Zhang et al. 2006c). The identified tea pre-miRNAs fold into hairpin secondary structures using minimum free energies, with an average -72.69 kcal/mol which was lower than the values of *Arabidopsis thaliana* pre-miRNAs and much lower than the folding free energies of tRNA (-27.5 kcal /mol) and rRNA (-33 kcal/ mol) (Khan Barozai et al. 2008).

Out of 13 newly identified miRNAs, 10 were from ESTs. There are several reports on miRNA identification from ESTs in various plant species (Zhang et al. 2006b, Qiu et al.2007). The source sequences of miRNAs show a probable link between expression of miRNAs and their tissues, organs or developmental stages to which it belongs. On that basis, it was recognized that csi-miR414f, cas-miRf10185 and cas-miR11590-akr were expressed in roots and the rest tea miRNAs were expressed in leaf tissues. Moreover, csi-miR1171 and csi-miR414d were found in leaf tissue under the stress of winter dormancy and pest infestation, respectively. Three miRNAs namely cas-miR1122, cja-miR2910 and csi-miR2914 were identified from the full length nt of

RNA polymerase second largest subunit (intron 23) and 18S ribosomal subunit, respectively. Plant miRNAs are highly conserved among distantly related plant species, both in terms of primary and mature miRNAs (Zhang et al. 2006b). This finding is also supported by present results where tea miRNAs were found conserved in diverse plant species from monocotyledonous to dicotyledonous plants. These results suggested that different miRNAs might evolve at different rates not only within the same plant species, but also in different ones.

The miRNA target gene identification is an important step for understanding the role of miRNAs in gene regulatory networks. In the present analysis of target genes for tea miRNAs revealed that more than one gene was regulated by individual miRNA. This result was similar to the recent findings in other plant species (Zhang et al. 2006b, Jones-Rhoades and Bartel 2004) which suggested that miRNA research should be focused on networks rather than individual connections between miRNA and strongly predicted targets. miRNAs may directly target transcription factors which affect plant growth and development, and also specific genes which control metabolism (Zhang et al. 2006a). In this study, a total of 37 potential targets were for the 7 identified miRNA families. The identified 37 potential target genes appeared to be related in diverse biological functions. Transcription factors such as MYB, translation intiation factor TIF3, important proteasome degrading pathway enzyme such as ubiquitin conjugating enzyme, different ion transporters such as copper ion binding protein, carbohydrate metabolism related enzyme such as aldose 1-epimerase, glutamate metabolism related enzymes such as glutamate semialdehyde dehydrogenase, important protein for nucleosome assembly such as histone as well as ribosomal proteins. In an earlier report, 20% target genes of transcription factors and 53% target genes of proteins related to diverse physiological processes were found in contrast to the present findings of 8% target genes of transcription factors and 48% target genes of physiological and metabolic processes, however in the previous investigation, total number of miRNAs was limited to only four (Prabu and Mandal 2010). It becomes clear that tea miRNAs targeted both trancription factors as well as specific genes.

Wet lab based validation of the miRNAs, their expression and functions have been remain as a task to be done in future, through which we can dissect and decipher

the miRNA world in tea. Nevertheless, these findings will considerably broaden the scope of miRNA research in this valuable plantation crop. Further, it shows a way for the prediction and analyses of miRNAs to those species whose genome is not available through bioinformatics tools.