A deeper view into the significance of simple sequence repeats in pre-miRNAs provides clues for its possible roles in determining the function of microRNAs

Nisha Joy*, Y. P. Maimoonath Beevi and E. V. Soniya*

Abstract

Background: The central tenet of ‘genome content’ has been that the ‘non-coding’ parts are highly enriched with ‘microsatellites’ or ‘Simple Sequence Repeats’ (SSRs). We presume that the presence and change in number of repeat unit (n) of SSRs in different genomic locations may or may not become beneficial, depending on the position of SSRs in a gene. Very few studies have looked into the existence of SSRs in the hair-pin precursors of miRNAs (pre-miRNAs). The interplay between SSRs and miRNAs is not yet clearly understood.

Results: Considering the potential significance of SSRs in pre-miRNAs, we analysed the miRNA hair-pin precursors of 171 organisms, which revealed a noticeable (29.8%) existence of SSRs in their pre-miRNAs. The maintenance of SSRs in pre-miRNAs even in the complex, highly evolved phyla like Chordata and Magnoliophyta shed light upon its diverse functions. Putative effects of SSRs in either regulating the biogenesis or function of miRNAs were more underlined based on computational and experimental analysis. A preliminary computational analysis to explore the relevance of such SSRs maintained in pre-miRNA sequences led to the detection of splicing regulatory elements (SREs) either in or near to the SSRs. The absence of SSRs correspondingly decreased the detection of SREs.

Conclusion: The present study is the first implication for the possible involvement of SSRs in shaping the SREs to undergo Alternative Splicing events to produce miRNA isoforms in accordance with different stress environments. This part of work well demonstrates the importance of studying such consistently maintained SSRs residing in pre-miRNAs and can enhance more and more research towards deciphering the exact function of SSRs in the near future.

Keywords: Non-coding RNAs, Tandem repeats, Pre-miRNAs, Alternative splicing, Arabidopsis thaliana
tandem repeats. They are tandem arrays of short (1–5 bp), repeated DNA sequences [2], that are commonly found in most genomes with a high mutation rate of $10^{-2}$ to $10^{-6}$ nucleotides per locus per generation [3] and hence utilized for fingerprinting studies. But once the discovery that change in tandem repeat unit of SSRs that fall in genes caused phenotypic changes, SSRs became more noticeable. The effect of SSRs were studied best across different plant species like rice [4, 5], common bean [6], barley [7], Arabidopsis [8] and also in humans [9].

But SSRs are poorly analysed in functional non-coding small regulatory RNAs like microRNAs (miRNAs). The importance of miRNAs (~20 nt) is that they play a major role in many biological processes and their biogenesis occurs from primary miRNA transcripts known as pri-miRNAs. The pri-miRNAs will adopt a stem-loop secondary structure known as the pre-miRNAs, from which a specific 21-nucleotide miRNA duplex is excised by a Dicer endonuclease [10]. Our previous experiments on transcriptome profiling revealed about the existence of SSRs in the non-coding transcripts of black pepper [11]. This true fact about the existence of SSRs in pre-miRNAs made us to ponder the possibility of SSRs in all the available pre-miRNAs across different taxa. To date, there is no lucid demonstration to prove the presence or pivotal functions of SSRs in hair-pin precursors of miRNAs, except for a few [11–13]. Hence our objective was to illustrate the exact incidence ratio of SSRs in all the available pre-miRNAs including plants, animals and viruses by performing a computational analysis in order to achieve a better understanding about the significance of such SSRs occurring in the pre-miRNAs. The preliminary observations revealed the significant incidence of SSRs and indicated the possible involvement of SSRs in Alternative Splicing (AS) events. AS also known as differential splicing is a regulated process that increases an organism’s transcriptome and proteome diversity [14]. One of the key regulators of AS are the cis acting Splicing Regulatory Elements (SREs) which are categorized into four classes like ESE (Exon Splicing Enhancer), ESS (Exon Splicing Silencer), ISE (Intron Splicing Enhancer) and ISS (Intron Splicing Silencer) depending on their location and its effect on splicing either as enhancers or silencers [15]. Here, the detection of SREs near SSRs in pre-miRNAs strongly suggests the possible involvement of SSRs in shaping episodes of AS.

Results

The distribution pattern of SSRs in hair-pin precursors of miRNAs

The SSRIT analysis of all the available miRNA precursors extracted from miRBase showed significant presence of SSRs which accounted to about 29.8%. The frequency and distribution pattern of SSRs varied extensively across different taxa analysed (see Additional file 1). SSR arrays were characterized as di, tri, tetra, penta or hexanucleotide based on the type of motif repeated in a sequence. Here, about 84.71% of SSRs were dinucleotide type of repeats, 12.5% were trinucleotide, 2.003% tetra, 0.544% penta and 0.181% were hexanucleotide type. When the relative count of SSRs bearing pre-miRNAs (the number of SSR bearing pre-miRNAs out of the total number of pre-miRNAs) were taken into consideration, Homo sapiens displayed the highest count, followed by Mus musculus, the least being Cunninghamia lanceolata, Macropus eugenii, Lemur catta, Marsupenaeus japonicas, Strigamia maritima, Glottidia pyramidata, Leucosolenia complicata, Sycon ciliatum, BK polyomavirus, Bandicoot papillomatisis carcinomatosis virus, Herpesvirus saimiri strain A11, JC polyomavirus, Merkel cell polyomavirus and Simian virus 40. The lesser count of SSR bearing pre-miRNAs may be due to lack of extensive miRNA characterisation studies in these organisms and hence these organisms cannot be completely demarcated. Similarly the relative count of SSR bearing pre-miRNAs were really scarce in a few organisms like Avicennia marina, Phytophthora sojae and Terebratulina retusa, still all of their pre-miRNAs revealed SSRs in their sequences showing 100% relative abundance of SSRs in miRNA precursors (Fig. 1). For a deeper and better understanding about the SSR motifs in the pre-miRNAs, we further focused our study in Arabidopsis thaliana, the model system. A closer examination of all the types of SSRs in the 325 reported pre-miRNAs of Arabidopsis thaliana exposed the significant presence of different types of SSR motifs. About 45% of pre-miRNAs in A. thaliana carried SSRs in their sequences of which 77% constituted dinucleotide type of SSRs, 19% trinucleotide, 3% tetrannucleotide and 1% pentanucleotide type of SSRs. The distribution pattern of SSR types identified is shown in Fig. 2. Out of the 45% of SSR bearing pre-miRNAs, 7.5% of SSR bearing pre-miRNA showed transcription factors like SBP, MYB, NAC, HLZ, ARF, GRAS, ZF, BZIP, bHLH and WRKY as corresponding targets. A comparative analysis between normal PCR with miRNA specific primers and deletion PCR with primers designed to avoid SSR regions revealed a difference in the size of the PCR products as shown in Fig. 3. Five sets of miRNAs were further chosen based on certain criteria like the type of SSR motif, miRNAs with transcription factors as targets, the length of SSR repeat unit etc. The PCR profile showed either an absence or difference in size of the amplicons, which indicated the possible deletions of SSR regions in pre-miRNAs. Deeper focused studies give way to open up the potential significant roles for SSRs in pre-miRNAs.
Fig. 1 (See legend on next page.)
Clues for SSR involvement in shaping SREs for alternative splicing events

The 149 SSR bearing pre-miRNAs identified in *A. thaliana* when subjected to RegRNA analysis, detected different functional RNA regulatory motifs. Among this, a most interesting and concurrent functional motif was Splicing Regulatory Element (SRE). The SREs were found to occur either in or near to the SSR motifs in pre-miRNA sequences. Out of the four SREs, the presence of Intronic Splicing Silencers (ISS) in most of the CT/TC SSR motif type were noticeable. Such CT/TC motifs were well sustained in most of the members of conserved miRNA families like miR156 and 157. In miR854 family members, a trinucleotide SSR type GGA was found to be conserved with Exon Splicing Enhancer (ESE) like activity. A striking existence of two different types of SSR motifs adjacent to SREs like Intronic Splicing Enhancers (ISE) [AG-ISE-CA] were also noticed among members of miR156 family. In addition to SREs, other functional RNA regulatory motifs associated with SSRs that were identified included Transcriptional Regulatory Motifs (TRM), Untranslated region motifs (UTRs), cis regulatory elements, noncoding RNA (ncRNA) hybridization regions, miRNA target sites etc. The AG motif was yet another SSR type which was conserved among miR8167 family members with potential function as TRMs. (see Additional file 2).

This together with the observation that SSR motifs are consistently maintained even in the highly evolved Chordata and Magnoliophyta increased the chances of promising functions for such SSRs. The significant matches of SSRs in pre-miRNAs with SREs made us to check the possible role of SSRs in determining SREs required for the process of AS. For this, a computational based deletion analysis of SSR motifs in sequences of pre-miRNAs was carried out. Such tailored pre-miRNAs, when subjected to SRE prediction showed that, in the absence of certain SSRs the corresponding SREs were not detected (Table 1). Out of the four different SREs like ESE, ESS, ISE and ISS, the ISS and ESE elements were found to be the most affected when the SSR motifs were deleted. Among the SSR motifs, the CT/TC type was found to be the most prominent and consistent which were predicted as ISS sites (Fig. 4). If the CT/TC motif were deleted, the ISS sites were not detected in the

---

**Fig. 1** Comprehensive Circos plot depicting the frequency and distribution pattern of tandem repeats occurring in the miRNA precursors across different organisms. Outermost circle (I): The names of individual organisms selected for the study whose details are given in Additional file 1. Subsequent Inner circle (II): Phyla based categorization of organisms: Chlorophyta (Chloro), Mycetozoa (Myce), Heterokontophyta (Hete), Embryophyta (Embr), Coniferophyta (Coni), Magnoliophyta (Magne), Porifera (Pori), Cnidaria (Cnid), Platyhelminthes (Plat), Nematoda (Nema), Annelida (Anne), Mollusca (Moll), Nemertea (Neme), Brachiopoda (Brac), Arthropoda (Arth), Deuterostoma (Deut), Hemichordata (Hemi), Echinodermata (Echi) and Chordata (Chor). Subsequent Inner circle (III): Kingdom based classification of organisms: Protista (Prot), Plantae (P), Animalia (A) and Viruses (V)

**Fig. 2** Distribution pattern of different types of SSR motif in the pre-miRNAs of *Arabidopsis thaliana*. The X-axis shows the different types of SSR motifs identified in the pre-miRNAs of *A. thaliana* and the Y-axis shows the relative count of each of the SSR motif identified.
corresponding pre-miRNA sequences. This initial result strengthened the possibility of SSR motifs to play a major role in shaping SREs to undergo AS.

Discussion
There exist different perspectives for SSRs like (1) hyper-variable molecular marker which is well addressed and demonstrated by its utility as molecular marker during fingerprinting studies; (2) biological effects of SSRs in genes and (3) interplay between SSRs and miRNAs. The distribution of SSRs in genes is non-random presumably because they are supposed to have a variety of putative functions. Reports suggest that SSRs present in both coding and non-coding regions can affect gene expression [1].
| Sl.no. | miRNA_Acc | SSR motif | ESE* | ESS* | ISE* | ISS* | ESE | ESS | ISE | ISS |
|-------|------------|-----------|------|------|------|------|-----|-----|-----|-----|
| 1     | ath-miR156a | ag        | no ISE |      |      | ag-ISE |
| 2     | ath-miR156b | ct        |      | no-ISS |      | ct |
| 3     | ath-miR156d | ct        |      | no-ISS |      | ct |
| 4     | ath-miR156e | ag        | no-ISE |      |      | ag-ISE-ca |
| 5     | ath-miR156f | ag        | no-ISE |      |      | ag-ISE-ca |
| 6     | ath-miR156g | ag        | no-ISE |      |      | ag-ISE-ca |
| 7     | ath-miR156h | ct        |      | no-ISS |      | ct |
| 8     | ath-miR156j | ag        | no ESE |      |      | ESE-ag |
| 9     | ath-miR157a | ct        |      | no-ISS |      | ct |
| 10    | ath-miR157b | ag        | no ESE |      |      | ag-ESE |
| 11    | ath-miR157c | at        | no-ESE |      |      | ESE-at |
| 12    | ath-miR157d | ct        |      | no-ISS |      | ct |
| 13    | ath-miR159b | ct        |      | no ISS |      | ct |
| 14    | ath-miR159c | ttc       | no-ESS |      |      | ttc |
| 15    | ath-miR163  | ct        |      | no-ISS |      | ct |
| 16    | ath-miR164b | ta        | no-ESE |      |      | ta-ESE |
| 17    | ath-miR165b | at        | no-ESE |      |      | at-ESE |
| 18    | ath-miR167c | ct        |      | no-ISE |      | ISE-ct |
| 19    | ath-miR167d | ca        | no-ESE |      |      | ca-ESE |
| 20    | ath-miR169b | ct        |      | no-ISS |      | at-ISE |
| 21    | ath-miR169j | ag        | no-ESE |      |      | ag-ISE |
| 22    | ath-miR169k | at        | no-ESE |      |      | ag-ISE |
| 23    | ath-miR169m | tc        |      | no-ISS |      | tc |
| 24    | ath-miR169n | ag        | no-ESE |      |      | ag |
| 25    | ath-miR170  | ct        |      | no-ISS |      | ct |
| 26    | ath-miR171a | ca        | no-ESE |      |      | ca |
| 27    | ath-miR173  | ag        | no-ESE |      |      | ag-ISE |
| 28    | ath-miR390b | ctt       |      | no-ESS |      | ctt |
| 29    | ath-miR391  | ga        | no-ESE |      |      | ga-ESE |
| 30    | ath-miR394a | at        |      | no-ISE |      | AT-ISE |
Table 1  Computational deletion analysis of SSR motifs in pre-miRNAs of *A.thaliana* identifies potential role for SSRs in shaping SRE elements (Continued)

| Sl.no. | miRNA_Acc | SSR motif | ESE* | ESS* | ISE* | ISS* | ESE | ESS | ISE | ISS |
|--------|-----------|-----------|------|------|------|------|-----|-----|-----|-----|
| 31     | ath-miR394b | ct        | no-ISE | no-ISS | at-ISE-ta |
| 32     | ath-miR398b | ac        | no-ESE | ac |
| 33     | ath-miR398c | ac        | no-ESE | ac |
| 34     | ath-miR399b | at        | no-ESE | ESE-at |
| 35     | ath-miR399f | ct        | no-ISE | ct |
| 36     | ath-miR408  | ga        | no-ESE | ga-ESE |
|        |            | tc        | no-ISE | tc |
| 37     | ath-miR777  | tc        | no-ISS | tc |
| 38     | ath-miR781b | at        | no-ISE | ta |
| 39     | ath-miR822  | aat       | no-ISS | aat |
| 40     | ath-miR824  | ct        | no-ISS | ct |
| 41     | ath-miR827  | ct        | ISE (ttcttttg) | no-ISS | tatt | ct |
| 42     | ath-miR838  | tct       | ESS (tatttatta) | tct |
| 43     | ath-miR847  | tct       | no ESS | tct |
| 44     | ath-miR855  | ta        | ISE (ttctttta) | ISE-ta |
| 45     | ath-miR857  | t'g       | ESS (tagacat) | tg |
| 46     | ath-miR858a | tc        | no ISS | tc |
| 47     | ath-miR858b | at        | no ESE | at |
| 48     | ath-miR859  | tc        | no ISS | tc |
| 49     | ath-miR862  | tg        | no ESE | tg-ESE |
| 50     | ath-miR863  | tatt      | no ESS | tatt |
| 51     | ath-miR1886.1 | ga       | no ISE | ga |
|        | ath-miR2111b | gaa      | ESE (ggataca) | gaa |
| 52     |            | tga      | ESE (ggataca) | tga-ESE |
| 53     | ath-miR2936 | ga        | ESE (aagaagct) | ga-ESE |
| 54     |            | tc        | no ISS | tc |
|        |            | tc        | no ISS | tc |
|        |            | ggcg      | ESE (aagaagct) | ggcg |
| 55     | ath-miR2937 | gt        | no ESE | gt-ESE |
| 56     | ath-miR3434 | tc        | no ISS | tc |
| 57     | ath-miR4240 | at        | no ISE | ISE-at |
| 58     | ath-miR5025 | at        | no ESE | at |
| 59     | ath-miR5027 | tg        | no ISE | tg |
| 60     | ath-miR5029 | ga        | no ESE | ga-ESE |
|        |            | tc        | no ISS | tc |
|        |            | atg       | no ESE | atg-ESE |
SSRs in the 5' UTR served as protein binding site thereby regulating translation [16]. SSR expansion in 3'UTR caused transcriptional slippage and produced expanded mRNA, which could accumulate as nuclear foci, and disrupt splicing, and other cellular function. Intronic SSR can affect gene transcription, mRNA splicing or export to cytoplasm.

Table 1: Computational deletion analysis of SSR motifs in pre-miRNAs of *A. thaliana* identifies potential role for SSRs in shaping SRE elements (Continued)

| Sl.no. | miRNA_Acc | SSR motif | ESE* | ESS* | ISE* | ISS* | ESE | ESS | ISE | ISS |
|--------|-----------|-----------|------|------|------|------|-----|-----|-----|-----|
| 61     | ath-miR5634 | tc        | no ISS | tc   |
| 62     | ath-miR5638a | ct        | no ISS | ct   |
|        |            | ac        | no ISS | ac-SS |
| 63     | ath-miR5638b | ga        | no ESS | ga-ESS |
|        |            | ct        | no ISS | ct   |
|        |            | ac        | no ISS | ac-SS |
|        |            | ct        | no ISS | ct   |
| 64     | ath-miR5640 | g₃        | no ESE | ga-ESE |
|        |            | ct        | no ISS | ct   |
| 65     | ath-miR5641 | tct       | no ESS | tct  |
| 66     | ath-miR5647 | ct        | no ISS | ct   |
|        |            | ct        | no ISS | ct   |
|        |            | ct        | no ISS | ct   |
| 67     | ath-miR5648 | ta        | ESE   | ESE-ta |
|        |            |          | (gaagaaa) |     |
| 68     | ath-miR5651 | at        | no ESE | at-ESE |
|        |            | at        | no ESE | ESE-at |
| 69     | ath-miR5652 | ga        | no ESE | ga-ESE |
| 70     | ath-miR5653 | gttga     | no ESS | ESS-gttga |
|        |            |           |       |       |
| 71     | ath-miR5655 | ag        | ESE   | ag   |
|        |            | ga        | ESE   | ga   |
|        |            | ag        | ESE   | ag   |
|        |            | tc        | ESE   | tc   |
|        |            | tgg       | ISE   | ISE-tgg |
|        |            | gga       | ESE   | gga  |
|        |            | cca       | ESE   | cca  |
| 72     | ath-miR5656 | tc        | no ISS | tc   |
| 73     | ath-miR5658 | ctt       | no ISE | ctt  |
|        |            | agg       | no ISS | ISS-agg |
| 74     | ath-miR5665 | tat       | no ISS | ISS-tat |
| 75     | ath-miR5997 | agg       | no ESE | agg  |
| 76     | ath-miR8166 | ga        | no ESE | ga   |
| 77     | ath-MIR8176 | tc        | no ISS | tc   |
| 78     | ath-miR8177 | ga        | no ESE | ga   |
| 79     | ath-miR8182 | ct        | no ISS | ct   |

Note: The SSR bearing pre-miRNA and the SSR deleted pre-miRNA (*) shows difference in the detection of SREs like ESE, ESS, ISE and ISS.
SSRs located in EST’s have a range of functions such as metabolic enzymes, structural and storage proteins, disease signaling, and transcription factors. A positive selection exists when the SSRs happen to occur near the transposons [17]. In Drosophila, a change in the 17-copy repeat of SSR in the period gene coding for Thr-Gly was found to affect the circadian rhythm maintenance [18]. Reporter assays predicted a putative enhancer-like function for TG repeats [19]. In rodents, the expression pattern of vasopressin 1a receptor (V1aR) is regulated by differences in SSR in the 5′ regulatory region which in turn affect the social behavior [20]. The expression of luciferase gene in reporter assays is directly proportional to the length of GA repeats thus can enhance the transcriptional output of a gene [21–23]. Repeats residing in the intronic regions also enhance gene expression like the TCAT repeat in the intron of Tyrosine Hydroxylase gene [24]. A significant enrichment of SSRs near the transcriptional start sites (TSS) (60 and 20% CCG and ACG found within 1 kb of TSS) was observed in humans. These examples indicate the importance and necessity of studying these SSRs and underline the fact that SSRs can be functional entities in the genome.

Out of the three perspectives, debates on the third perspective about the correlation between SSRs and miRNAs still exist. Very few reports suggest that SSRs are an important component of pre-miRNAs [10–12, 25, 26]. As the SSRs present in the genes are shown to have regulatory effects when associated with new miRNA candidates [27] and phenotypic effects [18, 20–24], we presume that the SSRs identified from pre-miRNAs in this study may have similar biological effects. Thus SSRs may be involved in the biological function of miRNAs or its biogenesis. Here, the presence of SSR bearing or SSR related miRNAs across different taxa was well demonstrated and the incidence ratio in each of the respective organism was portrayed. A miniature platform detailing each and every aspect of SSRs like extent of occurrence, the type of motif etc. was successfully generated from the study. Previous reports suggest that Repeat-related miRNAs (RrmiRs) are those miRNA genes having at least 50% of repetitive elements or 100% in one of the associated mature miRNA sequences [28]. Recently identified RrmiRs in fungi and animals include the piwi associated small interfering RNAs (rasiRNAs) and heterochromatic small RNAs (hcRNAs), processed from long double stranded RNA precursors [29, 30].

A comparative study on the type of SSRs across the taxa showed that dinucleotide SSRs were the predominant type (84.71%), whereas hexanucleotide SSRs (0.18%) were the least. Moreover the CT and AG dinucleotide SSR motifs were found to be consistent among the members of highly conserved miRNA families like miR854, miR156 etc. This preference for dinucleotide type of SSRs in pre-miRNAs
can be correlated to its probable functions. The frequencies of different repeats can vary considerably in different organisms. In humans, the A/T regions are more frequent and in Arabidopsis GA/CT repeats are more [31]. The 5′UTRs of Arabidopsis are reported to have relatively more number of AG/CT repeats, whereas the 3′UTRs of humans and catfish possess more number of AC/GT repeats [32]. The frequency of (A/T) n was high in the intronic regions of different species, (AC/GT) n was high in primates, rodentia, mammalian, vertebrata, arthropoda, fungi etc. and CG/GC repeats were more in Caenorhabditis elegans, yeast and embryophyta. Among the dinucleotides identified from the current study, the CT motif was found to be the most frequent SSR. This can be correlated to its function. Earlier reports suggest that there is an increased probability for CT repeats to play a major role in the transcrip-
tion of miRNA genes. CT repeats are reported to form non-B DNA that play important potential roles in gene transcription activation [33, 34]. Similar abundance of dinucleotide simple repeats like (CA) n and (TG) n were reported in the largest mir-467 family in mouse [27]. Each SSR generated might be the product of repeated mutations and cross-overs that might have occurred during the course of evolution. The resulting SSR type observed in a particular pre-miRNA may be the requisite of that particular pre-miRNA to undertake its specific function in the right way. We believe that ‘demand tunes the function by changing the sequence preference’. Also, the number (n) of times a particular type of SSR is repeated may or may not affect its putative function. One of the best reported examples is the fragile X syndrome (FXS), a triplet expansion disease (TRED), which is the most common neuropsychiatric and mental retardation disorder in humans [35]. When there is an expansion of a trinucleotide CGG repeat located in the 5′UTR of FMR 1 gene, to over 200 copies, it results in the deficiency of FMRP protein, which is required for normal neuronal development and plasticity.

The amplicon profile observed after deletion PCR is indeed a strong evidence of the study effects of deletion of SSRs in pre-miRNAs. This together with the computational identification of SREs either in or near to SSRs made us to presume that SSRs are involved in shaping AS to generate variant miRNAs during stress environments. SREs are sequences in exons and introns that are important for constitutive splicing as well as alternative splicing. They function either as splicing enhancers or suppressors and affect splice site choice [15]. Our preliminary identification of SRSs either adjacent to or as SREs with splicing activity is the first implication for likely involvement of SSRs in Alternative Splicing (AS). A possible explanation for the presence of SSRs in pre-miRNAs is that the SSRs may fine tune the Alternative Splicing (AS) events in pre-miRNAs which contributes to different isoforms of miRNAs. As the miRNAs are tissue specific and developmental stage specific, each miRNA formed in response to a stress factor has a specific role. About 61 and 95% of intron containing protein-coding genes in Arabidopsis and humans are reported to undergo AS [36, 37] and the stress responsive miRNAs were reported to be G/C rich in Arabidopsis [38]. The miRNAs with UGUGU sequences are said to activate the targets associated with carcinogenesis in humans [39]. The correlation between AS and miRNAs was well demonstrated [40], where competition between the splicing machinery and the miRNA processing machinery comes into play. It is assumed that when the splicing machinery does not recognise the internal exon, the miRNA processing components bind to pre-miRNA splice junction, thereby leading to the formation of pre-miRNA and a skipped isoform. Whereas when the splicing components recognise the internal exon, pre-miRNA is not formed, instead an isoform bearing the internal alternative exon is formed. It is known that miRNAs are generated either from intergenic or intronic regions of coding or noncoding genes [41] and splicing and processing of intronic miRNAs may affect each other [34, 42]. A characteristic GU dinucleotide at the 5′end and AG at the 3′end is noticed for the canonical U2 type introns; whereas AU and AC dinucleotides at the 5′ and 3′ ends were noticed for U12 type introns. This strengthens the intron retention process that may happen during AS events and this may also be another reason for occurrence of ‘tandem repeats’ in such hairpin precursor sequences.

Conclusions

The higher mutation rate of SSRs during recombination, polymerase slippage, DNA replication or repair, unequal crossing over etc. will finally end up with a change in number of repeat units of SSRs. This change may or may not become beneficial, depending on the incidence or position of these SSRs in a gene. The presence of SSRs in different locations that have an impact on genome strongly suggests that these SSRs should be considered significant and are not to be discarded as ‘nonsense’. Debates regarding the functional aspects of SSRs are never-ending. Those SSRs which are associated with miRNAs are speculated to have potential functions other than the conventional marker based assays. We speculate that there can be a tug of war between AS and miRNA biogenesis, which may in turn be affected, when there is a change in the number of repeat units (n) present in pre-miRNAs. All the three i.e. AS, SSRs and MIR genes are a complex interconnected network among which AS may be one of the crucial steps in miRNA biogenesis, which determine the formation of miRNAs.
in accordance with the external stress factors, whereas both AS and miRNA can be affected if a change in (n) occurs.

**Methods**

All the available miRNA precursors (miRBase v.21) of different taxa {Chromalveolata, Metazoa, Mycetozoa, Viridiplantae and Viruses (see Additional file 1)} were extracted for the current study from public database, miRBase (www.mirbase.org). Among the miRNA precursors of the same genus but different species, the species with more number of miRNAs were included. Simple Sequence Repeat Identification Tool (SSRIT) [43] (http://archive.gramene.org/db/markers/ssrtool markers/ssrtool) was used to study the frequency, type and distribution pattern of SSRs in each individual sequences.

To identify the presence of different functional RNA motifs including Splicing Regulatory Elements (SREs), the SSR bearing pre-miRNA transcripts were subjected to an integrated web server RegRNA 2.0 analysis [44]. The pre-miRNAs bearing SSRs with SRE activity were selected for further analysis. To figure out whether the SSR motif had any effect in determining SRE elements, a computational deletion was carried out manually for each of the SSR sequence motif in their corresponding pre-miRNAs. Then the trimmed pre-miRNAs (after the deletion of SSR motifs) were again subjected to RegRNA analysis to understand the effects of SSRs. An experimental approach to generate deletion constructs for SSRs in pre-miRNAs was carried out using normal PCR with miRNA specific forward and reverse primers and deletion PCR with primers designed to avoid the SSR region in the pre-miRNAs. Out of 15 sets of miRNAs, five sets of miRNAs were chosen for the study which included miR156b, miR164b, miR166f, miR167c and miR2936. The details of the primer sequences are given in Additional file 3. Total RNA was isolated from in vitro seedlings of *Arabidopsis thaliana* in Additional file 3. Total RNA was isolated from in

**Acknowledgements**

We thank Dr. Martin Krzywinski, Staff Scientist, Canada’s Michael Smith Genome Sciences Centre, for the Circos plot generated, Department of Science and Technology and Department of Biotechnology, Govt. of India for the financial support.

**Availability of data and materials**

The pre-miRNAs used for the analysis in this work were extracted from publicly available database of miRNAs –mirBase (www.mirbase.org).

**Authors’ contributions**

NJ and EVS conceived the experiment(s), NJ and M.B.Y.P conducted the experiment(s) and analysed the results. All authors reviewed the manuscript. All authors read and approved the final manuscript.

**Ethics approval and consent to participate**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Publisher’s Note**

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

**Received:** 11 December 2017 **Accepted:** 30 April 2018

**Published online:** 09 May 2018

**References**

1. Krishnan J, Mishra RK. Code in the non-coding. Proc Indian Nat Sci. 2015; 81(3):609–28.
2. Litt M, Luty JA. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am J Human Gen. 1989;44:397–401.

3. Kim PM, et al. Analysis of copy number variants and segmental duplications in the human genome: evidence for a change in the process of formation in recent evolutionary history. Gen Res. 2008;18:1865–74.

4. Ayers NM, McClung AM, Larkin PD, Bligh HF, Jones CA, Park WD. Microsatellites and a single nucleotide polymorphism differentiate apparent array-like alleles in an extended pedigree of US rice germplasm. Theor Appl Genet. 1997;94:773–81.

5. Bao J, et al. QTL for rice grain quality based on a DH population derived from parents with similar apparent amylase content. Euphytica. 2002;128:317–24.

6. Yaish MWF, Pérez De La Vega M. Isolation of (GA)n microsatellite sequences and description of a predicted MADS-box sequence isolated from common bean (Phaseolus vulgaris L.). Gen Mol Biol. 2003;26:337–42.

7. Li CD, Zang XQ, Eckstein P, Rossnagel BG, Scoles GJ. A polymorphic microsatellite in the limit dextrinase gene of barley (Hordeum vulgare L.). Mol Breed. 2003;3:569–77.

8. Fujimori S, et al. A novel feature of microsatellites in plants: a distribution gradient along the direction of transcription. FEBS Lett. 2003;554:17–22.

9. Grabcký E, Kumari D, Uslin K. Fragile X syndrome and Friedrich’s ataxia: two different paradigms for repeat induced transcript insufficiency. Brain Res Bull. 2001;56:367–73.

10. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell. 2004;116:281–300.

11. Joy N, Asha S, Mallika V, Soniya EV. An enhanced gene prediction method: ARTS: a transcriptome fingerprinting strategy to detect functionally relevant microsatellite mutations in cancer. Nucleic Acids Res. 2003;31:e33. https://doi.org/10.1093/nar/gng033.

12. Yuan Z, Sun X, Liu H, Xie J. MicroRNA genes derived from repetitive elements and expanded by segmental duplication events in mammalian genomes. PLoS One. 2011;6:e17666. https://doi.org/10.1371/journal.pone.0017666.

13. SanMiguel P, et al. Nested retrotransposons in the intergenic regions of the maize genome. Science. 1996;274:765–8.

14. Farazi TA, Jurank S, Tusch T. The growing catalog of small RNAs and their association with distinct Argonaute/Piwi family members. Development. 2008;135:1201–14.

15. Borchert GM, Lanier W, Davidson BL. RNA polymerase III transcribes human microRNAs. Nat Struct Mol Biol. 2006;13:1097–101.

16. Morgante M, Hanafey M, Powell W. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nat Genet. 2002;30:194–200.

17. Li YC, Konol AB, Falmia T, Nevo E. Microsatellites within genes: structure, function, and evolution. Mol Biol Evol. 2004;21:991–1007.

18. Yaish MWF, dela Vega MP. Isolation of (GA)n microsatellite sequences and description of a predicted MADS-box sequence isolated from common bean (Phaseolus vulgaris L.). Genet Mol Biol. 2003;26(3):337–42.

19. Yan K, et al. Stress-induced alternative splicing provides a mechanism for the regulation of microRNA processing in Arabidopsis thaliana. Mol Cell. 2012;48:521–31.

20. Kelley K, Chang SJ, Lin SL. Mechanism of repeat-associated microRNAs in fragile X syndrome. Neural Plasticity. 2012;2012:104796. https://doi.org/10.1155/2012/104796.

21. Marquez Y, Brown JW, Simpson C, Barta A, Kalyna M. Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis. Gen Res. 2012;22:1194–95.

22. Pan Q, et al. Quantitative microarray profiling provides evidence against widespread coupling of alternative splicing with nonsense-mediated mRNA decay to control gene expression. Gen Dev. 2006;20:153–8.

23. Mishra AK, Agarwal S, Jain CK, Rani V. High GC content: critical parameter in the regulation of microRNA processing in Arabidopsis thaliana. Bioinformation. 2009;4:151–4.

24. Rolle K. The sequence and structure determine the function of nature human microRNAs. PLoS One. 2016;11(5):e0151246. https://doi.org/10.1371/journal.pone.0151246.

25. Melamed ZE, et al. Alternative splicing regulates biogenesis of microRNAs located across exon-intron junctions. Mol Cell. 2013;52:869–81.

26. Sains HK, Griffiths-Jones S, Enright AJ. Genomic analysis of human microRNA transcripts. Proc Nat Acad Sci. 2007;104:17719–24.

27. Janas JM, et al. Feed-forward microprocessing and splicing activities at a microRNA–containing intron. PLoS Gen. 2011;7:1002330. https://doi.org/10.1371/journal.pgen.1002330.

28. Teinyk S, DeClerck G, Lukasheva A, Lipovich L, Cartinhour S, et al. Computational and experimental analysis of microsatellites in rice (Oryza sativa L.): frequency, length variation, transposition associations, and genetic marker potential. Gen Res. 2001;8:1441–52.

29. Chang TH, Huang HY, Hsu JB, Weng SL, Horng JT, Huang HD. An enhanced computational platform for investigating the roles of regulatory RNA and for identifying functional RNA motifs. BMC Bioinform. 2013;14(Suppl 2):S4.