We are IntechOpen, the world's leading publisher of Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

BOOK CITATION INDEX
CLARIVATE ANALYTICS
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Role of Next-Generation RNA-Seq Data in Discovery and Characterization of Long Non-Coding RNA in Plants

Shivi Tyagi, Alok Sharma and
Santosh Kumar Upadhyay

Additional information is available at the end of the chapter

Abstract

The next-generation sequencing (NGS) technologies embrace advance sequencing technologies that can generate high-throughput RNA-seq data to delve into all the possible aspects of the transcriptome. It involves short-read sequencing approaches like 454, illumina, SOLiD and Ion Torrent, and more advance single-molecule long-read sequencing approaches including PacBio and nano-pore sequencing. Together with the help of computational approaches, these technologies are revealing the necessity of complex non-coding part of the genome, once dubbed as "junk DNA." The ease in availability of high-throughput RNA-seq data has allowed the genome-wide identification of long non-coding RNA (lncRNA). The high-confidence lncRNAs can be filtered from the set of whole RNA-seq data using the computational pipeline. These can be categorized into intergenic, intronic, sense, antisense, and bidirectional lncRNAs with respect to their genomic localization. The transcription of lncRNAs in plants is carried out by plant-specific RNA polymerase IV and V in addition to RNA polymerase II and target the epigenetic regulation through RNA-directed DNA methylation (RdDM). lncRNAs regulate the gene expression through a variety of mechanism including target mimicry, histone modification, chromosome looping, etc. The differential expression pattern of lncRNA during developmental processes and different stress responses indicated their diverse role in plants.

Keywords: next-generation sequencing (NGS), high-throughput RNA-seq, long non-coding RNA (lncRNA), expression, development, stress

1. Introduction

Next-generation sequencing (NGS) technologies provide a new platform for the production of high-throughput sequencing data in less time at reduced cost. The tremendous

improvements in past years have allowed the sequencing of millions of DNA fragments in parallel. It has shifted the genomics to a newer edge by capturing the small details of DNA fragments. Earlier, Maxam and Gilbert's [1] and Sanger sequencing [2] techniques were leading approaches after the discovery of the DNA structure [3]. However, these techniques were time-consuming and limited to small-scale, dealing with few genes to the genome of simple organisms. But the necessity of sequencing the complex genome in short time and reduced cost have technologically advanced the sequencing approaches and evolved as NGS technologies. The NGS systems provide rapid, reproducible, and highly accurate sequencing techniques, and are based on the short-read sequencing approaches and a more advance single-molecule long-read sequencing [4]. The short-read sequencing approaches are dependent on sequencing by synthesis (SBS) and sequencing by ligation (SBL) methods. Further, these methods require pre-processing of DNA before directly proceeding to the sequencing steps, according to the requirement of different NGS platforms [4]. In SBS approach, the nucleotides are added by the polymerase into the elongating DNA strand and the signal is received in the form of fluorescence or ionic concentration change for every single nucleotide incorporated [5, 6]. Besides this, in SBL approach, probes having one- or two-base matching, bound to fluorophore, are ligated to the adjacent oligonucleotide on DNA fragments. The emitted fluorescent spectrum identifies the complementary bases of the probe at a specific position and reset primers are used to encrypt the complete DNA sequence [5]. Most of the short-read sequencing approaches require the clonal amplification of DNA on the solid surface such as bead-based, solid-state, and generation of DNA nanoball [5]. In all the methods, initially the DNA is fragmented and then ligated to a common set of adaptor for amplification and consequently ensue for DNA sequencing [5]. The short-read sequencing approaches include 454, illumina, SOLiD, and Ion Torrent platforms. Moreover, the *in-silico* approaches are used for the assembly of data generated by after these techniques [6]. The limitations in short-read sequencing approaches like *de novo* sequencing and the resolution of genomic variation leads to the development of more advance long-read sequencing approaches [6]. The long-read sequencing approaches are used for complex genomes with several long repetitive elements, structural variation, and alteration in copy number, which are significant for the occurrence of disease, and for evolution and adaptation [7–9]. It produces long reads of several kilobases and allows the higher resolution of the genome. In contrast to short reads, a single long-read can completely span the repetitive or complex region of genome, thus reducing the probability of vagueness in the size and positions of the genomic element [6]. Pacific Biosciences and Oxford Nanopore are commercially available sequencing technologies which provide the platform for sequencing the long reads with thousands of bases per read. These technologies are based on single-molecule sequencing, but have different methods of nucleotide detection. Oxford Nanopore is based on the detection through nanopores while Pacific Biosciences uses optical detection inside zero-mode waveguide [10]. Besides this, in synthetic approach, the data of short-read sequencing is combined with informatics and biochemical approaches for the construction of synthetic long reads. Long reads allow researchers for a deep transcriptomic study such as allele-specific transcription, alternative splicing, and in the identification of exact connectivity of exons and discern gene isoforms [6, 11, 12].

2. High-throughput RNA sequencing

Transcriptome consists of a whole set of transcripts present in a cell, and their expression level in particular developmental stage and cellular conditions. The detailed study of an organism at transcriptome level is necessary for revealing the molecular constituents involved in that particular stage or condition of the tissue. The high-throughput RNA-sequencing (RNA-seq) has emerged as an important technique in the field of transcriptomics for studying all the aspects of gene expression at large scale. It is one of the most commonly used techniques for quantification and mapping of transcriptomes. It involves the conversion of RNA into cDNA, followed by random sequencing of cDNAs fragments by using NGS platforms [13]. The generated millions of short reads were assembled by various bioinformatics approaches. Consequent mapping of these short reads reveals the position of gene transcribing the RNA on the reference genome or sets of a gene [13]. The high-throughput technologies also include direct RNA sequencing (DRS), in which the native RNA is directly sequenced without proceeding to the step of cDNA preparation. The technique is successful in sequencing native polyA+, where reverse transcription is undesirable. It is applicable in determining the precise sequence, identification of alternative polyadenylation sites, and deals with the small amount of nucleic acid [14]. In cap-assisted gene expression (CAGE) technique, RNAs with a 5′ cap are targeted. The short sequence tags are generated from 5′ ends of targeted RNAs with one tag per RNA molecule and allow the precise mapping of 5′ ends [15]. Series analysis of gene expression (SAGE) is another method for the sequencing of RNA molecules which target polyadenylated messages, and tags are generated near 3′ ends, typically one internal tag per RNA molecule [16]. Similarly, paired-end tags (PET) also targets polyadenylated RNA molecules, but the combined information on 5′ and 3′ ends of same RNA molecule generates the sequence tag [17]. Furthermore, rapid amplification of cDNA ends (RACE) is a PCR-based method used to identify the unknown sequences in conjunction with a known region. Together with the NGS technologies, it can be utilized for deep transcriptome sequencing of the particular locus [18]. Targeted RNA sequencing is also meant for a specific locus and by using tiling microarrays RNAs are selected and sequenced [19]. RNA profiling method by GRO-seq measures the steady-state levels of RNA and combined NGS analysis with the nuclear run-on experiments to generate information on RNA polymerase complexes competent with transcription [20]. This high-throughput RNA-seq is helpful in finding out the transcript (messenger RNAs, non-coding RNAs, and small RNAs) of species in short time and in determining the 5′ and 3′ splice sites, splicing patterns, and post-transcriptional modifications. The quantification of transcripts reveals the change in expression of genes in different conditions.

3. Long non-coding RNA (lncRNA)

3.1. Discovery and identification of lncRNA

In the era of NGS, the high-throughput RNA-seq data has lime lighted the necessity of non-coding part of the genome in the gene functioning. Non-coding RNAs (ncRNAs) are

transcribed from non-coding DNA, earlier called junk DNA. An extensive study on transcriptomes from multiple species indicated that about 90% of the genome can be transcribed, whereas only a small portion of such transcribed regions potentially codes for proteins [21]. The ncRNAs are categorized into housekeeping and regulatory ncRNAs on the basis their expression and role in different cells types. The expression of housekeeping ncRNAs (e.g., t-RNA, r-RNA, and snRNA) is prominent and has a structural role in all the cells [22]. While, the regulatory ncRNA shows temporal expression in specific cell types and includes microR NAs (miRNAs), small interfering RNAs (siRNAs), enhancer RNAs (eRNAs), promoter-associated RNAs (PARs), Piwi-interacting RNAs (piRNAs), and long non-coding (lncRNA). The criteria of >200 nt length are set for the identification of lncRNAs among all the organisms [23]. lncRNA comprises of a major group of ncRNAs and regulate various biological processes through different molecular mechanisms.
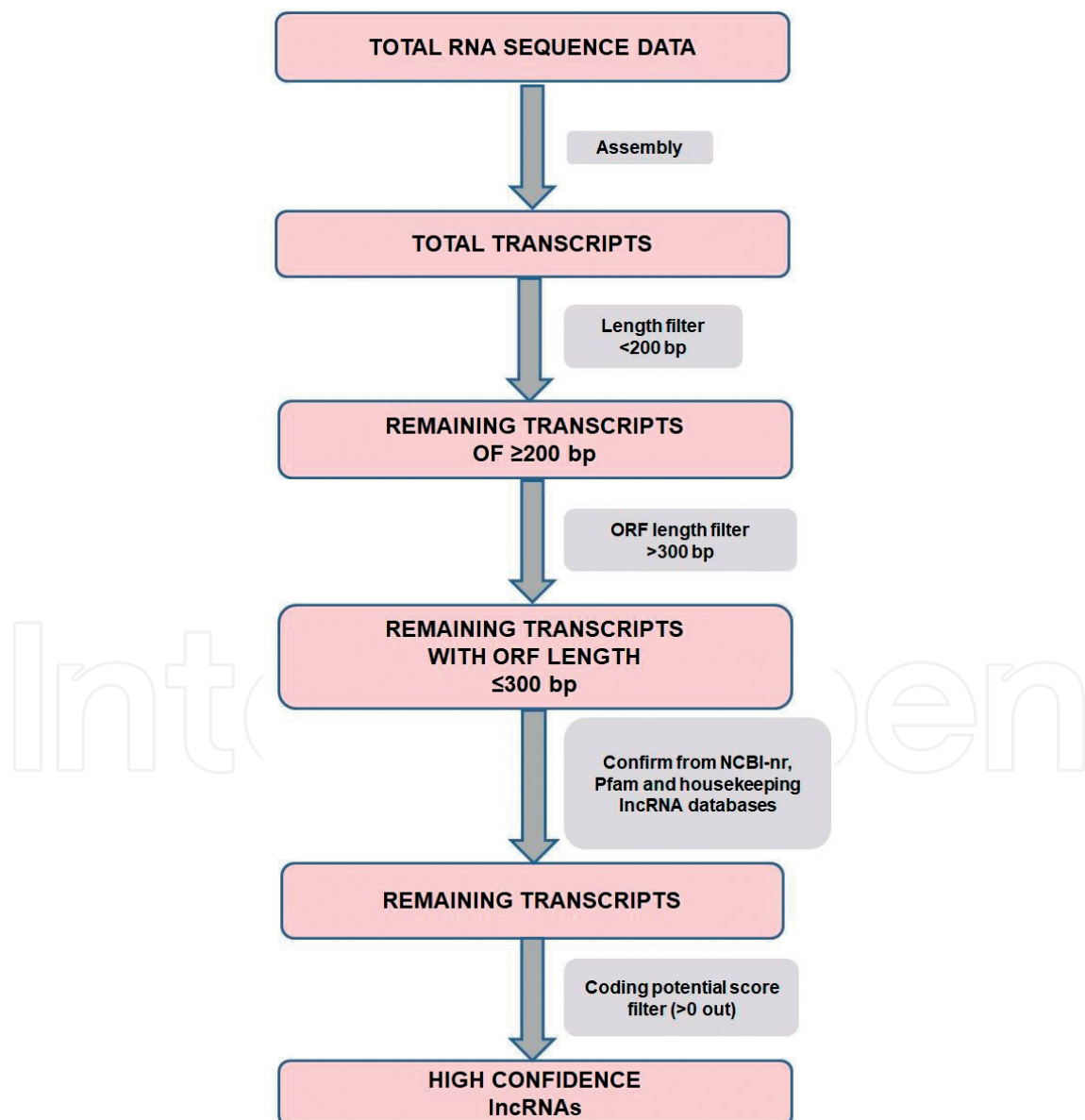

Figure 1. Pipeline for identification of long non-coding RNA.

In plants, the lncRNA was first reported in *Glycine max* [24], involved in changing the sub-cellular localization of a protein. In *Medicago truncatula* and *Oryza sativa*, *MtENOD40* and *OsENOD40* lncRNAs were discovered in nodule formation, respectively, and signify the involvement of lncRNA in biological roles [25, 26]. Likewise, in other plant species, for example, *COLD-INDUCED LONG-ANTISENSE INTRAGENIC RNA* (*COOLAIR*) and *COLD-ASSISTED INTRONIC NONCODING RNA* (*COLDAIR*), lncRNA in *Arabidopsis thaliana* [27, 28], involved in regulation of flowering, were identified and studied for their diverse function in the plant system. Furthermore, the exponential rise in high-throughput RNA-seq data have contributed to the discovery of lncRNA at genome-wide level, but the studies are limited in plants to some species. The amalgamation of experimental RNomics with the computational approaches has contributed to the identification of lncRNA and their function in wide-ranging biological processes [6]. The accurate identification and functional annotation is an ongoing challenge in

Sr. no.	Plant name	Number of lncRNAs	Tissues/organ/stress	Reference
1	*Amborella trichopoda*	2569	Tissue	[32]
2	*Arabidopsis thaliana*	~6480	Organ-specific and stress responsive	[22]
3	*Chlamydomonas reinhardtii*	2214	Cultured cells and synchronized vegetative cells	[32]
4	*Cicer arietinum*	2248	Three vegetative tissues and flower development	[30]
5	*Cucumis sativus*	3274	Fruit development and sex differentiation tissues	[33]
6	*Fragaria vesca*	1556	Floral, fruit tissue, and two vegetative tissues	[34]
7	*Medicago truncatula*	23,324	Control, osmatic, and salt stress in leaf and root tissues	[35]
8	*Oryza sativa*	2224	Development and reproductive organs	[36]
9	*Physcomitrella patens*	2711	Developmental stages	[32]
10	*Populus trichocarpa*	2542	Control and drought condition	[37]
11	*Setaria italica*	584	Drought stress	[38]
12	*Selaginella moellendorffii*	4422	Root, stem, and leaf	[32]
13	*Solanum lycopersicum*	10,774	In wild and ripening mutant	[39]
		1565	Tomato yellow-leaf curl virus stress	[40]
14	*Triticum aestivum*	44,698	Organ-specific and stress responsive	[31]
		283	Fungal-responsive lincRNAs	[41]
15	*Vitis vinifera*	4506	Organ-specific	[32]
16	*Zea mays*	1704	Different tissues	[42]
		664	Drought-stressed leaves	[43]
		7245	Leaves (under conditions of nitrogen deficiency and sufficiency)	[44]

Table 1. Occurrence of lncRNA in various plant species.

the field of bioinformatics for high-throughput RNA-seq data. The data of identified lncRNAs in plants is timely submitted to the different databases [29]. A pipeline with multiple filters has been designed for the assembly and identification of high confidence lncRNAs in **Figure 1** [30, 31]. The present status of most of the identified lncRNAs in different plant species are mentioned in **Table 1**.

3.2. Classification of lncRNAs

The biotypes of lncRNAs were identified with respect to their genomic localization, and were mainly categorized into intergenic, intronic, sense, antisense, and bidirectional lncRNAs. As the term suggest, the intergenic lncRNA are transcribed from the region amid two genes, while introns are the source of intronic lncRNA [45]. The sense and antisense lncRNAs are derived from overlapping region of exons on the sense and antisense strands, respectively [18], when the transcription of lncRNA is initiated in the juxtaposition of adjacent mRNA on complementary strand, termed as bidirectional lncRNA [45].

4. Molecular mechanisms of the functioning of lncRNAs

The dramatic change, in the past years about the knowledge of lncRNA in gene regulation mechanisms, has exponentially raised with high-throughput RNA-seq data. In plants, the studies are limited to small scale in comparison to animals, but the available reports suggested their different mechanisms as following.

4.1. lncRNA as target mimics of miRNA

Target mimicry is a mechanism of lncRNA for regulating the functions of miRNAs. They inhibit the interaction between the miRNA and their respective targets by binding to the target of miRNA via partial complementary sequence [46]. The novel mechanism of target mimicry was first discovered in *Arabidopsis*. In addition, *phosphate Starvation 1* (*IPS1*) was the first lncRNA identified as endogenous target mimic (eTM) of miR399 involved in phosphate homeostasis [46]. During phosphate starvation, the expression of miR399 is induced in companion cells and phloem [47]. Subsequently, the expression of *PHO2* gene, a target of miR399, is repressed [47–50]. This gene encodes UBC24 (E2 ubiquitin conjugase-related enzyme) and the reduction in its expression leads to the increased expression of *Pht1;8* and *Pht1;9* (phosphate transporter genes) in roots [47, 48]. Later, a similar mechanism was discovered in animals and humans suggesting target mimicry as the prevalent phenomenon [51, 52].

4.2. Histone modification

The lncRNAs are known to regulate gene expression through epigenetic changes. These epigenetic changes may result in alteration of gene expression in plants. Vernalization is the most common phenomenon of lncRNA mediated epigenetic regulation in plants. In *Arabidopsis*, FLOWERING LOCUS C (FLC) gene is the principal regulator of vernalization process and regulates the flowering time [53]. The expression of this gene is regulated by *COOLAIR* and *COLDAIR* lncRNAs through histone modifications [54].

4.3. Precursor lncRNA

lncRNAs constitute an important class of riboregulators by acting as a precursor in the synthesis of shorter ncRNAs, such as miRNAs and siRNAs. In this mechanism, some lncRNAs are processed to shorter ncRNAs or may directly act as a precursor [55]. The genes of primary miRNA transcripts (pri-miRNA) encoding miRNAs are transcribed by RNA polymerase II [56]. In plants, miRNA constitutes the modest portion in small regulatory ncRNA pool due to the presence of other complex small regulatory ncRNAs. In addition, they have plant-specific RNA polymerase IV/V involved in the transcription of siRNAs and endogenous siRNAs [57]. For example, in *Triticum aestivum*, 19 lncRNAs were predicted as a precursor of 28 miRNAs [31]. In *Arabidopsis*, the 24-nt sequence of several siRNAs were matched with five lncRNAs (npc34, npc351, npc375, npc520, and npc523), which was considered as potential precursor lncRNAs. The mapping of siRNAs on both the strands of lncRNAs also strengthened the findings [58].

4.4. RNA-dependent DNA methylation

The modification of chromatin is facilitated by recruitment of chromatin modifiers through lncRNA and small RNA (sRNA) into the specific locations in DNA. This RNA-dependent DNA methylation (RdDM) is a conserved process that recruits DNA methyltransferase and histone modifiers for DNA methylation and suppressive histone modification, respectively [59].

4.5. Chromosome looping

This mechanism is different from RdDM and histone modification, as it only involves the structural changes of chromatin. Thereby, it affects the binding potential of RNA polymerase and other transcription factors [60]. A persuasive example of chromosome looping mechanism in plants by *APOLO* lncRNA has been described in auxin transport by regulating the PID expression, an auxin transporter. When locus of *APOLO* lncRNA is transcribed by RNA Pol V and modified by RdDM, the expression of the locus is suppressed and loops to PID. This causes the inhibition of PID transcription. In contrast, when RNA Pol II carry out the transcription of APOLO lncRNA the looping of PID is restrained resulting in the expression of PID [60].

4.6. Protein re-localization

The mechanism of lncRNA in protein relocalization was first described in *G. max* and *Medicago sativa* [61, 62]. The symbiotic interactions among soil bacteria and leguminous plants are regulated by *Enod40* gene (early nodulin gene) which is induced by nitrogen-fixing bacteria in the pericycle and dividing cortical cells of roots [24, 63]. The diverse occurrence of *Enod40* lncRNA was suggested by its presence in non-leguminous plants, such as rice [26, 64]. The secondary structure of *Enod40* lncRNA is highly stable and has five highly conserved domains. The ORF of *Enod40* is very short and synthesis two short peptides. These short peptides regulate the biological activities of Enod40 and consequently help in nodulation [65, 66]. In *M. truncatula*, *Enod40* has been reported in the re-localization of MtRBP1 (*Medicago truncatula* RNA binding protein 1). *Enod40* showed direct interaction with MtRBP1 and re-localized the protein during nodulation process from nuclear speckles into cytoplasmic granules [25].

5. Expression profiling of lncRNAs

5.1. During developmental stages of different tissues

The expression of lncRNAs is regulated through different environmental and biological factors and delving into their diverse biological roles. They exhibit spatial and temporal expression during different developmental stages of various plant tissues. In contrast to the animals, a little is known about the functioning of lncRNAs in plants. The available reports reveal their role in nodule formation [26], lateral root development [67], vegetative and gametophytic development [68], cell-wall synthesis [69], flowering time [27, 54], and several others. The expression profiles developed using high-throughput RNA-seq data from various plants organs marks lncRNAs as an indispensable unit of the transcriptome. For instance, the expression profiles of lncRNAs from root, leaf, stem, spike, and grain in three developmental stages of *T. aestivum* have suggested the role in developmental processes. Furthermore, the lncRNAs show differential expression pattern comparable to the mRNA and highlight their function in related stages [31]. Besides this, the differential expression of lncRNAs in 11 different tissues of chickpea and 13 of maize also strengthens the findings [30]. These results also highlight the higher number of lncRNAs in actively dividing cells and reproductive tissues in comparison to the other [30, 33, 42, 43]. Depending on the expression values, they can be divided into different categories ranging from very low to very high expressing lncRNAs [30, 31]. Furthermore, fragments per kilobase of transcripts per million mapped reads (FPKM), reads per kilobase of transcripts per million mapped reads (RPKM) or transcripts per million (TPM) has to be determined for normalization and estimation of expression level [70]. The alteration in the expression level of various tissues within sundry plants can be correlated with the different genetic makeup and depth of transcriptome sequencing data. Tissue specificity index (TSI) is also calculated for studying the differential expression pattern of lncRNAs. The value of TSI ranges from zero to one, zero for housekeeping genes and one or near to one for sternly tissue-specific genes [31]. The criteria of TSI has revealed that lncRNAs are involved in flower and fruit development in *Fragaria vesca* [34], flower development in *Cicer arietinum* [30], development of fiber in *Gossypium arboreum* [71], and in development of root and floral tissues in *Morus notabilis* [72]. In addition to TSI, cell-type specificity can be interpreted for the expression of lncRNAs in specific cells [29]. For instance, in *Arabidopsis* cell-type specific lncRNAs have been identified in specialized cells but the expression was lower in comparison to mRNA [73]. The knowledge of lncRNAs is limited in plants, but the elevation in the survey of high-throughput RNA-seq data has allowed the prediction of their biological roles through expression profiling.

5.2. Expression under biotic and abiotic stresses

The expression of lncRNAs gets affected by biotic and/or abiotic factors in plants, but the mechanism remains poorly understood. Stress-responsive lncRNAs have emerged as an important component of plant defense machinery. The differential expression patterns in response to various stresses, including biotic and abiotic stresses, suggest the diverse function of lncRNAs at different intervals of stress exposure. For instance, the expression of 1832 lincRNAs gets remarkably affected after 2 h and/or 10 h of drought, salt, cold, and/or ABA (abscisic acid) treatments

in *Arabidopsis*. However, the expression of one of the candidate stress responsive lincRNA increased after treatment by elf18 (EF-Tu), which activates pathogen-associated molecular pat tern responses [22]. Likewise, in *T. aestivum*, 283 lncRNAs were identified as fungal-responsive lncRNAs, out of which 254 and 52 lincRNAs were specifically expressed after infestation with *Blumeria graminis* f. sp. tritici and *Puccinia striiformis* f. sp. tritici, respectively [41]. Later, a total of 44,698 lncRNAs were identified in *T. aestivum* consisting of both stress responsive and tissue-specific lncRNAs [31]. In response to tomato yellow-leaf curl virus, 1565 lncRNAs were expressed in *Solanum lycopersicum* [40]. In case of *Populus trichocarpa*, 2542 lncRNAs were expressed under drought stress condition [37]. The exploration of lncRNAs in various plant species in response to different stress conditions exhibit the dynamic role in plant defense.

6. Databases for lncRNAs

New high-throughput technologies have aided in the exponential rise of RNA-seq data from various plant species. A significant amount of lncRNAs has been identified and characterized for their diverse biological roles. Therefore, it is necessary to organize this data in web-based plat-forms or databases for further improvement, updates, and analysis [29]. Along with the aid of several computational tools, the data can be analyzed for phylogenetic relationships, expression patterns, molecular interactions, single nucleotide polymorphism, epigenetic variations, etc., and assist in understanding the lncRNAs in plants. The information in these databases can be managed specifically for single or numerous plant species. For instance, PLncRNAdb is specific for four plants species including *A. thaliana*, *A. lyrata*, *P. trichocarpa*, and *Z. mays* and consist of 5000 lncRNAs [74]. The information on 37 plants and 6 algae with data of >120,000 lncRNAs can be accessed on GreeNC database [75]. NONCODE v4 and PLncDB have information on 3853 and >13,000 lncRNA transcripts, respectively in Arabidopsis. Some databases cover the infor mation on both coding and non-coding transcripts like PlantNATsDB accumulating data of 70 plant species on NATs [76]. Besides this, some databases are plant-specific like TAIR10, PNRD, PlantNATsDB, etc., while certain databases (e.g., RNACentral, lncRNAdb v2.0, and NONCODE v4) consist of information from other organisms also in addition to plants [29]. These well-managed databases will allow the researchers to further study the lncRNA in more depth.

7. Biological roles of lncRNAs

The present knowledge on the function of lncRNAs is still limited in plants and a large por-tion of their function and mechanism is yet to be identified. In spite of this, the biological role of lncRNA has been studied in several plant species as discussed in **Table 2**. Some biological roles have been discussed here to highlight the importance of lncRNAs in plants.

7.1. lncRNA in plant fertility

The participation of lncRNAs in producing the male sterile lines in *O. sativa* is an important example of plant fertility. These male sterile lines are necessary for the hybridization and

Sr. no.	Species name	Annotated lncRNAs	Biological role	Regulatory mechanism	References
1	*Arabidopsis thaliana*	*APOLO*	Auxin-controlled development	Chromatin loop dynamics	[77]
		ASCO-lncRNA	Lateral root development	Alternative splicing regulators	[67]
		asHSFB2a	Vegetative and gametophytic development	Antisense transcription	[68]
		COLDAIR	Flowering time	Promoter interference	[27]
		COOLAIR	Flowering time	Histone modification	[28, 54, 78]
		HID1	Photomorphogenesis	Chromatin association	[79]
		IPS1	Phosphate homeostasis	Target mimicry	[46]
2.	*Glycine max*	*GmENOD40*	Nodule formation	Protein re-localization	[61]
3.	*Hordeum vulgare*	*HvCesA6 lnc-NAT*	Cell-wall synthesis	siRNA precursor	[69]
4.	*Medicago truncatula*	*MtENOD40*	Nodule formation	Protein re-localization	[66]
5.	*Oryza sativa*	*Cis-NATPHO1;2*	Phosphate homeostasis	Translational enhancer	[80]
		LDMAR (P/ TMS12-1	Fertility	Promoter interference	[81, 82]
		OsPI1	Phosphate homeostasis	Unknown	[83]
		OsENOD40	Nodule formation	Unknown	[26]
6.	*Petunia hybrid*	*SHO lnc-NAT*	Local cytokinin synthesis	dsRNA degradation	[84]
7.	*Solanum lycopersicum*	*TPS11*	Phosphate homeostasis	Unknown	[85]

Table 2. List of some annotated lncRNAs.

breeding processes. lncRNAs are known to induce photoperiod-sensitive genetic male sterility (PSMF) in *O. sativa* [82, 86], but the mechanism is not completely well understood. But according to the available reports, two different mechanisms of lncRNA can be possible [23]. In one mechanism, the high expression of the *long day (LD)-specific male-fertility-associated RNA* (*LDMAR*), a type of lncRNA, is important for the fertility of rice plant during long day (LD) conditions. During male sterility, the programme cell death (PCD) of anther cells occur due to lowered expression of *LDMAR* under LD conditions. The reduced expression of *LDMAR* is mediated by over expressing psi-LDMAR (a siRNA), transcribed in the promoter region of *LDMAR*. Enhanced expression of Psi-LDMAR caused methylation in promoter region through RdDM mechanism [81]. The other mechanism suggested the involvement of osa-sm R5864w (a 21-nt sRNA) which was formed from a unique ncRNA encoded by *LDMAR*. The

point mutation of C to G in osa-sm R5864w, resulting in the loss of function, leads to the production of light and temperature sensitive male sterile lines of rice [82].

7.2. Role in alternate splicing

Plant lncRNAs are known to increase the complexity of transcriptome and proteome by participating in alternative splicing. It was first reported in Arabidopsis, where lncRNA behaved as an alternative splicing competitor (ASCO) [67]. Together with the nuclear speckle RNA-binding protein (NSR), ASCO-lncRNA forms an alternative splicing regulatory module. The expression of AtNSR in primary and lateral root meristems regulates the development of lateral roots. The interaction of AtNSR with overexpressing ASCO-lncRNA affects the splicing pattern of mRNA targeted by NSR in transgenic plant [67, 87]. This indicates the role of lncRNA as a regulator of alternative splicing.

7.3. Plant lncRNAs in photomorphogenesis

Most of the plant growth and developmental processes are regulated by different climatic factors among them light is one of the most important factor [88]. The role of lncRNA in the regulation of photomorphogenesis is still an interesting area of research because most of the identified regulatory molecules are proteins. In *A. thaliana*, several light responsive lncRNAs have been identified associated with histone modifications [89]. Identification and functional characterization of *HIDDEN TREASURE 1* (*HID1*), a novel lncRNA, involved in photomorphogenesis have been accomplished [89]. It may control the process of photomorphogenesis by regulating the expression of PHYTOCHROME INTERACTING FACTOR 3 (PIF3), a transcription factor involved in light response [89]. It could negatively regulate the expression of *PIF3* gene by binding to its promoter directly or in association with chromatin [89]. The occurrence of *HID1* homologs has been described in other plant species exhibiting conserved functions. The findings also shed light on the involvement of other ncRNAs in light responses.

8. Limitations in computational analysis of lncRNAs

The selection of lncRNAs from the complete set of RNAs is broadly based on three criteria: (i) transcript length of ≥200 bp, (ii) small open reading frame with ≤300 bp, and (iii) transcripts without homology to known proteins. In addition to this, several other factors like the type of cDNA libraries or transcriptional sequence data, depth of sequencing, and coding potential of transcripts, also contribute in the screening of lncRNAs. The challenges during computational analysis come when some protein-coding gene which fulfill the basic selection criteria and encode a functional peptide. Besides this, the functional long non-coding transcript may have ORF >300 bp and share homology with known protein-coding genes will also produce hindrance in the identification [90]. Another challenge comes with the transcripts that not only function as an RNA molecule, but also encodes a peptide [91]. The advancement in computational approaches have been made to overcome these limitations and for more accurate

differentiation between coding and non-coding transcripts [92]. The use of support vector machines (SVMs) or other machine learning algorithms along with the computational methods have increased the confidence of disparity in between coding and non-coding transcripts [93]. However, the identity and function of computationally identified lncRNA needs to be validated individually by experimentation.

Acknowledgements

Author details

Shivi Tyagi, Alok Sharma and Santosh Kumar Upadhyay*

*Address all correspondence to: skupadhyay@pu.ac.in

Department of Botany, Panjab University, Chandigarh, India

References

[1]   Maxam AM, Gilbert W. A new method for sequencing DNA. PNAS. 1977;**74**:560-564

[2]   Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. Journal of Molecular Biology. 1975;**94**:441-448

[3]   Watson JD, Crick FH. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. Nature. 1953;**171**:737-738

[4]   Buermans HPJ, den Dunnen JT. Next generation sequencing technology: Advances and applications. Biochimica et Biophysica Acta. 2014;**1842**:1932-1941

[5]   Goodwin S, McPherson JD, Richard McCombie W. Coming of age: Ten years of next generation sequencing technologies. Nature Reviews Genetics. 2016;**17**:333-351

[6]   Levy SE, Myers RM. Advancements in next-generation sequencing. Annual Review of Genomics and Human Genetics. 2016;**17**:95-115

[7]   International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature. 2004;**431**:931-945

[8]   Kaper F, Swamy S, Klotzle B, Munchel S, Cottrell J, et al. Whole-genome haplotyping by dilution, amplification, and sequencing. PNAS. 2013;**110**:5552-5557

[9] Jain M, Fiddes IT, Miga KH, Olsen HE, Paten B, Akeson M. Improved data analysis for the MinION nanopore sequencer. Nature Methods. 2015;**12**:351-356

[10] Levene MJ, Korlach J, Turner SW, Foquet M, Craighead HG, Webb WW. Zero-mode waveguides for single-molecule analysis at high concentrations. Science. 2003;**299**:682-686

[11] Tilgner H, Grubert F, Sharon D, Snyder MP. Defining a personal, allele-specific, and single molecule long-read transcriptome. PNAS. 2014;**111**:9869-9874

[12] Treutlein B, Gokce O, Quake SR, Sudhof TC. Cartography of neurexin alternative splicing mapped by single-molecule long-read mRNA sequencing. PNAS. 2014;**111**:E1291-E1299

[13] St Laurent G, Wahlestedt C, Kapranov P. The landscape of long non-coding RNA classification. Trends in Genetics. 2015;**31**(5):239-251

[14] Ozsolak F, Platt AR, Jones DR, Reifenberger JG, Sass LE, et al. Direct RNA sequencing. Nature. 2009;**461**(7265):814-818

[15] Kodzius R, Kojima M, Nishiyori H, Nakamura M, Fukuda S, et al. CAGE: Cap analysis of gene expression. Nature Methods. 2006;**3**(3):211-222

[16] Phillipe N, Samra EB, Boureux A, Mancheron A, Ruffle F, et al. Combining DGE and RNA-sequencing data to identify new polyA+ non-coding transcripts in the human genome. Nucleic Acids Research. 2014;**42**(5):2820-2832

[17] Fullwood MJ, Wei CL, Ruan Y. Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. Genome Research. 2009;**19**(4):521-532

[18] Kapranov P, Drenkow J, Cheng J, Long J, Helt G, Dike S, Gingeras TR. Examples of the complex architecture of the human transcriptome revealed by Race and high-density tiling arrays. Genome Research. 2005;**15**(7):987-997

[19] Mercer TR, Daniel JG, Marcel ED, Joanna C, Cole T, et al. Targeted RNA sequencing reveals the deep complexity of the human transcriptome. Nature Biotechnology. 2012;**30**:99-104

[20] Hah N, Danko CG, Core L, Waterfall JJ, Siepel A, Lis JT, Kraus WL. A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. Cell. 2011;**145**:622-634

[21] Kim ED, Sung S. Long noncoding RNA: Unveiling hidden layer of gene regulatory networks. Trends in Plant Science. 2012;**17**:16-21

[22] Liu J, Jung C, Xu J, Wang H, Deng S, et al. Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in Arabidopsis. The Plant Cell. 2012;**24**:4333-4345

[23] Liu J, Wang H, Chua NH. Long noncoding RNA transcriptome of plants. Plant Biotechnology Journal. 2015;**13**:319-328

[24] Kouchi H, Hata S. Isolation and characterization of novel nodulin cDNAs representing genes expressed at early stages of soybean nodule development. Molecular & General Genetics. 1993;**238**:106-119

[25] Campalans A, Kondorosi A, Crespi M. Enod40, a short open reading frame-containing mRNA, induces cytoplasmic localization of a nuclear RNA binding protein in *Medicago truncatula*. The Plant Cell. 2004;**16**:1047-1059

[26] Kouchi H, Takane K, So RB, Ladha JK, Reddy PM. Rice ENOD40: Isolation and expression analysis in rice and transgenic soybean root nodules. The Plant Journal. 1999;**18**:121-129

[27] Heo JB, Sung S. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. Science. 2011;**331**:76-79

[28] Swiezewski S, Liu F, Magusin A, Dean C. Cold-induced silencing by long antisense transcripts of an *Arabidopsis* polycomb target. Nature. 2009;**462**:799-802

[29] Bhatia G, Goyal N, Sharma S, Upadhyay SK, Singh K. Present scenario of long noncoding RNAs in plants. Non Coding RNA. 2017;**3**:16

[30] Khemka N, Singh VK, Garg R, Jain M. Genome-wide analysis of long intergenic noncoding RNAs in chick pea and their potential role in flower development. Scientific Reports. 2016;**6**:33297

[31] Shumayla, Sharma S, Taneja M, Tyagi S, Singh K, Upadhyay SK. Survey of high throughput RNA-seq data reveals potential roles for lncRNAs during development and stress response in bread wheat. Frontiers in Plant Science. 2017;**8**:1019

[32] Szcześniak MW, Rosikiewicz W, Makalowska I. CANTATAdb: A collection of plant long non-coding RNAs. Plant and Cell Physiology. 2015;**57**:pcv201. DOI: 10.1093/pcp/pcv201

[33] Hao Z, Fan C, Cheng T, Su Y, Wei Q, et al. Genome-wide identification, characterization and evolutionary analysis of long intergenic noncoding RNAs in cucumber. PLoS One. 2015;**10**:e0121800

[34] Kang C, Liu Z. Global identification and analysis of long non-coding RNAs in diploid strawberry *Fragaria vesca* during flower and fruit development. BMC Genomics. 2015;**16**:815

[35] Wen J, Parker BJ, Weiller GF. *In silico* identification and characterization of mRNA-like noncoding transcripts in *Medicago truncatula*. In Silico Biology. 2007;**7**(4-5):485-505

[36] Zhang YC, Liao JY, Li ZY, Yu Y, Zhang JP, et al. Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. Genome Biology. 2014;**15**:512

[37] Shuai P, Liang D, Tang S, Zhang Z, Ye C, et al. Genome-wide identification and functional prediction of novel and drought-responsive lincRNAs in *Populus trichocarpa*. Journal of Experimental Botany. 2014;**65**:4975-4983

[38] Qi X, Xie S, Liu Y, et al. Genome-wide annotation of genes and noncoding RNAs of foxtail millet in response to simulated drought stress by deep sequencing. Plant Molecular Biology. 2013;**83**:459-473

[39] Scarano D, Rao R, Corrado G. *In silico* identification and annotation of non-coding RNAs by RNA-seq and De Novo assembly of the transcriptome of tomato fruits. PLoS One. 2017;**12**:e0171504

[40] Wang J, Yu W, Yang Y, Li X, Chen T, Liu T, et al. Genome-wide analysis of tomato long non-coding RNAs and identification as endogenous target mimic for microRNA in response to TYLCV infection. Scientific Reports. 2015;**5**:16946

[41] Zhang H, Hu W, Hao J, Lv S, Wang C. Genome-wide identification and functional prediction of novel and fungi-responsive lincRNAs in *Triticum aestivum*. BMC Genomics. 2016;**17**:238

[42] Li L, Eichten SR, Shimizu R, Petsch K, Yeh CT, Wu W, et al. Genome-wide discovery and characterization of maize long non-coding RNAs. Genome Biology. 2014;**15**:R40

[43] Zhang W, Han Z, Guo Q, Liu Y, Zheng Y, et al. Identification of maize long non-coding RNAs responsive to drought stress. PLoS One. 2014;**9**(6):e98958

[44] Lv Y, Liang Z, Ge M, Qi W, Zhang T, et al. Genome-wide identification and functional prediction of nitrogen-responsive intergenic and intronic long non-coding RNAs in maize (*Zea mays* L.). BMC Genomics. 2016;**17**:350

[45] Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. Annual Review of Biochemistry. 2012;**81**:145-166

[46] Franco-Zorrilla JM, Valli A, Todesco M, Mateos I, Puga MI, Rubio-Somoza I, et al. Target mimicry provides a new mechanism for regulation of microRNA activity. Nature Genetics. 2007;**39**:1033-1037

[47] Aung K, Lin SI, Wu CC, Huang YT, CL S, Chiou TJ. pho2, a phosphate overaccumulator, is caused by a nonsense mutation in a microRNA399 target gene. Plant Physiology. 2006;**141**:1000-1011

[48] Bari R, Datt Pant B, Stitt M, Scheible WR. PHO2, microRNA399, and PHR1 define a phosphate-signaling pathway in plants. Plant Physiology. 2006;**141**:988-999

[49] Chiou TJ, Aung K, Lin SI, Wu CC, Chiang SF, Su CL. Regulation of phosphate homeostasis by microRNA in *Arabidopsis*. Plant Cell. 2006;**18**:412-421

[50] Fujii H, Chiou TJ, Lin SI, Aung K, Zhu JKA. miRNA involved in phosphate-starvation response in *Arabidopsis*. Current Biology. 2005;**15**:2038-2043

[51] Rubio-Somoza I, Weigel D. MicroRNA networks and developmental plasticity in plants. Trends in Plant Science. 2011;**16**:258-264

[52] Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, et al. Integration of human large intergenic noncoding RNAs reveals global properties and specific sub-classes. Genes & Development. 2011;**25**:1925-1927

[53] Michaels SD, Amasino RM. FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. The Plant Cell. 1999;**11**:949-956

[54] Sun Q, Csorba T, Skourti-Stathaki K, Proudfoot NJ, Dean C. R-loop stabilization represses antisense transcription at the *Arabidopsis* FLC locus. Science 2013;**340**(6132):619-621

[55] Hirsch J, Lefort V, Vankersschaver M, Boualem A, Lucas A, et al. Characterization of 43 non-protein-coding mRNA genes in *Arabidopsis*, including the MIR162a-derived. Plant Physiology. 2006;**140**:1192-1204

[56] Lee Y, Kim M, Han J, Yeom KH, Lee S, et al. MicroRNA genes are transcribed by RNA polymerase II. The EMBO Journal. 2004;**23**:4051-4060

[57] Chen XM. Small RNAs and their roles in plant development. Annual Review of Cell and Developmental Biology. 2009;**25**:21-44

[58] Ben Amor B, Wirth S, Merchan F, Laporte P, d'Aubenton-Carafa Y, et al. Novel long non-protein coding RNAs involved in *Arabidopsis* differentiation and stress responses. Genome Research. 2009;**19**:57-69

[59] Castel SE, Martienssen RA. RNA interference in the nucleus: Roles for small RNAs in transcription, epigenetics and beyond. Nature Reviews. Genetics. 2013;**14**:100-112

[60] Böhmdorfer G, Wierzbicki AT. Control of chromatin structure by long noncoding RNA. Trends in Cell Biology. 2015;**25**(10):623-632

[61] Yang WC, Katinakis P, Hendriks P, Smolders A, de Vries F, et al. Characterization of GmENOD40, a gene showing novel patterns of cell-specific expression during soybean nodule development. The Plant Journal. 1993;**3**:573-585

[62] Crespi MD, Jurkevitch E, Poiret M, d'Aubenton-Carafa Y, Petrovics G, et al. Enod40, a gene expressed during nodule organogenesis, codes for a non-translatable RNA involved in plant growth. The EMBO Journal. 1994;**13**:5099-5112

[63] Compaan B, Yang WC, Bisseling T, Franssen H. ENOD40 expression in the pericycle precedes cortical cell division in Rhizobium-legume interaction and the highly conserved internal region of the gene does not encode a peptide. Plant and Soil. 2001;**230**:1-8

[64] Gultyaev AP, Roussis A. Identification of conserved secondary structures and expansion segments in enod40 RNAs reveals new enod40. Nucleic Acids Research. 2007;**35**: 3144-3152

[65] Rohrig H, Schmidt J, Miklashevichs E, Schell J, John M. Soybean ENOD40 encodes two peptides that bind to sucrose synthase. Proceedings of the National Academy of Sciences of the United States of America. 2002;**99**:1915-1920

[66] Sousa C, Johansson C, Charon C, Manyani H, Sautter C, et al. Translational and structural requirements of the early nodulin gene enod40, a short-open reading frame-containing RNA, for elicitation of a cell-specific growth response in the alfalfa root cortex. Molecular and Cellular Biology. 2001;**21**:354-366

[67] Bardou F, Ariel F, Simpson CG, Romero-Barrios N, Laporte P, et al. Long noncoding RNA modulates alternative splicing regulators in *Arabidopsis*. Developmental Cell. 2014;**30**:166-176

[68] Wunderlich M, Gross-Hardt R, Schof F. Heat shock factor HSFB2a involved in gametophyte development of *Arabidopsis thaliana* and its expression is controlled by a heat-inducible long non-coding antisense RNA. Plant Molecular Biology. 2014;**85**(6):541-550

[69] Held MA, Penning B, Brandt AS, Kessans SA, Yong W, et al. Small-interfering RNAs from natural antisense transcripts derived from a cellulose synthase gene modulate cell

wall biosynthesis in barley. Proceedings of the National Academy of Sciences of the United States of America. 2008;**105**:20534-20539

[70] Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, et al. A survey of best practices for RNA-seq data analysis. Genome Biology. 2016;**17**:13

[71] Zou C, Wang Q, Lu C, et al. Transcriptome analysis reveals long noncoding RNAs involved in fiber development in cotton (*Gossypium arboreum*). Science China. Life Sciences. 2016;**59**:164

[72] Song X, Sun L, Luo H, Ma Q, Zhao Y, et al. Genome-wide identification and characterization of long non-coding RNAs from mulberry (*Morus notabilis*) RNA-seq data. Genes. 2016;**7**:11

[73] Li S, Yamada M, Han X, Ohler U, Benfey PN. High-resolution expression map of the *Arabidopsis* root reveals alternative splicing and lincRNA regulation. Developmental Cell. 2016;**39**:508-522

[74] PLncRNAdb Ming Chen's Lab. Available online: http://bis.zju.edu.cn/PlncRNADB/index.php [Accessed: 25 December, 2016]

[75] Gallart AP, Pulido AH, de Lagrán IA, Sanseverino W, Cigliano RA. GREENC: A Wiki-based database of plant lncRNAs. Nucleic Acids Research. 2016;**44**:D1161-D1166

[76] Chen D, Yuan C, Zhang J, Zhang Z, Bai L, Meng Y, Chen LL, Chen M. PlantNATsDB: A comprehensive database of plant natural antisense transcripts. Nucleic Acids Research. 2012;**40**:D1187-D1193

[77] Ariel F, Romero-Barrios N, Jegu T, Benhamed M, Crespi M. Battles and hijacks: Non-coding transcription in plants. Trends in Plant Science. 2015;**20**(6):362-371

[78] Liu F, Marquardt S, Lister C, Swiezewski S. Targeted 3′ processing of antisense transcripts triggers *Arabidopsis* FLC chromatin silencing. Science. 2010;**327**(5961):94-97

[79] Wang H, Chung PJ, Liu J, Jang IC, Kean MJ, et al. Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in *Arabidopsis*. Genome Research. 2014;**24**:444-453

[80] Jabnoune M, Secco D, Lecampion C, Robaglia C, Shu QY, et al. A rice cis-natural antisense RNA acts as a translational enhancer for its cognate mRNA and contributes to phosphate homeostasis and plant fitness. The Plant Cell. 2013;**25**:4166-4182

[81] Ding J, Shen J, Mao H, Xie W, Li X, Zhang Q. RNA-directed DNA methylation is involved in regulating photoperiod-sensitive male sterility in rice. Molecular Plant. 2012;**5**:1210-1216

[82] Zhou H, Liu Q, Li J, Jiang D, Zhou L, Wu P, et al. Photoperiod- and thermo-sensitive genic male sterility in rice are caused by a point mutation in a novel noncoding RNA that produces a small RNA. Cell Research. 2012;**22**:649-660

[83] Wasaki J, Yonetani R, Shinano T, Kai M, Osaki M. Expression of the OsPI1 gene, cloned from rice roots using cDNA microarray, rapidly responds to phosphorus status. The New Phytologist. 2003;**158**:239-248

[84] Zubko E, Meyer P. A natural antisense transcript of the *Petunia hybrida* Sho gene suggests a role for an antisense mechanism in cytokinin regulation. The Plant Journal. 2007;**52**:1131-1139

[85] Liu C, Muchhal US, Raghothama KG. Differential expression of TPS11, a phosphate starvation-induced gene in tomato. Plant Molecular Biology. 1997;**33**:867-874

[86] Ding J, Lu Q, Ouyang Y, Mao H, Zhang P, et al. A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice. Proceedings of the National Academy of Sciences of the United States of America. 2012;**109**:2654-2659

[87] Kornblihtt AR. A long noncoding way to alternative splicing in plant development. Developmental Cell. 2014;**30**:117-119

[88] Jiao Y, Lau OS, Deng XW. Light-regulated transcriptional networks in higher plants. Nature Reviews. Genetics 2007;**8**:217-230

[89] Wang Y, Fan X, Lin F, He G, Terzaghi W, et al. *Arabidopsis* noncoding RNA mediates control of photomorphogenesis by red light. Proceedings of the National Academy of Sciences of the United States of America. 2014;**111**:10359-10364

[90] Dinger ME, Pang KC, Mercer TR, Crowe ML, Grimmond SM, Mattick JS. NRED: A database of long noncoding RNA expression. Nucleic Acids Research. 2009;**37**(Database issue):D122-D126

[91] Bardou F, Merchan F, Ariel F, Crespi M. Dual RNAs in plants. Biochimie. 2011;**93**(11): 1950-1954

[92] Castrignano T, Canali A, Grillo G, Liuni S, Mignone F, et al. CSTminer: A web tool for the identification of coding and noncoding conserved sequence tags through cross-species genome comparison. Nucleic Acids Research. 2004;**32**:W624-W627

[93] Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, et al. CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Research. 2007;**35**:W345-W349