Genomics 112 (2020) 3354–3364

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Uncloaking lncRNA-meditated gene expression as a potential regulator of CMS in cotton (*Gossypium hirsutum* L.)

Rasmieh Hamid[a,*], Feba Jacob[b], Hassan Marashi[a], Visha Rathod[c], Rukam S. Tomar[d]

[a] Department of Biotechnology and Plant Breeding, Ferdowsi University of Mashhad, Iran
[b] Centre for plant biotechnology and molecular biology, Kerala agricultural university, Thrissur, India
[c] Institute of Science, Nirma University, Ahmedabad, Gujarat, India
[d] Department of Biotechnology and Biochemistry, Junagadh Agricultural University, Junagadh, Gujarat, India

ARTICLE INFO

Keywords:
Long non-coding RNAs
Male sterility
Cotton
Micro RNA

ABSTRACT

Cytoplasmic male sterility is a well-proven mechanism for cotton hybrid production. Long non-coding RNAs belong to a class of transcriptional regulators that function in multiple biological processes. The cDNA libraries from the flower buds of the cotton CGMS, it's restorer (Rf) and maintainer lines were sequenced using high throughput NGS technique. A total of 1531 lncRNAs showed significant differential expression patterns between these three lines. Functional analysis of the co-expression network of lncRNA-mRNA using gene ontology vouchsafes that, lncRNAs play a crucial role in cytoplasmic male sterility and fertility restoration through pollen development, INO80 complex, development of anther wall tapetum, chromatin remodeling, and histone modification. Additionally, 94 lncRNAs were identified as putative precursors of 49 miRNAs. qRT-PCR affirms the concordance of expression pattern to RNA-seq data. These findings divulge the lncRNA driven miRNA-mediated regulation of gene expression profiling superintended for a better understanding of the CMS mechanisms of cotton.

1. Introduction

The inability of plants to generate viable pollens is referred to as cytoplasmic male sterility (CMS) which is the outcome of the collaboration between mitochondrial and nuclear genes or as genic male sterility (GMS) when encoded exclusively by nuclear genes [1]. Male-sterile lines are crucial for the production of hybrid seeds, as the requirement of laborious sterilization techniques, in a wide group of crops is eliminated. Scientists couldn't uncloak the detailed mechanism of CMS even after two decades of research. Understanding the molecular mechanism causing plant male sterility, is useful for hybrid breeding and thus is economically important [2]. The non-coding (nc) RNA molecules have a huge contribution to the regulatory pathways controlling anther and pollen development in plants. This aspect is being widely studied and documented [3].

The identification and characterization of the non-coding RNA molecules and the comprehensive understanding of their interactions with the coding RNAs can give an idea about the resulting biological processes. Non-coding RNA can be categorized as small or long non-coding RNAs based on their lengths. lncRNA is the most diverse regulatory group among all the non-coding RNAs. Long intergenic/intervening non-coding (linc) RNAs denote a class of long non-coding transcripts of approximately 200 bp length and are transcribed from the intergenic, non-coding regions. They are also capped and polyadenylated like mRNA [4,5], however, all the lncRNA are not polyadenylated [6]. The lncRNAs in plant control gene expression through various processes like i) interacting with nuclear speckle RNA-binding (NSR) protein to modulate alternative splicing [7] ii) and mRNA translation modulation [8], iii) mechanism of target mimicry and iv) masking of the splicing sites and signals by directly interacting with the target RNA and forming lncRNA–RNA duplex [9]. lncRNAs could suppress sense transcription by acting as antisense transcripts [10] as well as through chromatin remodeling [11]. There is an active transcription of lncRNAs in the reproductive organs of plants and animals, in tissue and cell-specific manner, especially in the male reproductive organs [12–15].

Our knowledge about lncRNA function has been greatly expanded by the global analyses of non-coding transcriptomes. LncRNAs were

Abbreviations: lncRNA, long non-coding RNA; miRNA, micro RNA; CGMS, cytoplasmic genic male sterility; FDR, False Discovery Rate; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; RT-qPCR, Quantitative real-time PCR

* Corresponding author at: Department of Biotechnology and Plant Breeding, Ferdowsi University of Mashhad, Iran.
E-mail address: rasihamid@gmail.com (R. Hamid).

https://doi.org/10.1016/j.ygeno.2020.06.027
Received 13 April 2020; Received in revised form 3 June 2020; Accepted 17 June 2020
Available online 20 June 2020

found to be master regulators as they control the expression of genes encoding various proteins in many scientific studies [7]. Moreover, many lncRNAs perform critical roles, with miRNAs as their targets or precursors [16]. For example in maize, lncRNAs behave as precursors for miRNAs and use a miRNA-dependent mechanism to regulate gene expression [17]. Wang et al. investigated the contribution of lncRNAs in regulating lignin metabolism in cotton fibers via miR397 [18]. Additionally, miR399 targets the mRNA PHO2 and thus regulates phosphate homeostasis in Arabidopsis. However, the role of miR399 is inhibited by lncRNA IPS1 which acts as a target mimic and prevents its binding to PHO2 [19,20]. Huang et al. predicted 15 lncRNAs as eTMs for 13 miRNAs in Brassica, of which two lncRNAs were shown to be functional eTMs for miR160 and to function in pollen development [21,22]. Although a lot of studies are being conducted for deciphering plant male sterility, the impact of lncRNA molecules on male sterility is yet to be uncovered in different crops. We can gain an insight into how the pollen sterility is caused due to damaged crosstalk between organelle and nuclear genomes, by decoding the role of lncRNAs in retrograde signaling.

In this study, the lncRNAs profile of CGMS line and its fertile lines were compared using RNA-seq in cotton. The differential expression patterns of cotton lncRNAs-mRNA were established. The identified lncRNAs were functionally annotated using an algorithm for network propagation. Furthermore, by analyzing the interactions between the cotton miRNAs and the identified lncRNAs, the potential regulatory traits of lncRNAs in pathways leading to cytoplasmic male sterility and fertility restoration can be revealed. Our study unveils the potential regulatory roles of lncRNAs in cotton and lays the groundwork for further research to reveal their precise functions in mechanism elucidating CMS in cotton.

2. Materials and methods

2.1. Plant materials, RNA isolation and sequencing

The three lines of cotton (*Gossypium hirsutum* L.) plants i.e. CGMS (line JA178), maintainer (line JB178), and restorer (line JR178) used for this study was provided by the Department of Agricultural Biotechnology, JAU, Gujarat, India. These plants were cultivated in a greenhouse with proper temperature and humidity control. After 2 months of plant growth, the flower buds at two developmental stages i.e. sporogenous cell stage (SS) (1.5–2.2 mm size) and microsporocyte stage (MS) (2.2–2.6 mm size) were carefully excised. To reduce variation between plants, samples from thirty plants from each line, were pooled together with duplicates for each plant. The pooled material was stored in liquid nitrogen, after careful removal of the external protecting part of the bud, to extract RNA.

The extraction of total RNA from the anthers of CGMS and its fertile lines at different stages was done using a previously described protocol [23]. The quantity and concentration of RNA were measured using NanoDrop2000 (NanoDrop Technologies, DE, USA) and the integrity was verified using 1% agarose gels. DNA contamination was eliminated by giving a treatment of DNase I (Qiagen, CA, USA). The quantification of total RNA isolated was done using Qubit 3.0 (Invitrogen, CA, USA) while the analysis of RNA integrity measured as RNA integrity number (RIN) was done using RNA Nano 6000 Lab chip (Agilent Technologies, CA, USA) on Bioanalyzer 2100 (Agilent Technologies, CA, USA). The Epicentre Ribo-zero RNA Removal Kit (Epicentre, Madison, USA) was used for removal of rRNA from the total RNA, following which RNA fragmentation and reverse transcription to double-stranded cDNA using random primers was performed. Subsequently, end repair, adenylating 3′ ends, ligating adapters, and enriching DNA fragments experiments were performed sequentially to generate RNA sequencing cDNA library. The quality of cDNA was evaluated using the Agilent Bioanalyzer 2100 system. Quant-iT™ PicoGreen® dsDNA Assay Kit (Thermo Fisher Scientific, MA, USA) was employed for the clustering of the sample. The cDNA library replicates of each sample were pooled in equimolar concentration for transcriptome sequencing. The library preparation, clonal amplification and RNA sequencing were done as described in our previous experiment [24].

2.2. Quality assessment and mapping of sequenced reads

The prinseq-lite.pl (v0.20.4) was used for filtering out raw reads with a Phred score < 25 (https://sourceforge.net/projects/prinseq/) followed by verification using the visualization tool FastQC (v0.11.7) (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The resultant reads with high quality from each sample were aligned using STAR aligner (v2.5.2a) to the cotton reference genome (http://mascotton.njau.edu.cn/info/1054/1118.html) [25]. Each mapped read from each sample was separately assembled using Cufflinks (v2.2.1) [26]. For running Cufflinks for lncRNAs, '-library_type_fr_firststrand' and 'minimum_fragments_per_transfrag = 0' with other default parameters. Read counts from each sample were combined using cuffmerge to produce a merged transcript information file while different lncRNAs were identified using cuffcompare based on class code annotated to each transcript.

2.3. Assessing coding potential for prognostication of lncRNA

The candidate lncRNAs transcripts were filtered by removal of transcripts having length ≤ 200 bp; with multiple exomes, and having ORF which code for > 100 amino acid long protein. Transcripts were categorized based on the cuffcompare class code according to its position on the reference genome. The transcripts were index into four categories viz., transcript fragment cascading within the reference genome (i), transcripts of intergenic origin (u), exon converging a ref_genome on the opposite strand (x), and an intron of the transcript fragment intersecting an intron of the ref_genome on the opposite strand (s). The protein-coding potential was also examined for these primarily selected lncRNAs, using three tools Coding Potential Calculator (CPC) (http://cpc.cbi.pku.edu.cn) [27], Coding-Potential Assessment Tools (CPAT) [28] and Predictor of Long non-coding RNAs and messenger RNAs based on an improved K-mer scheme (PLEK) [29,30]. Transcripts were considered as non-coding if they had a and PLEK score < 0, CPC score < 0 and a CPAT score ≤ 0.364 and were further screened using the Blast2GO tool, for any probable protein domain in Pfam database [31].

2.4. Expression target gene analysis of lncRNAs

The calculation of fragments per kilobase of transcript per million mapped (FPKMs) of both lncRNAs and coding genes was done using Cufflinks 2.1.1 software for every sample [32]. These calculated FPKM values were verified based on their log2fold change of FPKM values by Cuffdiff, and the samples showed significant differential expression for both mRNA and lncRNA when the absolute value of log2 (fold change) was ≥ 2 with p-value $\leq .05$ between two groups. The target genes 10 kb on both the upstream and downstream regions of lncRNAs were identified and their functions were analyzed. The GO seq R package (version 3.6, Melbourne, Australia) was used for the gene ontology (GO) enrichment analysis. The GO terms having ≤ 0.05 false discovery rate (FDR) value was used as the threshold to estimate significant enrichment [33]. The significant genes were allotted to the pathways where they function, by KEGG database (Kyoto Encyclopedia of Genes and Genomes) [34] Biological relationship between genes, lncRNA-mRNA, as well as genes-pathways, were deciphered by Cytoscape [35].

2.5. Prediction of putative targets and endogenous target mimics for miRNAs

The prediction of miRNA targets of lncRNAs or mRNAs was done

using the psRNATarget server (http://plantgrn.noble.org/psRNATarget/), for which less than three mismatches in targets and miRNA pairing regions were permitted. The eTMs for miRNAs were predicted by combining psRNATarget and application of the rules established by Wu et al. [19]. The miRNAs precursors were predicted by mapping the sequence of the mature miRNA to the putative lncRNAs sequence through local BLAST and no mismatch was allowed. Subsequently, the prediction of stem-loop structures was done using Mfold (http://unafold.rna.albany.edu/?q=mfold) as suggested by Jones-Rhoades [36].

2.6. Construction of miRNA-lncRNA-mRNA regulatory networks

The role of lncRNAs was understood by constructing a miRNA–lncRNA–mRNA network based on differentially expressed lncRNAs and miRNAs, and the target pairs of miRNAs–lncRNAs, miRNAs–mRNAs, and lncRNAs–mRNAs. The regulatory networks contained miRNAs, lncRNAs acting as miRNA targets, mRNAs acting as lncRNA targets, and mRNAs acting as miRNA targets. Cytoscape 3.7.1 software was used to visualize the regulatory networks of miRNA–lncRNA–mRNA [37].

2.7. Validation of lncRNAs, miRNAs, and mRNAs expression by RT-qPCR

The relative expression of lncRNAs, miRNAs, mRNA was validated through quantitative reverse transcription PCR (RT-qPCR) using specific primers as described previously [36]. Total RNAs and miRNAs were extracted from the same samples used previously for RNA-seq, and reverse-transcribed to cDNA using the TransScript® miRNA First-Strand cDNA Synthesis SuperMix kit (TransGen, Beijing, China) and PrimeScript— RT reagent kit (Takara, shiga, Japan) following the manufacturers' guidelines. For the qRT- PCR analysis, 7500 Fast real-time PCR system (Thermo Fisher Scientific, MA, USA) was used with three technical replicates for each of the two biological replicates as described [38] and SYBR green dye provided in KAPA SYBR® FAST qPCR kit (Kapa Biosystems, MA, USA) was used. The internal control gene used for normalization was the Elongation Factor-1 Alpha (EF1α) [38]. The relative expression of the selected miRNAs, lncRNAs, and target genes was calculated using the Livak's 2-ΔΔCt method. The primer sequences utilized for the analysis are listed in Additional file 1.

3. Results

3.1. Identification and characterization of lncRNAs in cotton

RNA-seq data obtained from 3 different lines i.e. cotton male sterile (line JA178), maintainer (line JB178) and restorer (line JR178) were used for the comprehensive identification of lncRNAs in cotton. More than 182,000,000 raw sequence reads were obtained through RNA-seq. An average of 94%, out of all the reads that were quality filtered reads, showed a Phred score of more than Q25 (Additional file 1). The STAR program was used to process the clean reads for each sample and the Cufflinks program was operated to process the clean reads for each sample using cotton (http://mascotton.njau.edu.cn/info/1054/1118) genome as a reference. Almost 86%, 87%, and 88% of the reads were aligned to the reference genome for JA178, JB178, and JR178, respectively (Additional file 2). A group of the non-overlapping and unique transcripts was retained using the transcriptome assembly obtained after Cuffmerge, resulting in a total of 63,565 transcripts with the average length of the transcripts being 238 bp.

The putative lncRNAs were scrutinized using filtering criteria including, single-exon transcripts located within 500 bp of other transcripts were first filtered out, followed by exclusion of transcripts with length ≤ 200 bp. Additionally, the transcripts were screened based on expression level (FPKM ≥0.5 for multi-exon transcripts, or FPKM ≥2 for single-exon transcripts), and finally, the known non-lncRNAs were filtered out. In toto, 3084 transcripts were taken as lncRNA candidates after these basic filtering processes. The remaining transcripts were evaluated for their coding potential through CPC (coding potential calculator) and Pfam protein domain analyses (Fig. 1). Those transcripts that had sequence similarity to the protein domains and annotated proteins, available in these databases were discarded. The 288 transcripts that showed the CPC score more than zero and 765 transcripts inside the 500 bp flanking regions of the predicted coding genes were also discarded (Fig. 1, and Supplementary Fig. 1). Thus, a total of 2031 lncRNAs were obtained in our study, including 1769 intergenic lncRNAs (87%), 185 intronic lncRNAs (9%), 77 antisense lncRNAs (3%), (Additional file 3, Fig. 2a).

A feature profiling of a total of 2031 identified lncRNAs (2031) was investigated and compared with those of mRNAs in cotton. The distribution pattern of lncRNAs and mRNAs was similar on the various chromosomes of the cotton genome (Fig. 2b). The lncRNAs varied in length from 201 bp to 33,098 bp and most of them (> 79%) had ≤1000 bp length (Fig. 2c). The average lncRNA length was 812 bp, which was quite less than the average length of mRNAs (3955 bp). the lower average length of predicted cotton lncRNAs in other plants, such as rice (800 bp) [39], cucumber (322 bp) [40] and Ganoderma (609 bp) [39] Almost 61% of the lncRNAs contained two exons, and 29% had multiple exons, whereas almost majority of the mRNAs contained more than two exons 75%. Comparing lncRNAs and mRNAs showed that the FPKM of lncRNAs was lower than that of mRNAs during anther development (Additional file 3).

3.2. Analysis of differentially expressed lncRNAs in JA178, JB178 and JR178 lines

Three comparisons of lncRNA expression levels were carried out: i) JA178 – JB178, having the isogenic nuclear genomes but different cytoplasm contain and fertility; ii) JA178– JR178, were different infertility and Rf1 alleles but had the same CMS cytoplasm; and iii) JB178– JR178, both isogenic as well as fertile but varied cytoplasmic contents with Rf1 alleles constitution. A total of 73, 117, and 40 unique lncRNAs were expressed specifically in the JA178, JB178, and JR178 lines, respectively, whereas, 1223 lncRNAs were expressed in common among all three lines (Fig. 2d). Based on the expression level, lncRNAs were considered to be differentially expressed in different samples if they had greater than two-fold change and a p-value < .01. A total of 518, 1167, and 1094 lncRNAs were identified to be differentially expressed in the JA178– JB178, JA178– JR178, and JB178– JR178 comparisons, respectively (Additional file 3). The differentially expressed lncRNAs represented the 1531 non-redundant lncRNAs that were scattered among the 26 chromosomes and different scaffolds (Additional file 2). Chromosome A11 had the highest number of lncRNA loci (109), followed by chromosome 5 (94) and chromosome 8 (67), whereas chromosome 2 had the lowest number of lncRNA loci (35) Attributing lncRNAs to subgenomes showed that there were 190 more lncRNAs in subgenome at than that in DD genome (Additional file 4).

3.3. GO and KEGG enrichment analysis of LncRNAs target genes (LTGs)

The annotation of the putative functions of lncRNAs is very crucial as they are observed to play major roles in several biological processes [18]. As lncRNAs are not conserved across different species, its functional prediction cannot be done using a sequence homology approach. A few approaches have been suggested for the functional annotation of lncRNAs recently [41]. The assumed function of lncRNAs involved in cotton cytoplasmic male sterility was inferred by studying the protein-coding genes that were less than 10 kb away from the differentially expressed lncRNAs and that were co-expressed with them [42]. On gene ontology enrichment analysis, it was observed that the enrichment this protein-coding was mainly in reproduction, pollen development,

chromatin remodeling, INO80 complex and cell wall organization (Fig. 3a)., Besides these, GO-term enrichments were also observed for cellular component (GO:0072669, tRNA-splicing ligase complex and GO:0003676, GO:0005840, ribosome), biological process (for example, GO:0048658, anther wall tapetum development, GO:0022402, cell cycle process; GO:0006355, regulation of transcription, DNA-templated; and GO:0016570, histone modification) and molecular function (GO:0001071, nucleic acid binding transcription factor activity; GO:0016772, transferase activity, transferring phosphorus-containing groups; and GO:0016301, kinase activity) (Additional file 5). The lncRNA targets were statistically enriched using the KEGG pathway analysis. There was an extensive-expression bias of lncRNAs in Cell cycle-G1/S transition, Ribosome, Starch and sucrose metabolism, RNA degradation and Wnt signaling (Additional file 5) (Fig. 3b).

3.4. LncRNAs as a potential regulator of miRNAs

LncRNAs and miRNAs, both belong to the non-coding RNA class and have pivotal roles in regulating various processes related to the development of plants [43]. Withal, lncRNAs is that it acts as endogenous target mimics (eTMs) of miRNAs, and thus regulates the miRNA activity, which in turn regulates gene expression [44]. Various biological processes like plant growth, differentiation and reproduction are regulated by the interaction between the lncRNAs and miRNAs [39]. We employed the software psRNATarget [42] and used the predicted lncRNAs to find out the potential miRNA targets and the results indicated that 88 lncRNAs were putative targets of 57 cotton miRNAs (Additional file 6). Most of miRNAs found interacting with lncRNAs were from ghr- miR-156, ghr-miR-169, miR-167, ghr-miR399b, ghr-miR-396, ghr-miR-319, ghr-miR-172 and ghr-miR159 families. As shown in Fig. 4, there were different interaction pattern, including one lncRNA with many miRNAs, many lncRNAs with one miRNA, and many lncRNAs with many miRNAs were identified (Fig. 4 and supplementary Fig. 2). The roles of these miRNA families in the regulation of anther development have been reported [45,46]. The major miRNA families regulating flowering time by controlling the transition from juvenile-to-adult vegetative phase are miR156, miR172 and miR390 while the families miR159, miR169, and miR399 along with miR172 help in the transition from vegetative to reproductive phase [47]. It was observed that there was an association between the GO terms of the cotton lncRNAs and the known function of their interacting miRNAs. For example, highly conserved miRNA families like miR-156, miR-399 and miR-169 had a significant role in plant reproduction [48]. The members of the families miR-156 and miR-399 targeted the lncRNAs TCONS_047188 and TCONS_046056, which showed association with GO terms, such as reproduction (GO:0000003) and specification of pollen development (GO:0009555).

The former reports also have revealed the importance of lncRNA-miRNA interactions in a variety of biological functions. In chickpea, it was observed that miR-159-a/b and miR-408b were involved in cleaving lncRNAs linked with the reproductive processes [49], while in rice, reproductive processes were under the regulation of miR-160 and miR-164 [12]. lncRNAs, through their interaction with miRNAs, disrupt the miRNA-mRNA interactions, thus regulating gene expression and biological processes. Hence, our work puts forward evidence for lncRNA-miRNA interactions in cotton and their significance in the development of reproductive organs.

The regulatory networks of lncRNAs involved in CMS and fertility restoration were explored by selecting the target genes of the differentially expressed lncRNAs and miRNAs based on the RNA-seq data and a putative miRNA–lncRNA–mRNA regulatory network was constructed by Cytoscape software (Fig. 5a, Additional file 7). The networks, thus constructed, comprised of 76 miRNAs, 172 lncRNAs, and 384 mRNAs and these predicted target genes regulated by miRNAs and lncRNAs, were divided into multiple groups, Several genes showed critical roles in tapetal differentiation and function [63] For example, the regulation

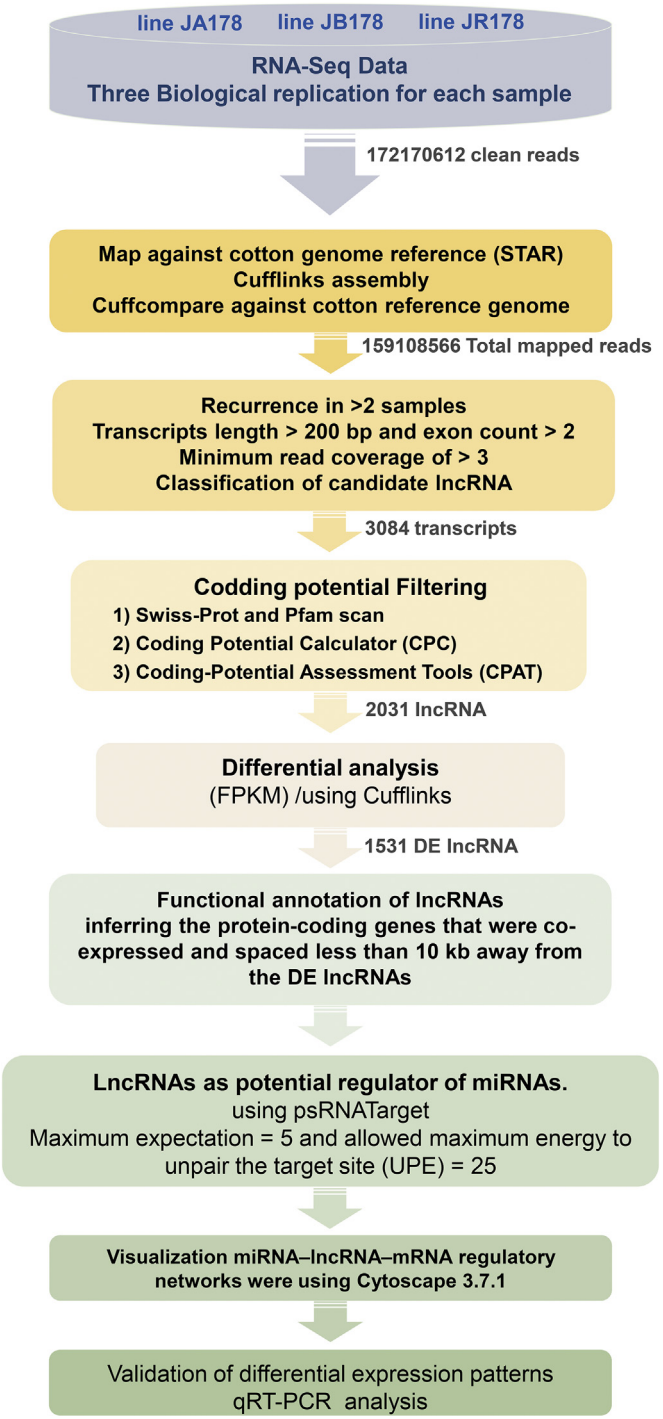

Fig. 1. A detailed workflow performed for identification of lncRNAs in CGMS, maintainer and restorer lines of *Gossypium hirsutum* L.

of BHLH transcription factor (Gh_A07G0112, Gh_A11G1614, Gh_A12G0337, Gh_D10G0179 and Gh_D12G0328), MYB transcription family (Gh_A05G2120, Gh_D07G2090 and Gh_A09G1458), AGAMOUS-like superfamily (Gh_A02G1782 and Gh_A12G2048) and ARF (Gh_D06G0360) transcription factors were controlled by the miRNAs and lncRNAs associated with regulation of gene expression processes. There was a down-regulation of all these genes in the sterile line in comparison to the fertile lines. These genes and several other functional genes from the expansin family and pentatricopeptiderepeat-containing (PPR) proteins regulated by miRNAs and lncRNAs are involved in the regulation of pollen maturation and maybe having vital roles in CMS

Fig. 2. Characteristics of long non-coding RNAs (lncRNAs) in CGMS comparision and its fertile lines. (a) Classification and percentage of different lncRNAs. (b) Chromosome-wise distribution of long non-coding RNAs (lncRNAs) in comparision with mRNA. (c) Length distribution of lncRNAs and mRNA. (d) The common and distinct differentially expressed lncRNAs identified in JA178 –JB178, JA178–JR178, and JB178–JR178 comparison.

and fertility restoration in cotton (Table 1). To scrutinize the relation between lncRNA, miRNA and mRNA we randomly selected three lncRNA–miRNA-mRNA triplex and used quantitative real-time polymerase chain reaction (RT-qPCR) to validate their expression patterns. The selected lncRNAs showed similar expression patterns for both RT-qPCR analysis and RNA-seq data (Fig. 5b).

4. Discussion

lncRNAs are the regulatory moieties in many important biological processes in plants like flowering and male fertility [64]. Plant lncRNAs undergo downstream processing before regulating gene expression. Several studies indicate the role of lncRNAs in complex regulatory network in plant male sterility [1–3]. Li et al. (2019), studied the genes involved in the abortion of cotton anthers and identified 17,897 differentially expressed (DE) mRNAs, and 865 DE long noncoding RNAs (lncRNAs). Among these DE genes, LTCONS_00105434, LTCONS_00004262, LTCONS_00126105, LTCONS_00085561, and LTCONS_00085561, corresponded to cis-target genes Ghir_A09G011050.1, Ghir_A01G005150.1, Ghir_D05G003710.2, Ghir_A03G016640.1, and Ghir_A12G005100.1, respectively, which are involved in anther development [6]According to various studies, there is an upregulation of lncRNAs during male reproductive development, For example, maize lncRNA (zm401) was expressed abundantly in developing male gametophyte and the mature pollen grains [65]. It was established that lncRNA zm401 was the prime regulator for the expression of genes crucial in pollen development like ZmC5, ZmMADS2 and MZm3–3. The result got supported by an experiment where transgenic plants with downregulated zm401 showed atypical formation of the tapetum and microspore which led to the production of infertile pollen [66]. Ding et al. studied that a 1236 bp long lncRNA, named long-day-specific male fertility associated RNA (LDMAR), when overexpressed in young panicles imparted fertility in long-day conditions in rice [67]. More recently, Stone and his collogues suggested that there may be an association between lncRNA of the mitochondrial genome and the CMS in *Silene vulgaris*, thus proposing the earliest proof of contribution of mt lncRNA in CMS [68]. In this study, more than 1500 non-redundant differentially expressed lncRNAs were identified through the analysis of more than 182,000,000 raw sequence reads based on stringent criteria. Although many lncRNAs may have been excluded, this is a unique and reliable set of lncRNAs. We performed a GO enrichment analysis to understand the function of the differentially

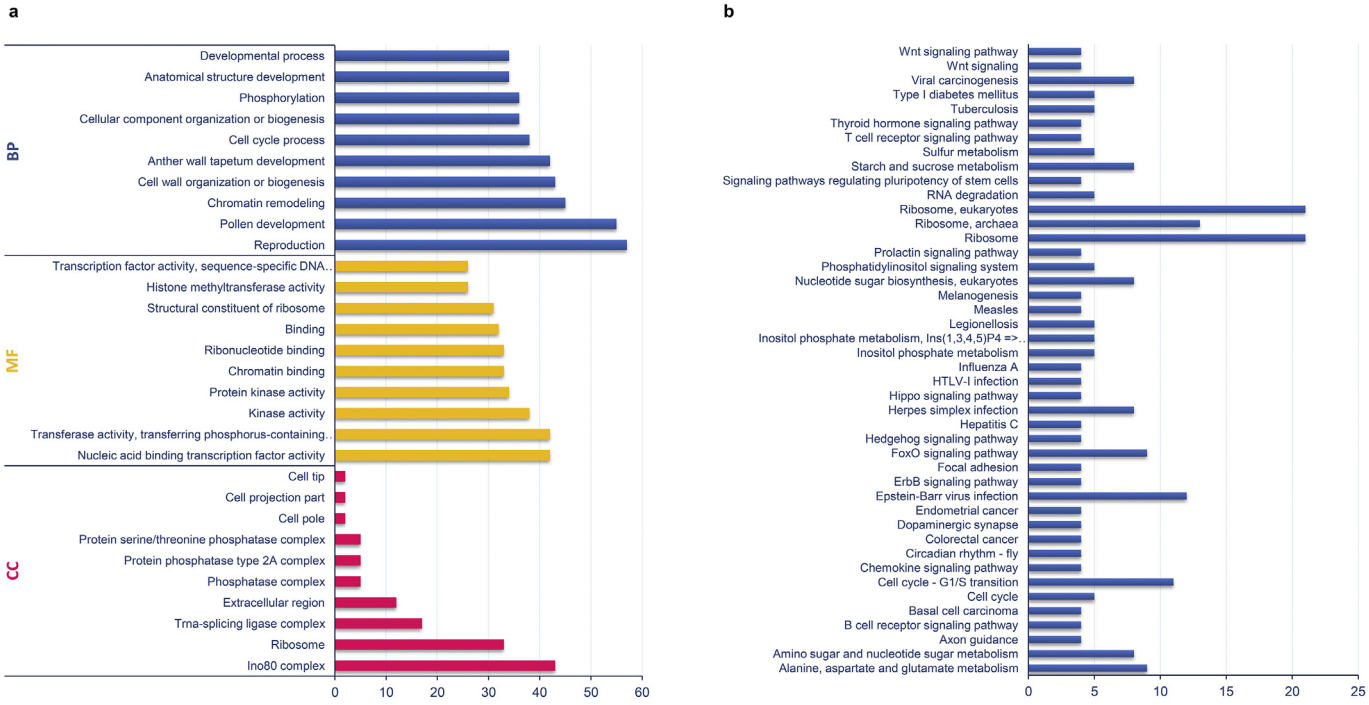

Fig. 3. Functional analysis of differentially expressed lncRNAs in JA178, JB178 and JR178 lines. (a) Gene ontology (GO) analysis of genes regulated by differentially expressed lncRNAs in JA178, JB178 and JR178 lines (b) Statistical KEGG enrichment of lncRNA-targets genes.

expressed lncRNAs. The GO terms that were most significantly enriched were reproduction, pollen development, chromatin remodeling, Ino80 complex, cell wall organization or biogenesis, anther wall tapetum development, and cell cycle process. The Ino80, belongs to the SNF2 family of ATPases and is a vital constituent of a multiunit ATP-dependent chromatin remodeling complex [69]. One of the recognized functions of lncRNAs, is to facilitate epigenetic changes at specific genomic loci by recruiting chromatin-remodeling complex [70].

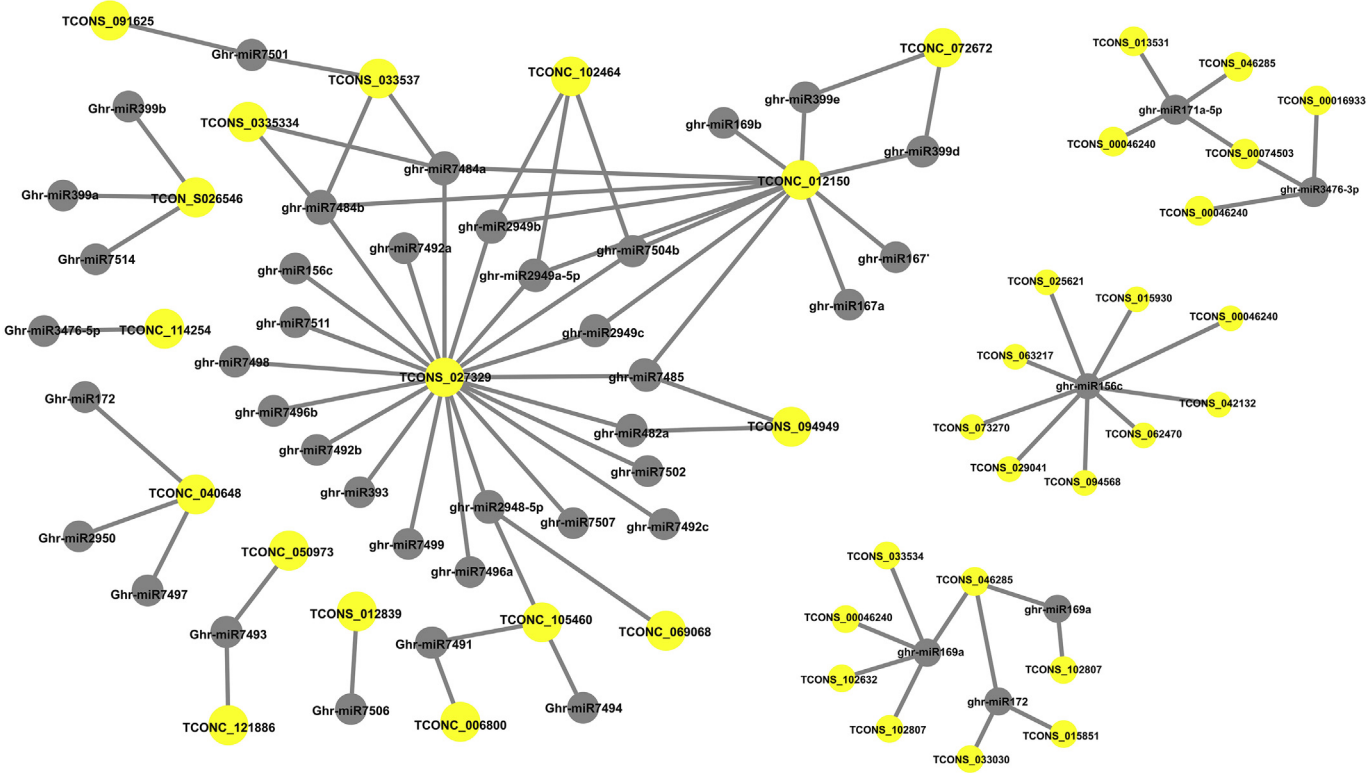

Fig. 4. LncRNA–miRNA interaction network. Yellow circle nodes represent lncRNAs, and gray circle nodes represent miRNAs. Examples of interactions are shown: one lncRNA with many miRNAs, one miRNA with many lncRNAs, and many lncRNAs with many. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

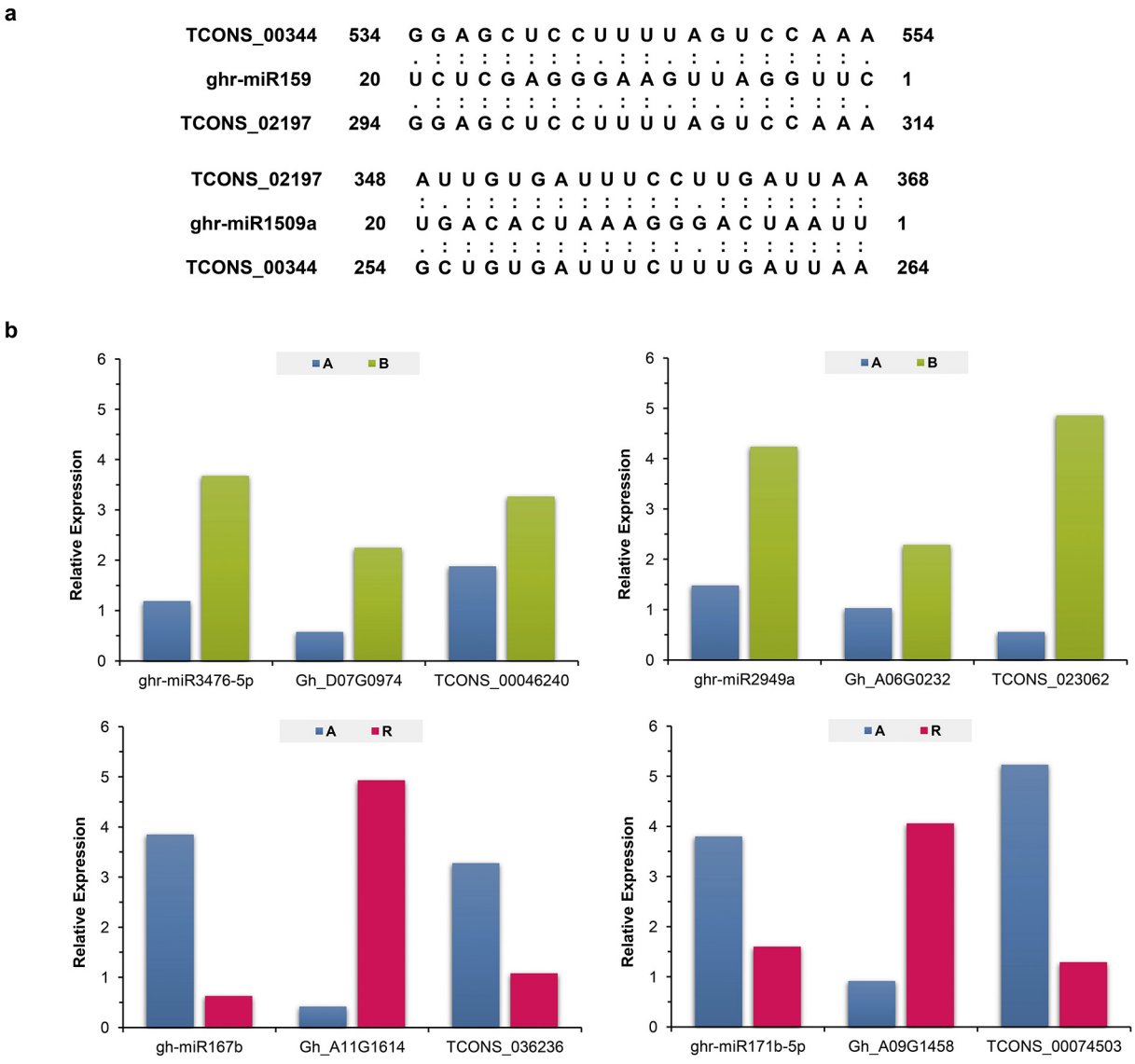

Fig. 5. Functional prediction of drought-responsive to lncRNAs as potential targets or target mimics of miRNAs. (a) The interaction miRNA–lncRNA was identified using psRNATarget webserver (https://plantgrn.noble.org/psRNATarget/analysis). (b) The relative expression levels of four lncRNA–miRNA pairs were validated by RT-qPCR. The expression levels of selected miRNAs and lncRNAs were calculated relative to their expression in the non-treated sample.

Table 1
A list of top thirteen lncRNA and their precursor miRNA and targeted genes with FPKM and role associated with reproductive development.

Sr. Number	Gene_id	miRNA	Gene	JA178 FPKM	JB178 FPKM	JR178 FPKM	Biological Role	References
1	TCONS_00046240	ghr-miR482a	AGL15	0.033	13.52	15.30	AGAMOUS-like 15	[50]
2	TCONS_037591	ghr-miR164	LAT52	0.026	2.79	16.65	Pollen extensin family protein	[51]
3	TCONS_016634	ghr-miR2949a	EXPA4	0.016	65.71	18.57	plant-type cell wall organization	[52]
4	TCONS_045705	ghr-miR2950	SF18	0.762	2.29	11.55	Anther maturation	[53]
5	TCONS_00062438	ghr-miR7484a	BHLH91	0.013	13.66	16.59	basic helix-loop-helix (bHLH)	[54]
6	TCONS_00092348	ghr-miR7493	SPL1	0.019	10.39	15.70	Anther maturation	[55]
7	TCONS_036236	ghr-miR167b	ARF8	0.532	10.53	14.27	Auxin Response Factors (ARFs)	[56]
8	TCONS_034293	ghr-miR167b	ARF6	0.092	7.22	12.93	Auxin Response Factors (ARFs)	[57]
9	TCONS_062470	ghr-miR156b	PPR	0.081	3.14	11.91	Pentatricopeptide Repeat Protein	[58]
10	TCONS_086156	ghr-miR160	GRAS	0.012	1.59	20.32	GRAS-domain transcription factors	[59]
11	TCONS_037534	ghr-miR162a	AMS	0.073	3.22	12.09	MYB transcription factor	[60]
12	TCONS_037591	ghr-miR164	DYT1	0.019	2.95	13.89	ARATH Transcription factor	[61]
13	TCONS_00185480	ghr-miR171b-5p	SPL16	0.575	3.30	41.17	Squamosa promoter-binding	[62]

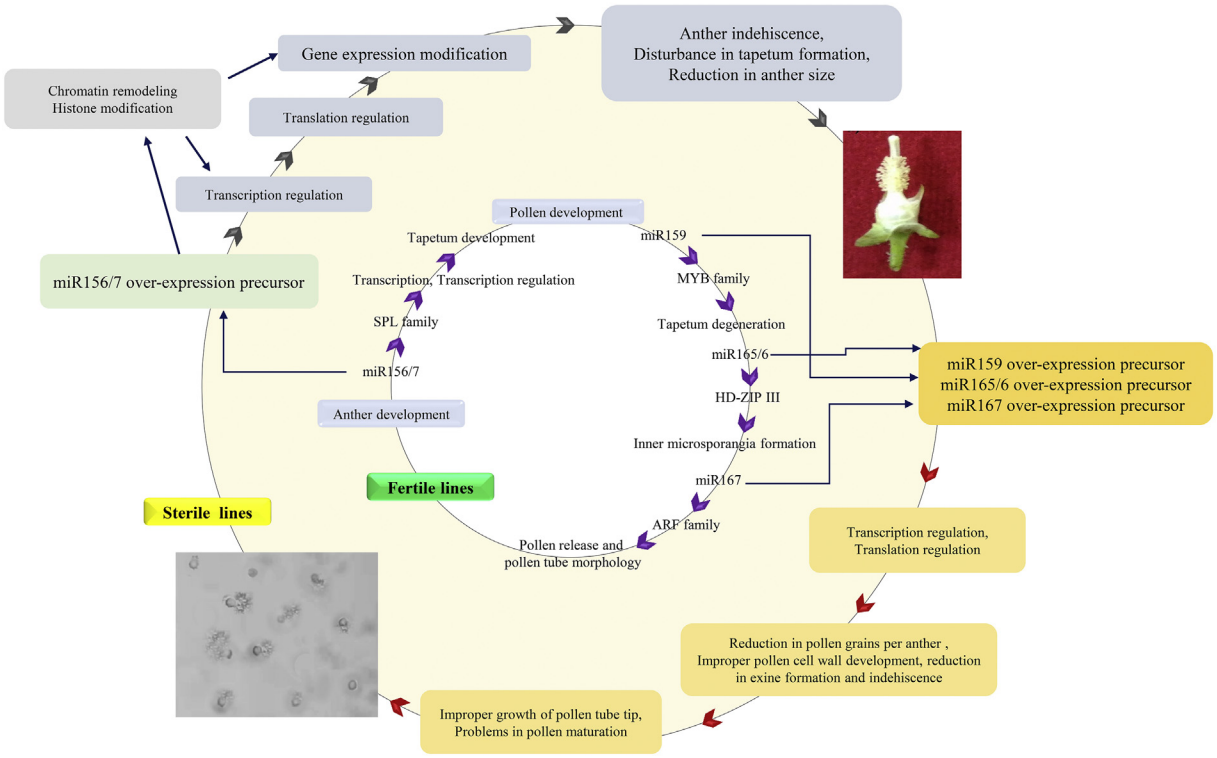

Fig. 6. miRNAs and lncRNAs network cotton plants result in cytoplasmic genic male sterilty. miR156 target SPL genes, which are involved in tapetum development, over-expression of miR156/7 precursors cause chromatin remodeling and histone modification which could negatively affect the transcription regulation and further brings about modifications in the expression of genes that function in anther indehiscence and cause disturbance in tapetum formation and degradation. While down regulation of these gene leads to reduction in anther size. The over-expression of precursors of miR159, miR165/6 and miR167 which target MYB family, HD-ZIP III and ARF family respectively will lead to transcription regulation and translation regulation of genes that act as in Tapetum degeneration, Inner microsporangia formation respectively which in turn leads to reduction in pollen grains per anther, improper pollen cell wall development, reduction in Exine formation, indehiscence, improper growth of pollen tube tip, problems in pollen maturation and thus results in male sterility.

Chromatin modifications are reported to regulate the levels of transcripts and/or the developmental stage specificity of expression. There are also reports in which the expression of FT and LFY are affected by chromatin modifiers [71,72]. FLC (Flowering Locus C) expression is regulated by the vernalization pathway and the autonomous pathway. FLC codes for a MADS-box transcription factor and behaves as a dominant repressor of flowering through its control of FT (Flowering Locus T) and SOC1 (Suppressor of Overexpression of CO1). FT and SOC1 are both targeted by the photoperiod pathway at the transcriptional level as well. Upon activation, FT encodes a protein, which is a key component of the systemic flowering signal “florigen” and so is crucial for the decision to flower. After integration of environmental and endogenous inputs at the leaves via transcription of FT, the FT protein travels to the apical meristem where it initiates floral development [73,74]. The plant specific transcription factor LFY activates many other target genes to evoke the program for flower formation. LFY plays a role as a “hub” by integrating signals from the GA and the aging pathways, respectively [75]. Anti-sense lncRNAs regulate the expression of sense transcripts through diverse mechanisms functioning at the transcriptional or post-transcriptional level. Prior studies have conveyed that chromatin remodeling at target loci by the anti-sense lncRNAs is an essential course of action they follow at the transcriptional level [46,47]. These outcomes indicated that the differentially expressed lncRNAs epigenetically controlled the transcription processes for fertility restoration in cotton. These lncRNAs represent functional candidates for CMS and fertility restoration for further investigation.

Previous studies indicate that lncRNAs may act as eTMs to prevent interaction between miRNAs and the target genes by competitively binding with the corresponding miRNAs [19,76]. For example, the Mt4 lncRNA in the *Medicago truncatula* acts as an eTM of miRNA-399 to regulate expression of PHO2 mRNA, involved in phosphate uptake [76]. In chickpea, Ca_linc_0964 consists of sites for binding of miR-159-a/b and miR-408b [49]. For the regulation of fruit development in hot pepper, 13 lncRNA was predicted to be eTMs for 9 miRNAs [76]. In the present study, 6 lncRNAs were put forward as potential eTMs for three miRNAs, of which *TCONS_049843* was a putative eTM for ghr-miR7493, ghr-miR7488, and ghr-miR7489, and *TCONS_089849* was a putative eTM for ghr-miR7499, ghr-miR7510b, and ghr-miR169b. Over and above their functions as eTMs of miRNAs, lncRNAs are also predicted to be precursors of miRNAs and the differential expression of lncRNAs may alter the expression of the corresponding mature miRNAs [16,72]. For instance, Cagirici et al. reported that a stress-responsive lncRNA was the precursor of miR1117 and miR1127a [16]. In the present study, 88 lncRNAs were identified as putative precursors of 57 miRNAs, of which four miRNAs (ghr-miR7494, ghr-miR390, ghr-miR156 and ghr-miR167) showed an expression level consistent with that of the precursor lncRNAs. For example, ghr-miR156, derived from TCONS_079241, was up-regulated in the A and B lines compared with the restorer line. miR156-targeted Squamosal Promoter-binding protein-like genes (SPLs, also known as SBPs) mediate the shift from the vegetative-to-reproductive. They are also key players in sustaining anther fertility in Arabidopsis [63,77], The overexpression of miR156 in the semi-sterile mutant background caused complete sterility, whereas the miR156-resistant transgenic form of SPLs partially mitigated the semi-sterile phenotype [78]. While miRNA ghr-miR167 which regulates an auxin response factor, *ARF8* (*Gh_D06G0360*), was down-regulated in the sterile line compared with the fertile lines. Several studies indicate that miR167/6 and *ARFs* families perform important roles during pollen maturation [79]. For instance, the overexpression of Arabidopsis miR167 in tomatoes, down-regulated ARF6 and ARF8 in tomato, which

subsequently resulted in a total abortion of the stylar trichomes, deterioration in the formation of pollen tube [80]. While in Arabidopsis, the same mechanism brings about changes like an undehisced anther, shorter filament and sterile pollen. Therefore, controlling ARF6 and ARF8 expression in Arabidopsis, showed that they regulate the development of stamen and gynoecium in immature flowers [81]. These results indicate that miRNA regulation of lncRNAs probably functions as a part of a complex regulatory pathway during plant development. For example, in the A–B comparison, a transcription factor AMS gene (Gh_D12G0328) and a GATA type zinc finger transcription factor family protein (Gh_D06G0196), regulated by the corresponding miRNAs and lncRNAs, were comparatively down-regulated in the A line than the B line. These genes might be involved in CMS during anther development (Fig. 6).

5. Conclusion

In this study, the transcriptome of the CGMS, its maintainer and restored lines were performed during two important anther developmental stages in cotton. Out of the total of 2031 unique lncRNAs identified, more than 1500 lncRNAs were differentially expressed between three lines. Enrichment analysis revealed that these lncRNAs were involved in reproduction, pollen development and chromatin remodeling. The putative relationship between lncRNAs and miRNAs were analyzed, which revealed that lncRNAs may act as miRNA precursors as well as miRNA eTMs. Eighty-eight lncRNAs were identified as putative precursors of 57 miRNAs. We found that ghr-miR156 most likely functions as a regulator of Gh_A08G0866 and Gh_D08G2688, which participate in anther cell wall development. RT-qPCR results showed concordance in expression profiling as reveled in RNA-seq data. To explore the functions of lncRNAs, we constructed putative miRNA–lncRNA–mRNA regulatory networks involved in CMS and fertility restoration. However, further functional analyses are needed for the precise elucidation of the regulatory networks. This study will offer widespread information about the lncRNAs in cotton and assist further studies focusing on the regulatory aspects of CGMS mechanism in cotton. Additional investigation on individual lncRNAs is required to clarify their precise function.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2020.06.027.

Author contribution

RH designed the research methodology, collected samples from the field, executed laboratory procedures of the project, performed data analysis, data visualization and diagram preparation, as well as drafted the manuscript, FJ drafted the manuscript as well as assisted in data analysis and diagram preparation, HM guided throughout the experiment, extended laboratory facility and helped in improving the manuscript, VR collected samples from the field, executed laboratory, as well as assisted in writing the initial draft of the manuscript, RST designed the research methodology, extended laboratory facility and participated in molecular genetic works analysis.

Key message

A floral buds RNA-seq analysis of CGMS, its maintainer and restorer lines in cotton identified 1531 lncRNAs related to pollen developmental pathways leading to CMS & fertility restoration.

Data availability statement

The datasets generated during and/or analyzed during the current study are available in the GenBank repository, under SRA archive with Bio Project ID: PRJNA383881: *Gossypium hirsutum* L. raw sequence reads and SRA ID SRX2786091, SRR5488111, SRR5487327, SRR5487041, SRR5485153, and SRR5483288. https://www.ncbi.nlm.nih.gov/bioproject/termPRJNA383881

Funding

This research was not granted by any agency.

Declaration of Competing Interest

None of the authors have any financial or personal relationships that could inappropriately influence or bias the content of the research paper.

Acknowledgments

We extend our heartfelt gratitude to the Junagadh Agricultural University for the provision of the laboratory facility for conducting the experiment and the Biotechnology and plant breeding Department, Ferdowsi University, Mashhad, Iran, for providing the facility for data analysis.

References

[1] S. Mei, T. Liu, Z. Wang, Comparative transcriptome profile of the cytoplasmic male sterile and fertile floral buds of radish (*Raphanus sativus* L.), Int. J. Mol. Sci. 17 (2016) 42.
[2] A. Bohra, U.C. Jha, P. Adhimoolam, D. Bisht, N.P. Singh, Cytoplasmic male sterility (CMS) in hybrid breeding in field crops, Plant Cell Rep. 35 (2016) 967–993.
[3] X. Yang, Y. Zhao, D. Xie, Y. Sun, X. Zhu, N. Esmaeili, Z. Yang, Y. Wang, G. Yin, S. Lv, Identification and functional analysis of microRNAs involved in the anther development in cotton genic male sterile line Yu98-8A, Int. J. Mol. Sci. 17 (2016) 1677.
[4] T. Lu, C. Zhu, G. Lu, Y. Guo, Y. Zhou, Z. Zhang, Y. Zhao, W. Li, Y. Lu, W. Tang, Strand-specific RNA-seq reveals widespread occurrence of novel cis-natural antisense transcripts in rice, BMC Genomics 13 (2012) 721.
[5] X. Wu, M. Liu, B. Downie, C. Liang, G. Ji, Q.Q. Li, A.G. Hunt, Genome-wide landscape of polyadenylation in Arabidopsis provides evidence for extensive alternative polyadenylation, Proc. Natl. Acad. Sci. 108 (2011) 12533–12538.
[6] Y. Wang, X. Wang, W. Deng, X. Fan, T.-T. Liu, G. He, R. Chen, W. Terzaghi, D. Zhu, X.W. Deng, Genomic features and regulatory roles of intermediate-sized non-coding RNAs in Arabidopsis, Mol. Plant 7 (2014) 514–527.
[7] F. Bardou, F. Ariel, C.G. Simpson, N. Romero-Barrios, P. Laporte, S. Balzergue, J.W. Brown, M. Crespi, Long noncoding RNA modulates alternative splicing regulators in Arabidopsis, Dev. Cell 30 (2014) 166–176.
[8] M. Jabnoune, D. Secco, C. Lecampion, C. Robaglia, Q. Shu, Y. Poirier, A rice cis-natural antisense RNA acts as a translational enhancer for its cognate mRNA and contributes to phosphate homeostasis and plant fitness, Plant Cell 25 (10) (2013) 4166–4182, https://doi.org/10.1016/j.ygeno.2020.06.027 (tpc. 113.116251).
[9] M. Beltran, I. Puig, C. Peña, J.M. García, A.B. Álvarez, R. Peña, F. Bonilla, A.G. de Herreros, A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial–mesenchymal transition, Genes Dev. 22 (2008) 756–769.
[10] S. Swiezewski, F. Liu, A. Magusin, C. Dean, Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target, Nature 462 (2009) 799.
[11] J.B. Heo, S. Sung, Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA, Science 331 (2011) 76–79.
[12] Y.-C. Zhang, J.-Y. Liao, Z.-Y. Li, Y. Yu, J.-P. Zhang, Q.-F. Li, L.-H. Qu, W.-S. Shu, Y.-Q. Chen, Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice, Genome Biol. 15 (2014) 512.
[13] A. Necsulea, M. Soumillon, M. Warnefors, A. Liechti, T. Daish, U. Zeller, J.C. Baker, F. Grützner, H. Kaessmann, The evolution of lncRNA repertoires and expression patterns in tetrapods, Nature 505 (2014) 635.
[14] C.-C. Hon, J.A. Ramilowski, J. Harshbarger, N. Bertin, O.J. Rackham, J. Gough, E. Denisenko, S. Schmeier, T.M. Poulsen, J. Severin, An atlas of human long non-coding RNAs with accurate 5′ ends, Nature 543 (2017) 199.
[15] M. Borg, F. Berger, Chromatin remodelling during male gametophyte development, Plant J. 83 (2015) 177–188.
[16] H.B. Cagirici, B. Alptekin, H. Budak, RNA sequencing and co-expressed long non-coding RNA in modern and wild wheats, Sci. Rep. 7 (2017) 10670.
[17] S. Boerner, K.M. McGinnis, Computational identification and functional predictions of long noncoding RNA in Zea mays, PLoS One 7 (2012) e43047.
[18] M. Wang, D. Yuan, L. Tu, W. Gao, Y. He, H. Hu, P. Wang, N. Liu, K. Lindsey, X. Zhang, Long noncoding RNA s and their proposed functions in fibre development of cotton (Gossypium spp.), New Phytol. 207 (2015) 1181–1197.
[19] H.-J. Wu, Z.-M. Wang, M. Wang, X.-J. Wang, Widespread long noncoding RNAs as endogenous target mimics for microRNAs in plants, Plant Physiol. 161 (2013) 1875–1884.
[20] J.L. Rinn, H.Y. Chang, Genome regulation by long noncoding RNAs, Annu. Rev. Biochem. 81 (2012) 145–166.

[21] S. Lin, X. Yue, Y. Miao, Y. Yu, H. Dong, L. Huang, J. Cao, The distinct functions of two classical arabinogalactan proteins BcMF8 and BcMF18 during pollen wall development in Brassica campestris, Plant J. 94 (2018) 60–76.
[22] L. Huang, H. Dong, D. Zhou, M. Li, Y. Liu, F. Zhang, Y. Feng, D. Yu, S. Lin, J. Cao, Systematic identification of long non-coding RNA s during pollen development and fertilization in Brassica rapa, Plant J. 96 (2018) 203–222.
[23] R. Hamid, R.S. Tomar, H. Marashi, S.M. Shafaroudi, B.A. Golakiya, M. Mohsenpour, Transcriptome profiling and cataloging differential gene expression in floral buds of fertile and sterile lines of cotton (*Gossypium hirsutum* L.), Gene 660 (2018) 80–91.
[24] R. Hamid, H. Marashi, R.S. Tomar, S.M. Shafaroudi, P.H. Sabara, Transcriptome analysis identified aberrant gene expression in pollen developmental pathways leading to CGMS in cotton (*Gossypium hirsutum* L.), PLoS One 14 (2019).
[25] A. Dobin, T.R. Gingeras, Mapping RNA-seq reads with STAR, Curr. Protoc. Bioinformatics 51 (2015) (11.14. 11-11.14. 19).
[26] C. Trapnell, A. Roberts, L. Goff, G. Pertea, D. Kim, D.R. Kelley, H. Pimentel, S.L. Salzberg, J.L. Rinn, L. Pachter, Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks, Nat. Protoc. 7 (2012) 562–578.
[27] L. Kong, Y. Zhang, Z.-Q. Ye, X.-Q. Liu, S.-Q. Zhao, L. Wei, G. Gao, CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine, Nucleic Acids Res. 35 (2007) W345–W349.
[28] L. Wang, H.J. Park, S. Dasari, S. Wang, J.-P. Kocher, W. Li, CPAT: coding-potential assessment tool using an alignment-free logistic regression model, Nucleic Acids Res. 41 (2013) e74.
[29] A. Li, J. Zhang, Z. Zhou, PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme, BMC bioinformatics 15 (2014) 311.
[30] P.H. Sabara, S.J. Jakhesara, K.J. Panchal, C.G. Joshi, P.G. Koringa, Transcriptomic analysis to affirm the regulatory role of long non-coding RNA in horn cancer of Indian zebu cattle breed Kankrej (Bos indicus), Functional & integrative genomics 20 (2020) 75–87.
[31] R.D. Finn, A. Bateman, J. Clements, P. Coggill, R.Y. Eberhardt, S.R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, Pfam: the protein families database, Nucleic Acids Res. 42 (2013) D222–D230.
[32] C. Trapnell, B.A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M.J. Van Baren, S.L. Salzberg, B.J. Wold, L. Pachter, Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation, Nat. Biotechnol. 28 (2010) 511.
[33] M.D. Young, M.J. Wakefield, G.K. Smyth, A. Oshlack, Gene ontology analysis for RNA-seq: accounting for selection bias, Genome Biol. 11 (2010) R14.
[34] M. Kanehisa, S. Goto, S. Kawashima, A. Nakaya, The KEGG databases at GenomeNet, Nucleic Acids Res. 30 (2002) 42–46.
[35] M. Kohl, S. Wiese, B. Warscheid, Cytoscape: Software for visualization and analysis of biological networks, Data Mining in Proteomics, Springer, 2011, pp. 291–303.
[36] M.W. Jones-Rhoades, D.P. Bartel, Computational identification of plant microRNAs and their targets, including a stress-induced miRNA, Mol. Cell 14 (2004) 787–799.
[37] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: a software environment for integrated models of biomolecular interaction networks, Genome Res. 13 (2003) 2498–2504.
[38] R. Garg, A. Sahoo, A.K. Tyagi, M. Jain, Validation of internal control genes for quantitative gene expression studies in chickpea (*Cicer arietinum* L.), Biochem. Biophys. Res. Commun. 396 (2010) 283–288.
[39] J. Li, B. Wu, J. Xu, C. Liu, Genome-wide identification and characterization of long intergenic non-coding RNAs in Ganoderma lucidum, PLoS One 9 (2014).
[40] C. Mu, R. Wang, T. Li, Y. Li, M. Tian, W. Jiao, X. Huang, L. Zhang, X. Hu, S. Wang, Long non-coding RNAs (lncRNAs) of sea cucumber: large-scale prediction, expression profiling, non-coding network construction, and lncRNA-microRNA-gene interaction analysis of lncRNAs in Apostichopus japonicus and Holothuria glaberrima during LPS challenge and radial organ complex regeneration, Mar. Biotechnol. 18 (4) (2016) 485–499.
[41] X. Guo, L. Gao, Q. Liao, H. Xiao, X. Ma, X. Yang, H. Luo, G. Zhao, D. Bu, F. Jiao, Long non-coding RNAs function annotation: a global prediction method based on bi-colored networks, Nucleic Acids Res. 41 (2013) e35.
[42] X. Lu, X. Chen, M. Mu, J. Wang, X. Wang, D. Wang, Z. Yin, W. Fan, S. Wang, L. Guo, Genome-wide analysis of long noncoding RNAs and their responses to drought stress in cotton (*Gossypium hirsutum* L.), PLoS One 11 (2016).
[43] S.H. Eom, H.J. Lee, J.H. Lee, S.H. Wi, S.K. Kim, T.K. Hyun, Identification and functional prediction of drought-responsive Long non-coding RNA in tomato, Agronomy 9 (2019) 629.
[44] A.M. Gurtan, P.A. Sharp, The role of miRNAs in regulating gene expression networks, J. Mol. Biol. 425 (2013) 3582–3600.
[45] C.E. Wong, Y.-T. Zhao, X.-J. Wang, L. Croft, Z.-H. Wang, F. Haerizadeh, J.S. Mattick, M.B. Singh, B.J. Carroll, P.L. Bhalla, MicroRNAs in the shoot apical meristem of soybean, J. Exp. Bot. 62 (2011) 2495–2506.
[46] P. Huijser, M. Schmid, The control of developmental phase transitions in plants, Development 138 (2011) 4117–4129.
[47] R. Schwab, Roles of miR156 and miR172 in reproductive development, MicroRNAs in Plant Development and Stress Responses, Springer, 2012, pp. 69–81.
[48] B. Zhang, X. Zhang, M. Zhang, L. Guo, T. Qi, H. Wang, H. Tang, X. Qiao, K. Shahzad, C. Xing, Transcriptome analysis implicates involvement of Long noncoding RNAs in cytoplasmic male sterility and fertility restoration in cotton, Int. J. Mol. Sci. 20 (2019) 5530.
[49] N. Khemka, V.K. Singh, R. Garg, M. Jain, Genome-wide analysis of long intergenic non-coding RNAs in chickpea and their potential role in flower development, Sci. Rep. 6 (2016) 33297.
[50] Q. Zheng, Y. Zheng, H. Ji, W. Burnie, S.E. Perry, Gene regulation by the AGL15 transcription factor reveals hormone interactions in somatic embryogenesis, Plant Physiol. 172 (2016) 2374–2387.
[51] P. Gerola, C. Mol, E. Newbigin, W. Lush, Regulation of LAT52 promoter activity during pollen tube growth through the pistil of Nicotiana alata, Sex. Plant Reprod. 12 (2000) 347–352.
[52] Z. Wang, L. Sun, W. Guan, C. Zhou, B. Tang, Y. Cheng, J. Huang, F. Xuan, De novo transcriptome sequencing and analysis of male and female swimming crab (*Portunus trituberculatus*) reproductive systems during mating embrace (stage II), BMC Genet. 19 (2018) 3.
[53] Y. Dattée, C. Dumas, A. Gallais, Reproductive Biology and Plant Breeding: Biologie de la Reproduction et Amélioration des Plantes, Springer Science & Business Media, 2012.
[54] Y. Zhang, The SBP-box gene SPL8 affects reproductive development and gibberellin response in Arabidopsis, in, Universität zu Köln, 2005.
[55] S. Xing, M. Salinas, P. Huijser, New players unveiled in early anther development, Plant Signal. Behav. 6 (2011) 934–938.
[56] Z.-F. Li, Y.-C. Zhang, Y.-Q. Chen, miRNAs and lncRNAs in reproductive development, Plant Sci. 238 (2015) 46–52.
[57] R. Tabata, M. Ikezaki, T. Fujibe, M. Aida, C.-e. Tian, Y. Ueno, K.T. Yamamoto, Y. Machida, K. Nakamura, S. Ishiguro, Arabidopsis auxin response factor6 and 8 regulate jasmonic acid biosynthesis and floral organ development via repression of class 1 KNOX genes, Plant Cell Physiol. 51 (2010) 164–175.
[58] Y.-J. Liu, X. Liu, H. Chen, P. Zheng, W. Wang, L. Wang, J. Zhang, J. Tu, A plastid-localized pentatricopeptide repeat protein is required for both pollen development and plant growth in rice, Sci. Rep. 7 (2017) 1–12.
[59] T. Ho Plágaro, N. Molinero Rosales, D. Fariña Flores, M. Villena Díaz, J.M. García-Garrido, Identification and expression analysis of GRAS transcription factor genes involved in the control of arbuscular mycorrhizal development in tomato, Front. Plant Sci. 10 (2019) 268.
[60] Z.B. Zhang, J. Zhu, J.F. Gao, C. Wang, H. Li, H. Li, H.Q. Zhang, S. Zhang, D.M. Wang, Q.X. Wang, Transcription factor AtMYB103 is required for anther development by regulating tapetum development, callose dissolution and exine formation in Arabidopsis, Plant J. 52 (2007) 528–538.
[61] J.N. Gu, J. Zhu, Y. Yu, X.D. Teng, Y. Lou, X.F. Xu, J.L. Liu, Z.N. Yang, DYT 1 directly regulates the expression of TDF 1 for tapetum development and pollen wall formation in a rabidopsis, Plant J. 80 (2014) 1005–1013.
[62] L.-M. Chao, Y.-Q. Liu, D.-Y. Chen, X.-Y. Xue, Y.-B. Mao, X.-Y. Chen, Arabidopsis transcription factors SPL1 and SPL12 confer plant thermotolerance at reproductive stage, Mol. Plant 10 (2017) 735–748.
[63] M. Liu, W. Li, G. Zhao, X. Fan, H. Long, Y. Fan, M. Shi, L. Zhang, X. Tan, New insights of salicylic acid into stamen abortion of female flowers in tung tree (*Vernicia fordii*), Front. Genet. 10 (2019) 316.
[64] A. Mishra, A. Bohra, Non-coding RNAs and plant male sterility: current knowledge and future prospects, Plant Cell Rep. (2018) 1–15.
[65] X.-Y. Dai, J.-J. Yu, Q. Zhao, D.-Y. Zhu, G.-M. Ao, Non-coding RNA for ZM401, a Pollen-Specific Gene of *Zea mays*, 46 ACTA BOTANICA SINICA-ENGLISH EDITION, 2004, pp. 497–504.
[66] J. Ma, B. Yan, Y. Qu, F. Qin, Y. Yang, X. Hao, J. Yu, Q. Zhao, D. Zhu, G. Ao, Zm401, a short-open reading-frame mRNA or noncoding RNA, is essential for tapetum and microspore development and can regulate the floret formation in maize, J. Cell. Biochem. 105 (2008) 136–146.
[67] J. Ding, Q. Lu, Y. Ouyang, H. Mao, P. Zhang, J. Yao, C. Xu, X. Li, J. Xiao, Q. Zhang, A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice, Proc. Natl. Acad. Sci. 109 (2012) 2654–2659.
[68] J.D. Stone, P. Kolouškova, D.B. Sloan, H. Štorchová, Non-coding RNA may be associated with cytoplasmic male sterility in Silene vulgaris, J. Exp. Bot. 68 (2017) 1599–1612.
[69] R.C. Conaway, J.W. Conaway, The INO80 chromatin remodeling complex in transcription, replication and repair, Trends Biochem. Sci. 34 (2009) 71–77.
[70] J.B. Heo, Y.-S. Lee, S. Sung, Epigenetic regulation by long noncoding RNAs in plants, Chromosom. Res. 21 (2013) 685–693.
[71] J. Adrian, S. Farrona, J.J. Reimer, M.C. Albani, G. Coupland, F. Turck, Cis-regulatory elements and chromatin state coordinately control temporal and spatial expression of FLOWERING LOCUS T in Arabidopsis, Plant Cell 22 (2010) 1425–1440.
[72] A. Yamaguchi, M. Abe, Regulation of reproductive development by non-coding RNA in Arabidopsis: to flower or not to flower, J. Plant Res. 125 (2012) 693–704.
[73] K.E. Jaeger, P.A. Wigge, FT protein acts as a long-range signal in Arabidopsis, Curr. Biol. 17 (2007) 1050–1054.
[74] M. Abe, Y. Kobayashi, S. Yamamoto, Y. Daimon, A. Yamaguchi, Y. Ikeda, H. Ichinoki, M. Notaguchi, K. Goto, T. Araki, FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex, Science 309 (2005) 1052–1056.
[75] M. Notaguchi, M. Abe, T. Kimura, Y. Daimon, T. Kobayashi, A. Yamaguchi, Y. Tomita, K. Dohi, M. Mori, T. Araki, Long-distance, graft-transmissible action of Arabidopsis FLOWERING LOCUS T protein to promote flowering, Plant Cell Physiol. 49 (2008) 1645–1658.
[76] L. Ou, Z. Liu, Z. Zhang, G. Wei, Y. Zhang, L. Kang, B. Yang, S. Yang, J. Lv, Y. Liu, Noncoding and coding transcriptome analysis reveals the regulation roles of long noncoding RNAs in fruit development of hot pepper (Capsicum annuum L.), Plant Growth Regul. 83 (2017) 141–156.
[77] A. Mishra, A. Bohra, Non-coding RNAs and plant male sterility: current knowledge and future prospects, Plant Cell Rep. 37 (2018) 177–191.
[78] S. Xing, M. Salinas, S. Höhmann, R. Berndtgen, P. Huijser, miR156-targeted and nontargeted SBP-box transcription factors act in concert to secure male fertility in Arabidopsis, Plant Cell 22 (2010) 3935–3950.

[79] Z. Tang, L. Zhang, C. Xu, S. Yuan, F. Zhang, Y. Zheng, C. Zhao, Uncovering small RNA-mediated responses to cold stress in a wheat thermosensitive genic male-sterile line by deep sequencing, Plant Physiol. 159 (2012) 721–738.
[80] N. Liu, S. Wu, J. Van Houten, Y. Wang, B. Ding, Z. Fei, T.H. Clarke, J.W. Reed, E. Van Der Knaap, Down-regulation of AUXIN RESPONSE FACTORS 6 and 8 by microRNA 167 leads to floral development defects and female sterility in tomato, J. Exp. Bot. 65 (2014) 2507–2520.
[81] P. Ru, L. Xu, H. Ma, H. Huang, Plant fertility defects induced by the enhanced expression of microRNA167, Cell Res. 16 (2006) 457–465.