

SHORT REPORT

Open Access



# Extrapolative microRNA precursor based SSR mining from tea EST database in respect to agronomic traits

Anjan Hazra<sup>1,2</sup>, Nirjhar Dasgupta<sup>1</sup>, Chandan Sengupta<sup>2</sup> and Sauren Das<sup>1\*</sup>

## Abstract

Tea (*Camellia sinensis*, (L.) Kuntze) is considered as most popular drink across the world and it is widely consumed beverage for its several health-benefit characteristics. These positive traits primarily rely on its regulatory networks of different metabolic pathways. Development of microsatellite markers from the conserved genomic regions are being worthwhile for reviewing the genetic diversity of closely related species or self-pollinated species. Although several SSR markers have been reported, in tea, the trait-specific Simple Sequence Repeat (SSR) markers, leading to be useful in marker assisted breeding technique, are yet to be identified. Micro RNAs are short, non-coding RNA molecules, involved in post transcriptional mode of gene regulation and thus effects on related phenotype. Present study deals with identification of the microsatellite motifs within the reported and predicted miRNA precursors that are effectively followed by designing of primers from SSR flanking regions in order to PCR validation. In addition to the earlier reports, two new miRNAs are predicting here from tea expressed tag sequence database. Furthermore, 18 SSR motifs are found to be in 13 of all 33 predicted miRNAs. Trinucleotide motifs are most abundant among all followed by dinucleotides. Since, miRNA based SSR markers are evidenced to have significant role on genetic fingerprinting study, these outcomes would pave the way in developing novel markers for tagging tea specific agronomic traits as well as substantiating non-conventional breeding program.

**Keywords:** Micro RNA, Simple sequence repeats, Tea quality, Trait specific marker

## Introduction

Tea (*Camellia sinensis*) is most popularly consumed beverage across the world and being a cash-crop, it receives much attention to the scientific community. The vibrant research interest basically stands for its massive demand to health conscious people for its antioxidant and a broad spectrum therapeutic potentiality [1, 2]. China is the largest tea producer as well as exporter preceding by India and Sri Lanka [3]. The fermented tea or black tea is the most common among all different types of tea consumed, although antioxidant and other health benefit properties lies maximum on non-fermented green tea [4]. As a result, demand for quality tea has increased much to

the end users and tea planters. Challenges for exploring superior cultivars with better agronomic traits are still point of interest to the researchers.

Molecular marker assisted technique in breeding programme for the selection or development of cultivars with desired trait from a large population is well established [5]. Among different markers used in crop improvement and molecular breeding technique, microsatellite markers are profoundly used for its reliability and time saving method. Moreover, due to being co-dominant, abundant, hyper-variable and co-operative to high-throughput analysis, microsatellite markers are considered as ideal for plant genetic linkage mapping, physical mapping, population studies, genotype identification and crop improvement [6]. Predominant of such markers like SSR, ISSR, EST-SSR have been effectively utilized in several crop improvement program [7–11]. MicroRNA precursor based SSR markers are very

\*Correspondence: sauren@isical.ac.in

<sup>1</sup> Agricultural and Ecological Research Unit, Indian Statistical Institute, 203, Barrackpore Trunk Road, Kolkata 700 108, India

Full list of author information is available at the end of the article

recently incorporated in this chapter and mostly utilized in marker trait association analysis in several species [12–15]. These micro RNAs are short, non-coding RNA molecules, involved in post transcriptional mode of gene regulation and [16] thus effects on related phenotype [17, 18]. There are a large number of molecular markers available for tea so far [19], however a very few are reported to be linked with some specific trait [20]. Therefore exploration and characterization of novel and already available markers are of prime point of interest. Considering the above, present study aims to identify the microsatellite motifs within the reported and predicted miRNA precursors that are effectively followed by designing of primers for PCR validation.

## Materials and methods

### Retrieval of data, filtering and trimming

Already predicted tea miRNA candidates were fished out from available literature [21–23]. Furthermore, to screen if any pre miRNA were within the tea EST database updated so far (November 2016), standard methodology of miRNA screening [24] were followed with some minor customization. All available reported miRNAs of viridiplantae from online repository miRBase v21.0 [25] and the entire EST collection of tea were retrieved from NCBI dbEST [26], followed by elimination of redundant sequences and trimming polyA tails using PRINSEQ v0.20.4 [27].

### Prediction of miRNAs

The set of published miRNAs were used for a homology search against tea EST collection and the best hits with a minimum length of 18 nucleotides and a maximum miRNA length cover up to 26 nucleotides and not more than 3 mismatches were taken for further analysis. After elimination of protein coding transcripts utilizing BLASTx [28], the remaining candidates were subjected to the prediction of stem-loop structure using Mfold [29] to check possibility of their pre-miRNA existence. The potential miRNA was mined considering the criteria: (a) position of mature miRNA on arm of the hairpin, (b) minimum paired residues in miRNA = 14 and unpaired residues not more than = 5, (c) maximum number of G–U pairs in miRNA = 5, (d) maximum bulge size of 3nt, (e) the negative minimal folding free energy (MFE) is low ( $\leq -18$  kcal/mol) [22], and (f) minimal folding free energy index ( $MFEI = [(MFE/\text{length of the RNA sequence}) * 100]/(G+C)\%$ ) is high ( $>0.85$ ) [30–32].

### MicroRNA target predictions and their function

The exclusively predicted miRNAs were analyzed for their putative target genes employing the psRNATarget server [33] with default parameters. Subsequently, to

recognize the functions of such predicted targets, they were undergone BLAST programme in NCBI. Finally a complete list was prepared taking all previously and presently reported tea miRNAs with their putative function.

### Exploration of SSRs within predicted microRNAs

The simple sequence repeat motifs within all available and reported pre-miRNA sequences were investigated by the Websat online program [34]. The parameters were set for identifying perfect di-, tri-, tetra-, penta-, and hexanucleotide motifs with minimum repeat numbers of 6, 4, 3, 3 and 3 respectively.

### Designing primers from SSR flanking region

The primer pairs from SSR flanking regions were designed with BatchPrimer3 server [35]. For the same, parameters were set as follows: length range = 18–23 nucleotides with 21 as optimum; PCR product size range = 100–400 bp [36]; optimum annealing temperature = 55 °C; and GC content 40–60%, with 50% as optimum.

## Result and discussion

In the present study, previously reported miRNA sequences were utilized to find their homolog ones from tea as it is already known that plant mature miRNAs are highly conserved within the plant kingdom, and miRNA genes in one species may exist as orthologs or homologs in other species [30, 37]. With the help of this hypothesis, known miRNAs were utilized to discover novel potential miRNAs in tea. All 8442 miRNAs reported from viridiplantae so far were utilized and after elimination of 3676 exact duplicates using bioinformatics tool, a non-redundant collection was taken for further analysis. Similarly, tea EST collection of 49,670 sequences were made into a non-redundant 40,686 numbers. The BLAST search could fish out 52 number of ESTs with required level of homology i.e. minimum 18nt length similarity with not more than 3 mismatches. A total 15 of them already reported earlier as pre-miRNA [21–23]. Leftover candidates were analyzed through Mfold program as the miRNA precursors should be able to form stem-loop hairpin in their secondary structure for processing by Dicer enzyme [38] and subsequently possible false miRNA precursors were manually removed. In present study two more potential pre-miRNA (Fig. 1) were identified with accession JK478587.1 and FS955851.1 for having miR1533 and miR8002-3p respectively.

The identified pre-miRNAs belonging to two families had sequence lengths of 419 and 787 bp respectively (Table 1). Length variation was also evident in previous report [22, 39, 40]. Beside, commonly studied parameter in miRNA prediction is the minimum free energy (MFE)



**Table 1 List of predicted microRNAs and occurrence of SSR motifs with designed primer**

Predicted miRNA	Precursor id	MFE (-kCal)	References	SSR start	SSR end	SSR motif	SSR length	Primer F	Primer R	Product size
miR164	CV013669	20	[21]					GGAGTGAAGGAGGGAAGA	CCGAGAGAATAAAGGAGGA	205
miR169	GE650220	7.8	[21]	81	93	CGC	3			
miR1846	DN976181	40.5	[21]					AGTGATTAATGGTGGTGTC	TCAACCAATCAACAAGTC	152
miR1863	GH623864	19.2	[21]	224	236	GAG	3	CTGTACTGCAGCTTAACCAA	AAATATGCTGCTCATTTCAAC	152
miR 408	GD254786.1	120.1	[22]	519	541	TA	2	AGAAGATGCTCAGGGAAGAG	ACACAAGTATGTCACCCAGCTC	177
csi-miR1171	FE943069.1	45.97	[22]	142	154	GGC	3	GAACCTTTCCTCCAGAATTTA	TCACATTTAGCTTTTCACTCC	162
csi-miR414a	GD254734.1	63.18	[22]	280	295	GTGGA	5			
csi-miR414d	GW342817.1	37.72	[22]	170	188	GATGAC	6	ACGATGGTGGTTATGAATATG	GGTTTTTGTAAAGTTGTCA	139
csi-miR414f	GE651542.1	18.5	[22]	99	111	TTC	3	AGACAAAAACCAAGGCTAGAT	CATCTGTGCAGATCTCAGTT	149
cas-miR1122	EU849076.1	91.23	[22]	121	139	ATC	3			
csi-miR414 g	CV013826.1	25.2	[22]	504	519	CTC	3	TGATGATGAGGAGGAGATAA	TTGCTTTAGTGAACAACACTCC	136
csi-miRf10132-akr	CV014169.1	69.5	[22]							
cja-miR2910	U42815.1	-91	[22]							
csi-miR2914	AB120309.1	20.9	[22]							
cas-miRf10185-akr	GH623933.1	51.1	[22]							
cas-miR11590-akr	GE651674.1	23.8	[22]							
csi-miR414 h	GW863581.1	86.83	[22]							
miR156	HS396956.1	45.8	[23]	113	125	AAC	3	CGTTAGGCTATTTTGTTCAA	TTCTGCAATCATCCAATTC	159
miR171 a	F5948108.1	39.2	[23]	115	127	TC	2	TACTTCCAACCAACACAAAGT	TAGCTTACCACCTCAATCAAA	168
miR171 b	F5948109.1	40.4	[23]							
miR397	CV699725.1	39.2	[23]							
miR399	F5958856.1	52.8	[23]							
miR2863	F5950435.1	21.5	[23]	57	69	TCTA	4	CTCCTGTACACTCTCTCTCTCC	GATGAACAGCATAGGTATCCA	127
miR2911a	JK476023.1	65.4	[23]							
miR2911b	F5953337.1	69.2	[23]							
miR5021a	GW690847.1	71.9	[23]	72	84	GGA	3	AGACACAGGCAGACATAGAGA	AAGATGCGATGAGATCAGATA	165
miR5021b	GE651759.1	72.2	[23]	239	254	TTC	3	AGATACACATTTGGGAAGAGG	TAGAACTTGCAGAGAGAACC	149
miR5368a	GE653011.1	76.1	[23]	252	264	TC	2			
miR5368b	F5945766.1	68.5	[23]	60	78	GA	2	TAGAACTTGCAGAGAGAATCG	TACACATTTGGGAAGAAGA	152
miR6483a	HS398296.1	29.8	[23]	75	90	AGA	3			
miR6483b	JK714410.1	30	[23]							
miR1533	JK478587.1	89.99	Reported	403	419	AC	2			
miR8002-3p	F5955851.1	155.67	Reported							

genes encoded transcriptional factors, involved in stress response, transmembrane transport, and signal transduction and transcription regulation. Target genes encoding transcription factors and cell integrity maintenance machinery during stress response was also mentioned by Prabu and Mandal [21]. Multiple target of a single miRNA is the system biology network, when they control expression of different transcription factors which in turn regulates specific genes for different metabolism [48].

Microsatellite markers are widely used tool for estimating the genetic variation and especially used for construction linkage map, understanding of marker trait association, identification of disease resistant loci [14, 49]. The distribution of simple sequence repeats or SSRs in the genome is inherently unstable and therefore highly polymorphic [50]. People assumed that increment of the repeat unit and repeat tracts gives rise the chances of the mutation rate [51]. Some reports have already established the fact that SSR expansions or contractions within genome sequences can affect functions of these sequences and even lead to phenotypic changes [17]. Evidences have shown the effect of SSR unit variation within protein coding regions. However, the consequences of the same in non-coding transcripts are less studied. A very few reports demonstrated the significance of analysis of SSRs in non-coding miRNA [12–15, 49]. In current study, 13 of 33 total predicted pre-miRNAs had one or more SSR motifs (Table 1). A total of 9 sequences had SSR motif in a single region, whereas 3 sequences contained SSR motifs in two locations and 1 sequence was with 3 different SSR motifs. Trinucleotide motifs were most abundant among all followed by dinucleotides. There was one each of penta and hexa-nucleotide motifs. Forward and reverse primers from the each microsatellite motifs flanking region could be generated in all members excluding only one where SSR motif present toward the terminal end. The microsatellite motifs may be conserved among group or become signature [13] which might be used in studies of genetic fingerprinting work of tea. Some traits rely on specific repeats of microsatellites and their numbers. Such advantage has been efficiently employed by rice researchers when miRNA-SSR markers were employed to differentiate the salt tolerant and susceptible genotypes [15]. They found more repeat variation of the salt responsive miRNA genes among the susceptible rice genotypes than tolerant one. Such extensive work can be applied in tea to distinguish the cultivars with varying agronomic traits on the basis of miRNA-SSR polymorphism.

Finally, newly predicted microRNAs in tea would enrich the assemblage in absence of whole genomic information of tea and make easier subsequent studies for experimental validation. Some of them might be related

to certain metabolic functions thereby phenotypes as well. Excavating of microsatellite motifs from predicted microRNA precursors and designing primers from SSR flanking regions would pave the way in developing novel markers for tagging tea specific agronomic traits as well as accelerating non-conventional breeding program. This can freely be followed by genetic diversity assessment of tea cultivars with varying characters.

#### Authors' contributions

AH and SD have conceived the work, designed methodology, interpreted data and written the manuscript, NDG and CS participated in data interpretation and manuscript writing. All authors read and approved the final manuscript.

#### Author details

<sup>1</sup> Agricultural and Ecological Research Unit, Indian Statistical Institute, 203, Barrackpore Trunk Road, Kolkata 700 108, India. <sup>2</sup> Department of Botany, University of Kalyani, Nadia, Kalyani 741235, India.

#### Acknowledgements

AH is thankful to National Tea Research Foundation, Tea Board, India for providing research fellowship.

#### Competing interests

The authors declare that they have no competing interests.

#### Availability of data and materials

Data retrieved from public database, not deposited any.

#### Funding

National Tea Research Foundation, Tea Board, India.

#### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 31 January 2017 Accepted: 28 June 2017

Published online: 06 July 2017

#### References

- Hayat K, Iqbal H, Malik U, Bilal U, Mushtaq S. Tea and its consumption: benefits and risks. *Crit Rev Food Sci Nutr*. 2015;55(7):939–54.
- Khan N, Mukhtar H. Tea and health: studies in humans. *Curr Pharm Des*. 2013;19(34):6141–7.
- Basu Majumder A, Bera B, Rajan A. Tea statistics: global scenario. *Inc J Tea Sci*. 2010;8(1):121–4.
- Carloni P, Tiano L, Padella L, Bacchetti T, Customo C, Kay A, et al. Antioxidant activity of white, green and black tea obtained from the same tea cultivar. *Food Res Int*. 2013;53(2):900–8.
- Bang H, Kim S, Leskovaar D, King S. Development of a codominant CAPS marker for allelic selection between canary yellow and red watermelon based on SNP in lycopene  $\beta$ -cyclase (LCYB) gene. *Mol Breed*. 2007;20(1):63–72.
- Morgante M, Olivieri A. PCR-amplified microsatellites as markers in plant genetics. *Plant J*. 1993;3(1):175–82.
- Roychowdhury R, Taoutaou A, Hakeem KR, Gawwad MRA, Tah J. Molecular marker-assisted technologies for crop improvement. In: Roychowdhury R, ed. *Crop improvement in the era of climate change*; 2013: p. 241–58.
- Kumar S, Rajendran K, Kumar J, Hamwieh A, Baum M. Current knowledge in lentil genomics and its application for crop improvement. In: Kumar S, editor. *Crop breeding: bioinformatics and preparing for climate change*. USA: CRC Press; 2016. p. 309–27.
- Varshney RK, Graner A, Sorrells ME. Genic microsatellite markers in plants: features and applications. *Trends Biotechnol*. 2005;23(1):48–55.

10. Singh RB, Srivastava S, Rastogi J, Gupta GN, Tiwari NN, Singh B, et al. Molecular markers exploited in crop improvement practices. *Res Environ Life Sci*. 2014;7(4):223–32.
11. Kesawat MS, Kumar BD. Molecular markers: its application in crop improvement. *J Crop Sci Biotechnol*. 2009;12(4):169–81.
12. Wang X, Gui S, Pan L, Hu J, Ding Y. Development and characterization of polymorphic microRNA-based microsatellite markers in *Nelumbo nucifera* (Nelumbonaceae). *Appl Plant Sci*. 2016;4(1):1500091.
13. Nithin C, Patwa N, Thomas A, Bahadur RP, Basak J. Computational prediction of miRNAs and their targets in *Phaseolus vulgaris* using simple sequence repeat signatures. *BMC Plant Biol*. 2015;15(1):140.
14. Ganie SA, Mondal TK. Genome-wide development of novel miRNA-based microsatellite markers of rice (*Oryza sativa*) for genotyping applications. *Mol Breed*. 2015;35(1):51.
15. Mondal TK, Ganie SA. Identification and characterization of salt responsive miRNA-SSR markers in rice (*Oryza sativa*). *Gene*. 2014;535(2):204–9.
16. Großhans H, Filipowicz W. Molecular biology: the expanding world of small RNAs. *Nature*. 2008;451(7177):414–6.
17. Fondon JW, Garner HR. Molecular origins of rapid and continuous morphological evolution. *Proc Natl Acad Sci*. 2004;101(52):18058–63.
18. Kashi Y, King DG. Simple sequence repeats as advantageous mutators in evolution. *Trends Genet*. 2006;22(5):253–9.
19. Mukhopadhyay M, Mondal TK, Chand PK. Biotechnological advances in tea (*Camellia sinensis* [L.] O. Kuntze): a review. *Plant Cell Rep*. 2016;35(2):255–87.
20. Elangbam M, Misra A. Development of CAPS markers to identify Indian tea (*Camellia sinensis*) clones with high catechin content. *Genet Mol Res*. 2016;15(2):1–13.
21. Prabu G, Mandal A. Computational identification of miRNAs and their target genes from expressed sequence tags of tea (*Camellia sinensis*). *Genom Proteom Bioinform*. 2010;8(2):113–21.
22. Das A, Mondal TK. Computational identification of conserved microRNAs and their targets in tea (*Camellia sinensis*). *Am J Plant Sci*. 2010;1(02):77.
23. Q-w Zhu, Y-p Luo. Identification of miRNAs and their targets in tea (*Camellia sinensis*). *J Zhejiang Univ Sci B*. 2013;14(10):916–23.
24. Zhang B, Pan X, Anderson TA. Identification of 188 conserved maize microRNAs and their targets. *FEBS Lett*. 2006;580(15):3753–62.
25. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. miRBase: tools for microRNA genomics. *Nucleic Acids Res*. 2008;36(suppl 1):D154–8.
26. Boguski MS, Lowe TM, Tolstoshev CM. dbEST—database for ‘expressed sequence tags’. *Nat Genet*. 1993;4(4):332–3.
27. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics*. 2011;27(6):863–4.
28. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403–10.
29. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*. 2003;31(13):3406–15.
30. Zhang B, Pan X, Cannon CH, Cobb GP, Anderson TA. Conservation and divergence of plant microRNA genes. *Plant J*. 2006;46(2):243–59.
31. Li X, Hou Y, Zhang L, Zhang W, Quan C, Cui Y, et al. Computational identification of conserved microRNAs and their targets from expression sequence tags of blueberry (*Vaccinium corybosum*). *Plant Signal Behav*. 2014;9(9):e29462.
32. Zhang B, Pan X, Cox S, Cobb G, Anderson T. Evidence that miRNAs are different from other RNAs. *Cell Mol Life Sci*. 2006;63(2):246–54.
33. Dai X, Zhao PX. psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res*. 2011;39(suppl 2):W155–9.
34. Martins WS, Lucas DCS, Neves Kds, Bertoli DJ. WebSat—a web software for microsatellite marker development. *Bioinformatics*. 2009;3(6):282–3.
35. You FM, Huo N, Gu YQ, M-c Luo, Ma Y, Hane D, et al. BatchPrimer3: a high throughput web application for PCR and sequencing primer design. *BMC Bioinform*. 2008;9(1):253.
36. Yu Y, Yuan D, Liang S, Li X, Wang X, Lin Z, et al. Genome structure of cotton revealed by a genome-wide SSR genetic map constructed from a BC 1 population between *Gossypium hirsutum* and *G. barbadense*. *BMC Genom*. 2011;12(1):15.
37. Weber MJ. New human and mouse microRNA genes found by homology search. *FEBS J*. 2005;272(1):59–73.
38. Kurihara Y, Watanabe Y. Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. *Proc Natl Acad Sci USA*. 2004;101(34):12753–8.
39. Biswas S, Hazra S, Chattopadhyay S. Identification of conserved miRNAs and their putative target genes in *Podophyllum hexandrum* (Himalayan Mayapple). *Plant Gene*. 2016;6:82–9.
40. Patanun O, Lertpanyasampath M, Sojikul P, Viboonjun U, Narangajavana J. Computational identification of microRNAs and their targets in cassava (*Manihot esculenta* Crantz.). *Mol Biotechnol*. 2013;53(3):257–69.
41. Bonnet E, Wuyts J, Rouzé P, Van de Peer Y. Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences. *Bioinformatics*. 2004;20(17):2911–7.
42. Zhang BH, Pan XP, Wang QL, George PC, Anderson TA. Identification and characterization of new plant microRNAs using EST analysis. *Cell Res*. 2005;15(5):336–60.
43. Pandey B, Gupta OP, Pandey DM, Sharma I, Sharma P. Identification of new stress-induced microRNA and their targets in wheat using computational approach. *Plant Signal Behav*. 2013;8(5):e23932.
44. Mallory AC, Vaucheret H. Functions of microRNAs and related small RNAs in plants. *Nat Genet*. 2006;38:S31–6.
45. Wang X-J, Reyes JL, Chua N-H, Gaasterland T. Prediction and identification of *Arabidopsis thaliana* microRNAs and their mRNA targets. *Genome Biol*. 2004;5(9):R65.
46. Chen X. Small RNAs and their roles in plant development. *Annu Rev Cell Develop*. 2009;25:21–44.
47. Chen R, Hu Z, Zhang H. Identification of microRNAs in wild soybean (*Glycine soja*). *J Integr Plant Biol*. 2009;51(12):1071–9.
48. Zhang B, Pan X, Cobb GP, Anderson TA. Plant microRNA: a small regulatory molecule with big impact. *Develop Biol*. 2006;289(1):3–16.
49. Chen M, Tan Z, Zeng G, Peng J. Comprehensive analysis of simple sequence repeats in pre-miRNAs. *Mol Biol Evol*. 2010;27(10):2227–32.
50. Heesacker A, Kishore VK, Gao W, Tang S, Kolkman JM, Gingle A, et al. SSRs and INDELS mined from the sunflower EST database: abundance, polymorphisms, and cross-taxa utility. *Theor Appl Genet*. 2008;117(7):1021–9.
51. Katti MV, Ranjekar PK, Gupta VS. Differential distribution of simple sequence repeats in eukaryotic genome sequences. *Mol Biol Evol*. 2001;18(7):1161–7.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

