



Long Non-coding RNA in Plants in the Era of Reference Sequences

Hikmet Budak^{1*}, Sezgi Biyiklioglu Kaya² and Halise Busra Cagirci²

¹ Montana BioAgriculture, Inc., Bozeman, MT, United States, ² Engineering and Natural Sciences, Molecular Biology, Genetics and Bioengineering Program, Sabancı University, Istanbul, Turkey

OPEN ACCESS

Edited by:

Jacqueline Batley,
The University of Western Australia,
Australia

Reviewed by:

Matthew R. Willmann,
Cornell University, United States
Xiaojun Nie,
Northwest A&F University, China

*Correspondence:

Hikmet Budak
hikmet.budak@icloud.com

Specialty section:

This article was submitted to
Plant Breeding,
a section of the journal
Frontiers in Plant Science

Received: 06 December 2018

Accepted: 21 February 2020

Published: 12 March 2020

Citation:

Budak H, Kaya SB and
Cagirci HB (2020) Long Non-coding
RNA in Plants in the Era of Reference
Sequences. *Front. Plant Sci.* 11:276.
doi: 10.3389/fpls.2020.00276

The discovery of non-coding RNAs (ncRNAs), and the subsequent elucidation of their functional roles, was largely delayed due to the misidentification of non-protein-coding parts of DNA as “junk DNA,” which forced ncRNAs into the shadows of their protein-coding counterparts. However, over the past decade, insight into the important regulatory roles of ncRNAs has led to rapid progress in their identification and characterization. Of the different types of ncRNAs, long non-coding RNAs (lncRNAs), has attracted considerable attention due to their mRNA-like structures and gene regulatory functions in plant stress responses. While RNA sequencing has been commonly used for mining lncRNAs, a lack of widespread conservation at the sequence level in addition to relatively low and highly tissue-specific expression patterns challenges high-throughput *in silico* identification approaches. The complex folding characteristics of lncRNA molecules also complicate target predictions, as the knowledge about the interaction interfaces between lncRNAs and potential targets is insufficient. Progress in characterizing lncRNAs and their targets from different species may hold the key to efficient identification of this class of ncRNAs from transcriptomic and potentially genomic resources. In wheat and barley, two of the most important crops, the knowledge about lncRNAs is very limited. However, recently published high-quality genomes of these crops are considered as promising resources for the identification of not only lncRNAs, but any class of molecules. Considering the increasing demand for food, these resources should be used efficiently to discover molecular mechanisms lying behind development and abiotic stress responses. As our understanding of lncRNAs expands, interactions among ncRNA classes, as well as interactions with the coding sequences, will likely define novel functional networks that may be modulated for crop improvement.

Keywords: wheat, barley, whole genome sequencing, computational identification, long non-coding RNA

INTRODUCTION

Since the realization of regulatory information contained within the non-protein-coding parts of DNA, efforts to identify non-coding RNA molecules have greatly accelerated. Advances in RNA sequencing technology have contributed to this acceleration and the discovery of non-coding RNAs, including lncRNAs, which elucidated their structures and functions. As our understanding of the regulatory roles of lncRNAs has improved, the importance of these non-coding molecules has

become more apparent. However, there is still much to discover about the functions of lncRNAs in cellular pathways.

A step further to understand both coding and non-coding elements was taken recently for wheat and barley: high-quality reference sequences have been published (Mascher et al., 2017; IWGSC, 2018). Wheat and barley are two of the most consumed and cultivated crops; thus, increasing the yield have been the ultimate goal for breeders and scientists to overcome the effects of population growth and climate change. Having a reference genome in hand improved the accuracy of the analyzes to find the origins of favorable traits and regulatory mechanisms that control the expression of the genes responsible for those traits. Therefore, wheat and barley reference sequences have opened a new era in the field of multi-omics research, allowing more accuracy and robustness toward the lightening of the undiscovered mechanisms within these important crops.

BIOGENESIS OF lncRNAs

Long non-coding RNAs (lncRNAs) are defined as transcripts longer than 200 bp that cannot construct a full-length protein (Kapranov et al., 2007). The lack of discernable coding potential is what mainly differentiates lncRNAs from mRNAs.

Similar to mRNAs, most lncRNAs are transcribed by RNA polymerase II and are subject to 5'-end capping, alternative splicing, and the addition of 3' poly-A tails (Chekanova, 2015). Plant lncRNAs can be transcribed by two additional polymerases; RNA Pol IV or RNA Pol V (Wierzbicki et al., 2008). Unlike Pol II transcripts, these lncRNAs are less characterized and possess some structural differences such as lack of poly-A tails (Zhou and Law, 2015). Identification of RNA Pol IV or PolV transcribed lncRNAs is particularly challenging due to their extremely low expression and instability (Rai et al., 2018). However, these transcripts are the major players driving RNA-mediated DNA methylation (RdDM). Plants have evolved a highly sophisticated RNA interference-dependent RdDM mechanism to ensure genomic stability (Matzke and Mosher, 2014). Briefly, in this pathway, an lncRNA transcribed by RNA polymerase IV is later processed into 24-nt small interfering RNAs (siRNAs) (You et al., 2013). lncRNA transcribed by RNA polymerase V is recognized by the siRNA-AGO complex and drives this complex to the chromatin target site together with chromatin modifying enzymes. Following interaction with the AGO complex, additional proteins and methyltransferases are recruited to cytosine residues at the target region to initiate gene silencing (Wierzbicki et al., 2008).

RNA polymerase IV transcripts reportedly act mostly as siRNA precursors, whereas RNA polymerase V and some RNA polymerase II transcripts are siRNA targets. RNA polymerase IV and V transcripts have mostly been studied in *Arabidopsis thaliana*, where a recent study identified 10s of 1000s of RNA polymerase IV-dependent lncRNAs using an RNA polymerase IV mutant (Li et al., 2015).

INFLUENCE OF RNA SEQUENCING TECHNOLOGIES ON THE DISCOVERY OF lncRNAs

A general method for identifying and functionally characterizing transcripts is shown in **Figure 1**. Improvements in RNA sequencing technology paved the way for expanding our understanding of RNA. Previous attempts to uncover transcriptomes relied mostly on microarray technology, which is inefficient and limited in coverage of the whole transcriptome, whereas next-generation DNA and RNA sequencing applications are readily available on many platforms, offering better and more consistent quality (Denoëud et al., 2008; Ozsolak and Milos, 2011). Together with the development of computational tools, the most striking and unexpected evidence has been collected from the non-coding parts of the genome, revealing the transcription of numerous non-coding RNA molecules in various structures and roles. Of all the RNA species discovered to date, lncRNAs are the most unclear class of molecules and might still hide many unknown features. To reveal the secrets of lncRNAs and other non-coding RNA species, new RNA sequencing applications have been developed. For example, while conventional RNA sequencing allowed sequencing of up to 600 nucleotides at a time, the deep sequencing approach has enabled sequencing of longer reads at high accuracy (Malone and Oliver, 2011; Chu et al., 2015). RNA capture sequencing detects targeted RNA molecules with low abundance in the transcriptome (Mercer et al., 2011; Clark et al., 2015), and was designed to overcome obstacles in conventional RNA sequencing in detecting low-abundance lncRNAs.

To study the functions of lncRNAs, several immunoprecipitation-based methods have been developed that reveal the interacting RNA partners of specific proteins, together with high-throughput sequencing. ChIRP-seq is one of these methods, and involves precipitation of *in vivo* cross-linked RNA-DNA and RNA-protein hybrids by a biotin-streptavidin interaction and then sequencing of the RNA and DNA molecules that appear in the precipitated hybrids (Chu et al., 2011, 2015).

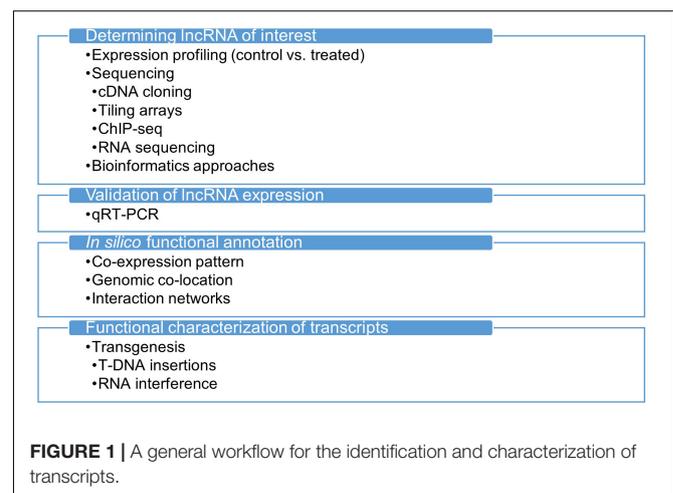


FIGURE 1 | A general workflow for the identification and characterization of transcripts.

CLIP-seq is another immunoprecipitation-based technique that has been used to explore miRNA–lncRNA interactions (Murigneux et al., 2013; Li J. et al., 2014). As demonstrated by these examples, RNA sequencing technology can be improved and modified according to the needs of the study. Improvements in the efficiency and accuracy of RNA and DNA sequencing techniques, not only for the identification of lncRNAs, but also for other RNA species, will lead to a more complete understanding of the secrets of the cellular mechanisms and their regulators.

In silico predictions based on sequencing have revealed many lncRNAs with expression patterns that remain to be confirmed. qRT-PCR allows the detection and quantification of expression in real time and is therefore widely used to verify the expression of *in silico*-predicted lncRNAs (Shuai et al., 2014). lncRNAs have been functionally annotated based on co-expression patterns, interaction networks, or both. Functions of lncRNAs can be predicted based on co-expressed protein-coding genes and/or genomic co-localization of genes (Guttman et al., 2009; Liao et al., 2011). For example, the lncRNAs COOLAIR and COLDAIR are expressed at the *FLC* locus and control *FLC* expression (Heo and Sung, 2011). Moreover, lncRNAs can serve as sRNA targets, preventing interaction between the sRNA and its protein-coding target, thereby enhancing the function of a particular protein-coding gene (Britton et al., 2014; Shuai et al., 2014).

These interaction networks between lncRNA, miRNA, and mRNAs suggest that some lncRNAs function as endogenous target mimics (Franco-Zorrilla et al., 2007; Chen et al., 2013). lncRNAs can also serve as sRNA precursors, with the downstream patterns of the corresponding sRNA revealing the involvement of lncRNAs in various molecular pathways (Matzke and Mosher, 2014; Ariel et al., 2015). Potential functions of lncRNAs can be confirmed by construction of transgenic lines with either downregulation or overexpression of genes. T-DNA insertions can be used for either gain-of-function or loss-of-function mutagenesis (Radhamony et al., 2005) whereas RNAi interference results in loss-of-function. For example, Zhu et al. (2014) identified lncRNAs in *Arabidopsis thaliana* that were differentially expressed during infection with *Fusarium oxysporum* and confirmed antifungal activity of 10 lncRNAs using T-DNA insertion and RNAi lines. Identification and confirmation of the interactions and functions of these non-coding RNAs is critical for the characterization of important molecular pathways.

lncRNA ANNOTATION FROM RNA SEQUENCING DATA

When using RNA sequencing data to annotate lncRNAs, computational procedures commonly begin with the alignment of sequencing reads on the reference genome, if available, and the assembly of transcript models from the mapped reads using computational tools that can be chosen from a wide range of software and algorithms based on their features and computational requirements (Ilott and Ponting, 2013). When a reference genome is lacking for the species of interest, the

assembly can be accomplished *de novo* although this strategy is more error-prone by being more sensitive to sequencing errors and chimeric molecules, and requiring more coverage in sequencing (Martin and Wang, 2011). After this point, the assembled transcripts should be evaluated to distinguish lncRNAs from a variety of non-coding RNAs and protein-coding mRNAs. Although complex and unclear features of lncRNAs have led researchers to adopt different methods and tools for the identification process, they seem to agree on a few basic criteria to select lncRNAs from other RNAs, such as minimum length. Many studies assume a 200-nucleotide length threshold to separate lncRNAs from snRNAs. Even though the presence of lncRNAs below this threshold has not been fully disproven, it is useful to eliminate snRNAs from the data (Ma et al., 2013). However, this criterion is mostly arbitrary and, alone, cannot define lncRNAs. In addition, this criterion does not distinguish between lncRNAs and mRNAs, since both types of RNA are commonly longer than 200 nucleotides (Milligan and Lipovich, 2015).

Therefore, for sequences that pass the first criterion, researchers usually assess open reading frame (ORF) content and length. Since transcripts containing long ORFs are assumed to be translated into full-length proteins, lncRNAs are expected to lack an ORF, or at least a long ORF (Boerner and McGinnis, 2012). Previous studies have speculated that most lncRNAs contain a short ORF (Banfai et al., 2012; Lv et al., 2013; Ruiz-Orera et al., 2014) and can occupy ribosomes, with contradictory conclusions about whether they encode protein products (Guttman et al., 2013; Ruiz-Orera et al., 2014; Popa et al., 2016). Despite lacking a clear explanation of the translational features of lncRNAs, these conflicting findings agree on another arbitrarily determined criterion, that is, an ORF size threshold of encoding 100 amino acids (Ilott and Ponting, 2013; Musacchia et al., 2015). After eliminating transcripts containing ORFs above the threshold, transcripts that satisfy the ORF size criterion are often examined to determine whether the remaining ORFs potentially encode any functional proteins. Several methods are used to calculate coding potentials and various algorithms can be used to assess candidate transcripts in terms of ORF presence, quality, intactness, and similarities to sequences encoding known proteins (Boerner and McGinnis, 2012; Mattick and Rinn, 2015). As this step is highly dependent on the quality of RNA sequencing reads and alignments on reference genomes, low-quality sequencing or alignment data, or lack of a reference genome, increases the chances of misleading coding potential calculations.

The use of machine learning techniques alone has increased the accuracy of coding potential calculations to over 90% (Kong et al., 2007; Hoff and Stanke, 2013; Sun et al., 2013). Nonetheless, due to slight differences in the approaches of conventional coding potential calculation tools, combining several of these tools may increase the stringency of the identification pipeline (Pauli et al., 2012).

The final criterion applied in many lncRNA identification pipelines involves exclusion of candidate transcripts that exhibit homology to known coding sequences, proteins, or protein domains. Similar to coding potential, homology can be assessed by several methods that use different databases for transcript comparisons (Jia et al., 2010; Pauli et al., 2012). However, a caveat

of this criterion is the loss of these exonic lncRNAs, leaving only lncRNAs expressed from intronic or intergenic spaces that do not overlap with the exons of any protein-coding genes (Housman and Ulitsky, 2016). Therefore, a balance between sensitivity and robustness must be properly maintained while designing the pipeline with elimination thresholds tailored to the aim of the study.

STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF lncRNAs

lncRNAs can be classified with respect to their genomic location and the direction of transcription (**Figure 2**), including intergenic, intronic, or exonic regions in the sense and antisense directions (Mattick and Rinn, 2015). The most controversial class was exonic lncRNAs that transcribed in the sense orientation. The lncRNA transcripts intersecting with the exons of protein coding genes had been eliminated until the latest release of GENCODE v7 catalog of human long non-coding RNAs (Derrien et al., 2012). However, some non-coding transcripts may arise from alternative splicing or truncation of first or last exons of protein coding genes. For example, SRA1 gene encodes for a lncRNA transcript [steroid receptor RNA activator (SRA)] as well as a protein coding transcript (SRAP) by alternative splicing (Sheng et al., 2018). Functional characterization of SRA have been performed well in both human and mouse (Nam et al., 2016). In fact, functions of SRAP has been less studied when compared to SRA. Although, there are currently not exonic lncRNAs with known functions available in plants yet, exonic lncRNAs have been reported in several plant species but without functional characterization (Liu et al., 2013; Quattro et al., 2017). Broadly, plant lncRNAs with known functions are classified as long intergenic non-coding RNAs (lincRNAs), intronic non-coding RNAs (intrcRNAs), and natural antisense transcripts (NATs) (**Table 1**).

lncRNAs transcribed outside of protein-coding genes are loosely classified as lincRNAs. Most research on plant lncRNAs has focused on lincRNAs, leading to the identification of

several lncRNAs with well-studied functions, such as LDMAR (Ding et al., 2012), APOLO (Ariel et al., 2014), IPS1 (Franco-Zorrilla et al., 2007), and Enod40 (Campalans, 2004). lncRNAs transcribed from intronic regions in the sense direction are called intrcRNAs. COLDAIR, transcribed from the first intron of *Flowering Locus C (FLC)*, is the best-known plant lncRNA (Heo and Sung, 2011). lncRNAs transcribed from the antisense direction to a protein-coding gene are classified as NATs. Well-studied examples of plant NATs include COOLAIR (Csorba et al., 2014) and HID1 (Wang et al., 2014). Recently, an antisense transcript of HvCesA6, which acts as a precursor to small RNA (sRNA) targeting the CesA6 gene, was shown to be involved in regulating cell wall synthesis in barley (*Hordeum vulgare*) (Held et al., 2008). Several plant NATs with newly characterized functions include *cis*-NAT PHO1;2 (Jabnونة et al., 2013), TL (Liu et al., 2018), and LAIR (Wang et al., 2018). The functions of the best-studied plant lncRNAs are listed in **Table 1**.

lncRNAs can also be classified based on their function, such as a decoy, scaffold, guide, signal, or signal enhancer. Decoys, such as IPS1, delay protein function by mimicking specific regions of the protein's target (Franco-Zorrilla et al., 2007). Scaffolds help to bring multiple proteins and RNAs together to form functional machineries, and recruit RNAs or proteins to a target region, as in RdDM (Matzke and Mosher, 2014). Signals, such as COLDAIR, are expressed under specific conditions to mediate biological processes (Heo and Sung, 2011). However, a single function model does not always apply to lncRNA function. An lncRNA might exhibit several functions which are usually linked. For example, in RdDM, an lncRNA transcribed by RNA polymerase V can act as a guide for the siRNA-AGO complex to the chromatin target site and as a scaffold for chromatin modifying enzymes and proteins.

An alternative model for classifying lncRNA functions is based on their structural features and the types of interactions they have with their targets, such as DNA interactions or protein interactions (Kung et al., 2013). As in the example of highly complex RdDM pathway, lncRNAs can be expected to have certain secondary structures to bring different

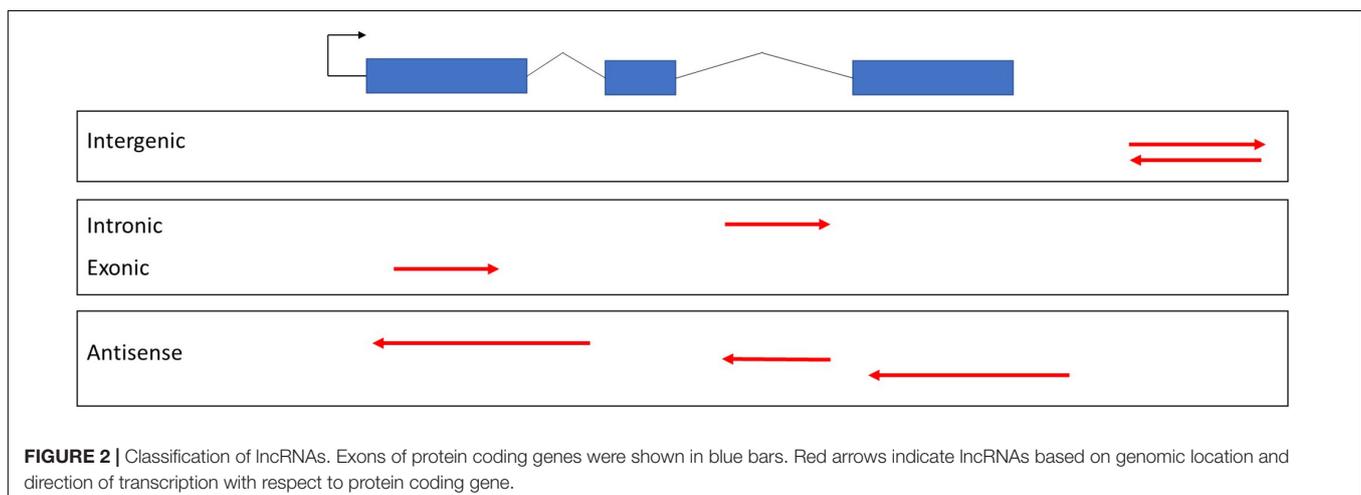


TABLE 1 | Examples of lncRNAs with known functions in plants.

lncRNAs	Species	Class	Biological function	Mechanism of action	Reference
Plant lncRNAs with best-studied functions					
COOLAIR/ COLDAIR	<i>A. thaliana</i> <i>B. distachyon</i>	NATs/ lncRNAs	Mediate flowering process	Histone modification	Heo and Sung, 2011; Csorba et al., 2014; Shafiq et al., 2015
APOLO	<i>A. thaliana</i>	lncRNAs	Modulates polar auxin transport	Chromatin-loop formation	Ariel et al., 2014
LDMAR	<i>Rice</i>	lncRNAs	Regulates photoperiod-sensitive male sterility	RNA-dependent DNA methylation (RdDM)	Ding et al., 2012
IPS1	<i>A. thaliana</i>	lncRNAs	Balances phosphate homeostasis	Endogenous target mimicry	Franco-Zorrilla et al., 2007
HID1	<i>A. thaliana</i>	NATs	Enhances photomorphogenesis	Chromatin association	Wang et al., 2014
Enod40	<i>Medicago</i> , rice, maize, legumes, <i>A. thaliana</i> , soybean	lncRNAs	Regulates nodulation	Protein relocalization	Yang et al., 1993; Rohrig et al., 2002; Campalans, 2004; Gultyaev and Roussis, 2007
Plant lncRNAs with less characterized functions					
WSGAR	Wheat	N/K*	Modulates seed germination	sRNA target and sRNA precursor	Guo et al., 2018
HvCesA6	Barley	NATs	Regulates cell wall synthesis	sRNA precursor	Held et al., 2008
<i>cis</i> -NAT PHO1;2	Rice	NATs	Phosphate homeostasis and plant fitness	Translation enhancer	Jabnourne et al., 2013
LAIR	Rice	NATs	Increases grain yield	Uncharacterized	Wang et al., 2018
Twisted lead (TL)	Rice	NATs	Regulates leaf morphology development	Chromatin modifications	Liu et al., 2018

*N/K: not known.

chromosome regions or proteins in close proximity. At the end, mechanisms of action include formation of chromosome looping between enhancer and promoter regions, modulation of gene activation and regulation, recruitment of chromatin modifying factors, enhancement of DNA methylation, and chromosome inactivation (Liu et al., 2012).

In some other cases, expression of an lncRNA, rather than the lncRNA itself, is important to initiate a biological process. For example, in mice, rather than the action of lncRNA Airn, its transcription induces *Igf2r* gene silencing (Latos et al., 2012). Airn is an antisense lncRNA to *Igf2r* gene whose promoter lies between Airn transcript in the opposite orientation (Santoro et al., 2013). Airn transcribing RNA polymerase prevents assembly of transcription initiation complex at the *Igf2r* promoter, thus prevents its expression. In another study, mutant lines, of *Arabidopsis*, with an enhanced promoter inside the T-DNA region resulted in a strong expression of a long transcript extending over the promoters of neighboring genes in the same orientation. Similarly, initiation of transcription from an intergenic T-DNA insertion halted expression of a downstream gene in *Arabidopsis* (Hedtke and Grimm, 2009), by making its promoter site inaccessible by transcription initiation complex. In diverse species, polymerase activity extending over the promoter of another gene halted the expression of downstream genes in either opposite or same orientation, indicating that this mechanism is likely to be conserved between species. These studies also emphasize the challenges of functional characterization of lncRNAs.

DEVELOPMENTAL STAGE-RELATED lncRNAs

Many lncRNAs function in developmental pathways in plants. One of the best-characterized examples of this regulation was discovered in *Arabidopsis* at the transition from the vegetative to generative stage. *FLC* is a regulator of flowering time in *Arabidopsis* that represses the induction of flowering (Csorba et al., 2014). An antisense lncRNA to *FLC* gene, COOLAIR, was discovered as upregulated at the beginning of vernalization (Shafiq et al., 2015). COOLAIR is involved in *FLC* repression by both autonomous and vernalization pathways, which leads to flowering in spring. Homology-based search was performed to find *FLC* locus and antisense *FLC* transcripts in other monocots, and the results showed that although there is no sequence conservation between antisense *FLC* transcripts and *Arabidopsis* COOLAIR lncRNA, the locations of these transcripts were conserved in six grass species including *T. aestivum* (Jiao et al., 2019). COLDAIR was identified as another lncRNA potentially regulating *FLC* expression. It is also transcribed in response to cold; however, in contrast to COOLAIR, COLDAIR is oriented in the sense direction of *FLC* (Heo and Sung, 2011). COLDAIR has been suggested to maintain vernalization by repressing *FLC* (Yamaguchi and Abe, 2012). Both lncRNAs serve as signals that determine the developmental stage of the plant, but much remains to be discovered on their exact functions, interactions and their presence in grass species such as wheat and barley.

Another lncRNA regulating developmental pathways is long-day-specific male-fertility-associated lincRNA (LDMAR). LDMAR expression below a certain level affects pollen development in rice under long-day conditions. Mutations causing reduced expression of LDMAR result in photoperiod-sensitive male sterility in plants grown under long-day conditions (Ding et al., 2012; Zhang and Chen, 2013). Again, the mechanism by which LDMAR regulates pollen development and whether it is expressed in cereals is unclear.

Recently, Guo et al. (2018) identified a novel lncRNA, Wheat Seed Germination Associated RNA (WSGAR), that modulates wheat seed germination. The proposed mechanism of action starts with a wheat-specific miRNA (miR9678) targeting WSGAR, which in turn is processed into phasiRNA and interferes with seed germination. Even though being not well-characterized, another study identified 177 lncRNAs that were responsive to a drug that blocked Ca²⁺ channels in wheat roots. They also observed that lengths of the roots were significantly decreased and root growth was prevented with increasing amounts of drug. Therefore, these 177 lncRNA identified was suggested to be related to root growth in wheat (Ma et al., 2018).

STRESS-RESPONSIVE AND OTHER lncRNAs IN WHEAT, BARLEY, AND RELATIVES

lncRNAs have been identified in many species from mammals to plants, including model organisms and economically important crop species, as more transcriptomic and genomic data have become available. One of these classes of crops is the Triticeae tribe, which includes cereal species such as wheat and barley important sources of nutrition in the human diet (Moore et al., 1995). Unraveling cellular mechanisms responsible for gene expression under stress conditions is the objective of ongoing research, in efforts to breed cultivars better able to withstand abiotic and biotic stresses (Pieri et al., 2018). For this purpose, the lncRNA repertoires of two of the three diploid wild ancestors of bread wheat (*Triticum aestivum*, AABBDD), *Triticum urartu* (AA) and *Aegilops tauschii* (DD), whose draft and reference genomes were recently published (Jia et al., 2013; Ling et al., 2013; Luo et al., 2017), were examined. Identified lncRNAs, 13,993 lncRNAs from *T. urartu* and 20,338 from *Ae. tauschii*, were also compared to bread wheat and tetraploid wild emmer wheat (*Triticum turgidum* ssp. *dicoccoides*, AABB), a wild subspecies of *T. turgidum* (AABB), the tetraploid ancestor of bread wheat (Pieri et al., 2018). Comparative analyses using RNA sequencing data suggested that the conservation between lncRNA repertoires decreased as the evolutionary distance increased (Pieri et al., 2018). Wild emmer wheat has long been a promising resource for exploration and exploitation of stress responses, due to the remarkable genetic diversity its wild populations retain. Akpinar et al. (2018) predicted lncRNA genes in the *T. turgidum* ssp. *dicoccoides* genome and investigated potential lncRNA-miRNA-mRNA networks. The results of this study revealed 89,623 lncRNAs where 23,713 were identified as potential miRNA targets (Akpinar et al., 2018). Another study identified lncRNAs in two

cultivars of wild emmer wheat, Kiziltan and TR39477, and one durum wheat (*T. turgidum* ssp. *durum*, AABB), a domesticated subspecies of *T. turgidum*, revealing 63,773, 61,823, and 43,932 lncRNAs in Kiziltan, TR39477 and durum wheat, respectively. This study reported that 3% of the identified Kiziltan lncRNAs, 6% of the identified TR39477 lncRNAs, and 4% of the durum wheat lncRNAs were differentially expressed in response to drought and called as 'drought-responsive' lncRNAs, with most only expressed under drought (Cagirici et al., 2017). Moreover, lncRNAs were identified from the transcriptome of durum wheat cultivar Svevo concurrently with the assembly of its genome. 115,437 lncRNAs were identified and chromosome 3B contained the highest number of lncRNA genes (Maccaferri et al., 2019).

As its ancestors, the bread wheat genome and transcriptome were investigated for lncRNA expression patterns under various biotic and abiotic stress conditions. An analysis of lncRNAs in bread wheat genotypes revealed 77 that were responsive to heat stress, 71 to fungal infection, and 23 to both conditions (Xin et al., 2011). A more comprehensive study identified lncRNAs from 52 sets of RNA sequencing data obtained under heat and drought stress, concluding that 29% of the lncRNAs were responsive to these abiotic stress conditions. Furthermore, the same study explored lncRNA expression under salt stress and identified two lncRNA groups showing distinct expression patterns; one was upregulated in the first hours after exposure and downregulated later, and the second group showed the opposite pattern (Shumayla et al., 2017).

Barley is another economically important species consumed worldwide, and has been studied for a better understanding of response mechanisms to stress (Gozukirmizi and Karlik, 2017). One study examined the barley transcriptome for lncRNAs and their expression patterns under excess boron (Karakulah and Unver, 2017). A second study observed differential expression patterns of two specific lncRNAs in cultivars exposed to salinity; one of those lncRNAs, AK372814, was upregulated under salinity stress (Karlik and Gozukirmizi, 2018), providing a clue to gene regulatory elements involved in responses to salinity. These results give a broad perspective of expression patterns and abundance of lncRNAs in genomes, suggesting that lncRNAs function in cellular mechanisms that are regulated under various stress conditions. However, specific lncRNAs that are involved in stress response pathways largely remain to be identified. Even though next-generation sequencing has provided insight into many species' genomes and transcriptomes, it will be a long path to narrow down these findings and identify the cellular pathways responsible for stress resistance and regulatory molecules.

STRESS-RESPONSIVE lncRNAs IN OTHER CROPS

Maize is another important crop and perhaps one of the plant species that has been most extensively studied for lncRNAs. Maize lncRNAs are mostly single exonic and found in intergenic regions, whereas only a small portion coincide with protein-coding genes on the genome (Li L. et al., 2014). Attempts to find lncRNAs responsive to drought revealed 664 lncRNAs

that were differentially expressed under drought stress, and were also identified as potential precursors for small non-coding RNA (snRNA) species such as miRNAs, siRNAs, and shRNAs (Zhang et al., 2014). In addition to drought, differentially expressed lncRNAs were predicted in maize under nitrogen deficiency, with most being downregulated. These nitrogen deficiency responsive lncRNAs were examined for co-expression with protein-coding transcripts; 32 were co-expressed with 239 protein-coding transcripts in functional annotation categories including NADPH/NADH dehydrogenation, indicating that these lncRNAs are potential regulators of nitrogen assimilation and photosynthesis since elevated NADH/NADPH consumption is associated with nitrogen assimilation and since photosynthesis reactions are the most important NADPH resources (Lv et al., 2016). In rice (*Oryza sativa*), lncRNAs were investigated under drought and cadmium stress. Under drought stress, 98 lncRNAs were differentially regulated (Chung et al., 2016). Under cadmium stress, 122 of the differentially-expressed transcripts were defined as lncRNAs (He et al., 2015). However, the functions of these lncRNAs are unclear. As in cereals, attempts to discover stress responsive lncRNAs in other crops are still in progress.

DRAWBACKS IN lncRNA IDENTIFICATION AND TARGET PREDICTION

Current methods used to identify lncRNAs are not sufficiently accurate or comprehensive. In the absence of a standardized set of selection criteria, researchers must design their own pipelines and decide on the thresholds and tools to use, which may cause incorrect and conflicting results to accumulate in the literature and in databases. Despite continuous efforts to identify lncRNAs from many species, methods developed to date are far from complete, especially due to the complex and unclear nature of these molecules.

In contrast to mRNAs, lncRNAs rarely show evolutionary sequence conservation among species (Ponjavic et al., 2007; Diederichs, 2014). Therefore, instead of directly selecting transcripts that show sequence similarities with lncRNAs of closely related species, lncRNA identification pipelines highly depend on the elimination of RNAs that exhibit mRNA-like and snRNA-like features and classification of the remaining transcripts as lncRNAs. However, the precise identification of the whole lncRNA repertoire for an organism seems impossible due to transcripts that are short and protein coding, and transcripts that are non-coding with long ORFs. Therefore, researchers should also be cautious when considering novel protein-coding transcripts; some transcripts that do not show homology to known sequences stored in public databases might represent undiscovered short protein-coding sequences that could be misannotated as lncRNAs.

Although identifying conserved lncRNA sequences has proven challenging, studies of plant and animal transcriptomes have suggested better sequence conservation at lncRNA promoter sites of vertebrates than the sequence conservation at lncRNA transcripts, particular gene structures and locations

around protein-coding genes (Kutter et al., 2012; Johnsson et al., 2014; Nitsche et al., 2015; Deng et al., 2018; Singh et al., 2018), as well as at the structural and functional levels (Kashi et al., 2016). Such positional information and gene structure characteristics such as splice sites will reveal lncRNA genes in other organisms and guide researchers toward more accurate lncRNA identification; however, this approach requires high-quality reference genomes and transcriptomes. Moreover, the features of lncRNAs when folded into secondary and tertiary structures and the relationship between conformation and function suggest another promising opportunity for better lncRNA prediction. However, *in silico* RNA folding algorithms are usually more inaccurate as the transcript length increases (Mathews and Turner, 2006). Even though the relationship between structure and function has been examined for few lncRNAs, studies evaluating the complete folding process of lncRNAs have identified domains that might be important for functional interactions and have compared the folding characteristics of lncRNA with other RNA species computationally (Yang and Zhang, 2014; Liu et al., 2017). Considering that lncRNA secondary and tertiary structures might be important for their interactions and cellular activities (Johnsson et al., 2014) and considering that even lncRNAs that are not conserved can still adopt the same secondary structures (Diederichs, 2014), gaining more information about lncRNA folding might contribute to lncRNA identification by facilitating searches for evolutionary conservation in secondary and tertiary structures instead of in the primary sequences.

The low expression of lncRNAs and expression profiles that are tissue- or developmental stage-specific have further hindered their discovery (Tsoi et al., 2015). Expression profiles of lncRNAs might also provide clues for the prediction of new lncRNAs. However, transcripts with low abundance are usually harder to capture with conventional RNA sequencing applications (Clark et al., 2015; Kashi et al., 2016). Tissue- or developmental stage-specific lncRNAs are also difficult to detect. The time or conditions of sample collection can directly affect which lncRNAs appear in the sequencing results and exclude others expressed at different stages, in different tissues, or under different conditions.

FUTURE PERSPECTIVES AND CONCLUSION

De novo assembled partial transcripts used to cause trouble in the identification of any molecule, leading to false annotations or underestimation of transcriptomes. Especially in the case of lncRNAs, these erroneous annotations become very hard to distinguish due to the fact that lncRNAs lack sequence conservation. For that reason, obtaining a well-assembled transcriptome data and having chance to locate the annotated lncRNAs will greatly advance the lncRNA identification procedures.

Now that we have the high-quality reference genomes of wheat and barley, it is now time to use them as efficient as possible. To do that, both breeders, biologists and bioinformaticians should undertake responsibilities and work for better tools and

methods. Drawbacks that has been encountered in currently-used lncRNA identification strategies should be overcome for a better understanding of mechanisms lying behind important traits to be used for developing more resistant and more yielding cultivars. Despite the fact that it is challenging, machine learning approaches give promising outcomes in terms of the identification of a group of non-conserved molecules, lncRNAs. Further development of these approaches may lead us to discover other features of lncRNAs that are conserved, such as location, folding characteristics or function. For instance, development of better algorithms that assess folding of lncRNA transcripts would provide clues on their interaction interfaces and thus, on their interacting partners. Similarly, gaining more idea about the

interacting partners of a lncRNA would direct us to its function in molecular pathways. Altogether, even though we still have a long way to go until perfectness in lncRNA identification, wheat and barley reference sequences provides a more precise perspective. Better understanding the world of lncRNAs by the help of reference sequences would lead us to the development better cultivars to feed the planet.

AUTHOR CONTRIBUTIONS

HB conceived and designed the study. HB, SK, and HC wrote the article.

REFERENCES

- Akpinar, B. A., Biyiklioglu, S., Alptekin, B., Havrankova, M., Vrana, J., and Dole, J. (2018). Chromosome-based survey sequencing reveals the genome organization of wild wheat progenitor *Triticum dicoccoides*. *Plant Biotechnol. J.* 2018, 1–11. doi: 10.1111/pbi.12940
- Ariel, F., Jegu, T., Latrasse, D., Romero-Barrios, N., Christ, A., Benhamed, M., et al. (2014). Noncoding transcription by alternative rna polymerases dynamically regulates an auxin-driven chromatin loop. *Mol. Cell* 55, 383–396. doi: 10.1016/j.molcel.2014.06.011
- Ariel, F., Romero-Barrios, N., Jégu, T., Benhamed, M., and Crespi, M. (2015). Battles and hijacks: Noncoding transcription in plants. *Trends Plant Sci.* 20, 362–371. doi: 10.1016/j.tplants.2015.03.003
- Banfai, B., Jia, H., Khatun, J., Wood, E., Risk, B., William, G. Jr., et al. (2012). Long noncoding RNAs are rarely translated in two human cell lines. *Genome Res.* 22, 1646–1657. doi: 10.1101/gr.134767.111
- Boerner, S., and McGinnis, K. M. (2012). Computational identification and functional predictions of long noncoding RNA in *Zea mays*. *PLoS One* 7:e43047. doi: 10.1371/journal.pone.0043047
- Britton, C., Winter, A. D., Gillan, V., and Devaney, E. (2014). MicroRNAs of parasitic helminths - Identification, characterization and potential as drug targets. *Int. J. Parasitol. Drugs Drug Resist.* 4, 85–94. doi: 10.1016/j.ijppdr.2014.03.001
- Cagirici, H. B., Alptekin, B., and Budak, H. (2017). RNA sequencing and co-expressed long non-coding RNA in modern and wild wheats. *Sci. Rep.* 7, 1–16. doi: 10.1038/s41598-017-11170-8
- Campalans, A. (2004). Enod40, a short open reading frame-containing mRNA, Induces cytoplasmic localization of a nuclear RNA binding protein in *Medicago truncatula*. *Plant Cell Online* 16, 1047–1059. doi: 10.1105/tpc.019406
- Chekanova, J. A. (2015). Long non-coding RNAs and their functions in plants. *Curr. Opin. Plant Biol.* 27, 207–216. doi: 10.1016/j.pbi.2015.08.003
- Chen, J., Liang, Z., Liang, Y., Pang, R., and Zhang, W. (2013). Conserved microRNAs miR-8-5p and miR-2a-3p modulate chitin biosynthesis in response to 20-hydroxyecdysone signaling in the brown planthopper, *Nilaparvata lugens*. *Insect Biochem. Mol. Biol.* 43, 839–848. doi: 10.1016/j.ibmb.2013.06.002
- Chu, C., Qu, K., Zhong, F., Artandi, S., and Chang, H. (2011). Genomic maps of lincRNA occupancy reveal principles of RNA-chromatin interactions. *Mol. Cell* 44, 667–678. doi: 10.1016/j.molcel.2011.08.027
- Chu, C., Spitale, R. C., and Chang, H. Y. (2015). Technologies to probe functions and mechanisms of long noncoding RNAs. *Nat. Struct. Mol. Biol.* 22, 29–35. doi: 10.1038/nsmb.2921
- Chung, P. J., Jung, H., Jeong, D., Ha, S., Choi, Y., and Kim, J. (2016). Transcriptome profiling of drought responsive noncoding RNAs and their target genes in rice. *BMC Genom.* 17:563. doi: 10.1186/s12864-016-2997-3
- Clark, M. B., Mercer, T. R., Bussotti, G., Leonardi, T., Haynes, K. R., Crawford, J., et al. (2015). Quantitative gene profiling of long noncoding RNAs with targeted RNA sequencing. *Nat. Methods* 12, 339–342. doi: 10.1038/nmeth.3321
- Csorba, T., Questa, J. I., Sun, Q., and Dean, C. (2014). Antisense COOLAIR mediates the coordinated switching of chromatin states at FLC during vernalization. *PNAS* 111, 16160–16165. doi: 10.1073/pnas.1419030111
- Deng, P., Liu, S., Nie, X., Weining, S., and Wu, L. (2018). Conservation analysis of long non-coding RNAs in plants. *Sci. China Life Sci.* 61, 190–198. doi: 10.1007/s11427-017-9174-9
- Denoeud, F., Aury, J., Silva, C. D., Noel, B., Rogier, O., Delledonne, M., et al. (2008). Annotating genomes with massive-scale RNA sequencing. *Genome Biol.* 9:R175. doi: 10.1186/gb-2008-9-12-r175
- Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* 22, 1775–1789. doi: 10.1101/gr.132159.111
- Diederichs, S. (2014). The four dimensions of noncoding RNA conservation. *Trends Genet.* 30, 121–123. doi: 10.1016/j.tig.2014.01.004
- Ding, J., Lu, Q., Ouyang, Y., Mao, H., Zhang, P., Yao, J., et al. (2012). A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice. *PNAS* 109, 2654–2659. doi: 10.1073/pnas.1121374109/
- Franco-Zorrilla, J. M., Valli, A., Todesco, M., Mateos, I., Puga, M. I., Rubio-Somoza, I., et al. (2007). Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat. Genet.* 39, 1033–1037. doi: 10.1038/ng2079
- Gozukirmizi, N., and Karlik, E. (2017). *Barley Hordeum vulgare L. Improvement Past, Present and Future*. London: IntechOpen.
- Gulyaev, A. P., and Roussis, A. (2007). Identification of conserved secondary structures and expansion segments in enod40 RNAs reveals new enod40 homologues in plants. *Nucleic Acids Res.* 35, 3144–3152. doi: 10.1093/nar/gkm173
- Guo, G., Liu, X., Sun, F., Cao, J., Huo, N., Wuda, B., et al. (2018). Wheat miR9678 affects seed germination by generating phased siRNAs and modulating abscisic acid/gibberellin signaling. *Plant Cell* 30, 796–814. doi: 10.1105/tpc.17.00842
- Guttman, M., Amit, I., Garber, M., French, C., Lin, M. F., Feldser, D., et al. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458, 223–227. doi: 10.1038/nature07672
- Guttman, M., Russell, P., Ingolia, N. T., Weissman, J. S., and Lander, E. S. (2013). Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* 154, 240–251. doi: 10.1016/j.cell.2013.06.009
- He, F., Liu, Q., Zheng, L., Cui, Y., Shen, Z., and Zheng, L. (2015). RNA-seq analysis of rice roots reveals the involvement of post-transcriptional regulation in response to cadmium stress. *Front. Plant Sci.* 6:1136. doi: 10.3389/fpls.2015.01136
- Hedtke, B., and Grimm, B. (2009). Silencing of a plant gene by transcriptional interference. *Nucleic Acids Res.* 37, 3739–3746. doi: 10.1093/nar/gkp241
- Held, M. A., Penning, B., Brandt, A. S., Kessans, S. A., Yong, W., Scofield, S. R., et al. (2008). Small-interfering RNAs from natural antisense transcripts derived from a cellulose synthase gene modulate cell wall biosynthesis in barley. *Proc. Natl. Acad. Sci. U.S.A.* 105, 20534–20539. doi: 10.1073/pnas.0809408105
- Heo, J. B., and Sung, S. (2011). Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science* 331, 76–79. doi: 10.1126/science.1197349
- Hoff, K. J., and Stanke, M. (2013). WebAUGUSTUS—a web service for training augustus and predicting genes in eukaryotes. *Nucleic Acids Res.* 41, W123–W128. doi: 10.1093/nar/gkt418

- Housman, G., and Ulitsky, I. (2016). Methods for distinguishing between protein-coding and long noncoding RNAs and the elusive biological purpose of translation of long noncoding RNAs. *Biochim. Biophys. Acta Gene Regul. Mech.* 1859, 31–40. doi: 10.1016/j.bbagr.2015.07.017
- Ilott, N. E., and Ponting, C. P. (2013). Predicting long non-coding RNAs using RNA sequencing. *Methods* 63, 50–59. doi: 10.1016/j.ymeth.2013.03.019
- IWGSC (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361:eaar7191. doi: 10.1126/science.aar7191
- Jabnoun, M., Secco, D., Lecampion, C., Robaglia, C., Shu, Q., and Poirier, Y. (2013). 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, 4166–4182. doi: 10.1105/tpc.113.116251
- Jia, H. U. I., Osak, M., Bogu, G. K., Stanton, L. W., Johnson, R., and Lipovich, L. (2010). Genome-wide computational identification and manual annotation of human long noncoding RNA genes. *RNA J.* 16, 1478–1487. doi: 10.1261/rna.1951310.4
- Jia, J., Zhao, S., Kong, X., Li, Y., Zhao, G., He, W., et al. (2013). Aegilops tauschii draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature* 496, 91–95. doi: 10.1038/nature12028
- Jiao, F., Pahwa, K., Manning, M., Dochy, N., and Geuten, K. (2019). Cold induced antisense transcription of flowering locus C in distant grasses. *Front. Plant Sci.* 10:72. doi: 10.3389/fpls.2019.00072
- Johnsson, P., Lipovich, L., Grandér, D., and Morris, K. V. (2014). Evolutionary conservation of long non-coding RNAs; sequence, structure, function. *Biochim. Biophys. Acta Gen. Subj.* 1840, 1063–1071. doi: 10.1016/j.bbagen.2013.10.035
- Kapranov, P., Cheng, J., Dike, S., Nix, D. A., Duttaputra, R., Willingham, A. T., et al. (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* 316, 1484–1488. doi: 10.1126/science.1138341
- Karakulah, G., and Unver, T. (2017). Barley long non-coding RNAs and their tissue-specific co-expression pattern with coding-transcripts. *bioRxiv* [Preprint], doi: 10.1101/229559
- Karlik, E., and Gozukirmizi, N. (2018). Expression analysis of lncRNA AK370814 involved in the barley vitamin B6 salvage pathway under salinity. *Mol. Biol. Rep.* 45, 1597–1609. doi: 10.1007/s11033-018-4289-2
- Kashi, K., Henderson, L., Bonetti, A., and Carninci, P. (2016). Discovery and functional analysis of lncRNAs: methodologies to investigate an uncharacterized transcriptome. *Biochim. Biophys. Acta Gene Regul. Mech.* 1859, 3–15. doi: 10.1016/j.bbagr.2015.10.010
- Kong, L., Zhang, Y., Ye, Z. Q., Liu, X. Q., Zhao, S. Q., Wei, L., et al. (2007). CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* 35, 345–349. doi: 10.1093/nar/gkm391
- Kung, J. T. Y., Colognori, D., and Lee, J. T. (2013). Long noncoding RNAs: Past, present, and future. *Genetics* 193, 651–669. doi: 10.1534/genetics.112.146704
- Kutter, C., Watt, S., Stefflova, K., Wilson, M. D., Goncalves, A., Ponting, C. P., et al. (2012). Rapid turnover of long noncoding RNAs and the evolution of gene expression. *PLoS Genet.* 8:e1002841. doi: 10.1371/journal.pgen.1002841
- Latos, P. A., Pauley, F. M., Koerner, M. V., Senegin, H. B., Hudson, Q. J., Stocsits, R. R., et al. (2012). Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science* 338, 1469–1472. doi: 10.1126/science.1228110
- Li, J., Liu, S., Zhou, H., Qu, L., and Yang, J. (2014). miRNA-ncRNA and protein – RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* 42, 92–97. doi: 10.1093/nar/gkt1248
- Li, L., Eichten, S. R., Shimizu, R., Petsch, K., Yeh, C., Wu, W., et al. (2014). Genome-wide discovery and characterization of maize long non-coding RNAs. *Genome Biol.* 15, 1–15. doi: 10.1186/gb-2014-15-2-r40
- Li, S., Vandivier, L. E., Tu, B., Gao, L., Won, S. Y., Li, S., et al. (2015). Detection of pol IV/RDR2-dependent transcripts at the genomic scale in Arabidopsis reveals features and regulation of siRNA biogenesis. *Genome Res.* 25, 235–245. doi: 10.1101/gr.182238.114
- Liao, Q., Liu, C., Yuan, X., Kang, S., Miao, R., Xiao, H., et al. (2011). Large-scale prediction of long non-coding RNA functions in a coding-non-coding gene co-expression network. *Nucleic Acids Res.* 39, 3864–3878. doi: 10.1093/nar/gkq1348
- Ling, H.-Q., Zhao, S., Liu, D., Wang, J., Sun, H., Zhang, C., et al. (2013). Draft genome of the wheat A-genome progenitor *Triticum urartu*. *Nature* 496, 87–90. doi: 10.1038/nature11997
- Liu, F., Somarowthu, S., and Pyle, A. M. (2017). Visualizing the secondary and tertiary architectural domains of lncRNA RepA. *Nat. Chem. Biol.* 13, 282–289. doi: 10.1038/nchembio.2272
- Liu, J., Jung, C., Xu, J., Wang, H., Deng, S., Bernad, L., et al. (2012). Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in *Arabidopsis*. *Plant Cell* 24, 4333–4345. doi: 10.1105/tpc.112.102855
- Liu, T. T., Zhu, D., Chen, W., Deng, W., He, H., He, G., et al. (2013). A global identification and analysis of small nucleolar RNAs and possible intermediate-sized non-coding RNAs in rice. *Mol. Plant* 6, 830–846. doi: 10.1093/mp/sss087
- Liu, X., Li, D., Zhang, D., Yin, D., Zhao, Y., Ji, C., et al. (2018). A novel antisense long noncoding RNA, TWISTED LEAF, maintains leaf blade flattening by regulating its associated sense R2R3-MYB gene in rice. *PLoS Genet.* 8:e1002841. doi: 10.1111/nph.15023
- Luo, M., Gu, Y. Q., Puiu, D., Wang, H., Twardziok, S. O., Deal, K. R., et al. (2017). Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*. *Nat. Publ. Gr.* 551, 498–502. doi: 10.1038/nature24486
- Lv, J., Liu, H., Huang, Z., Su, J., He, H., Xiu, Y., et al. (2013). Long non-coding RNA identification over mouse brain development by integrative modeling of chromatin and genomic features. *Nucleic Acids Res.* 41, 10044–10061. doi: 10.1093/nar/gkt818
- Lv, Y., Liang, Z., Ge, M., Qi, W., Zhang, T., Lin, F., et al. (2016). Genome-wide identification and functional prediction of nitrogen-responsive intergenic and intronic long non-coding RNAs in maize (*Zea mays* L.). *BMC Genomics* 17:350. doi: 10.1186/s12864-016-2650-1
- Ma, K., Shi, W., Xu, M., Liu, J., and Zhang, F. (2018). Genome-wide identification and characterization of long non-coding RNA in wheat roots in response to Ca²⁺ + Channel blocker. *Front. Plant Sci.* 9:244. doi: 10.3389/fpls.2018.00244
- Ma, L., Bajic, V. B., and Zhang, Z. (2013). On the classification of long non-coding RNAs. *RNA Biol.* 10, 924–933. doi: 10.4161/rna.24604
- Maccaferri, M., Harris, N. S., Twardziok, S. O., Pasam, R. K., Gundlach, H., Spannagl, M., et al. (2019). Durum wheat genome highlights past domestication signatures and future improvement targets. *Nat. Genet.* 51, 885–895. doi: 10.1038/s41588-019-0381-3
- Malone, J. H., and Oliver, B. (2011). Microarrays, deep sequencing and the true measure of the transcriptome. *BMC Biol.* 9:34. doi: 10.1186/1741-7007-9-34
- Martin, J. A., and Wang, Z. (2011). Next-generation transcriptome assembly. *Nat. Rev. Genet.* 12:671. doi: 10.1038/nrg3068
- Mascher, M., Gundlach, H., Himmelbach, A., Beier, S., Twardziok, S. O., Wicker, T., et al. (2017). A chromosome conformation capture ordered sequence of the barley genome. *Nat. Publ. Gr.* 544, 427–433. doi: 10.1038/nature22043
- Mathews, D. H., and Turner, D. H. (2006). Prediction of RNA secondary structure by free energy minimization. *Curr. Opin. Struct. Biol.* 16, 270–278. doi: 10.1016/j.sbi.2006.05.010
- Mattick, J. S., and Rinn, J. L. (2015). Discovery and annotation of long noncoding RNAs. *Nat. Publ. Gr.* 22, 5–7. doi: 10.1038/nsm.2942
- Matzke, M. A., and Mosher, R. A. (2014). RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat. Rev. Genet.* 15, 570–570. doi: 10.1038/nrg3794
- Mercer, T. R., Gerhardt, D. J., Dinger, M. E., Crawford, J., Trapnell, C., Jeddell, J. A., et al. (2011). Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nat. Biotechnol.* 30, 99–104. doi: 10.1038/nbt.2024
- Milligan, M. J., and Lipovich, L. (2015). Pseudogene-derived lncRNAs: Emerging regulators of gene expression. *Front. Genet.* 6:476. doi: 10.3389/fgene.2014.00476
- Moore, G., Devos, K. M., Wang, Z., and Gale, M. D. (1995). Grasses, line up and form a circle. *Curr. Biol.* 5, 737–739. doi: 10.1016/s0960-9822(95)00148-5
- Murigneux, V., Saulière, J., Crollius, H. R., and Hir, H. (2013). Transcriptome-wide identification of RNA binding sites by CLIP-seq. *Methods* 63, 32–40. doi: 10.1016/j.ymeth.2013.03.022
- Musacchia, F., Basu, S., Petrosino, G., Salvemini, M., and Sanges, R. (2015). Annotript: A flexible pipeline for the annotation of transcriptomes able to identify putative long noncoding RNAs. *Bioinformatics* 31, 2199–2201. doi: 10.1093/bioinformatics/btv106
- Nam, J.-W., Choi, S.-W., and Bo-Hyun, Y. (2016). Incredible RNA: dual functions of coding and noncoding. *Mol. Cells* 39, 367–374. doi: 10.14348/molcells.2016.0039
- Nitsche, A., Rose, D., Fasold, M., Reiche, K., and Stadler, P. F. (2015). Comparison of splice sites reveals that long noncoding RNAs are evolutionarily well conserved. *RNA J.* 21, 801–812. doi: 10.1261/rna.046342.114

- Ozsolak, F., and Milos, P. (2011). RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet.* 12, 87–98. doi: 10.1038/nrg2934.RNA
- Pauli, A., Valen, E., Lin, M. F., Garber, M., Vastenhouw, N. L., Levin, J. Z., et al. (2012). Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res.* 22, 577–591. doi: 10.1101/gr.133009.111.2011
- Pieri, A., Pe, M. E., and Bertolini, E. (2018). Long non-coding RNAs in wild wheat progenitors. *bioRxiv* [Preprint]. Available online at: <https://www.biorxiv.org/content/10.1101/301804v1> (accessed 29 September 2019).
- Ponjavic, J., Ponting, C. P., Lunter, G., Ponjavic, J., Ponting, C. P., and Lunter, G. (2007). Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Res.* 17, 556–565. doi: 10.1101/gr.6036807
- Popa, A., Lebrigand, K., Barbry, P., and Waldmann, R. (2016). Pateamine A-sensitive ribosome profiling reveals the scope of translation in mouse embryonic stem cells. *BMC Genomics* 17:52. doi: 10.1186/s12864-016-2384-0
- Quattro, C. D., Enrico, P. M., and Bertolini, E. (2017). Long noncoding RNAs in the model species *Brachypodium distachyon*. *Sci. Rep.* 200, 618–627. doi: 10.1038/s41598-017-11206-z
- Radhamony, R. N., Prasad, A. M., and Srinivasan, R. (2005). T-DNA insertional mutagenesis in *Arabidopsis*: a tool for functional genomics. *Electron. J. Biotechnol.* 8, 82–106. Available online at: <http://www.ejbiotechnology.info/index.php/ejbiotechnology/article/view/219>
- Rai, M. I., Alam, M., Lightfoot, D. A., Gurha, P., and Afzal, A. J. (2018). Genomics Classification and experimental identification of plant long non-coding RNAs. *Genomic* 111, 997–1005. doi: 10.1016/j.jgeno.2018.04.014
- Rohrig, H., Schmidt, J., Miklashevichs, E., Schell, J., and John, M. (2002). Soybean ENOD40 encodes two peptides that bind to sucrose synthase. *Proc. Natl. Acad. Sci. U.S.A.* 99, 1915–1920. doi: 10.1073/pnas.022664799
- Ruiz-Orera, J., Messegue, X., Subirana, J. A., and Alba, M. M. (2014). Long non-coding RNAs as a source of new peptides. *eLife* 3, 1–24. doi: 10.7554/eLife.03523
- Santorio, F., Mayer, D., Klement, R. M., Warczok, K. E., Stukalov, A., Barlow, D. P., et al. (2013). Imprinted Igf2r silencing depends on continuous Airn lncRNA expression and is not restricted to a developmental window. *Development* 140, 1184–1195. doi: 10.1242/dev.088849
- Shafiq, S., Li, J., and Sun, Q. (2015). Functions of plants long non-coding RNAs. *Biochim. Biophys. Acta* 1008, 133–154. doi: 10.1016/j.bbagr.2015.06.009
- Sheng, L., Ye, L., Zhang, D., Cawthorn, W. P., and Xu, B. (2018). New insights into the long non-coding RNA SRA: physiological functions and mechanisms of action. *Front. Med.* 5:244. doi: 10.3389/fmed.2018.00244
- Shuai, P., Liang, D., Tang, S., Zhang, Z., Ye, C. Y., Su, Y., et al. (2014). Genome-wide identification and functional prediction of novel and drought-responsive lincRNAs in *Populus trichocarpa*. *J. Exp. Bot.* 65, 4975–4983. doi: 10.1093/jxb/eru256
- Shumayla, S. S., Taneja, M., Tyagi, S., and Singh, K. (2017). Survey of high throughput RNA-Seq data reveals potential roles for lncRNAs during development and stress response in bread wheat. *Front. Plant Sci* 8:1019. doi: 10.3389/fpls.2017.01019
- Singh, P. K., Ganguli, S., and Pal, A. (2018). Functions of long non-coding RNAs in plants: a riddle to explore. *Nucleus* 61, 261–272. doi: 10.1007/s13237-018-0249-4
- Sun, L., Luo, H., Bu, D., Zhao, G., Yu, K., Zhang, C., et al. (2013). Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.* 41, e166. doi: 10.1093/nar/gkt646
- Tsoi, L. C., Iyer, M. K., Stuart, P. E., Swindell, W. R., Gudjonsson, J. E., Tejasvi, T., et al. (2015). Analysis of long non-coding RNAs highlights tissue-specific expression patterns and epigenetic profiles in normal and psoriatic skin. *Genome Biol.* 16:24. doi: 10.1186/s13059-014-0570-4
- Wang, Y., Fan, X., Lin, F., He, G., Terzaghi, W., Zhu, D., et al. (2014). Arabidopsis noncoding RNA mediates control of photomorphogenesis by red light. *Proc. Natl. Acad. Sci.* 111, 10359–10364. doi: 10.1073/pnas.1409457111
- Wang, Y., Luo, X., Sun, F., Hu, J., Zha, X., Su, W., et al. (2018). Overexpressing lincRNA LAIR increases grain yield and regulates neighbouring gene cluster expression in rice. *Nat. Commun.* 9:3516. doi: 10.1038/s41467-018-05829-7
- Wierzbicki, A. T., Haag, J. R., and Pikaard, C. S. (2008). Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell* 135, 635–648. doi: 10.1016/j.cell.2008.09.035
- Xin, M., Wang, Y., Yao, Y., Song, N., Hu, Z., Qin, D., et al. (2011). Identification and characterization of wheat long non-protein coding RNAs responsive to powdery mildew infection and heat stress by using microarray analysis and SBS sequencing. *BMC Plant Biol.* 11:61. doi: 10.1186/1471-2229-11-61
- Yamaguchi, A., and Abe, M. (2012). Regulation of reproductive development by non-coding RNA in *Arabidopsis*: to flower or not to flower. *J. Plant Res.* 125, 693–704. doi: 10.1007/s10265-012-0513-7
- Yang, J., and Zhang, J. (2014). Human long noncoding RNAs are substantially less folded than messenger RNAs. *Mol. Biol. Evol.* 32, 970–977. doi: 10.1093/molbev/msu402
- Yang, W. C., Katinakis, P., Hendriks, P., Smolders, A., Devries, F., Spee, J., et al. (1993). Characterization of Gmenod40, a gene showing novel patterns of cell-specific expression during soybean nodule development. *Plant J.* 3, 573–585. doi: 10.1046/j.1365-313X.1993.03040573.x
- You, W., Lorkovic, Z. J., Matzke, A. J. M., and Matzke, M. (2013). Interplay among RNA polymerases II, IV and V in RNA-directed DNA methylation at a low copy transgene locus in *Arabidopsis thaliana*. *Plant Mol. Biol.* 82, 85–96. doi: 10.1007/s11103-013-0041-4
- Zhang, W., Han, Z., Guo, Q., Liu, Y., Zheng, Y., Wu, F., et al. (2014). Identification of maize long non-coding RNAs responsive to drought stress. *PLoS One* 9:e098958. doi: 10.1371/journal.pone.0098958
- Zhang, Y. C., and Chen, Y. Q. (2013). Long noncoding RNAs: new regulators in plant development. *Biochem. Biophys. Res. Commun.* 436, 111–114. doi: 10.1016/j.bbrc.2013.05.086
- Zhou, M., and Law, J. A. (2015). RNA Pol IV and V in gene silencing: rebel polymerases evolving away from Pol II's rules. *Curr. Opin. Plant Biol.* 27, 154–164. doi: 10.1016/j.pbi.2015.07.005
- Zhu, Q.-H., Stephen, S., Taylor, J., Helliwell, C. A., and Wang, M.-B. (2014). Long noncoding RNAs responsive to *Fusarium oxysporum* infection in *Arabidopsis thaliana*. *New Phytol.* 201, 574–584. doi: 10.1111/nph.12537

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Budak, Kaya and Cagirici. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.