

SCIENTIFIC REPORTS



OPEN

The sRNAome mining revealed existence of unique signature small RNAs derived from 5.8SrRNA from *Piper nigrum* and other plant lineages

Srinivasan Asha & E. V. Soniya

Small RNAs derived from ribosomal RNAs (srRNAs) are rarely explored in the high-throughput data of plant systems. Here, we analyzed srRNAs from the deep-sequenced small RNA libraries of *Piper nigrum*, a unique magnoliid plant. The 5' end of the putative long form of 5.8S rRNA (5.8S_LrRNA) was identified as the site for biogenesis of highly abundant srRNAs that are unique among the *Piperaceae* family of plants. A subsequent comparative analysis of the ninety-seven sRNAomes of diverse plants successfully uncovered the abundant existence and precise cleavage of unique rRF signature small RNAs upstream of a novel 5' consensus sequence of the 5.8S rRNA. The major cleavage process mapped identically among the different tissues of the same plant. The differential expression and cleavage of 5'5.8S srRNAs in *Phytophthora capsici* infected *P. nigrum* tissues indicated the critical biological functions of these srRNAs during stress response. The non-canonical short hairpin precursor structure, the association with Argonaute proteins, and the potential targets of 5'5.8S srRNAs reinforced their regulatory role in the RNAi pathway in plants. In addition, this novel lineage specific small RNAs may have tremendous biological potential in the taxonomic profiling of plants.

The advent of next-generation sequencing technologies has led to the discovery of functional small RNAs such as microRNAs and siRNAs from diverse organisms. In addition, the novel class of regulatory sRNAs originating from non-coding RNAs, including small nucleolar RNAs (snoRNAs)¹ and transfer RNAs (tRNAs)², were recently reported. Ribosomal RNAs (rRNA) are essential housekeeping genes, with integral roles in protein synthesis in all domains of life. The high copy number of ribosomal DNA (rDNA) genes in the genome leads to an abundance of rRNA transcripts that accounts for ~60% of the total RNA in the cell. The gene regulatory role of rRNA is under exploration. The qiRNAs (QDE-2-interacting small RNAs) generated from the rDNA locus have been reported to be induced during DNA damage in the filamentous fungi, *Neurospora crassa*³. The expression of specific ribosomal RNA derived fragments (rRFs) of unknown origin and function were reported from the rice embryogenic callus⁴. A novel set of rRNA derived chloroplast sRNAs were reported to be induced during heat response in the Chinese cabbage (*Brassica rapa* ssp. *chinensis*)⁵. Moreover, a deeply conserved non-canonical microRNA was recently identified from the rDNA internal transcribed spacer1 (ITS1) of the dipteran species⁶.

Multiple copies of ribosomal RNA are encoded in the genome, and the evolution of the rRNA copy number is related to cell size. In eukaryotes, the small (40S) ribosomal subunits comprise an assembly of 18SrRNA and 33 ribosomal proteins (RPs), whereas the large ribosomal units contain discrete rRNA components (5S, 5.8S and 25/28S rRNAs) and 46 ribosomal proteins. Except for 5S rRNA, all other rRNAs (18S, 5.8S, 25–28S rRNAs) are transcribed by RNA polymerase I (Pol I) as a single polycistronic unit; this unit undergoes complex post transcriptional processing into two external transcribed spacers (5'ETS and 3'ETS) and two internal transcribed spacers (ITS1 and ITS2), which flank and separate the mature rRNA components⁷. This precursor rRNA processing occurs in the nucleolus and nucleus by multiple endonucleolytic and exonucleolytic cleavage steps to produce mature rRNAs (Fig. 1). Maturation of 5.8SrRNA is one of the most complicated forms of pre-ribosomal

Plant Molecular Biology, Rajiv Gandhi Centre for Biotechnology, Thiruvananthapuram, Kerala, 695014, India. Correspondence and requests for materials should be addressed to E.V.S. (email: evsoniya@rgcb.res.in)

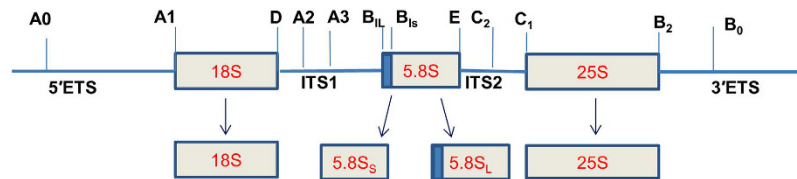


Figure 1. The cleavage points during the processing of 5.8S rRNA from pre-ribosomal RNA.

Library	Pn_CL	Pn_IL	Pn_IR
Total reads	18,629,003	12,621,368	12,948,539
rRNA mapped reads (srRNAs)	750,269	1,288,184	1,401,971
Total Unique srRNAs	50,926	60,277	122,031
Unique srRNAs (≤ 10 reads)	7,788	9,557	13,149

Table 1. Summary of ribosomal RNA derived small RNAs in the *P. nigrum* sRNA libraries of black pepper.

RNA processing, and two alternate pathways give rise to the short (5.8S_s) and long forms (5.8S_L) of 5.8SrRNA that differ by the presence of additional 7 or 8 nucleotides at the 5' end^{7,8}. The occurrence of two forms of 5.8S rRNA with slightly different 5' ends is common in eukaryotes, even though the ratio between the two forms vary from one organism to another. The shorter form of 5.8S rRNA (5.8S_s) represents 80% of the total⁹ and is generated by endonucleolytic cleavage at the A3 site of the internal transcribed spacer 1 (ITS1) by RNase MRP and subsequent 5' → 3' exonucleolytic trimming of the 27S-A3 pre-rRNA down to site B_{1S} by Rat1 and XRN1^{10,11}. The direct endonucleolytic cleavage gives rise to the longer 5.8S_L in eukaryotes and does not rely on the cleavage at A3 site¹². The 3' end of both variants is generated by cleavage at site C2 in the ITS2 by cofactors of the nuclear exosome, Rrp6¹¹. In plants, the exact mechanism by which the long and short forms of 5.8S rRNA are produced is not fully known. Only Rat1/XRN2-like members have been reported in plants, and three XRN exoribonuclease homologs, XRN2, XRN3 and XRN4, were identified in *Arabidopsis*¹³. Among these, XRN2, XRN3 and XRN4 act as endogenous PTGS suppressors; XRN2 and XRN3 are nuclear exoribonucleases, while XRN4 localizes to the cytoplasm¹⁴. In *Arabidopsis*, it has been shown that AtXRN2, similar to yeast nuclear Rat1, is responsible for cleavage at the sites A2 and A3 in ITS1¹⁵. In several model systems, RNase MRP (Mitochondrial RNA Processing) has also been reported to participate in the cleavage of the internal transcribed spacer 1 (ITS1) of the rRNA precursor at the specific site A3, producing the mature 5' end of the 5.8S rRNA¹⁶. Land plants do not encode any recognizable genes for RNase P enzymes or proteins specific for ribonucleoprotein RNase P¹⁷. The protein-only RNase P enzymes (PRORPs), which are present in the organelles and the nucleus, have entirely replaced the ribonucleoproteins to exert RNase P functions in plants¹⁸. AtMTR4, an exosome cofactor in the nucleolus, cooperates in several steps of 5.8S rRNA maturation and the removal of rRNA maturation by-products¹⁹. HEN 2, a plant specific RNA helicase that is absent in other eukaryotic lineages, targets different subsets of nuclear RNA substrates for degradation by the exosome²⁰.

Our study demonstrated the existence of highly expressed, unique small RNAs generated from the 5' terminus of 5.8SrRNA in black pepper, a unique basal angiosperm plant, as well as in other representative plant species ranging from green algae to eudicots. Until now, very little has been known about functional small RNAs from the 5.8S ribosomal RNAs in plants. Such a lineage specific small RNAs can be a novel candidate for use in molecular systematics of plants.

Results

Identification and analysis of unique ribosomal small RNAs from black pepper. The high throughput small RNA sequence data of control leaf (Pn_CL), *P. capsici* infected leaf (Pn_IL) and infected root (Pn_IR) of *Piper nigrum* were used for the analysis of sRNAs originating from ribosomal RNAs. Among the Pn_CL, Pn_IL and Pn_IR sRNA libraries of *Piper nigrum*, 750,269 (4.0%), 1,288,184 (10.2%) and 1,401,971 (10.8%) reads, respectively, were mapped to rRNA (Table 1). The mapped unique sRNA candidates were distributed throughout the rRNA transcript, and the small RNA reads that mapped to internal transcribed spacers (ITS1 and ITS2) were negligible (Fig. 2A). The abundance analysis of srRNAs (≥ 100 RPM) revealed that most of the prominent srRNAs clustered around the 5.8SrRNA region (Fig. 2B). Among the total srRNAs of all the three libraries, an average of 40.94 percent of the sequences could be categorized as 5.8S rRNA-derived small RNAs (Fig. 2C), of which the largest subset were sRNAs of 23 nt in length (Fig. 2D). The size distribution of srRNAs showed that 21 nt was the most common length of sRNAs in the Pn_CL and Pn_IR sRNA libraries, whereas the largest percentage of reads in the Pn_IL library were 23 nt long (Supplementary Figure 1). Because the sRNAome might also contain a large portion of rRNA degradation products, it was difficult to identify the functional srRNA from the large dataset of small RNAs. Generally, rRNA reads were neglected in the sRNA characterization and were assumed to be degradation products. The RIN values > 8.0 and 28S:18S ratio > 1.5 of the samples indicated a high degree of RNA integrity, thus excluding the possibility of extensive degradation (Supplementary Figure 2). The small RNAs that mapped to the 5' terminus of 5.8SrRNA were found to be extremely abundant (Fig. 3A), and many of the variants contained 2–3 upstream nucleotides (Table 2). The proportion of small RNAs from the

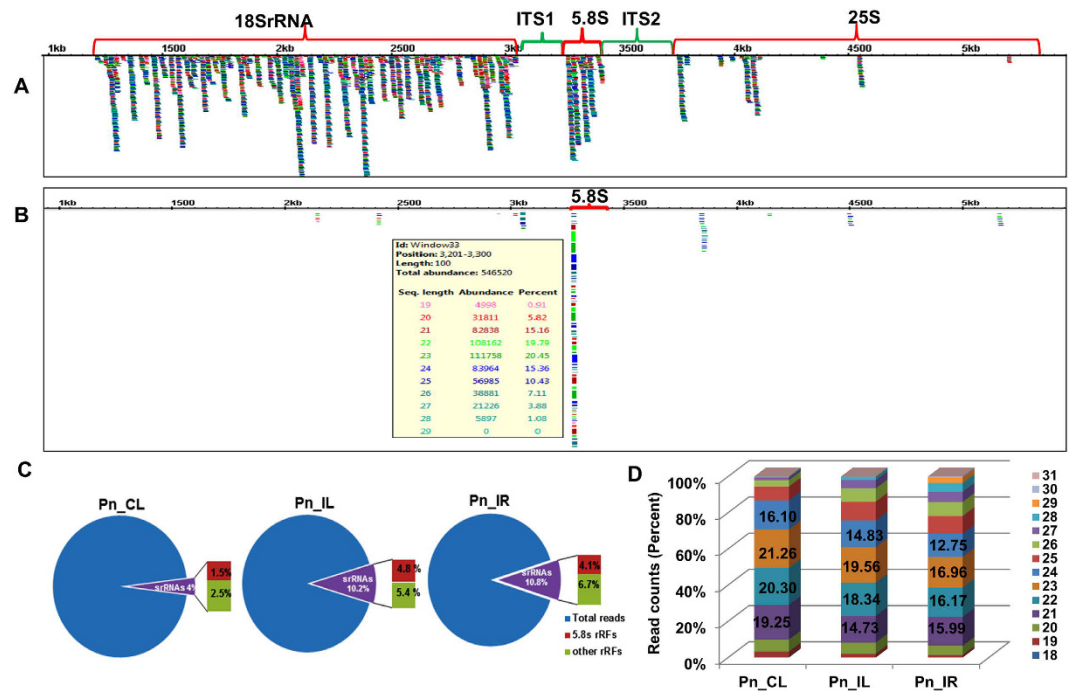


Figure 2. Distribution and abundance pattern of ribosomal RNA derived small RNAs in *Piper nigrum* small RNA libraries (A) Distribution of unique srRNA candidates against black pepper rRNA. (B) Abundant srRNA from Pn_CL (upto 100 rpm) mapped against rRNA. Most of the abundant reads were mapped at the 5.8S rRNA region. (C) The proportion of 5.8S rRFs in the Pn_CL, Pn_IL and Pn_IR libraries. (D) The length categorisation of 5.8S rRFs among the three libraries.

internal and 3' terminal regions of 5.8SrRNA was much lower than the proportion of sRNAs from the 5' end (Fig. 3A). 5.8S rRNA with few upstream nucleotides may potentially indicate the presence of 5.8S₁rRNA. Notably, the 5' sequences of 5.8S rRNA near ITS1 are highly conserved among the different species of Piperaceae family (Fig. 3B). Also, a major share (>99%) of srRNAs were mapped at the 5' end of the sense strand of 5.8S rRNA (Fig. 3C). The positional distribution and high abundance of 5.8S rRFs indicated that sRNAs are generated precisely from the 5.8S rRNA.

Apart from the three conserved motifs previously described in viridiplants (Fig. 4), GACUCUCGGCAACGGGAUAUCUC was observed as the novel 5' consensus sequence present in all the seed plants (Spermatophytes). The 5'5.8S rRNA of different *Piper* species possessed a unique seven nucleotide sequence, UCAGUAC, upstream of the 5.8S rRNA consensus sequence GACUCUCGGCAACGGGAUAUCUC. The different 5' length variants of the 5.8S srRNAs from the *P. nigrum* library were represented as sequence logo (Fig. 5A). An analysis of the 5' terminal nucleotides showed that cytosine was the terminal nucleotide in 25–28% of the 5'5.8S rRF sequences in the Pn_IL and Pn_IR sRNA libraries, indicating major cleavage at C (U|CAGUACG). In contrast, major cleavage (23.35%) was observed at the U position in the Pn_CL library (U|CAGUACG) (Fig. 5B). The length distribution pattern of 5'5.8S rRFs with the C terminal nucleotide showed that the major distribution of the small RNA variants ranged from 21 to 25 nt in size (Fig. 5C).

Signature small RNAs in diverse plant lineages. The signature srRNAs were detected in different lineages of plants by analyzing their high throughput small RNA sequence datasets (Supplementary Figure 3A–E). A close examination of the 5' end of 5.8SrRNA revealed very distinct sequence diversity among different plant lineages (Fig. 6A), and the most abundant rRFs from the small RNA libraries mapped to 5.8S rRNA, especially at the 5' end (Fig. 6B). Furthermore, these small RNAs contained the conserved 5' motif. Therefore, it is possible that the cleavage at the 5' end of the 5.8S rRNA is a conserved mechanism of small RNA biogenesis in plants. An analysis of the small RNAome confirmed that the precise cleavage of unique rRF signature small RNAs were prevalent across diverse plant lineages. The monotypic genus *Amborella* possessed a single nucleotide polymorphism at the consensus motif, GACUCUCGCAACGGGAUAUC, whereas in the Pteridophyte *Marselia quadrifolia*, a G to A transition in the 8th position of the consensus sequence GACUCUCAGCAACGGGAUAU was observed (Table 3, Fig. 6A). Among the monocot group, the consensus sequence of CACGACUCUCGGCAACGGGAUAU was observed in all the plants belonging to the order Poales except *Miscanthus sp.*, which showed a C to U transition at the third nucleotide position. Within the Poacea family, the closely related plants *Oryza sativa* (Sub family Ehrhartoideae), *Hordeum vulgare* (Pooideae), *Triticum aestivum* (Pooideae), and *Setaria italica* (panicoideae) shared the distinct 5' sequence of CACAC. In *Zea mays*, a transversion mutation resulted in CGACAC at the 5', and in *Panicum virgatum*, AACAC was the 5' terminal nucleotide sequence. *Musa accuminata*, a monocot plant belonging to the Zingiberales order, exhibited a very distinct 5' nucleotide sequence, UUGGAU, which was

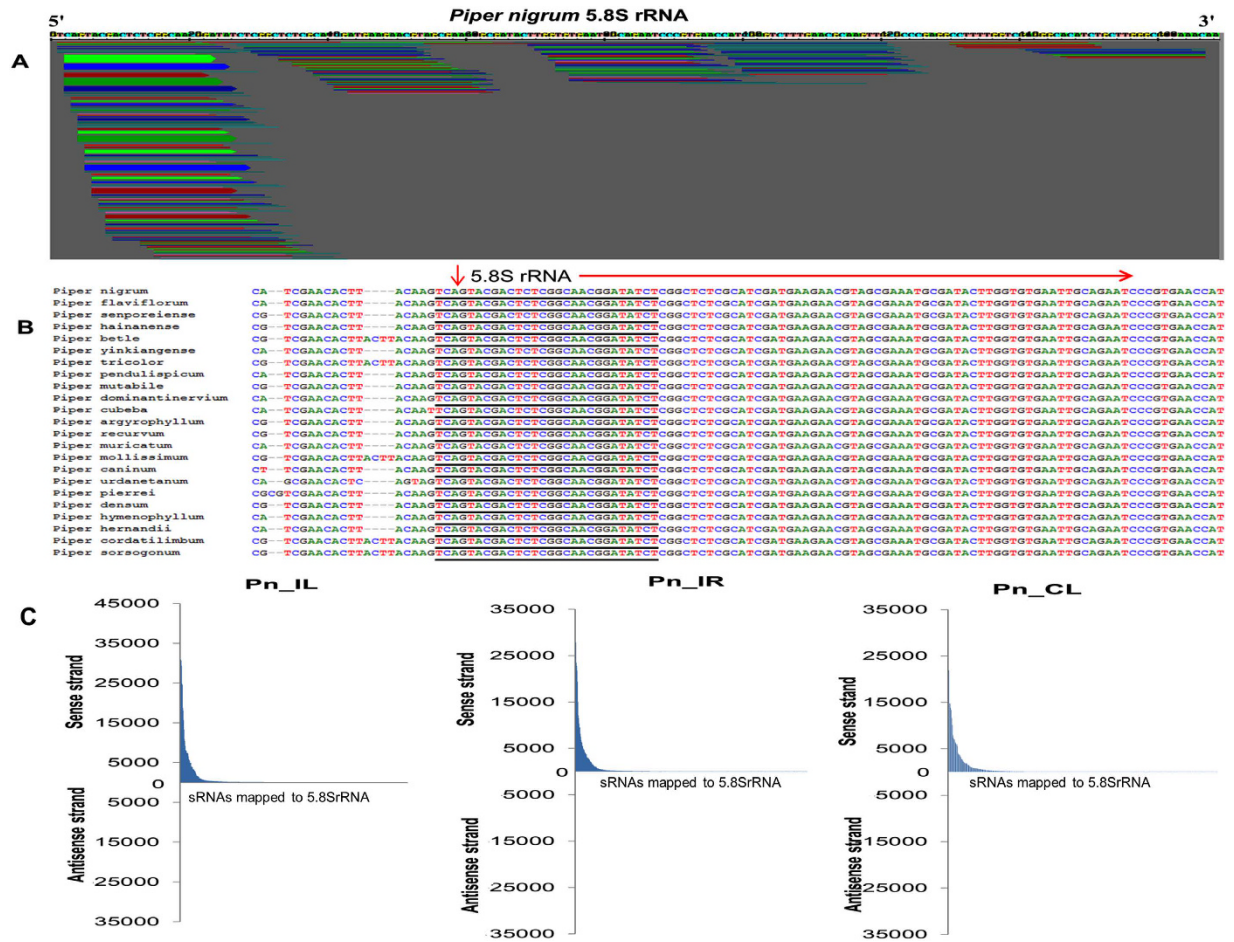


Figure 3. Abundance analysis of 5.8S rRFs. (A) The 5' 5.8S rRFs were the most abundant category of srRNAs in *Piper nigrum*. srRNAs with read numbers > 10 were analyzed. (B) The 5' conserved sequences from different species of *Piper*. (C) The mapped srRNAs to 5.8S rRNA.

terminal to the consensus sequence GACUCUCGGCAACGGAUUAU in the rRFs. Even though the rRF consensus sequence CACGACUCUCGGCAACGGAUUAU observed in the Poacea plants was present in *Cycas rumphii* and *Ginkgo biloba*, they both possessed distinct 5' terminal nucleotides in the 5.8S rRFs, such as AUGCAC and CCUCAC. AACGACUCUCGGCAACGGAUUAU was identified as the consensus sequence in all the dicot plants studied except the fabid plants *Medicago truncatula* and *Glycine max*, in which a C to U transition was observed in the 3rd nucleotide base. However, each plant species possessed a unique patch of 5.8S rRNA; these signatures were present as abundant 5.8SrRNA in the sRNAome. The three plant species studied from the magnoliid clade had the consensus sequence of ACGACUCUCGGCAACGGAUUAUCU, and the plants belonging to each family possessed distinct 5' sequences in the unique 5.8SrRFs, such as CAGU in *Piper nigrum* (Piperaceae), UAAC in *Aristolochia fimbriata* (Aristolochiaceae), and UUAG in *Persea americana* (Lauraceae). Even though a similar consensus sequence was identified in the closely related basal angiosperm plant *Nuphar advena*, it possessed the well-defined 5' sequence CAAA. Despite the nucleotide variation present in the 5' consensus sequence, *Amborella sp.* had AAAC nucleotides upstream of the consensus sequence.

Small RNA libraries of lower plants also possessed the rRF signature small RNAs with similar processing and cleavage events, but with lower abundances. We identified unique 5.8SrRF signature sequences from the *Chara sp.*, a green algae species among the most closely related to land plants. The distinct nucleotide variations of the 5.8SrRF signatures and their low expression levels compared with the land plants were detected in *Chara sp.* Meanwhile, signature small RNAs were also detected in green algae such as *Chlamydomonas reinhardtii* and *Volvox carteri*, which represent the oldest ancestors of plants, but the read counts were below detectable levels (RPM < 10).

Analysis of length distribution of the 5' 5.8S rRFs and the 5' cleavage patterns. To determine whether any unique cleavage and length patterns were present in each species, we analyzed the sRNA datasets from different tissues. An analysis of 5.8S srRNAs from different sRNA libraries of the same plant showed similar cleavage patterns in all the plant tissues (Supplementary Figure 4A–E). In a majority of the plant species, the exact 5' nucleotide sequences of 5.8S rRNA were not clear from the database sequences. Therefore, the 5' cleavage patterns of the rRF sequences were studied from the GACUCUC consensus sequence. The major cleavage was

Sequence	Length	Read count in sRNA library		
		Pn_CL	Pn_IL	Pn_IR
CAGUACGACUCUCGGCAACGGGA	22	13752	38322	25775
CAGUACGACUCUCGGCAACGGGAU	23	13380	33983	23518
ACGACUCUCGGCAACGGGAUUCU	23	22983	30890	21796
GUACGACUCUCGGCAACGGGAU	23	10125	30640	27894
CAGUACGACUCUCGGCAACGGGAU	24	7075	29330	22808
UACGACUCUCGGCAACGGGAUUCU	24	21900	27486	19404
CAGUACGACUCUCGGCAACGGGAU	25	4069	24862	19808
ACGACUCUCGGCAACGGGAU	21	11698	24572	30275
UACGACUCUCGGCAACGGGAU	22	14729	18815	14578
GACUCUCGGCAACGGGAUUCU	21	7915	18098	11482
CAGUACGACUCUCGGCAACGGG	21	14807	16489	10380
GUACGACUCUCGGCAACGGGAU	22	6009	15753	15029
ACGACUCUCGGCAACGGGAUUC	22	6496	13332	10836
GUACGACUCUCGGCAACGGGAUUCU	25	7360	10919	12142
GACUCUCGGCAACGGGAUUCUCGGCU	26	1866	10466	8135
GUACGACUCUCGGCAACGGGAU	21	6218	10348	9506
ACGACUCUCGGCAACGGGAUUCUC	24	5945	9207	6776
AGUACGACUCUCGGCAACGGGAU	24	2110	8104	5672
GUACGACUCUCGGCAACGGGAUUCUC	26	1889	8091	9523
UACGACUCUCGGCAACGGGAU	21	6991	7961	9531

Table 2. The list of abundant srRNAs identified from *P. nigrum* sRNA libraries.

observed in the C nucleotide in all species belonging to the Poaceae family except *Panicum virgatum*, in which major cleavage was observed at the 'A' nucleotide (Supplementary Figure 4C). Other monocot orders, such as Zingiberales, showed a very distinct cleavage pattern at the U nucleotide in all the *Musa accuminata* sRNA libraries. We also observed that the srRNAs cleaved from these unique 5.8S rRNA loci were predominantly 23 to 25 nt in length, and the major cleavage event occurred at the 3rd nucleotide upstream of the GACUCUC consensus sequence in most of the monocot plants. The major cleavage was mapped to the 3rd and 6th C nucleotide upstream of the consensus sequence in *Triticum aestivum* and *Zea mays*, respectively. By contrast, in both sRNA libraries of the Pteridophytic plant *Marselia sp.*, the major cleavage was mapped to the 2nd nucleotide upstream of the GACUCUC consensus, in which the 5'5.8S rRFs were predominantly 21 nt in length (Supplementary Figures 4A & 5E).

Interestingly, in all the dicot plants except *Mimulus guttatus*, major cleavage was observed at the 'A' nucleotide upstream of the consensus sequence (Supplementary Figure 4B). Among the core dicots, such as Vitales and Malvids, the most prevalent cleavage was mapped to the 6th and 5th nucleotides upstream of the GACUCUC consensus, respectively. The cleavage at the 5th nucleotide was also observed in Cucurbitales and Fabales of the Fabids group, whereas in *Populus trichocarpa* (Malpighiales), the major cleavage was mapped to the 3rd nucleotide upstream of the consensus sequence. The cleavage at the 6th nucleotide was predominant in all the Solanales plants. The length categorization of small RNAs from each cleavage position upstream of the GACUCUC consensus sequence showed that the small RNAs produced at the major cleavage site were predominantly 23–24nt in length (Supplementary Figure 5A–E). The 5'5.8S rRFs constituted an average of approximately 4.33 percent of the total reads in the small RNA libraries (Supplementary Figure 6). Relative occurrence of the most prominent 5'5.8SrRF, 'AACGACUCUCGGCAACGGGAUUCU' from the small RNA datasets of Arabidopsis next gen sequence database (<https://mpss.danforthcenter.org>) revealed a read count greater than 100 RPM from eighty nine sRNA libraries (Supplementary Figure 7). While thirteen sRNA libraries, that represents the 8.5% of the total libraries had shown the read count (RPM) less than 10. This shows the higher occurrence of the 5'5.8S srRNAs in the high throughput sRNA datasets.

Analysis of the association between 5'5.8S rRFs and the Argonaute complexes. In order to study the potential role of rRFs in the RNAi silencing pathways, their occurrence in the AGO complexed sRNAome of model plant species such as *Arabidopsis thaliana* and *Oryza sativa* was analyzed. It was proposed that each AGOs have distinct binding affinities towards the small RNAs with different 5' terminal nucleotides²¹. The functional diversification of ten different AGOs was deduced from the associated sRNAs and genetic analysis in Arabidopsis²². AGO1 was reported to associate with 21-nucleotide sRNAs with a 5' U, AGO2 binds 21 and 22 nucleotides with a 5' A, while AGO4 predominantly associates with 24-nucleotide sRNAs with a 5' A^{21–23}. While AGO7 mainly binds MIR390, an initiator of tasiRNA production; AGO4 binds repeat and heterochromatin associated small RNAs. The association of tRNA derived RNA fragments with AGO proteins were also reported in plants²⁴. The sequence data of argonaute coimmunoprecipitated sRNAome of Arabidopsis reported previously by Wang *et al.*, 2011, Montgomery *et al.*, 2008 and Havecker *et al.*, 2010 were analyzed to identify the association of rRF candidates to AGOs. The 5'5.8S rRF variants were detected in the sRNAome that co-immunoprecipitated

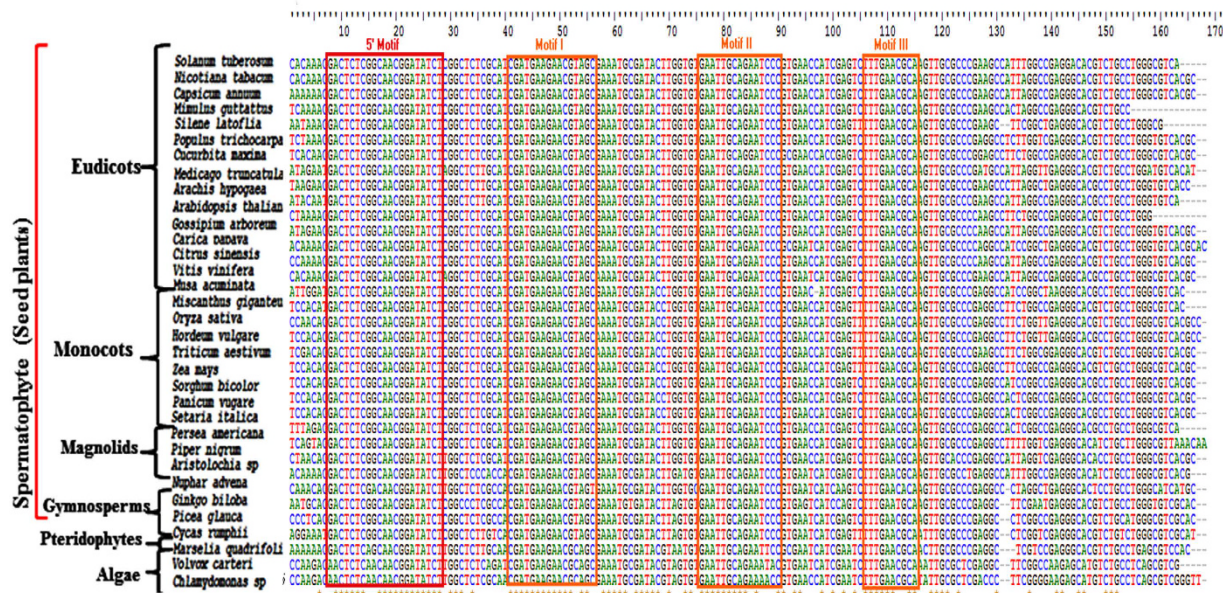


Figure 4. Multiple sequence alignment of rRNAs from different plant species. GACTCTCGGCAACGGATAT is the 5'5.8S rRF mapped 5' motif found in all the lineages of spermatophyte plants.

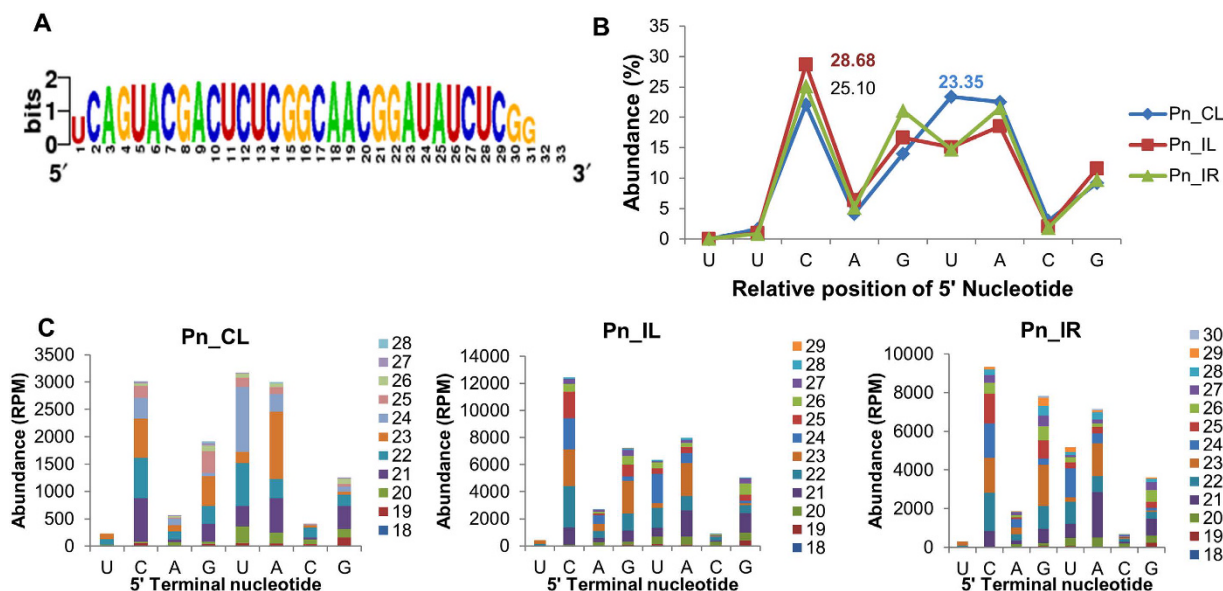


Figure 5. Analysis of isoforms and cleavage pattern of 5/5.8S rRFs from black pepper (A) Sequence logo generated from the different isoforms of 5/5.8S rRFs. (B) The cleavage pattern of 5.8S rRFs assessed upstream of the GACUCUC consensus (C) The length distribution of 5/5.8S rRFs.

with different argonaute complexes, such as AGO1 to AGO9 (Fig. 7). The 5/5.8S rRFs displayed a strong preference for AGO1 in *Arabidopsis thaliana* (Fig. 7A). By comparing the total normalized reads among the AGO1 precipitated small RNA libraries from different tissues of *Arabidopsis*, 5/5.8S rRFs were found to be most prominent in the flower sRNA library, in which the relative abundance of FLAG AGO1 protein was high. In contrast, the AGO4 associated 5/5.8S rRFs were highly abundant in the root tissue, which has a higher relative FLAG AGO expression²⁵. The abundances of 5.8S rRFs associated with AGO1 or AGO4 were very low in the leaf compared to other tissues. The length distribution of 5/5.8S rRFs also showed variations in the different AGO complexed sRNA libraries. The AGO4 sRNA datasets showed that the 5/5.8S rRFs were predominantly of 21 nt in length (Fig. 7B). In the AGO1, AGO2 and AGO7 complexed sRNA libraries, the 5/5.8S rRFs were predominantly 19 nt in length (Fig. 7C), while the 5/5.8S rRFs of 21 and 24 nt in length constituted the majority of the 5/5.8S rRFs in the AGO6 and AGO9 sRNA libraries, respectively (Fig. 7D). The relative abundance of normalized read counts

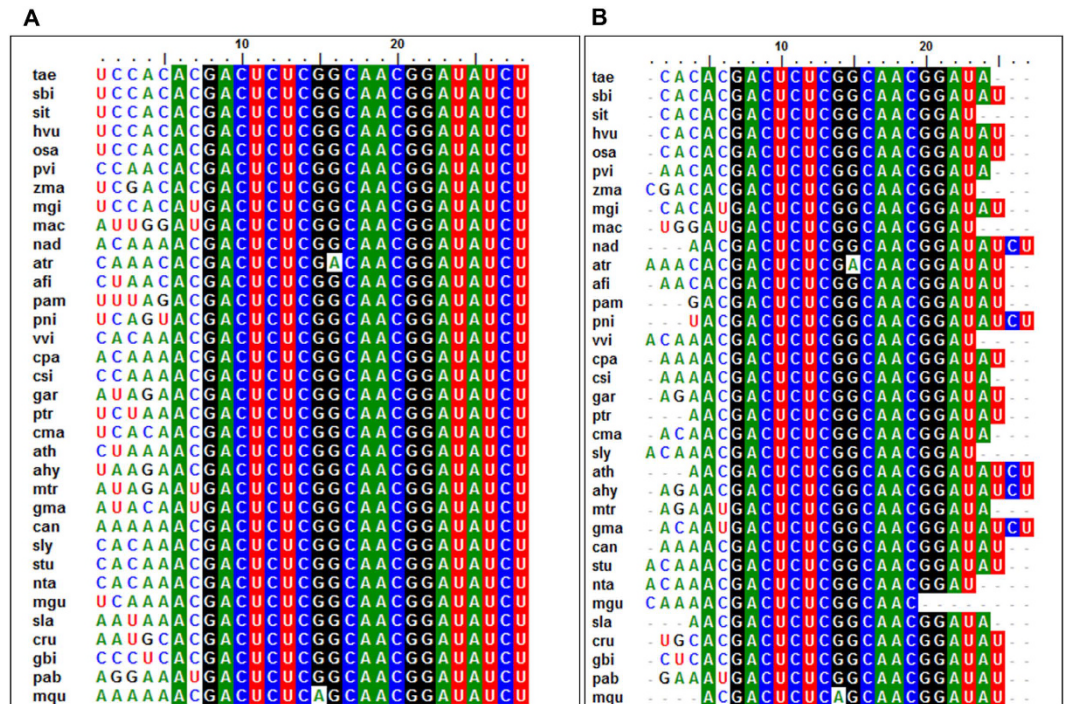


Figure 6. Unique sequence variation at the 5' ends of 5.8S rRNAs and the prominent 5.8S rRFs in plants. (A) Sequence diversity at the 5' termini of 5.8S rRNA in different plant species (B) Multiple sequence alignment of 5.8S rRF from different plant lineages. The most predominant 5.8S rRFs from each plant species were aligned using ClustalW. (tae: *T.aestivum*, sbi: *S. bicolor*, sit: *S. italica*, hvu: *H. vulgare*; osa: *O. sativa*; pvi: *P. virgatum*, zma: *Z. mays*, mgi: *M. giganteus*, mac: *M. acuminata*, nad: *N. advena*, atr: *A. trichopoda*, afi: *A. fimbriata*, pam: *P. americana*, pni: *P. nigrum*, vvi: *V. vinifera*, cpa: *C. papaya*, csi: *C. sinensis*, gar: *G. arboretum*, ptr: *P. trichocarpa*, cma: *C. maxima*, ath: *A. thaliana*, ahy: *A. hypogaea*, mtr: *M. truncatula*, gma: *G. max*, can: *C. annuum*, sly: *S. lycopersicum*, stu: *Solanum tuberosum*, nta: *N.tabacum*, mgu: *M. guttattus*, sla: *S. latifolia*, cru: *C. rumphii*, gbi: *G. biloba*, pab: *P. abies*, mqu: *Marsilea quadrifolia*.

of 5.8S rRFs among the different AGO complexed sRNA library from the *Arabidopsis* flower showed their high occurrence in the AGO1 sRNA library (Supplementary Figure 8).

When we compared the 5.8S rRFs associated with different components of Argonaute 1, such as AGO1a, AGO1b, and AGO1c, in the AGO1 knock-down *Oryza sativa* seedlings, the 19nt 5.8S rRFs showed increased accumulation; this is similar to observations in *Arabidopsis*, and among the three AGO1 complexes, AGO1b predominantly bound the 19 nt 5.8S rRFs (Fig. 7E). The detailed analysis of sRNA libraries of *Arabidopsis Ago1-25* mutant (GSM343002, GSM343003, GSM343004) revealed the high expression of 5.8S rRFs compared with the control plants (Fig. 7F). This suggests the functional specification of 5.8S rRFs according to their interactions with different Argonautes as well as their biological roles; namely, 5.8S rRFs probably play a similar role as microRNAs in the gene regulatory mechanisms of different physiological processes.

Do rRFs have a non-canonical miRNA-like precursor? We assessed the ability of the 5.8S rRNA region to form non-canonical microRNA-like stem loop precursor structures. The *in silico* analysis predicted that the secondary structure of the *P. nigrum* 5.8S rRNA consisted of two hairpins joined by a hinge, with an MFE value of -50.50 kcal/mol (Fig. 8A). A similar hairpin structure was predicted for the 5.8S rRNA from *Arabidopsis thaliana*, with an MFE value of -44.70 kcal/mol. An alternate, non-canonical miRNA-like short hairpin precursor was also predicted, which had a sequence length of 110 nucleotides and an MFE value of -28.40 kcal/mol (Fig. 8B). 5.8S rRF could be mapped to the stem region of this short hairpin. Our results also indicated the second most abundant class of small RNAs from 5.8S rRNA that could be mapped to the same precursor structure. Furthermore, the 5' modified RLM RACE experiments revealed cleavage at the C nucleotide (Supplementary Figure 9), which was identified as the major cleavage position of 5.8S rRFs during pathogenic infection. Thus, it is possible that the precursor 5.8S rRNA located 6 or 7 nucleotides upstream of the GACUCUC consensus sequence was expressed during the processing of 5.8SrRNA in pathogen-stressed black pepper plants. Similar non-canonical short hairpin structures with MFE values within the range of -28.5 KCal/mol were predicted for most of the plants, even if their 5' ends were different (Supplementary Figure 10).

Analysis for the Dicer Dependency of 5.8S rRFs. The sRNA datasets of flower tissues from the partial-loss-of-function *dcl-1* mutant and *dcl-234* triple mutants (GSE44622) showed high expression of 5.8S rRF in the mutant as compared to the wild type plants (Fig. 9). Notably, the 24 nt long 5.8S rRF 'AACGACUCUCGGCAACGGGAUAUCU' showed high read count in all the sRNA libraries of *dcl-1*, *dcl-234*

Plant group	Scientific Name	Major Cleavage site	Relative Position of major cleavage from 5' consensus
Dicots	<i>Vitis vinifera</i>	C↓ACAAACG	-6
	<i>Carica papaya</i>	C↓AAAACG	-6
	<i>Citrus sinensis</i>	C↓AAAACG	-6
	<i>Gossipium arboreum</i>	U↓AGAACG	-6
	<i>Medicago truncatula</i>	U↓AGAAUG	-5
	<i>Arachis hypogea</i>	A↓AGAACG	-5
Dicots	<i>Glycine max</i>	U↓ACAAUG	-5
	<i>Cucurbita maxima</i>	C↓ACAACG	-5
	<i>Populus trichocarpa</i>	CUA↓AACG	-3
	<i>Arabidopsis thaliana</i>	UAA↓AACG	-3
	<i>Silene latifolia</i>	AUA↓AACG	-3
	<i>Mimulus guttatus</i>	C↓CAAAACG	-6
	<i>Solanum lycopersicum</i>	C↓ACAAACG	-6
	<i>Solanum tuberosum</i>	C↓ACAAACG	-6
	<i>Capsicum annuum</i>	A↓AAAACG	-5
	<i>Nicotiana tabaccum</i>	C↓ACAAACG	-6
Monocots	<i>Oryza sativa</i>	C↓CACACG	-5
	<i>Triticum aestivum</i>	C↓CACACG	-5
	<i>Hordeum vulgare</i>	C↓CACACG	-5
	<i>Setaria italica</i>	C↓CACACG	-5
	<i>Zea mays</i>	U↓CGACACG	-5
	<i>Sorghum bicolor</i>	C↓CACACG	-5
	<i>Panicum virgatum</i>	C↓AACACG	-5
	<i>Miscanthus giganteus</i>	C↓CACAUG	-5
	<i>Musa acuminata</i>	U↓UGGAUG	-5
Magnoliids	<i>Piper nigrum</i>	UCAG↓UACG	-3
	<i>Aristolochia fimbriata</i>	U↓AACACG	-5
	<i>Persea americana</i>	UUA↓GACG	-3
Other basal angiosperms	<i>Amborella sp</i>	↓AAACACG	-6
	<i>Nuphar advena</i>	CAAA↓ACG	-3
Gymnosperms	<i>Cycas rumphii</i>	A↓UGCACG	-5
	<i>Ginkgo biloba</i>	C↓CUCACG	-5
	<i>Picea abies</i>	G↓GAAAUG	-5
Pteridophyte	<i>Marsilea quadrifolia</i>	AAAA↓ACG	-3
Chlorophyta	<i>Chara corallina</i>	A↓GAACUG	-5

Table 3. The relative position of major cleavage of 5.8S srRNAs from different plant species.

mutant compared to Col-1 wild type plants. The expression in the *dcl-234* triple mutant plant suggests the enrichment of 5.8S rRFs, similar to miRNAs.

The potential functional role of 5'5.8S rRFs. The preferential breakdown of 5'5.8S rRFs and their prominence in the plant system suggests their stability and functional relevance. The variations in the expression of twenty abundant 5'5.8S srRNAs in the leaf and root sRNA libraries of *P. capsici* infected *Piper nigrum* plants (Fig. 10) showed the higher expression of many of the rRF variants. Because these small RNAs may potentially function as microRNAs, targets of the abundant 5'5.8S rRFs were predicted from black pepper mRNA transcripts using psRNATarget by following the miRNA target criteria. Genes encoding the 40S ribosomal protein, adenosylhomocysteinase, diaminopimelate epimerase, shikimate O-hydroxycinnamoyltransferase, diphthamide synthase and cytochrome P450 were the predicted targets of the 5'5.8S rRF variants from black pepper (Table 4). Even though mRNA cleavage and translational repression were predicted as the target regulatory mechanisms, cleavage was identified as the most prominent mechanism among the predicted targets. Potential target mRNAs were also predicted from other plants, such as *Arabidopsis thaliana*, *Oryza sativa*, *Carica papaya*, *Citrus sinensis*, *Cucumis sativus*, *Gossipium arboreum*, *Glycine max*, *Medicago truncatula*, *Populus trichocarpa*, *S. tuberosum*, *S. lycopersicum* and *Nicotiana tabaccum*. Transposable element genes, mRNAs encoding RNI-like proteins, xyloglucan endotransglucosylase, and several other functionally relevant proteins were predicted as the putative 5'5.8S rRF targets in *Arabidopsis thaliana*, while homeobox domain containing proteins and S-adenosyl-L-methionine:benzoic acid/salicylic acid carboxyl methyltransferase were identified as the rRF targets in *Oryza sativa* (Supplementary Table 1). Adenosylhomocysteinase, the competitive inhibitor of S-adenosylmethionine (SAM)-dependent methyl transferase reactions, was identified as a putative common

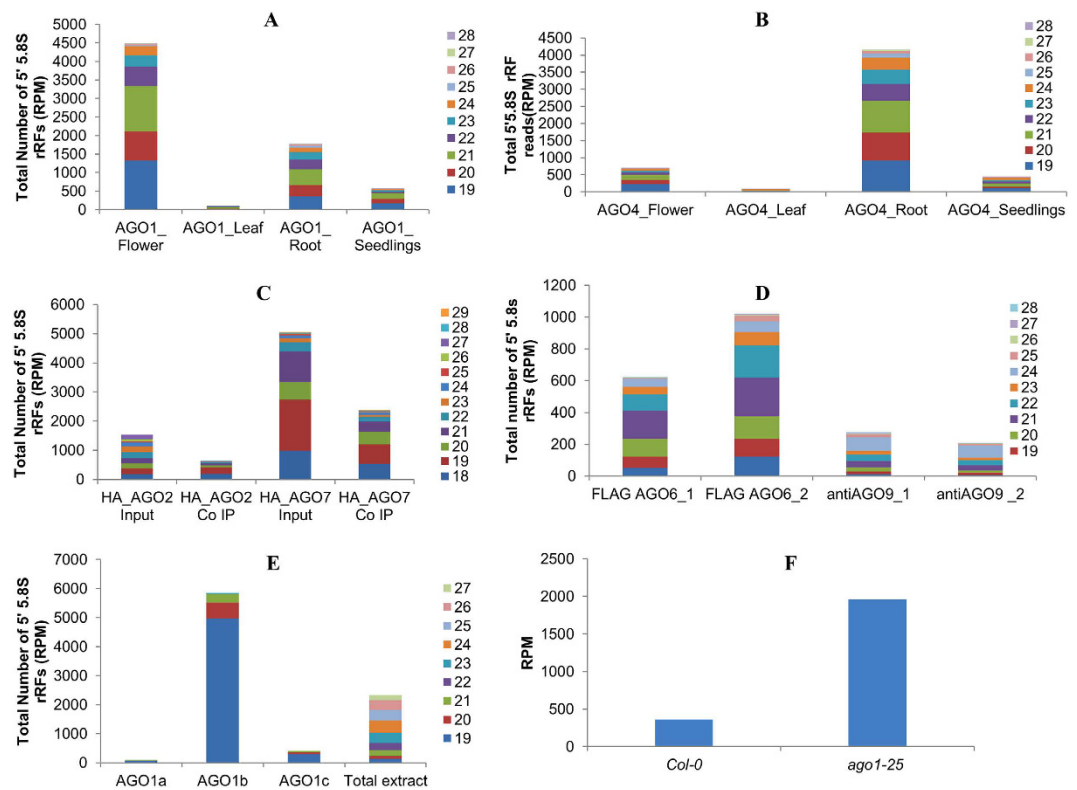


Figure 7. Association of 5'/5.8S rRFs with different Argonaute proteins. 5'/5.8S rRFs identified from AGO immuno-precipitated sRNA libraries of *A. thaliana*. (A) AGO1 (B) AGO4 (C) AGO2 & AGO7 (D) AGO6 & AGO9 (E) 5'/5.8S rRFs identified from different homologs of AGO1 immuno-precipitated sRNA libraries of *O. sativa*. The sRNA data (GSM707682, GSM707686, GSM304284, GSM415789 and GSM415791)^{22,23,39} from the studies of Wang *et al.*, 2011, Montgomery *et al.*, 2008 and Havecker *et al.*, 2010 were analysed and the total read counts of 5'/5.8S rRFs was represented as the normalized read count (RPM) from the corresponding sRNA library. (F) 5'/5.8S rRFs detected from the *Ago1-5* mutants.

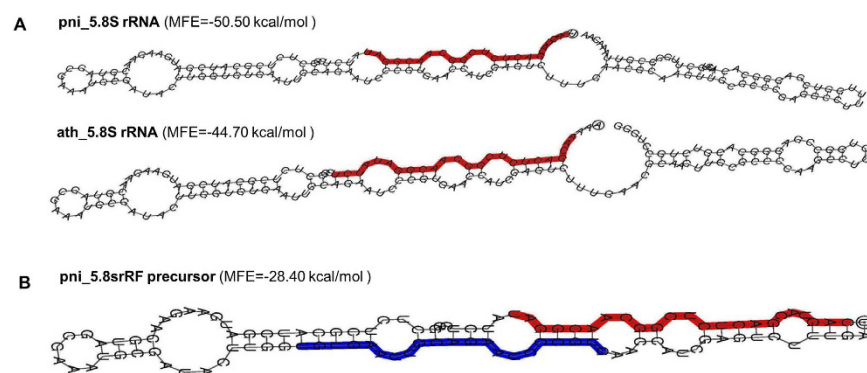


Figure 8. The predicted secondary structure of 5.8S rRNA of black pepper. The secondary structure of (A) *P. nigrum* and (B) *A. thaliana* 5.8S rRNA as predicted from the RNAfold. (C) The non-canonical short hairpin precursor of 5'/5.8S rRFs. The 5'/5.8S rRF was highlighted in red color. The next abundant rRF identified from the sRNA library was highlighted in blue in the opposite strand.

target in the monocot plants *Hordeum vulgare*, *Triticum aestivum* and *Sorghum bicolor*. SAM methyl transferase and adenosyl homocysteinase are involved in methionine synthesis and methylation recycling in plants. In many plant species, proteins involved in rRNA biogenesis and ribosome assembly were predicted as 5'/5.8S rRF targets. The 5' modified RLM RACE experiments in black pepper further mapped the cleavage events to the 5'/5.8S rRF aligned region of the transcript encoding 40S ribosomal protein S13, which is the RNA binding domain (Fig. 11). Four variants of RPS13 mRNA transcripts were identified from the assembled stress responsive transcriptome data from black pepper. The details of the FPKM normalized expression of the RPS13 mRNA isoforms from the

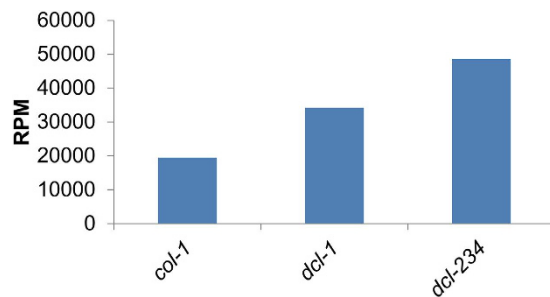


Figure 9. Involvement of Dicer-like activity on the biogenesis of srRNAs. The expression of 5'/5.8S rRFs detected from the *dcl-1* and *dcl-234* mutant Arabidopsis plants.

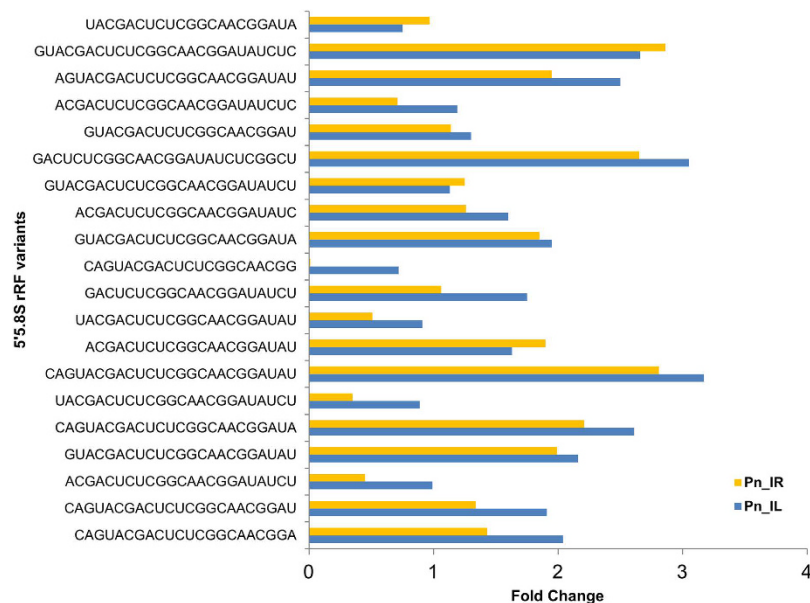


Figure 10. The expression of 5'/5.8S rRF variants in the Pn_IL and Pn_IR libraries compared to the control leaf library (Pn_CL).

P. nigrum transcriptome were shown in Supplementary Table 5. All the four variants of RPS13 mRNA transcripts showed down regulation in the pathogen infected tissues compared to the control (Supplementary Figure 11).

The sRNA mediated cleavage at the predicted target gene encoding ribosomal L5P family protein (AT5G45775.1) was detected from Arabidopsis degradome data (Fig. 12). The multiple cleavages (Supplementary Figure 12) were detected at the sRNA aligned region that encodes the conserved domain for the ribosomal protein L5. The degradome analysis also mapped the cleavage from other targets such as glutamate:glyoxylate aminotransferase (AT1G23310.2), N₂,N₂-dimethylguanosine tRNA methyltransferase (AT5G15810.1), myb domain protein (AT4G37260.1), N-terminal nucleophile aminohydrolases (AT5G61540.1) from Arabidopsis with comparatively less read count (<10) (Supplementary Figure 13). These findings demonstrated the precise processing of 5'/5.8S rRFs and their potential functional role in gene regulation in plants.

Discussion

The higher abundance of srRNAs that mapped to the 5' terminus of 5.8S rRNA of *Piper nigrum* indicated that these srRNAs are precisely processed rather than randomly degraded. In a eukaryotic cell, ribosomal RNAs are prone to fast degradation, which eliminates rRNAs that were defective in length processing, folding and assembly²⁶. The 5' end of 5.8S rRNAs might be protected from degradation and may function as sites for biogenesis of functional srRNAs through unknown processing events. The length variation and differential density has been previously reported as a peculiar characteristic of rRNA-derived small RNAs²⁷. In agreement with the findings from *Piper nigrum*, similar unique 5'/5.8S srRNA signatures were detected from plants belonging to different phylogenetic groups. Even in well-studied model plants, these small RNAs remained hidden, as they generated from the rRNA. The deep evolutionary history of the plant can be assessed from the unique srRNAs that mapped to the syntenic rRNA regions in the genome, as these srRNAs may act in the ancestral gene regulatory apparatus for the biogenesis of evolutionarily labile srRNAs. The widespread, multi-laterally adapted RNAi system orchestrates several cellular processes and includes potentially lineage-specific variations in the srRNAome²⁸. Thus, it would

(motif III), were identified in Viridiplantae²⁹. Sequence motifs in small RNAs can potentially regulate their stability and intracellular localization. The multiple sequence alignment of 5.8S_L rRNA consisting of 6–7 nt upstream sequences and a LocARNA-P analysis confirmed GACUCUCGRCAACGGAUUCU as the novel, functionally active, 5.8S rRNA motif present among the seed plants (spermatophytes) (Supplementary Figure 14). The abundant srRNAs spans this novel, conserved motif in all the lineages of the Viridiplantae. The lower plant groups, such as Pteridophytes and green algae, possessed distinct nucleotide variation in this consensus sequence. Because the sequence variations that contribute to the lineage differences are highly evident in the rRNA, the characterization of 5′/5.8S rRFs is an important model to study the evolution of plants³⁰. Within-family, the 5′/5.8S rRFs were generally similar, while among the distant lineages, the differences in the 5′ nucleotides were apparent. Therefore, the study of the evolutionary dynamics of these unique srRNAs can give informative insights into their functional importance in plant development. As previously described³¹, the presence of all the three Viridiplantae motifs in the selected 5.8S rRNA and the ability to form a helix at the motif III (TTTGAyGCA) indicated that no pseudogenes were involved in the dataset. The genomic locations of 5′/5.8S rRFs were mapped to Chr2:5784–5808 (+strand) and Chr3:14199755–14199779 (+strand) in *Arabidopsis thaliana* and to Chr2: 28715492–28715515, Chr9: 4632–465 and Chr9: 12560–12583, 20488–20511, 28416–28439, 36344–36367 in *Oryza sativa*. Above all, the close examination of small RNAs mapped to the 5′ and 3′ ends of 5.8S rRNA could accurately define the termini of the 5.8S rRNAs from each species, which is difficult using other methods.

The regulatory structural RNAs were capable of forming a great variety of secondary structures, rendering them for a classical DICER processing system³². The non-canonical pathways involved in the biogenesis of small RNAs may have been predicted from the noncoding structural RNAs such as tRNAs³³ and snoRNAs¹. Moreover, recent research has demonstrated their biologically functional roles^{33–35} and genesis by a specific RNase under stress conditions³⁶. The alternate secondary structure of 5.8S_L rRNA from *Piper nigrum* consisted of two hairpins and a hinge in the third motif (the TTTGAyGCA region), which was similar to the previously reported novel box H/ACA in snoRNAs, in which small RNAs are formed from the conserved motifs box C, box D, and box H/ACA¹. Similarly, the 5.8S srRNAs at the conserved motif III (TTTGA) can also form a non-canonical short hairpin structure. Interestingly, we observed that most of the plant 5.8S_L rRNAs could potentially form similar short hairpin structures even though sequence variation existed in the 5′ ends. The potential of forming secondary stem loop structures indicated that other than the conventional RNA processing by endonucleolytic and exonucleolytic cleavage, a possible non-canonical biogenesis pathway may be involved in their precise excision as microRNAs. Similar to tRNA genes, the 5.8S_L rRNA loci also has the intrinsic potential to form an alternative short hairpin structure, giving rise to mature RNAs or srRNAs. Apart from the usual rRNA recycling events for eliminating the degraded and defective rRNAs, the occurrence of functional srRNAs in the sRNAome of an organism has also been reported²⁷. Therefore, many or some of the core pathways for producing miRNAs may function in the biogenesis of these small RNAs.

The abundant srRNAs identified from all the plant species were mapped to the 5′ end of the 5.8S rRNA. Even though the number of isoforms of 5′/5.8S srRNAs was very large, the cleavage patterns were identical in different tissues of the same plant. The major cleavage events at the 5′ ends of 5.8S rRNAs were similar within a single plant group. The sequence artifacts that constituted the substitution variants, which differed by one or few nucleotides and were expressed at low levels, were excluded from the datasets. Regulatory RNAs are usually influenced by environmental signals and have potential functional roles during the stress response in plants. The normalization of each srRNA variant allowed for a more accurate estimation of their relative abundance. During pathogen-induced stress responses in black pepper, the major cleavage events were shifted by a few nucleotides upstream compared to the control leaf. The alteration of the 5′ end sequences can have a strong effect on its function and can potentially target an entirely different mRNAs during the stress response. A similar shift in the nucleotide cleavage site during the stress response was observed in *Hordeum vulgare* under pathogen infection and *Panicum virgatum* under drought compared to control tissues. It has also been reported that the rRNA-derived small RNAs are highly expressed during oxidative³⁷ and heat stress³⁸. The presence of the unique 5′/5.8S rRFs and their association with different AGO proteins were identified in the AGO-associated sRNA profiles from *Arabidopsis thaliana*^{23,39} and *Oryza sativa*³⁹. AGO1 harbors miRNA or sRNA, with a 5′ terminal uridine nucleotide, and change in the 5′ nucleotide redirects the miRNA or sRNA to a different argonaute complex^{40,41}. Most of the 5′/5.8S rRFs from plants do not possess the typical U nucleotide at the 5′ end, and *Arabidopsis* and Rice 5′/5.8S rRFs detected from the AGO1 complexes raise the question of whether AGO1 has a preference for these unique small RNAs. Whereas the high expression of 5′/5.8S rRFs were detected from the inflorescence of *ago 1–25* mutant, the fertile mutants with hypomorphic *ago1* alleles compared with the wild type plants. It may suggest the biogenesis of 5′/5.8S rRFs not affected by the single nucleotide transition in the 3′ portion of the gene leading to the Gly-758 → Ser (19–3) substitutions²⁵. The precisely processed, stable srRNAs and their occurrence in the AGO protein complexes further suggested that these srRNAs function like microRNAs or siRNAs in post-transcriptional gene silencing events.

An in-depth analysis of small RNA libraries from the flower tissues of *dcl-1*, *dcl-234* mutants, and (Col-0) wild-type plants revealed the abundance of 5′/5.8S rRFs. The expression of 5.8S srRNAs was found high in *dcl-234* triple mutant compared with the *dcl-1* mutant and wild type plants. The *dcl1* mutant is known to amplify the proportion of siRNAs and reduce the expression of miRNAs in the library⁴²; whereas *dcl-234* triple mutant is known to reduce the proportion of siRNAs and significantly enriches the miRNA population⁴³. Consistent with previous reports⁴⁴, the high expression of the small RNAs might be due to the known residual activity and the insensitivity to certain miRNAs in the partial-loss-of-function *dcl1–7* mutant allele.

SAM-dependent carboxyl methyltransferase was identified as a potential target of 5′/5.8S rRNAs in rice. S-Adenosyl methionine is the universal biological cofactor found in all branches of life, and it plays critical roles in the methylation of various biomolecules such as DNA, proteins and small molecule secondary metabolites. Methyl transferases, which are the largest group of SAM-dependent enzymes⁴⁵ in plants, have key roles in stress responses and in primary and secondary metabolism⁴⁶. Adenosyl homocysteinase acts by regulating DNA methylation, which affects the expression of certain genes involved in seed dormancy⁴⁷. Certain common targets, such

as genes involved in methionine biosynthesis and recycling, were predicted in the monocot plants, whereas the genes involved in ribosome biosynthesis, such as the 40S ribosomal protein, were identified as common targets among the dicot plants. Modified 5'RLM RACE experiments mapped the cleavage site to the highly conserved RPS13 domain on the predicted target mRNAs of the 40S ribosomal protein in black pepper. S13, which has been reported to form part of the amino acyl tRNA binding domain in the ribosome, also functions along with other ribosomal proteins to form ribonucleoprotein complexes with the 5.8S ribosomal RNA⁴⁸. In plants, multiple expressed copies of the gene for the ribosomal protein S13 exist; these transcripts have been proposed to act through the gene dosage model⁴⁹, as their proteins are required for optimum growth during different stages of the plant life cycle. It is possible that the unique ribosomal small RNAs may function in the gene dosage mechanisms that precisely regulate the expression of RPS13 in the different cell types and in the transitions between different developmental stages in plants. Also, analysis on Arabidopsis degradome data revealed the sRNA mediated cleavage on the predicted target mRNA encoding ribosomal L5P protein. The cleavage sites mapped at the sRNA aligned target region indicates the potential functional role of 5'/5.8S rRFs in plants. In conclusion, our results showed that 5'/5.8S rRFs are the unique, lineage-specific small RNAs with critical functional roles in plants.

Materials and Methods

Library preparation and Bioinformatic analysis of small RNAs from *Piper nigrum*. In this study, two high throughput sRNA library were prepared from leaf and root from *Phytophthora capsici* infected black pepper plants, and analyzed along with the previously sequenced control uninfected leaf sRNA library of black pepper⁵⁰. The pathogen infection was artificially given to the plants using mycelia disc of virulent *P. capsici*, as mentioned in ref. 51. The total RNA enriched with the small RNA fraction (≤ 200 nt) was extracted from the leaves and roots of *P. capsici* infected black pepper plants at 24 hour after pathogen infection, using mirVana miRNA isolation kit (Ambion) according to the manufacturer's instructions. After the quality and quantity analysis of RNA using NanoDropTM 1000 spectrophotometer (Thermo Scientific, Wilmington, DE) and Agilent Bioanalyzer, the library of sRNAs (~18–30 nt) was constructed and sequenced⁵² on an Illumina Genome Analyzer at Beijing Genomics Institute (Shenzhen, Guangdong Province, China). The raw reads were processed to remove the low quality reads, and clean reads ranging from 18–30 nt length were selected. The small RNAs derived from rRNA, tRNA, snoRNA etc. were annotated from the black pepper small RNAome using BGI small RNA analysis pipeline procedure that analyze the known non-coding RNAs deposited in the Rfam and NCBI GenBank databases. The rRNA mapped reads were selected from the entire three libraries and analyzed further. The relative position of srRNAs was determined by mapping the reads to *P. nigrum* ribosomal RNA using PatMaN in VisSR Tool of UEA small RNA workbench (srna-workbench.cmp.uea.ac.uk)⁵³. The full-length, perfect matches were accepted as hits. The rRNA-derived small RNAs (srRNAs) from all the three libraries were further analyzed for their read frequencies, size distribution and 5' terminal nucleotide preferences. The srRNAs reads were normalized by dividing hit counts by the number of redundant reads in corresponding sRNA library and srRNAs with read numbers > 10 were selected, and their distributions and frequencies within the black pepper rRNA were analyzed. The cleavage pattern of the terminal nucleotides of the 5'/5.8S rRFs were assessed upstream of the GACUCU consensus sequence. The individual read counts of twenty abundant 5'/5.8S rRF variants were normalized (RPM) and their expression levels were compared to the expression levels of these variants in the control leaf sRNA library. The sRNA profiles of black pepper were deposited to the NCBI GEO database under the accession numbers GSM1606153 (Pn_CL), GSM1606154 (Pn_IL) and GSM1606155 (Pn_IR).

Comparative analysis of sRNA datasets from diverse plant groups. The high throughput small RNA datasets of plants belonging to different lineages, including green algae (Chlorophyta), vascular (angiosperms and gymnosperms) and nonvascular plants (Pteridophyta), that were generated by Chavez Montes *et al.*⁵⁴ were downloaded from the NCBI gene expression omnibus. The study group consisted of a total of 97 small RNA library datasets from thirty-seven distinct plant species^{21,55}. Brief information about the sRNA datasets used in the study is provided in Supplementary Table 2. The sequence accession numbers of ribosomal RNAs from different plant species are given in Supplementary Table 3. The predominant group of 5.8S rRNA-derived small RNAs from most of the plant species possessed few upstream nucleotides compared to the database-curated 5.8S rRNAs. We selected a sequence boundary located 7 nt upstream of the GACUCUC motif in the 5.8S rRNA of all the plant species, and the consensus secondary structure was analyzed using locARNA-P, a software package that computes STARS (the structure based alignment reliabilities)⁵⁶. The abundance and the distribution of 5'/5.8S rRNAs were analyzed using the sequence alignment and VisSR tools in the UEA sRNA workbench. The 5' cleavage pattern of 5'/5.8S srRNAs were checked based on the 5' conserved motif GACUCUCG, and the corresponding length distribution of isoforms were analyzed. The genomic locations of the 5'/5.8S rRF sequences were further analyzed using the annotated plant genomes from the Phytozome database (<http://phytozome.jgi.doe.gov/pz/portal.html>). The occurrence of the 5'/5.8SrRF was checked from the 153 small RNA datasets of Arabidopsis next gen sequence database (<https://mpss.danforthcenter.org>).

In silico analysis of the non-canonical short hairpin precursors and experimental mapping of 5'/5.8S₁ rRNA. 5.8S ribosomal RNA of black pepper was analyzed *in silico* for its potential to form non-canonical, miRNA-like hairpin precursors. The secondary structures were predicted using the RNA folding program in the UEA small RNA workbench. Because the exact 5' end of the 5.8S rRNA could not be accurately determined, the selected sequence of 5.8S rRNA contained an additional seven nucleotides upstream of the 5' motif. The 5' end of the 5.8S rRNA in black pepper was mapped using modified 5' RLM RACE experiments. In brief, the total RNA isolated from black pepper was ligated at the 5' end with the Gene RacerTM RNA oligo (Gene RACER kit, Invitrogen, Carlsbad, CA). The ligated RNA was purified using phenol:chloroform and precipitated. The reverse transcription was performed using the rRNA specific reverse primer

(5'CAACTTGC GTTCAAAGACTCGATGGT3'), and PCR amplification was carried out using the Gene Racer 5' primer and 5.8S rRNA-specific reverse primer. The amplicons of ~150bp in length were eluted from the gel and cloned into the PCR4 TOPO cloning vector (Invitrogen) and sequenced. The 5.8S rRNAs from different plant lineages were multiple aligned and a consensus structure was predicted. The alternate secondary structures were predicted for the 5.8S_L rRNA (including 7 upstream nt) from all the plant species studied. The cleavage patterns were assessed upstream of the GACTCT consensus and the length distribution at each upstream nucleotide position were analyzed. The sequence logo of abundant 5'/5.8S rRFs (≥ 100 RPM) were generated using RNALogo (rnalogo.mbc.nctu.edu.tw/)⁵².

Prediction and validation of molecular functions of 5'/5.8S rRFs. To study the role of rRFs in the RNAi pathway, high-throughput sequencing datasets of small RNAs were co-immunoprecipitated with AGO proteins from *Arabidopsis thaliana* and *Oryza sativa*. The details of the sRNA sequence dataset used were shown in Supplementary Table 4. The datasets consisted of immunoprecipitated small RNAs from AGO1 (GSM707682- GSM707685)³⁸, AGO4 (GSM707686- GSM707689)³⁸, AGO2 (GSM304282 and GSM304283)²³, AGO7 (GSM304284- GSM304285)²³, AGO6 (GSM415789 and GSM415790)²² and AGO9 (GSM415791 and GSM415792)²² from *A. thaliana*, and with different AGO1 homologs (GSM455962- GSM455965) from *O. sativa*⁴⁰. The studies of Wu *et al.*, 2009 generated AGO1 knock down rice plants with an RNAi approach that consisted of an inverted repeat (IR) construct (AGO1IR) for targeting the four rice AGO1 homologs such as AGO1a, b, c, and d. The relative position of srRNAs in the selected libraries was determined by mapping the reads to ribosomal RNA using PatMaN in SiLoCo and VisSR Tool of UEA small RNA workbench (srna-workbench.cmp.uea.ac.uk)⁵³. The srRNAs were analyzed for their length distribution pattern and cleavage around the unique signature motif. The read counts of different length variants were normalized and represented in RPM (Reads Per Million). The higher abundance and biogenesis from the short hairpin structure and the association with AGO indicated the potential functional role of 5'/5.8S rRFs, which is similar to microRNAs in the plant system. The dependency of the rRF candidates to Dicer-like (DCL) proteins were checked from GSE44622 that corresponds to sRNA datasets of partial loss of function *dcl-1* and *dcl-234* triple mutants⁴². The most prevalent 5.8S srRNA isoforms were selected and their targets were predicted from the black pepper mRNA transcriptome using psRNATarget (<http://plantgrn.noble.org/psRNATarget/>)⁵⁷. The targets were also predicted for the 5'/5.8S rRFs from other plant species from the psRNATarget. To further validate the predicted cleavage on the sRNA–target pairs, the degradome library was searched for the cleavage product using the dRNA mapper tool from SoMART web server (<http://somart.ist.berkeley.edu>). The cleavage from the predicted target of *A. thaliana* was analysed from the PARE (Parallel Analysis of RNA Ends) sequencing data from Col-0 inflorescence (GSM280226) and xrn4 mutant inflorescence (GSM280226).

Validation of rRF mediated cleavage of the target mRNAs of the Ribosomal Protein RPS13.

The 5'/5.8S rRF mediated cleavage of the target mRNAs of RPS13 was experimentally mapped using modified 5'RLM RACE experiments⁵¹. Total RNA was isolated from black pepper using mirVana miRNA isolation kit (Ambion), according to the manufacturer's instructions. The GeneRacer™ RNA oligo adapter (Invitrogen) (5'CGACUGGAGCAGGAGGACACUGACAUGGACUGAAGGAGUAGAAA3') was ligated to the 5' end of the cleaved mRNAs, without an alkaline phosphatase treatment. The reverse transcription of the ligated RNA was carried out at 50 °C for 60 min using GeneRacer™ oligo dT primer and SuperScript™ III Reverse Transcriptase (Gene racer kit, Invitrogen). Followed by the inactivation of RT reaction at 70 °C for 15 min, the PCR amplification of the cDNA ends were performed at a thermal profile: 94 °C/2 min (1 cycle); 94 °C/30 s, 72 °C/1 min (5 cycles); 94 °C/30 s, 70 °C/1 min (5 cycles); 94 °C/30 s, 65 °C/30 s, 72 °C/1 min (25 cycles); 72 °C/10 min (1cycle), using a forward primer specific to 5' adapter sequences (5'CGACTGGAGCAGGAGGACACTGA3') and a reverse primer specific to the RPS13 gene (5'TGGAAAGGCGACATCACACGA3'). Amplicons of the required size were extracted from 1.5% gel using S.N.A.P.™ Columns (Invitrogen) and cloned into the pCR4-TOPO cloning vector (Invitrogen). After transformation into competent *E. coli* strain DH5-Alpha, the plasmid DNA was isolated from the positive clones and sequenced with BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems), using M13 forward primer (5'GTAAAACGACGGCCAG-3') and M13 reverse primer (5'CAGGAAACAGCTATGAC3') in ABI PRISM® 3700 DNA Analyzer (ABI). The expression of the RPS13 mRNAs were further analyzed from the stress responsive mRNA transcriptome data (NCBI Sequence Read Archive under accession no. SRX853366 and SRX856639) (unpublished data) using FPKM method (Fragments Per kb per Million fragments)⁵⁸.

References

- Chen, H. & Wu, S. Mining small RNA sequencing data : a new approach to identify small nucleolar RNAs in Arabidopsis. *Nucleic Acids Res.* **37**, e69 (2009).
- Hackenberg, M., Gustafson, P., Langridge, P. & Shi, B. J. Differential expression of microRNAs and other small RNAs in barley between water and drought conditions. *Plant Biotechnol. J.* **13**, 2–13, doi: 10.1111/pbi.12220 (2014).
- Lee, H. C. *et al.* qiRNA is a new type of small interfering RNA induced by DNA damage. *Nature* **459**, 274–277 (2009).
- Chen, C. J. *et al.* Genome-wide discovery and analysis of microRNAs and other small RNAs from rice embryogenic callus. *RNA Biol.* **8**, 538–547 (2011).
- Wang, L. *et al.* A novel class of heat-responsive small RNAs derived from the chloroplast genome of Chinese cabbage (*Brassica rapa*). *BMC Genomics* **12**, 289 (2011).
- Chak, L., Mohammed, J., Lai, E. C., Tucker-kellogg, G. & Okamura, K. A deeply conserved, noncanonical miRNA hosted by ribosomal DNA. *RNA* **21**, 1–10, doi: 10.1261/rna.049098.114 (2015).
- Faber, A. W., Vos, H. R., Vos, J. C. & Raúé, H. A. 5' -End formation of yeast 5.8S rRNA is an endonucleolytic event. *Biochem. Biophys. Res. Commun.* **345**, 796–802 (2006).

8. Henry, Y. *et al.* The 5' end of yeast 5.8S rRNA is generated by exonucleases from an upstream cleavage site. *EMBO J.* **13**, 2452–2463 (1994).
9. Fatica, A. & Tollervey, D. Making ribosomes. *Curr. Opin. Cell Biol.* **14**, 313–318 (2002).
10. Henras, A. K., Plisson-Chastang, C., O'Donohue, M. F., Chakraborty, A. & Gleizes, P. E. An overview of pre-ribosomal RNA processing in eukaryotes. *Wiley Interdiscip. Rev. RNA* **6**, 225–242 (2015).
11. Strunk, B. S. & Karbstein, K. Powering through ribosome assembly. *RNA* **15**, 2083–2104 (2009).
12. Granneman, S., Petfalski, E. & Tollervey, D. A cluster of ribosome synthesis factors regulate pre-rRNA folding and 5.8S rRNA maturation by the Rat1 exonuclease. *EMBO J.* **30**, 4006–4019 (2011).
13. Kastenmayer, J. P. & Green, P. J. Novel features of the XRN-family in Arabidopsis: evidence that AtXRN4, one of several orthologs of nuclear Xrn2p/Rat1p, functions in the cytoplasm. *Proc. Natl. Acad. Sci. USA* **97**, 13985–13990 (2000).
14. Gy, I. *et al.* Arabidopsis FIERY1, XRN2, and XRN3 are endogenous RNA silencing suppressors. *Plant Cell* **19**, 3451–3461 (2007).
15. Zakrzewska-Placzek, M., Souret, F. F., Sobczyk, G. J., Green, P. J. & Kufel, J. Arabidopsis thaliana XRN2 is required for primary cleavage in the pre-ribosomal RNA. *Nucleic Acids Res.* **38**, 4487–4502 (2010).
16. Esakova, O. & Krasilnikov, A. S. Of proteins and RNA: the RNase P/MRP family. *RNA* **16**, 1725–1747 (2010).
17. Hartmann, E. & Hartmann, R. K. The enigma of ribonuclease P evolution. *Trends Genet.* **19**, 561–569 (2003).
18. Gutmann, B., Gobert, A. & Giegé, P. PRORP proteins support RNase P activity in both organelles and the nucleus in Arabidopsis. *Genes Dev.* **26**, 1022–1027 (2012).
19. Lange, H., Sement, F. M. & Gagliardi, D. MTR4, a putative RNA helicase and exosome co-factor, is required for proper rRNA biogenesis and development in Arabidopsis thaliana. *Plant J.* **68**, 51–63 (2011).
20. Lange, H. *et al.* The RNA Helicases AtMTR4 and HEN2 Target Specific Subsets of Nuclear Transcripts for Degradation by the Nuclear Exosome in Arabidopsis thaliana. *PLoS Genet.* **10**, e1004564 (2014).
21. Jeong, D. H. *et al.* Massive analysis of rice small RNAs: mechanistic implications of regulated microRNAs and variants for differential target RNA cleavage. *Plant Cell* **23**, 4185–4207 (2011).
22. Havecker, E. R. *et al.* The Arabidopsis RNA-Directed DNA Methylation Argonautes Functionally Diverge Based on Their Expression and Interaction with Target Loci. *Plant Cell* **22**, 321–334 (2010).
23. Montgomery, T. A. *et al.* Specificity of ARGONAUTE7-miR390 Interaction and Dual Functionality in TAS3 Trans-Acting siRNA Formation. *Cell* 128–141, doi: 10.1016/j.cell.2008.02.033 (2008).
24. Loss-Morais, G., Waterhouse, P. M. & Margis, R. Description of plant tRNA-derived RNA fragments (tRFs) associated with argonaute and identification of their putative targets. *Biol. Direct* **8**, 6, doi: 10.1186/1745-6150-8-6 (2013).
25. Morel, J. *et al.* Fertile Hypomorphic ARGONAUTE (ago1) Mutants impaired in post-transcriptional gene silencing and virus resistance. *Plant Cell* **14**, 629–639 (2002).
26. Zundel, M. A., Basturea, G. N. & Deutscher, M. P. Initiation of ribosome degradation during starvation In *Escherichia coli*. *RNA* **15**, 977–983, doi: 10.1261/rna.1381309.identify (2009).
27. Wei, H. *et al.* Profiling and Identification of Small rDNA-Derived RNAs and their potential biological functions. *PLoS One* **8**, e56842, doi: 10.1371/journal.pone.0056842 (2013).
28. Morris, K. V. & Mattick, J. S. The rise of regulatory RNA. *Nat. Rev. Genet.* **15**, 423–37 (2014).
29. Harpke, D. & Peterson, A. 5.8S motifs for the identification of pseudogenic ITS regions. *Botany* **86**, 300–305 (2008).
30. Fahlgren, N. *et al.* MicroRNA gene evolution in Arabidopsis lyrata and Arabidopsis thaliana. *Plant Cell* **22**, 1074–1089 (2010).
31. Harpke, D. & Peterson, A. Extensive 5.8S nrDNA polymorphism in Mammillaria (Cactaceae) with special reference to the identification of pseudogenic internal transcribed spacer regions. *J. Plant Res.* **121**, 261–270 (2008).
32. Röther, S. & Meister, G. Biochimie Small RNAs derived from longer non-coding RNAs. *Biochimie* **93**, 1905–1915 (2011).
33. Lee, Y. S., Shibata, Y., Malhotra, A. & Dutta, A. A novel class of small RNAs : tRNA-derived RNA fragments (tRFs) A novel class of small RNAs : tRNA-derived RNA fragments (tRFs). *Genes Dev.* **23**, 2639–2649 (2009).
34. Sobala, A. & Hutvagner, G. Transfer RNA-derived fragments: Origins, processing, and functions. *Wiley Interdiscip. Rev. RNA* **2**, 853–862 (2011).
35. Wang, Q. *et al.* Identification and functional characterization of tRNA-derived RNA fragments (tRFs) in respiratory syncytial virus infection. *Mol. Ther.* **21**, 368–79 (2013).
36. Yamasaki, S., Ivanov, P., Hu, G. F. & Anderson, P. Angiogenin cleaves tRNA and promotes stress-induced translational repression. *J. Cell Biol.* **185**, 35–42 (2009).
37. Thompson, D. M., Lu, C., Green, P. J. & Parker, R. tRNA cleavage is a conserved response to oxidative stress in eukaryotes. *RNA* **14**, 2095–2103 (2008).
38. Wang, H. *et al.* Deep sequencing of small RNAs specifically associated with Arabidopsis AGO1 and AGO4 uncovers new AGO functions. *Plant J.* **67**, 292–304 (2011).
39. Wu, L. *et al.* Rice MicroRNA Effector Complexes and Targets. *Plant Cell* **21**, 3421–3435 (2009).
40. Colaïacovo, M. *et al.* A survey of microRNA length variants contributing to miRNome complexity in peach (*Prunus persica* L.). *Front. Plant Sci.* **3**, 1–17(2012).
41. Mi, S. J. *et al.* Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide. *Cell* **133**, 116–127 (2008).
42. Jeong, D.H. *et al.* Massive analysis of rice small RNAs: mechanistic implications of regulated microRNAs and variants for differential target RNA cleavage. *Plant Cell* **23**, 4185–4207 (2011).
43. Henderson, I. R. *et al.* Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nat. Genet.* **38**, 721–725 (2006).
44. Allen, E. *et al.* Evolution of microRNA genes by inverted duplication of target gene sequences in Arabidopsis thaliana. *Nat. Genet.* **36**, 1282–1290 (2004).
45. Struck, A. W., Thompson, M. L., Wong, L. S. & Micklefield, J. S-Adenosyl-Methionine-Dependent Methyltransferases: Highly Versatile Enzymes in Biocatalysis, Biosynthesis and Other Biotechnological Applications. *ChemBioChem* **13**, 2642–2655 (2012).
46. Zubieta, C. *et al.* Structural basis for substrate recognition in the salicylic acid carboxyl methyltransferase family. *Plant Cell* **15**, 1704–1716 (2003).
47. Pawłowski, T. A. Proteome analysis of Norway maple (*Acer platanoides* L.) seeds dormancy breaking and germination: influence of abscisic and gibberellic acids. *BMC Plant Biol.* **9**, 48 (2009).
48. Chan, Y., Ulbrich, N., Ackerman, J. & Todokoro, K. The Binding of Transfer Ribonucleic Acids to 5 S and 5.8 S eukaryotic ribosomal ribonucleic acid-protein complexes. *J. Biol. Chem.* **19265**, 2522–2527 (1982).
49. Ito, T., Kim, G. T. & Shinozaki, K. Disruption of an Arabidopsis cytoplasmic ribosomal protein S13-homologous gene by transposon-mediated mutagenesis causes aberrant growth and development. *Plant J.* **22**, 257–264 (2000).
50. Asha, S., Sreekumar, S. & Soniya, E. V. Unravelling the complexity of microRNA-mediated gene regulation in black pepper (*Piper nigrum* L.) using high-throughput small RNA profiling. *Plant Cell Rep.* **35**, 53–63, doi: 10.1007/s00299-015-1866-x (2016).
51. Asha S. & Soniya E. V. Transfer RNA derived small RNAs targeting defense responsive genes are induced during *Phytophthora capsici* infection in Black Pepper (*Piper nigrum* L.). *Front Plant Sci.* **7**, 767, doi: 10.3389/fpls.2016.00767 (2016).
52. Chang, T. H., Horng, J. T. & Huang, H. Da. RNALogo: a new approach to display structural RNA alignment. *Nucleic Acids Res.* **36**, 91–96 (2008).

53. Stocks, M. B. *et al.* The UEA sRNA workbench : a suite of tools for analysing and visualizing next generation sequencing microRNA and small RNA datasets. *Bioinformatics* **28**, 2059–2061 (2012).
54. Chavez Montes, R. A. *et al.* Sample sequencing of vascular plants demonstrates widespread conservation and divergence of microRNAs. *Nat. Commun.* **5**, 3722 (2014).
55. Zhai, J. *et al.* MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs. *Genes Dev.* **25**, 2540–2553 (2011).
56. Will, S., Joshi, T., Hofacker, I. V. O. L., Stadler, P. F. & Backofen, R. LocARNA-P: Accurate boundary prediction and improved detection of structural RNAs. *RNA* **18**, 900–914, doi: 10.1261/rna.029041.111.10 (2012).
57. Dai, X. & Zhao, P. X. PsRNATarget: A plant small RNA target analysis server. *Nucleic Acids Res.* **39**, 1–5 (2011).
58. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNASeq. *Nat Methods.* **5**, 621–628 (2008).

Acknowledgements

AS greatly acknowledge the Senior Research Fellowship from Council of Scientific and Industrial Research (CSIR), New Delhi, India. EVS greatly acknowledge financial support from Department of Biotechnology, Government of India. All authors declare that no conflict of interest exists.

Author Contributions

A.S. performed the experiments and data analysis. E.V.S. supervised the project. A.S. and E.V.S. designed the project and wrote the paper.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Asha, S. and Soniya, E. V. The sRNAome mining revealed existence of unique signature small RNAs derived from 5.8SrRNA from *Piper nigrum* and other plant lineages. *Sci. Rep.* **7**, 41052; doi: 10.1038/srep41052 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017