RUSSIAN JOURNAL OF GENETICS, 2025, Vol. 61, No. 1, pp. 3–23

REVIEW AND THEORETICAL ARTICLES

THE ROLE OF LONG NON-CODING RNAs IN PLANTS

© 2025 A. Yu. Pronozin[a,b,*], D. A. Afonnikov[a,b,c]

[a]*Institute of Cytology and Genetics, Siberian Branch, Russian Academy of Sciences, Novosibirsk, Russia*

[b]*Kurchatov Genomic Center, Institute of Cytology and Genetics, Siberian Branch, Russian Academy of Sciences, Novosibirsk, Russia*

[c]*Novosibirsk National Research State University, Novosibirsk, Russia*

[*]*e-mail: pronozinartem95@gmail.com*

Received April 19, 2024
Revised July 29, 2024
Accepted August 15, 2024

Abstract. Long non-coding RNAs (lncRNAs) are a class of linear or circular RNA molecules longer than 200 nucleotides without open reading frames. Experimental studies have shown the involvement of lncRNAs in the regulation of resistance to cold, salt, and heat stress, and in fruit, root, and leaf development. However, experimental methods are labor-intensive and costly approaches and cannot yet be used for genome-wide mass studies of lncRNAs. For this purpose, bioinformatic approaches that aim at large-scale recognition of lncRNA sequences in genomes and transcriptomes have been applied. However, despite the growing number of studies devoted to the structural and functional analysis of lncRNAs, this type of molecule remains poorly understood. This is due to the many factors that need to be considered when identifying lncRNAs. The use of pan-genomes and pan-transcriptomes will improve the efficiency of the study and the total number of predicted lncRNAs compared to using the genome of a single species representative. This review focuses on describing the molecular and biological functions of lncRNAs, experimental and bioinformatic methods of identification, patterns of evolution, detection and analysis of lncRNAs at the scale of pan-genomes and pan-transcriptomes.

Keywords: *long non-coding RNAs, microRNAs, transcription regulation, pan-genome*

DOI: 10.31857/S00166758250101e1

INTRODUCTION

Non-coding RNAs (ncRNAs) perform a number of essential functions in plant genomes related to the regulation of gene expression and homeostasis of plant physiological parameters. Knowledge of the functional characteristics of ncRNAs allows for targeted improvement of various plant properties [1–3], which drives interest in studying these molecules. Non-coding RNAs can be divided into two groups: structural non-coding (ribosomal RNAs, transfer RNAs, small nuclear RNAs, small nucleolar RNAs) and regulatory non-coding RNAs (small non-coding RNAs and long non-coding RNAs) [4]. The least studied of all these molecular classes are long ncRNAs [5, 6].

Long non-coding RNAs (lncRNAs) are an important class of non-coding RNAs in living organisms and represent linear or circular RNA molecules containing more than 200 nucleotides without open reading frames [4, 6]. The functions and structure of lncRNAs are poorly studied compared to protein-coding genes and small non-coding RNA genes. However, the involvement of lncRNAs has been identified in gene expression regulation [7], formation of macromolecular complex structures [8], interaction with proteins [9], and pathogenesis [10]. Long ncRNAs are most actively studied experimentally and bioinformatically in humans and other model species (mouse, rat, drosophila, etc.). These studies allow identification of lncRNA sequences in genomes and transcriptomes, deciphering important features of their structure, expression regulation and functioning, and classifying them by structure and function. To date, these methods have made it possible to identify a large number of lncRNA sequences in animals. According to the GENCODE database [11], about 13,000 lncRNA genes are described in the mouse genome. The human genome contains more than 18,000 lncRNA

genes [12]. The study of lncRNAs in animals has allowed their classification according to molecular mechanism of functioning (promoter, enhancer, telomeric, signaling, etc.) [13–15], genomic position (intergenic, intronic, antisense, etc.) [7, 16–18]. These studies lead to the creation of specialized databases dedicated to lncRNAs, containing sequences of tens of thousands of lncRNAs and their structural and functional annotation [19, 20]. Due to the absence of clearly defined functional segments in the primary structure of lncRNAs (such as domains in protein sequences), predicting the function of lncRNAs based on primary structure is currently a complex task [15, 21, 22].

It should be noted that obtaining experimental information about the structure and function of lncRNAs is a labor-intensive process and requires expensive experiments. Therefore, bioinformatics methods are actively used to study this class of RNAs. In particular, they are aimed at large-scale recognition of lncRNA sequences in genomes and transcriptomes [23–25], prediction of their interactions with proteins [26], with mRNAs and other non-coding RNAs [27], regulation of expression [7], evolution [28, 29], and prediction of their structure and functions [30].

Long ncRNAs are also known in plants. It is interesting to note that, despite a number of specific features, the functional and structural characteristics of animal and plant lncRNAs have much in common [31]. It is known that in plants, lncRNAs are involved in regulating resistance to cold, salt, and heat stress [1–3], affect resistance to hypoxia [32], and participate in the development of fruits, roots, and leaves [17, 33]. Nevertheless, our knowledge of plant lncRNAs is still less complete than that of animal lncRNAs. The studies [34, 35] show that lncRNA sequences rapidly accumulate substitutions during evolution, so that they have weak homology even with sequences from closely related species. The application of the pan-genome concept, which implies coverage of sequences subject to structural variation and possibly absent in the reference sequence of each species representative, will increase the efficiency of research and the total number of predicted lncRNAs [36–40].

This review is devoted to plant lncRNAs and includes a description of their molecular and biological functions, experimental and bioinformatic identification methods, patterns of evolution, identification and analysis of lncRNAs on the scale of pan-genomes and pan-transcriptomes.

CLASSIFICATION OF lncRNAs BY GENOMIC LOCATION

Long ncRNAs are classified according to many parameters, but the most commonly used classification is based on their genomic location relative to known protein-coding genes. According to this classification, lncRNAs are divided into four major classes [35] (see Fig. 1):

1) intronic — overlap with the intron of a gene;
2) antisense — oriented against the transcription direction of a protein-coding gene;
3) intergenic — located between two gene loci;
4) sense — oriented in the direction of transcription of a protein-coding gene.

Initiation of transcription of intronic lncRNAs occurs in the direction of transcription of the target gene, termination occurs in the region overlapping with the exon of the target gene [41].

Transcription of antisense lncRNAs begins at the 3` end of the protein-coding gene. These lncRNAs are divided into two groups depending on the cis-trans geometric isomerism. Transcription of cis isomers of antisense lncRNA (cis-lncRNA) occurs from the opposite end of the protein-coding gene locus with which this transcript overlaps. Due to this, the lncRNA has a high level of complementarity with this gene. Transcription of trans isomers of antisense lncRNA occurs from the locus of the neighboring protein-coding gene with which this transcript overlaps. Thus, such transcripts can overlap with several genes, but have a low level of complementarity with them [42].

The length of intergenic lncRNAs exceeds 200 nt, these transcripts do not overlap with protein-coding genes. Transcript features of this class: high conservation, pronounced tissue specificity, and high stability. In the work [43] bioinformatic methods demonstrated the interaction of intergenic lncRNAs with mobile elements. The conducted research provides a systematic assessment of the contribution of mobile elements to the composition, evolutionary origin, and regulation of lncRNAs.

Sense lncRNAs are transcribed in the direction of transcription of protein-coding genes. This makes their classification as a separate class of lncRNAs or even their identification as lncRNAs quite controversial. The structural characteristics of such RNAs (length, number of exons and introns) may coincide with the structure of a protein-coding gene. For example, in the work [17] it is suggested that transcripts of this class of lncRNAs may completely overlap with protein-coding genes. However, in the work [41] examples of lncRNAs that arose due to alternative splicing or shortening of the first or last exon of genes are described. The authors suggest that these genes may represent isoforms of proteins encoded by these genes, rather than lncRNAs. For example, in the works [34, 41] RNAs that overlap with the exon of a protein-coding gene are not considered lncRNAs. Few such transcripts have been

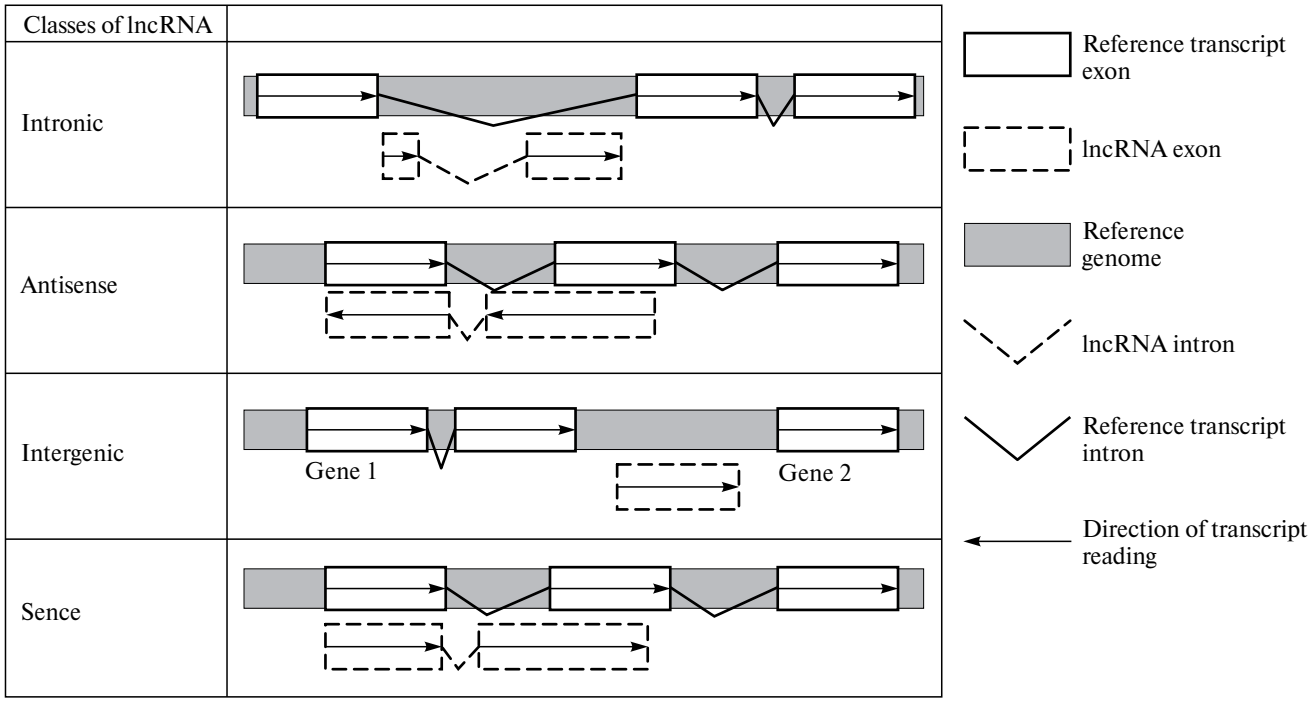

Fig. 1. Classification of lncRNAs based on location and orientation relative to the nearest or overlapping protein-coding genes

found for plants, and none of them have functional annotation [41].

MOLECULAR FUNCTIONS OF lncRNAs

Long ncRNAs have several common characteristics with protein-coding genes, which makes their identification difficult. For instance, most lncRNAs are transcribed by RNA polymerase II. They have similar epigenetic profiles, the presence of splicing and polyadenylation signals, as well as similar sizes of exons and introns [44, 45]. In addition, some lncRNAs have a length exceeding 1,000 nt. Thus, they may contain open reading frames (ORFs). The sizes of such ORFs can vary, some of them can be quite short. For example, in the work [46] it is mentioned that lncRNAs on average may contain several translated ORFs with a length of 50 codons.

A distinctive feature of lncRNA is the ability to be transcribed by RNA polymerase IV and V, especially in plants [47]. Most lncRNAs are localized both in the cell nucleus and in the cytoplasm [48]. Long ncRNAs also demonstrate a lower level of expression, while being expressed more specifically for certain tissues [49]. This is also observed when analyzing the conservatism of lncRNA structure. Comparison of lncRNAs from different plant species showed that the proportion of similar sequences among different species is small. However, lncRNAs of the same species demonstrate a fairly high level of conservatism in primary, secondary, and tertiary structure [15].

In works [14, 15, 50] it is indicated that the functioning of lncRNAs is also influenced by their intracellular localization. In works [4, 15, 50, 51] 5 main functions of lncRNAs are identified (see Fig. 2).

Signal lncRNAs

Long ncRNAs are most often expressed in specific tissues. Some studies note that lncRNAs are specifically expressed in response to various stimuli [4, 52–54]. In this regard, it can be assumed that the expression of some lncRNAs is under the control of transcription factors [55], occurs at a specific time and in a specific tissue, and can serve as a signal to trigger regulatory processes in the cell [55]. Such specific expression is characteristic of lncRNAs that are called signal lncRNAs (see Fig. 2*a*). Signal lncRNAs can act as regulatory elements. It is known, for example, that signal lncRNAs interact with chromatin-modifying enzymes, such as histone methyltransferases, to suppress the expression of their target genes through the formation of heterochromatin [56]. Such interactions are part of the regulatory and signaling functions of this type of lncRNA.

Decoy or trap lncRNAs

Decoy lncRNAs bind proteins and thereby prevent their interaction with regulatory regions of target genes [55]. These lncRNAs indirectly repress transcription by blocking microRNAs, modifying complexes (chromatin subunits), catalytic proteins, and transcription factors

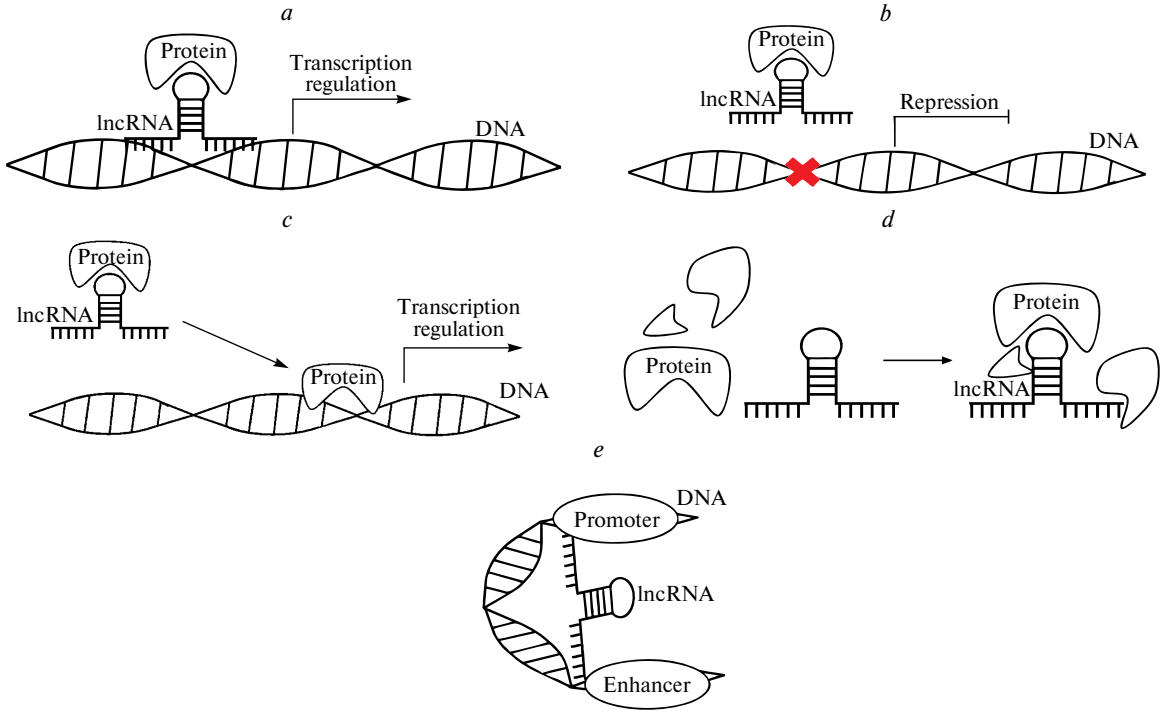

Fig. 2. Molecular functions of lncRNAs: *a* – signaling, lncRNA receives a signal to interact with chromatin-modifying enzymes to regulate transcription; *b* – decoys, bind proteins and prevent their attachment to target genes; *c* – guides, direct transcription factors to their functional sites, promote the necessary chromatin state; *d* – scaffolds, combine proteins into complexes and initiate various biological processes; *e* – enhancers, can transcriptionally activate neighboring genes

[55, 57]. Long ncRNAs also act as traps for microRNAs; this interaction mechanism is described in detail in the subsection "*Mechanisms of lncRNA and microRNA interaction*".

Guide lncRNAs

Long ncRNAs also function in grouping/localizing transcription factors in specific genomic loci. Such lncRNAs are called guide lncRNAs (see Fig. 2*b*). Long ncRNAs direct chromatin modifiers and protein complexes or transcription factors to specific target genes. They facilitate the localization of regulatory molecules in specific loci to implement transcriptional regulation [56]. Long ncRNAs have cis-trans geometric isomerism. Cis isomers direct chromatin changes, acting as complementary targets for small RNAs (such as microRNAs, small interfering RNAs) [55]. Trans isomers direct DNA targets through RNA-DNA interactions [58].

Long ncRNA scaffolds

Long ncRNA scaffolds play an important role in the assembly of macromolecular protein complexes and the initiation of various biological processes [59] (see Fig. 2*d*). Such lncRNAs can connect multiple proteins into a single ribonucleic complex, which allows suppression or activation of transcription depending on the presence and nature of the involved lncRNAs and proteins [55].

Enhancer lncRNAs

Long ncRNAs transcribed from enhancer loci are called enhancer lncRNAs (see Fig. 2*e*). Like all lncRNAs, their sequences are read in one direction, have a poly(A) tail, and are spliced [60]. They should be distinguished from enhancer RNAs (eRNAs), which are products of bidirectional transcription, are not spliced, and do not have a poly(A) tail. Enhancer lncRNAs are currently poorly studied. However, according to the hypothesis [61], the synthesis of enhancer lncRNAs may serve as a way to maintain an open chromatin state in enhancer regions to facilitate their interaction with promoters through loop formation. Some studies show the involvement of enhancer lncRNAs in the process of activating target gene expression [62, 63]. Several studies also show decreased expression of target genes upon knockdown of the nearest enhancer lncRNAs [63, 64].

Mechanisms of interaction between lncRNAs and microRNAs

Special attention should be paid to the interaction between lncRNAs and microRNAs. The work [65]

indicates the participation of these interactions in regulating resistance to diseases, cold, heat, biotic, and abiotic stress in plants.

MicroRNAs are small non-coding RNAs 18–25 nt in length, transcribed from introns, non-coding regions of genes, or intergenic spacers (DNA regions between genes) [66]. Currently, 4 types of interactions between lncRNAs and microRNAs are identified [67] (see Fig. 3). Below we will discuss these interaction mechanisms in more detail.

Degradation of lncRNA during microRNA incorporation

MicroRNAs function in complex with proteins from the argonaute (AGO) family. The microRNA-AGO complex is incorporated into the RNA-induced gene expression suppression complex either by blocking protein synthesis from mRNA or leading to mRNA degradation [68]. A similar mechanism of functioning is observed for lncRNA. MicroRNA, through RNA-binding proteins, is incorporated into the complementary region of lncRNA, causing degradation of these molecules [69, 70]. In this case, lncRNA acts as a competitor to mRNA and thereby adsorbs some of the microRNA molecules, reducing their impact on potential target molecules (mRNA). Importantly, this also results in the degradation of the lncRNA itself (see Fig. 3*a*).

Long ncRNA targets for microRNA

For the first time, lncRNA targets with a mimicry effect or competitive endogenous lncRNAs (competitive endogenous RNAs, ceRNAs) were discovered in plant genomes [71], (see Fig. 3*b*). The mechanism of mimicry ("sponge effect") consists in blocking the interaction of microRNA with the target mRNA by binding this microRNA to the lncRNA-decoy through partially complementary sequences [72].

To date, the influence of endogenous lncRNAs on the regulation of phosphate synthesis in *Arabidopsis thaliana* [72], *Zea mays* [73], and *Medicago truncatula* [74] through binding with microRNAs has been experimentally confirmed.

Competition between lncRNA and microRNA

The mechanisms of interaction between microRNAs and endogenous and competitive lncRNAs are similar. Long ncRNA acts as a target and reduces the concentration of microRNA (see Fig. 3*c*). However, lncRNA can also compete with microRNA. This mechanism of lncRNA functioning arises when they bind to mRNA through complementary bases, preventing microRNA from interacting with these regions of mRNA. Such a mechanism allows preventing the degradation of mRNA with which lncRNAs bind [75–77].

lncRNA precursor of microRNA

Some lncRNAs as primary transcripts can serve as precursors of specific microRNAs (see Fig. 3*d*). Small RNAs can be produced through the splicing mechanism and lead to enhanced post-transcriptional regulation of corresponding mRNA targets [35]. Such a mechanism of lncRNA functioning has been experimentally confirmed in humans [78]. It has been shown that they are responsible for regulating tumor development and progression by controlling the transcription of microRNA suppressors [79, 80]. For plants, lncRNA precursors have been discovered using bioinformatic methods. In works [81, 82] most of the identified microRNAs align with the terminal regions of the lncRNA sequence, indicating that these lncRNAs are their precursors.

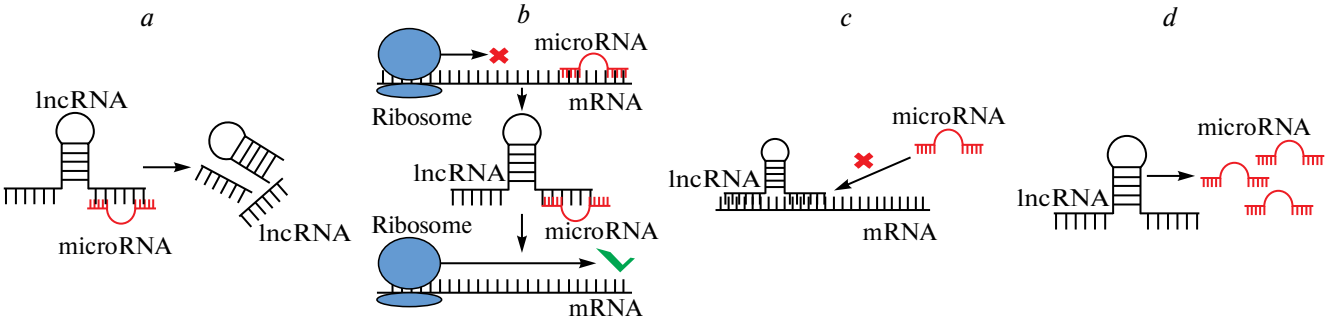

Fig. 3. Mechanisms of lncRNA-microRNA interactions: *a* – degradation of lncRNA when microRNA is incorporated; *b* – lncRNA acts as a target (sponge) for microRNA; *c* – lncRNA competes with microRNA to regulate mRNA expression levels; *d* – lncRNA acts as a precursor to microRNA (spliced into microRNA)

LONG ncRNA EXPRESSION

To date, an increasing number of works on lncRNA analysis in various organisms are emerging, in which expression research methods are applied for these molecules. It is generally accepted that lncRNAs demonstrate lower expression levels compared to mRNAs, while being expressed more specifically for certain tissues [34]. However, recently, studies have begun to emerge demonstrating lncRNAs with expression levels equal to those of protein-coding genes [16, 83]. Such cases are isolated, and their results require experimental confirmation.

Three directions of lncRNA expression research are distinguished: analysis of lncRNA expression in tissues and in response to a specific stimulus, analysis of differential expression and co-expression with protein-coding genes, and analysis of tissue specificity of lncRNAs. These approaches are applied both for large-scale transcriptome analysis and for analysis of small sequence samples.

Analysis of lncRNA expression in tissues is the most common research direction. Most works begin their research with this direction, gradually delving into differential expression or co-expression.

Analysis of tissue specificity of lncRNA is the least studied, as there is no precise understanding of which method for determining the specificity of gene expression in tissues is the most accurate [84]. In many works, a gene is considered specific if it is expressed only in one tissue [16, 85, 86]. One of the first approaches is counting the number of tissues in which a given gene is expressed [87]. However, this method allows only the study of genes with high expression, which makes it difficult to choose a threshold for filtering sequences. There is also the problem of closely related tissues, such as parts of the stem or human brain.

With the development of understanding about the structure and functions of lncRNA, more approaches to solving this problem began to emerge. Existing methods are divided into two groups. In the first group, the specificity indicator is a number that determines how specifically the studied gene is expressed: 1 — if the gene has specific expression, 0 — if it is expressed in several tissues — Tau [88], Gini [89], tissue specificity index (TSI) [90]. In the second group, the specificity of gene expression is calculated separately for each tissue — z-score, SPM [91], expression enrichment (EE) [92], with the level of specificity varying between different metrics.

Basically, all metrics for calculating the specificity of gene expression (Tau, EE score, TSI, Gini coefficient) are similar: the calculation uses the level of gene expression in a particular tissue and the number of experiments to which this tissue belongs. For example, the formula for calculating Tau:

$$\tau = \frac{\sum_{i=1}^{n}\left(1-\widehat{x_i}\right)}{n-1}; \quad (1)$$

$$\widehat{x_i} = \frac{x_i}{\max\limits_{1\le i\le n}\left(x_i\right)}, \quad (2)$$

where, x_i is the gene expression in the tissue, n is the number of experiments with the tissue.

In addition to the metrics listed above, there are works using metrics based on Shannon entropy, Jensen-Shannon (JS) [93]:

$$Hg = \sum_{1\le t\le N} -p_{t|g}\log_2\left(p_{t|g}\right); \quad (3)$$

$$p_{t|g} = \frac{\omega_{g,t}}{\sum_{1\le t\le N}\omega_{g,t}}, \quad (4)$$

where $p_{t|g}$ is the distribution of gene expression in the tissue and $\omega_{g,t}$ is the gene expression. Thus $p_{t|g}$ shows the ratio of gene expression in a particular tissue to the sum of expressions of this gene in all tissues.

$$Q_{g|t} = Hg - \log_2\left(p_{t|g}\right), \quad (5)$$

where Hg is the entropy that shows how tissue-specific a gene is. If the entropy value is close to or equal to 0, then the gene is specific to a certain tissue. $Q_{g|t}$ shows the expression entropy of the studied gene in one tissue. The higher the entropy value, the higher the specificity of gene expression in a particular tissue.

BIOLOGICAL ROLE OF lncRNA IN PLANTS

Previous sections have shown the diversity and importance of molecular functions of lncRNAs in plants. Long ncRNAs can act as genetic and epigenetic regulators of gene expression. Depending on their geometric isomerism, they can function as cis-regulatory elements, interacting with closely located genes and directly affecting genes on the same strand, or trans-regulatory elements, acting far from the site of synthesis. Long ncRNAs can prevent transcription factors from binding to the promoter region, acting as transcription repressors. As a result of these molecular processes, plant functions are controlled at a higher level of organization: tissues, organs, and the entire organism. Examples of processes affected by lncRNA regulation in plants are shown in Fig. 4. These include the following: plant resistance to various types of stress, development of plant organs, vernalization, and fruit development.

Long ncRNA in the vernalization process

Flowering is one of the most important adaptive traits that ensures the transition to reproductive growth and development, occurring under favorable conditions during the plant's life cycle. Vernalization is an important mechanism controlling flowering in some plant species that grow in a vegetative state during cold winter seasons and begin to flower in spring [94]. Vernalization is also the most studied regulatory process in plants involving lncRNAs. Long ncRNAs participate in the regulation of the FLOWERING LOCUS C (FLC) gene, which is responsible for *Arabidopsis thaliana* resistance to cold stress [3]. To date, two classes of lncRNAs have been discovered in the *FLC* locus: COOLAIR — lncRNA that is transcribed in the antisense direction relative to the *FLC* gene promoter; COLDAIR — lncRNA that is transcribed from the first intron of the FLC gene in the sense direction. COOLAIR lncRNAs bind to the FLC locus in the cis position and form an R-loop between the promoterand the first intron of the *FLC* gene, which leads to suppression of the expression of this target gene. This process occurs in both autonomous and vernalized pathways [94].

Long ncRNA and influence on plant yield

The development of normal pollen grain is crucial for fertilization and seed development. The anther acts as the parent tissue of the pollen grain and regulates its development. Abnormal pollen grain can lead to the development of plants with male sterility [95]. Experimental studies [96, 97] have shown the involvement of lncRNA in the process of yield control. For example, lncRNA (LDMAR) of rice (*Oryza sativa japonica*) causes programmed cell death of anthers under long day conditions (LD, long day), through splicing into microRNA osa-smR5864. A single nucleotide mutation of microRNA osa-smR5864 leads to male sterility of the plant [96]. Decreased expression of lncRNA (BcMF11) in *Brassica campestris* prevents tapetum degradation and leads to interruption of pollen grain formation [97]. Also, using bioinformatics methods in the work [33] for five stages of anther development in *Brassica rapa*, 14 lncRNAs were identified that co-express with 10 known protein-coding genes, which play a crucial role in pollen development. According to the authors, the identified lncRNAs are of interest for further experimental studies of lncRNA involvement in plant reproductive processes.

Long ncRNA in the process of photomorphogenesis

The molecular signaling mechanism of photomorphogenesis has been widely studied in various plant species. During this phenomenon, a number of proteins are associated as primary and secondary signaling molecules. Among them, a family of transcription factors with the structure of a DNA-binding domain helix–loop–helix (bHLH) is distinguished. Among the family members, phytochrome-interacting factors (PIF) are identified that function to suppress photomorphogenesis of seedlings during the dark time of day [98]. In the work

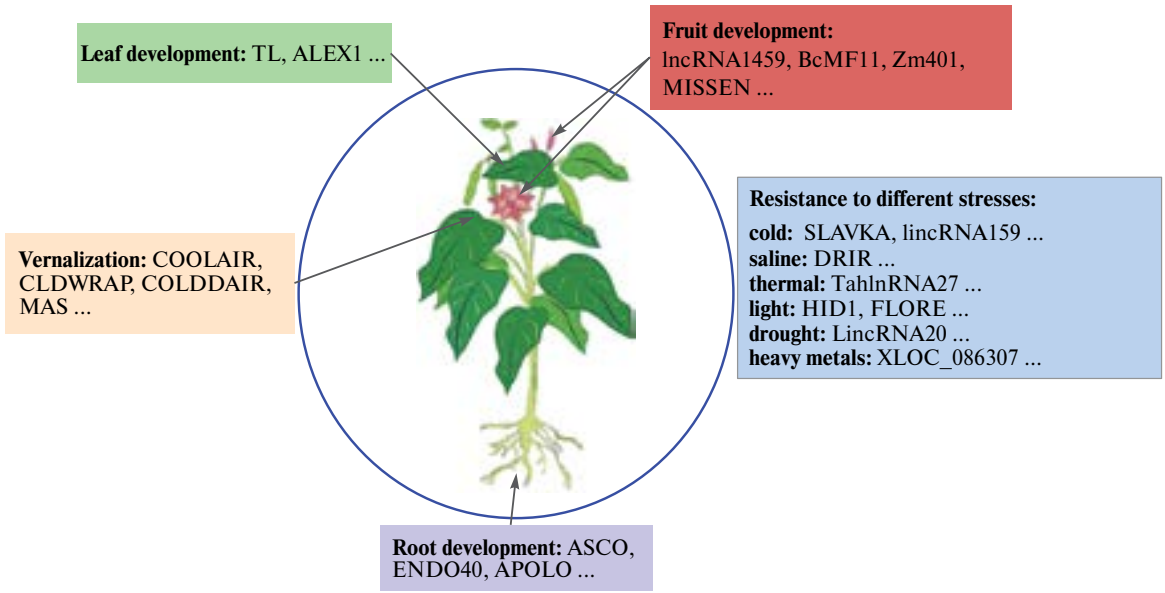

Fig. 4. Biological role of lncRNAs in plants. Long ncRNAs participate in regulating genes responsible for plant resistance to various types of stress (blue rectangle), also affect plant fertility (red rectangle), participate in leaf and root development (green, purple rectangle) and in vernalization processes (orange rectangle)

[77] lncRNA (HID1) was discovered, which functions as a trans-regulatory element, interacting with the promoter of the *PIF3* gene (a key signaling molecule of photomorphogenesis that negatively regulates responses to red light) and inhibits its transcription. As a result, elongation of the embryonic stem (hypocotyl) is observed in seedlings of *Arabidopsis thaliana*.

Long ncRNA in resistance processes to various types of stress

In nature, plants are subjected to various stress factors, such as salt, drought, cold, heat, as well as various viral infections. This leads to problems in plant growth and fertility. To adapt and survive in such adverse conditions, plants use multiple gene regulation mechanisms to restore cellular homeostasis. Recent studies demonstrate the significant role of lncRNA in regulating gene expression in response to stress conditions [99].

The involvement of lncRNA in resistance processes to abiotic stress types is demonstrated in *A. thaliana* [81], *B. rapa* [100], grape (Cabernet Sauvignon) [101], *O. sativa* [102] and many other plants. For example, in lncRNA of *A. thaliana*, the following mechanisms of cold stress resistance regulation are identified. Chromatin modulation and chromatin reconstruction (such as the lncRNA COOLAIR) are described above. Alternative splicing of lncRNA TAS1 leads to the formation of a transcript variant with an unspliced intron, which is more resistant to cold stress than the spliced variant of this lncRNA [103]. Transcriptional regulation of genes, for example, lncRNA SVALKA suppresses the transcription of the CBF1 gene (responsible for triggering a cascade of reactions aimed at increasing plant resistance to low temperatures) through transcriptional collision [104].

The lncRNA973 in *A. thaliana* regulates plant resistance to salt stress by modulating the expression of genes responsible for reactive oxygen species (SOD, CAT, and POD), transcription factors (MYB5, WRKY46, NAC29, and ERF62), and ionic balance (NHX7) [105].

Overexpression of lncRNA At5NC056820 in *A. thaliana* increases plant drought resistance [106]. Long ncRNA TCONS_00021861 in *O. sativa*, through a sponge effect, interacts with miR528–3p and prevents its interaction with the target gene YUCCA-7 (which is involved in tryptophan-dependent auxin biosynthesis and increases drought resistance), which activates the synthesis of indole-3-acetic acid (IAA), enhancing drought resistance [107].

A similar mechanism is observed for lncRNA TCONS_00048391, which acts as a "sponge" for microRNA bra-miR164a, suppressing the expression of target gene NAC1 (responsible for heat stress resistance in *B. rapa*) [100].

Thus, even within a single organism, lncRNAs have various mechanisms for regulating resistance to abiotic stress types.

EVOLUTION OF LNCRNA

Evolutionary processes lead to variability in the structure and functions of lncRNAs. There are several directions of active research in the field of lncRNA evolution. First of all, these are issues related to the emergence of new lncRNAs in the genomes of organisms. One such mechanism is gene duplication. The appearance of an additional copy of a gene in the genome opens up freedom for the evolution of one of the gene copies. Mutations occurring in one of the two copies and weakening the original function of the gene will not be eliminated by selection because the second copy can retain its former functionality. Thus, "pseudogenes" (inactive regions), isoforms, and lncRNA transcripts arise through duplication [28].

There are several hypotheses for the emergence of lncRNAs through duplication. The first is based on the loss of protein-coding ability by one of the isoforms of a protein-coding gene (during duplication). This is explained by the presence of regulatory elements, splicing sites, exon sequences, and sequences with poly(A) tails in lncRNAs. The most well-known example of an lncRNA that arose through duplication of a protein-coding gene is the human lncRNA Xist [29]. The second hypothesis is based on the duplication of non-coding RNAs. However, such cases are not as common as duplication of protein-coding genes, due to the high evolutionary lability of lncRNAs [108].

Also, one of the hypotheses for the origin of lncRNA is associated with the integration of mobile elements (MEs) into the transcript sequence. Mobile elements can facilitate gene expression regulation during development and adaptation, and can also serve as major sources of genetic variations in genome evolution. MEs are present in exons, transcription start sites, polyadenylation sites (poly(A) tails), or in some combinations of these parts. The study conducted on 8 plants (including *A. thaliana,* maize, tomato, poplar) [109] showed that 59.7% of lncRNAs found in all 8 organisms have at least one exon with a partially embedded ME. The study [43] demonstrates the involvement of MEs not only in the emergence of lncRNAs but also in the regulation and variability of lncRNAs.

To assess the conservation of lncRNAs between genomes of different organisms, their sequences are compared [110]. For example, when aligning 67 thousand

rice lncRNAs to 39 thousand maize lncRNAs, the proportion of matching positions of lncRNA homologs is 20 % [111]. When comparing lncRNAs for evolutionarily more distant species, for example, *A. thaliana* and *O. sativa,* the level of similarity drops significantly to the point of absence of homology [109]. Only a small number of lncRNA sequences demonstrate a level of conservation comparable to that of protein-coding genes [112]. One such lncRNA is IPS, involved in phosphorus (P_i) homeostasis in *A. thaliana*, *O. sativa, S. lycopersicum,* and *S. moellendorffii*. IPS contains a conserved motif of 24 nucleotides in the listed species, which serves as a target for miR399. IPS initially appeared in *S. moellendorffii*, a representative of ancient spore-bearing higher plants. During evolutionary development, the function of IPS has not changed, which may indicate the role of this lncRNA in adaptation to land during the migration from aquatic environment.

However, most lncRNAs show high variability in their primary structure, which is confirmed in the study [113], dedicated to the analysis of lncRNA conservation among 25 flowering plants. The analysis showed that only 6.79 % of lncRNAs are conserved in all studied plants. In the work [114], the analysis of rice lncRNA (intergenic) conservation relative to 7 plant species was conducted. The results demonstrated low conservation of lncRNAs compared to mRNAs, but higher conservation compared to mRNAs containing MEs. The analysis of the conservation of structural elements of lncRNAs showed high conservation of lncRNA exons compared to introns. These studies show the rapid evolution of lncRNAs, which leads to low sequence conservation. It is also noted that the percentage of conserved lncRNAs is higher among related species (within the same genus), with conservation rapidly decreasing as evolutionary distance increases [113].

Despite existing theories about the origin and evolution of lncRNAs, they should be treated with appropriate skepticism due to the lack of experimental confirmation.

LONG ncRNAs IN PAN-GENOMES AND PAN-TRANSCRIPTOMES

In nature, there is high genomic and morphological diversity among species caused by structural variations. These include insertion, duplication, inversion, translocation, as well as variations in gene copy number and their presence or absence. These variations can affect DNA segments several megabases in length and, in some cases, change the number of chromosomes. As a result, a single reference genome cannot cover all the gene content of one species. For example, reference plant genomes lack a number of genes important for agricultural research. To address this issue, a pan-genome is used.

The concept of a pan-genome implies coverage of sequences subject to structural variation and possibly absent in the reference sequence of each representative of the species. Today, there are several formulations of the term pan-genome, which are presented in works [36–40].

In this review, the pan-genome is defined as a complete non-redundant set of genes represented in the genome of at least one individual belonging to a given taxonomic unit [115]. Note that this concept can be extended to taxonomic units of any order [116]. For example, there was an attempt to create a common bacterial pan-genome based on 573 previously sequenced bacterial genomes [117]. However, currently, the concept of a pan-genome is most often used in relation to species.

Genes in the pan-genome can be divided into two groups:

1. Conserved genes — present in the genomes of all representatives of a taxonomic unit.

2. Variable genes — not present in the genomes of all species representatives. These are further divided into two categories: genes represented in two or more genomes, and unique genes represented in only one individual genome.

As the number of variable genes included in a specific pan-genome increases, the total number of genes in that pan-genome will increase, while the number of conserved genes will gradually decrease. Moreover, each of these samples will eventually reach a plateau, after which it will not significantly change with the addition of new genomes. It is noted that variable genes are on average evolutionarily younger than conserved ones [36]. This result seems logical in light of the fact that most conserved genes are housekeeping genes. The functions of conserved and variable genes differ. The functions of conserved genes are related to essential cellular processes, while variable genes are associated with the environment and protective reactions, receptor and antioxidant activity, gene regulation, and signal transmission.

Using conserved and variable genes, pan-genomes can be subdivided into two categories — open and closed. Closed pan-genomes are those that quickly reach a plateau in terms of gene count, meaning that sequencing genomes of new representatives of the studied species does not introduce new genes. Open pan-genomes, conversely, increase in size each time new genomes are added to their composition.

Another method for analyzing the gene composition in several representatives of a taxon is the analysis of its transcriptomes. The nucleotide sequences of transcripts

allow assessing the presence of genes in the genome only if they are expressed in some tissue or organ of the plant. Thus, the definition of a pan-transcriptome can be formulated as follows: a pan-transcriptome reflects the totality of all RNA molecules present in a given species or individual and consists of conserved and variable transcripts (unique to one species or individual).

In addition to single nucleotide polymorphisms (SNPs) and multiple structural variations, lncRNAs play an important role. For example, complete transcriptome sequencing has shown that an extremely large portion of genes in higher eukaryotes is transcribed. However, more than 90 % of all transcripts are not translated into proteins [4, 44, 78, 118] and are non-coding sequences. As described above, lncRNAs perform important functions in many molecular and biological processes in plants. However, despite the growing number of studies dedicated to the structural and functional analysis of lncRNAs, this type of molecule still remains poorly understood. This is due to many factors that need to be considered when identifying lncRNAs. The size of lncRNAs is similar to the size of some protein-coding genes, which leads to prediction errors. Studies also show that lncRNA sequences undergo rapid evolution, the patterns of which are still unclear. Typically, during evolution, these sequences rapidly accumulate substitutions so that they have weak homology even with sequences from closely related species [113]. Thus, the application of pan-genomes and pan-transcriptomes will increase the efficiency of research and the total number of predicted lncRNAs compared to using the genome of a single species representative. The use of a pan-genome will also increase the accuracy and completeness of the studied dataset, as the pan-genome is the totality of all coding and non-coding sequences of all species representatives. Moreover, lncRNAs predicted for different representatives of the same species can be aligned to one pan-genome of that species, which will allow for further investigation of variable genes, help determine the exact number, type, and functions of lncRNAs contributed by each representative of the studied species. This approach will also eliminate the need to align, compare, and combine results for genomes of different representatives of the same species, which will significantly reduce the time, labor, and computational resources required for this research. For example, a study of the pan-transcriptome of alfalfa (*Medicago sativa L.*) [119] in the tissues of leaves, stem, and roots revealed 11,677 new lncRNAs. This study identified 2,105 highly expressed and 1,894 low expressed lncRNAs, and the gene ontology analysis showed the involvement of these lncRNAs in the regulation processes of alfalfa resistance to drought and salt stress. In the work [120], analysis of the maize (*Zea mays*) pan-transcriptome revealed 644 new lncRNAs. However, to date, works devoted to the study of pan-genomes and pan-transcriptomes are mainly aimed at identifying and studying new protein-coding genes. While there are not so many works devoted to the study of lncRNAs on the scale of pan-genomes and pan-transcriptomes, especially for plants.

EXPERIMENTAL METHODS FOR lncRNA IDENTIFICATION

Until recently, the identification of lncRNAs was considered a by-product of research aimed at identifying protein-coding genes. However, over the past decade, it has become evident how important the role of lncRNAs is in molecular and biological processes in various organisms. This has led to the development of experimental technologies specifically aimed at identifying lncRNAs. Today, the following experimental methods are widely used to identify and annotate lncRNAs [121]:

1) RNA sequencing analysis — identification of new lncRNA transcripts.

2) Microarrays designed for lncRNAs.

3) SAGE, CAGE — identification of new lncRNAs.

4) PARS, SHAPE — prediction of lncRNA structure.

However, before starting the experiment, it is necessary to prepare lncRNA. Various purification methods are used for this. The application of standard RNA purification methods, such as denaturation by heating, can negatively affect the structure of lncRNA [122]. Thus, new methods have been developed to avoid RNA denaturation. Most of these approaches use affinity tags that participate in the immobilization of target RNAs and ribozymes, providing higher specificity during elution [123]. The method proposed by [124] allows the detection of lncRNA using ultrafiltration and gel filtration. This purification protocol allows the analysis of lncRNA with preserved co-transcriptional folding patterns, and also maintains potentially functional structural elements [124].

RNA sequencing (RNA-seq) analysis is the most common method for identifying and annotating new lncRNAs [121]. RNA sequencing (1) does not require prior knowledge of the sequences being studied, which means it can be used to study the transcriptomes of non-model organisms, (2) provides more accurate estimates of expression levels, and (3) requires less RNA to conduct the experiment [125]. Modifications of RNA sequencing analysis, such as strand-specific RNA sequencing, allow for the separation of sense and antisense classes of lncRNA [95].

Microchips were developed for transcriptome analysis, as well as quantitative and comparative analysis of mRNA expression between one or two samples of different origins. Standard microarray platforms do not contain lncRNA annotations and are not suitable for their detection and analysis. However, thanks to new advances in microarray technology, it is now possible to create custom oligonucleotide microarrays [126]. Microarrays designed for lncRNAs are used to study the expression profile of lncRNAs in cells and tissues [121]. For example, [127] used a microarray containing 3,478 intergenic and intronic sequences in human, mouse, and rat genomes to search for lncRNAs. The researchers discovered 55 new highly expressed lncRNAs, of which the expression of eight was confirmed in mouse tissues using Northern blot analysis.

Both molecular genetic methods — Serial Analysis of Gene Expression (SAGE) and Cap Analysis of Gene Expression (CAGE) — are based on obtaining and reading short (usually 27 nucleotides long) RNA sequence segments. This allows simultaneous quantitative and qualitative characterization of the expression of thousands of different genes by analyzing their transcripts [128]. Unlike the similar method of Serial Analysis of Gene Expression (SAGE, superSAGE), in which fragments come from other parts of transcripts, CAGE is mainly used to determine the exact location of transcription start sites in the genome. This knowledge, in turn, allows the researcher to determine the promoter structure necessary for gene expression [128]. For example, the SAGE method was applied in the work [96] dedicated to studying the molecular pathogenesis and biomarkers of precancerous lesions in the oral cavity. The study identified 10 differentially expressed lncRNAs, among which lncRNA NEAT1, which is involved in cancer metastasis, had the highest expression.

SHAPE and PARS methods [129, 130] allow obtaining data on the mode of action and interactions of RNA with other regulatory molecules [121]. The SHAPE (Selective 2'-hydroxyl acylation analyzed by primer extension) method is based on the use of N-methylisatoic anhydride (NMIA) and its derivatives to study flexible regions in the secondary structure of RNA. The SHAPE method has been applied to study the secondary structure of lncRNAs such as COOLAIR [131] and Braveheart [132]. The PARS method uses two nucleases simultaneously: S1 and V1. The first nuclease (S1) cleaves single-stranded DNA or RNA, while the second one (V1) cleaves double-stranded molecules. Thus, by conducting two experiments in parallel, information about both unpaired and paired nucleotides can be obtained [130].

Despite the described variety of experimental methods for identifying lncRNAs, this type of molecules remains poorly studied. This is primarily due to the shared characteristics of lncRNAs with protein-coding genes, as described in the "Molecular Functions of lncRNAs" section. Therefore, using only experimental methods for identifying lncRNAs is an inefficient approach for their investigation. For their mass study on a genome-wide scale, it is better to apply bioinformatic methods in combination with experimental ones.

BIOINFORMATIC METHODS FOR IDENTIFICATION OF lncRNA

Bioinformatic approaches are based on solving the following tasks: data preprocessing, lncRNA prediction, and structural-functional annotation [16, 24, 136]. However, today, due to the application of next-generation sequencing technologies, the number of total RNA sequencing (RNA-seq) experiments is rapidly increasing. In order to improve the efficiency of transcriptome analysis for large amounts of processed data, computational pipelines are being actively developed, which allow performing standardized data processing steps in the mode that does not require user control until the final result is obtained. The main advantages of existing pipelines include the ability to assemble the transcriptome for the studied organism and analyze differential expression of lncRNAs. A more detailed comparison of the functionality of bioinformatic pipelines is presented in Table S1.

Bioinformatic pipelines allow automating and accelerating the stages of lncRNA analysis listed above. Thus, data preprocessing includes quality control of reads using standard programs and alignment to the genome.

Prediction of lncRNAs is carried out by programs such as CPC2 [137], CPAT [138], PLEK [136]. These programs are based on the use of machine learning methods (support vector machines) to determine the coding potential of the studied transcript. The main criteria by which the coding potential is evaluated are features of open reading frames (ORFs). There are cases where ORF features of lncRNAs resemble those of protein-coding genes, however, the coverage or integrity of ORFs in lncRNAs is usually worse.

Annotation of lncRNA includes analysis of structural features (exon-intron structure, chromosomal position), expression, co-expression, homologous sequences, and conservation. Analysis of lncRNA structural features mainly focuses on transcript length, number of exons/introns, and chromosomal location. LncRNA conservation analysis in research can be divided into 2 groups: conservation within a single organism and conservation between different organisms. Long ncRNAs with a homology level above 50 % with lncRNAs of other organisms are considered conservative, while non-conservative ones have no homology with other organisms [137].

However, despite the advantages of pipelines, they have their disadvantages. Most pipelines are developed based on transcriptomic data from model organisms (human, mouse, fly). As a result, to study another organism, one either has to use existing models, which can introduce errors in lncRNA prediction, or additionally train a prediction model for the given organism and then incorporate this data into the pipeline code, which may lead to program malfunction.

It should be noted that despite common features, there are differences between the structure and functions of plant and animal lncRNAs [31]. For example, in the work [139] on the analysis of lncRNA prediction programs, it was shown that programs developed for humans poorly predict plant lncRNAs and vice versa, indicating differences in structural features of lncRNAs between humans and plants. However, there are currently no experimental studies aimed at comparing the structure of plant and animal lncRNAs.

Therefore, for more accurate identification of lncRNAs, additional sequence filters are necessary, and not all pipelines provide them. Today, modern bioinformatic pipelines address the disadvantages mentioned above. For example, ICAnnoLncRNA [140] and lncRNA-screen [141] allow training prediction models for any studied species. Additionally, ICAnnoLncRNA provides several sequence filters (alignment to reference genome, removal of mobile elements, filtering of short isoforms, and classification), which allows obtaining the most accurate number of identified lncRNAs.

In the bioinformatic analysis of lncRNA, newly sequenced transcripts are compared with already known lncRNA sequences. Databases serve as a source of information about these sequences. Various sources are used to replenish them: results of bioinformatic processing of transcriptomic experiments and lncRNA predictions, experimental results obtained from literature analysis, and combined use of both bioinformatic methods and literature sources. Databases that describe the structure and functions of plant lncRNAs are listed in Table S2.

CONCLUSION

To date, various studies demonstrate the importance and indispensability of lncRNAs in many cellular processes, such as activation and suppression of gene expression, imprinting and DNA demethylation, RNA interference, chromatin remodeling, etc. [142]. In plants, lncRNAs participate in the regulation of resistance to various types of stress [1–3], influence resistance to hypoxia [32], and participate in the development of fruits, roots, and leaves. Thus, there is an increasing interest in identifying lncRNAs in various species. For this purpose, methods of large-scale recognition of sequences in genomes and transcriptomes are used. Of particular interest is the identification of lncRNAs in pan-transcriptomes and pan-genomes. Many works [119, 120, 143] devoted to the investigation of pan-genomes note more complete and accurate information about identified gene sequences. However, despite the growing number of studies focused on lncRNAs, their identification still remains a problematic task. Various experimental and bioinformatic approaches are being developed to solve these problems.

ACKNOWLEDGMENTS

This work was supported by the budget project FWNR-2022–0020.

DECLARATION OF INTEREST STATEMENT

The authors have no financial interest in the materials or methods presented.

STATEMENT OF COMPLIANCE WITH ETHICS REQUIREMENTS

This article does not contain any studies using humans and animals as objects.

CONFLICT OF INTERESTS

The authors declare that they have no conflict of interest.

REFERENCES

1. *Deng P., Liu S., Nie X., Wu L.* Conservation analysis of long non-coding RNAs in plants // Sci. China Life Sci. Springer. 2018. Vol. 61, pp. 190–198.
2. *Wu H.-J., Wang Z-M., Wang M., Wang X-J.* Widespread long noncoding RNAs as endogenous target mimics for microRNAs in plants // Plant Physiol. American Society of Plant Biologists. 2013. Vol. 161, No. 4, pp. 1875–1884.
3. *Zhu Q.-H., Wang M.-B.* Molecular functions of long non-coding RNAs in plants // Genes. MDPI. 2012. Vol. 3, No. 1, pp. 176–190.
4. *Nazipova N.N.* Diversity of non-coding RNAs in eukaryotic genomes // Mathematical Biology and Bioinformatics. Institute of Mathematical Problems of Biology, Russian Academy of Sciences. 2021. Vol. 16, No. 2, pp. 256–298.
5. *Joshi A., Romanowska J.* Recent advances in computational-based approaches in epigenetics studies // Epigenetics Methods. Elsevier. 2020. pp. 569–590.
6. *Kim E.-D., Sung S.* Long noncoding RNA: unveiling hidden layer of gene regulatory networks // Trends Plant Sci. Elsevier. 2012. Vol. 17, No. 1, pp. 16–21.
7. *Karlik E., Ari S., Gozukirmizi N.* lncRNAs: genetic and epigenetic effects in plants // Biotechnol.

Biotechnol. Equip. 2019. Vol. 33, No. 1, pp. 429–439. DOI: 10.1080/13102818.2019.1581085

8. *Tsai M.-C., Manor O., Wan Y. et al.* Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes // Science. 2010. Vol. 329, No. 5992, pp. 689–693. DOI: 10.1126/science.1192002
9. *Sousa C., Johansson C., Charonet 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 // Mol. Cell. Biol. 2001. Vol. 21, No. 1, pp. 354–366. DOI: 10.1128/MCB.21.1.354-366.2001
10. *Medvedeva Y.A., Lennartsson A., Ehsani R. et al.* EpiFactors: a comprehensive database of human epigenetic factors and complexes // Database. Oxford University Press. 2015. Vol. 2015. P. bav067.
11. *Frankish A., Diekhans M., Jungreis I. et al.* GENCODE 2021 // Nucleic Acids Res. Oxford University Press. 2021. Vol. 49, No. D1, pp. D916–D923.
12. *Gordiuk V.V.* Long non-coding RNAs – tuning forks in the regulation of cellular processes // Ukr. Biochem. J. National Academy of Sciences of Ukraine, Palladin Institute of Biochemistry. 2014. No. 86, No. 2, pp. 5–15.
13. *Zhao X., Li J., Lian B. et al.* of MAF4 by a natural antisense RNA // Nat. Commun. Nature Publishing Group UK London. 2018. Vol. 9, No. 1, p. 5056.
14. *Li X., Wu Z., Fu X. et al.* lncRNAs: insights into their function and mechanics in underlying disorders // Mutat. Res. Mutat. Res. Elsevier. 2014. Vol. 762, pp. 1–21.
15. *Ahmad P., Bensaoud C., Mekki I. et al.* Long Non-Coding RNAs and their potential roles in the vector–host–pathogen triad // Life. MDPI. 2021. Vol. 11, No. 1, p. 56.
16. *De Quattro C.*, *Pè M.E.*, *Bertolini E.* Long noncoding RNAs in the model species Brachypodium distachyon // Sci. Rep. Nature Publishing Group UK London. 2017. Vol. 7, No. 1, p. 11252.
17. *Ma L.*, *Bajic V.B.*, *Zhang Z.* On the classification of long non-coding RNAs // RNA Biol. 2013. Vol. 10, No. 6, pp. 924–933. DOI: 10.4161/rna.24604
18. *Chen L.*, *Zhu Q.-H.*, *Kaufmann K.* Long non-coding RNAs in plants: emerging modulators of gene activity in development and stress responses // Planta. 2020. Vol. 252, No. 5, p. 92. DOI: 10.1007/s00425-020-03480-5
19. *Griffiths-Jones S.* miRBase: The MicroRNA Sequence Database // MicroRNA Protocols. New Jersey: Humana Press. 2006. Vol. 342, pp. 129–138.
20. *Amaral P.P.*, *Mattick J.S.* Noncoding RNA in development // Mamm. Genome. Springer. 2008. Vol. 19, pp. 454–492.
21. *Beyerli O.A., Gareev I.F.* Long non-coding RNAs: what are the prospects? // Preventive Medicine. Limited Liability Company Media Sfera Publishing House. 2020. Vol. 23, No. 2, pp. 124–128.
22. *Blythe A.J.*, *Fox A.H.*, *Bond C.S.* The ins and outs of lncRNA structure: How, why and what comes next? // Biochim. Biophys. Acta BBA-Gene Regul. Mech. Elsevier. 2016. Vol. 1859, No. 1, pp. 46–58.
23. *Bryzghalov O.*, *Makałowska I.*, *Szcześniak M.W.* lncEvo: automated identification and conservation study of long noncoding RNAs // BMC Bioinformatics. 2021. Vol. 22, No. 1, p. 59. DOI: 10.1186/s12859-021-03991-2.
24. *Zhao Q., Sun Y., Wang D. et al.* LncPipe: A Nextflow-based pipeline for identification and analysis of long non-coding RNAs from RNA-Seq data // J. Genet. Genomics. 2018. Vol. 45, DOI: 10.1016/j.jgg.2018.06.005.
25. *Talyan S., Filipów S., Ignarski M. et al.* CALINCA—A Novel Pipeline for the Identification of lncRNAs in Podocyte Disease // Cells. MDPI. 2021. Vol. 10, No. 3, pp. 692.
26. *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 // Plant Cell. American Society of Plant Biologists. 2004. Vol. 16, No. 4, pp. 1047–1059.
27. *Unver T.*, *Tombuloglu H.* Barley long non-coding RNAs (lncRNA) responsive to excess boron // Genomics. Elsevier. 2020. Vol. 112, No. 2, pp. 1947–1955.
28. *Khorkova O.*, *Hsiao J.*, *Wahlestedt C.* Basic biology and therapeutic implications of lncRNA // Adv. Drug Deliv. Rev. Elsevier. 2015. Vol. 87, pp. 15–24.
29. *Duret L., Chureau C., Samain S. et al.* The *Xist* RNA Gene Evolved in Eutherians by Pseudogenization of a Protein-Coding Gene // Science. 2006. Vol. 312, No. 5780, pp. 1653–1655. DOI: 10.1126/science.1126316.
30. *Graf J.*, *Kretz M.* From structure to function: Route to understanding lncRNA mechanism // BioEssays. 2020. Vol. 42, No. 12, pp. 2000027. DOI: 10.1002/bies.202000027
31. *Golicz A.A.*, *Singh M.B.*, *Bhalla P.L.* The long intergenic noncoding RNA (LincRNA) landscape of the soybean genome // Plant Physiol. American Society of Plant Biologists. 2018. Vol. 176, No. 3, pp. 2133–2147.

32. *Cheng F., Wu J., Fang L., Wang X. et al.* Syntenic gene analysis between Brassica rapa and other Brassicaceae species // Front. Plant Sci. Frontiers. 2012. Vol. 3, p. 30895.
33. *Huang L., Dong H., Zhou D. et al.* Systematic identification of long non-coding RNA s during pollen development and fertilization in *Brassica rapa* // Plant J. 2018. Vol. 96, No. 1, p. 203–222. DOI: 10.1111/tpj.14016
34. *Derrien T., Johnson R., Bussotti G. et al.* The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression // Genome Res. Cold Spring Harbor Lab. 2012. Vol. 22. No. 9, pp. 1775–1789.
35. *Ponting C.P.*, *Oliver P.L.*, *Reik W.* Evolution and functions of long noncoding RNAs // Cell. Elsevier. 2009. Vol. 136, No. 4, pp. 629–641.
36. *Golicz A.A., Bayer P.E., Barker G.C. et al.* The pangenome of an agronomically important crop plant Brassica oleracea // Nat. Commun. Nature Publishing Group UK London. 2016. Vol. 7, No. 1, pp. 13390.
37. *Meile L., Croll D., Brunner P.C. et al.* A fungal avirulence factor encoded in a highly plastic genomic region triggers partial resistance to septoria tritici blotch // New Phytol. 2018. Vol. 219, No. 3, pp. 1048–1061. DOI: 10.1111/nph.15180
38. *Alcaraz L.D., Moreno-Hagelsieb G., Eguiarte L.E. et al.* Understanding the evolutionary relationships and major traits of Bacillus through comparative genomics // BMC Genomics. 2010. Vol. 11, No. 1, p. 332. DOI: 10.1186/1471-2164-11-332
39. *Rasko D.A., Rosovitz M.J., Myers G.S.A. et al.* The Pangenome Structure of *Escherichia coli* : Comparative Genomic Analysis of *E. coli* Commensal and Pathogenic Isolates // J. Bacteriol. 2008. Vol. 190. No. 20, pp. 6881–6893. DOI: 10.1128/JB.00619-08
40. *Merot-L'anthoene V., Tournebize R., Darracq O. et al.* Development and evaluation of a genome-wide Coffee 8.5K SNP array and its application for high-density genetic mapping and for investigating the origin of *Coffea arabica* L. // Plant Biotechnol. J. 2019. Vol. 17, No. 7, pp. 1418–1430. DOI: 10.1111/pbi.13066
41. *Budak H.*, *Kaya S.B.*, *Cagirici H.B.* Long non-coding RNA in plants in the era of reference sequences // Front. Plant Sci. Frontiers. 2020. Vol. 11, p. 441273.
42. *Britto-Kido S. de A., Neto* J.R.C.F.*, Pandolfi V. et al.* Natural antisense transcripts in plants: a review and identification in soybean infected with Phakopsora pachyrhizi SuperSAGE Library // Sci. World J. Hindawi. 2013. Vol. 2013.
43. *Kapusta A., Kronenberg Z., Lynch V.J. et al.* Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs // PLoS Genet. Public Library of Science San Francisco, USA. 2013. Vol. 9, No. 4, p. e1003470.
44. Lukina S.S., Burdenny A.M., Filippova E.A. et al. The role of long non-coding RNA and DNA methylation in the pathogenesis of ovarian cancer // Pathological Physiology and Experimental Therapy. 2022. Vol. 66, No. 4, pp. 143–156.
45. *Ulitsky I.*, *Bartel D.P.* lincRNAs: genomics, evolution, and mechanisms // Cell. Elsevier. 2013. Vol. 154, No. 1, pp. 26–46.
46. *Dinger M.E., Pang K.C., Mercer T.R. et al.* Differentiating protein-coding and noncoding RNA: challenges and ambiguities // PLoS Comput. Biol. Public Library of Science San Francisco, USA. 2008. Vol. 4, No. 11, p. e1000176.
47. *Xie C., Yuan J., Li H., et al.* NONCODEv4: exploring the world of long non-coding RNA genes // Nucleic Acids Res. Oxford University Press. 2014. Vol. 42, No. D1, pp. D98–D103.
48. *Pal D.*, *Rao M.R.S.* Long Noncoding RNAs in Pluripotency of Stem Cells and Cell Fate Specification // Long Non Coding RNA Biology / ed. Rao M.R.S. Singapore: Springer Singapore. 2017. Vol. 1008, pp. 223–252.
49. *Tatosyan K.A., Zinevich L.S., Demin D.E., Schwartz A.M.* Functional features of long non-coding RNAs containing sequences of mobile genetic elements // Molecular Biology. Federal State Budgetary Institution "Russian Academy of Sciences". 2020. Vol. 54, No. 5, pp. 718–724.
50. *Zhao Z., Zang S., Zou W. et al.* Long non-coding RNAs: new players in plants // Int. J. Mol. Sci. MDPI. 2022. Vol. 23, No. 16, p. 9301.
51. *Kopp F.*, *Mendell J.T.* Functional classification and experimental dissection of long noncoding RNAs // Cell. Elsevier. 2018. Vol. 172, No. 3, pp. 393–407.
52. *Huarte M.*, *Rinn J.L.* Large non-coding RNAs: missing links in cancer? // Hum. Mol. Genet. Oxford University Press. 2010. Vol. 19, No. R2, pp. R152–R161.
53. *Hung T., Wang Y., Lin M.F. et al.* Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters // Nat. Genet. Nature Publishing Group. 2011. Vol. 43, No. 7, pp. 621–629.
54. *Loewer S., Cabili M.N., Guttman M. et al.* Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem

cells // Nat. Genet. Nature Publishing Group US New York. 2010. Vol. 42, No. 12, pp. 1113–1117.
55. *Wang K.C.*, *Chang H.Y.* Molecular mechanisms of long noncoding RNAs // Mol. Cell. Elsevier. 2011. Vol. 43, No. 6, pp. 904–914.
56. *Chen J.*, *Wang H.*, *Yao Y.* Experimental study of nonlinear ultrasonic behavior of soil materials during the compaction // Ultrasonics. Elsevier. 2016. Vol. 69, pp. 19–24.
57. *Kotake Y., Nakagawa T., Kitagawa K. et al.* Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15INK4B tumor suppressor gene // Oncogene. Nature Publishing Group. 2011. Vol. 30, No. 16, pp. 1956–1962.
58. *Ye X., Wang S., Zhao X. et al.* Role of lncRNAs in *cis* - and *trans*- regulatory responses to salt in *Populus trichocarpa* // Plant J. 2022. Vol. 110, No. 4, pp. 978–993.
DOI: 10.1111/tpj.15714
59. *Yang L.*, *Froberg J.E.*, *Lee J.T.* Long noncoding RNAs: fresh perspectives into the RNA world // Trends Biochem. Sci. Elsevier. 2014. Vol. 39, No. 1, pp. 35–43.
60. *Lam M.T.Y. , Cho H., Lesch H.P. et al.* Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription // Nature. Nature Publishing Group UK London. 2013. Vol. 498, No. 7455, pp. 511–515.
61. *Kim T.-K.*, *Hemberg M.*, *Gray J.M.* Enhancer RNAs: a class of long noncoding RNAs synthesized at enhancers // Cold Spring Harb. Perspect. Biol. Cold Spring Harbor Lab. 2015. Vol. 7. No. 1, pp. a018622.
62. *Kim T.-K., Hemberg M., Gray J.M. et al.* Widespread transcription at neuronal activity-regulated enhancers // Nature. Nature Publishing Group UK London. 2010. Vol. 465, No. 7295, pp. 182–187.
63. *Lai F., Orom U.A., Cesaroni M. et al.* Activating RNAs associate with Mediator to enhance chromatin architecture and transcription // Nature. Nature Publishing Group UK London. 2013. Vol. 494, No. 7438, pp. 497–501.
64. *Melo C.A., Drost J., ijchers P.J. et al.* eRNAs are required for p53-dependent enhancer activity and gene transcription // Mol. Cell. Elsevier. 2013. Vol. 49, No. 3, pp. 524–535.
65. *Zhang P., Meng J., Luan Y. et al.* Plant miRNA–lncRNA interaction prediction with the ensemble of CNN and IndRNN // Interdiscip. Sci. Comput. Life Sci. Springer. 2020. Vol. 12, pp. 82–89.
66. *Chen K.*, *Rajewsky N.* The evolution of gene regulation by transcription factors and microRNAs // Nat. Rev. Genet. Nature Publishing Group UK London. 2007. Vol. 8, No. 2, pp. 93–103.
67. *Jin Q., Zhao Z., Zhao Q. et al.* Long noncoding RNAs: emerging roles in pulmonary hypertension // Heart Fail. Rev. Springer. 2020. Vol. 25, pp. 795–815.
68. *Huntzinger E.*, *Izaurralde E.* Gene silencing by microRNAs: contributions of translational repression and mRNA decay // Nat. Rev. Genet. Nature Publishing Group UK London. 2011. Vol. 12, No. 2, pp. 99–110.
69. *Yoon J.-H., Abdelmohsen K., Kim J. et al.* Scaffold function of long non-coding RNA HOTAIR in protein ubiquitination // Nat. Commun. Nature Publishing Group UK London. 2013. Vol. 4, No. 1, p. 2939.
70. *Mukherjee N., Corcoran D.L., Nusbaum J.D. et al.* Integrative regulatory mapping indicates that the RNA-binding protein HuR couples pre-mRNA processing and mRNA stability // Mol. Cell. Elsevier. 2011. Vol. 43, No. 3, pp. 327–339.
71. *Thomson D.W.*, *Dinger M.E.* Endogenous microRNA sponges: evidence and controversy // Nat. Rev. Genet. Nature Publishing Group. 2016. Vol. 17, No. 5, pp. 272–283.
72. *Franco-Zorrilla J.M., Valli A., Todesco M. et al.* Target mimicry provides a new mechanism for regulation of microRNA activity // Nat. Genet. Nature Publishing Group US New York. 2007. Vol. 39, No. 8, pp. 1033–1037.
73. *Du Q., Wang K., Zou C. et al.* The PILNCR1-miR399 regulatory module is important for low phosphate tolerance in maize // Plant Physiol. American Society of Plant Biologists. 2018. Vol. 177, No. 4, pp. 1743–1753.
74. *Wang T., Zhao M., Zhang X. et al.* Novel phosphate deficiency-responsive long non-coding RNAs in the legume model plant Medicago truncatula // J. Exp. Bot. Oxford University Press UK. 2017. Vol. 68, No. 21–22, pp. 5937–5948.
75. *Faghihi M.A., Zhang M., Huang J. et al.* Evidence for natural antisense transcript-mediated inhibition of microRNA function // Genome Biol. 2010. Vol. 11, No. 5, p. R56. DOI: 10.1186/gb-2010-11-5-r56
76. *Kimura T., Jiang S., Nishizawa M. et al.* Stabilization of human interferon-α1 mRNA by its antisense RNA // Cell. Mol. Life Sci. 2013. Vol. 70, No. 8, pp. 1451–1467. DOI: 10.1007/s00018-012-1216-x
77. *Wang Y., Pang W.J, Wei N. et al.* Identification, stability and expression of Sirt1 antisense long non-coding RNA // Gene. Elsevier. 2014. Vol. 539, No. 1, pp. 117–124.
78. *Cai X.*, *Cullen B.R.* The imprinted H19 noncoding RNA is a primary microRNA precursor // Rna.

Cold Spring Harbor Lab. 2007. Vol. 13, No. 3, pp. 313–316.
79. *Augoff K., McCue B., Plow E.F. et al.* miR-31 and its host gene lncRNA LOC554202 are regulated by promoter hypermethylation in triple-negative breast cancer // Mol. Cancer. 2012. Vol. 11. No. 1, p. 5. DOI: 10.1186/1476-4598-11-5.
80. *Kallen A.N., Zhou X.B., Xu J. et al.* The imprinted H19 lncRNA antagonizes let-7 microRNAs // Mol. Cell. Elsevier. 2013. Vol. 52, No. 1, pp. 101–112.
81. *Amor B.B., Wirth S., Merchan F. et al.* Novel long non-protein coding RNAs involved in Arabidopsis differentiation and stress responses // Genome Res. Cold Spring Harbor Lab. 2009. Vol. 19, No. 1, pp. 57–69.
82. *Hirsch J., Lefort V., Vankersschaver M. et al.* Characterization of 43 non-protein-coding mRNA genes in Arabidopsis, including the MIR162a-derived transcripts // Plant Physiol. American Society of Plant Biologists. 2006. Vol. 140, No. 4, pp. 1192–1204.
83. *Lamin-Samu A.T., Zhuo S., Ali M., Lu G.* Long non-coding RNA transcriptome landscape of anthers at different developmental stages in response to drought stress in tomato // Genomics. Elsevier. 2022. Vol. 114, No. 4, pp. 110383.
84. *Kryuchkova-Mostacci N.*, *Robinson-Rechavi M.* A benchmark of gene expression tissue-specificity metrics // Brief. Bioinform. Oxford University Press. 2017. Vol. 18, No. 2, pp. 205–214.
85. *Li L., Eichten S.R., Shimizu R. et al.* Genome-wide discovery and characterization of maize long non-coding RNAs // Genome Biol. 2014. Vol. 15, No. 2, p. R40.
DOI: 10.1186/gb-2014-15-2-r40.
86. *Han L., Mu Z., Luo Z. et al.* New lncRNA annotation reveals extensive functional divergence of the transcriptome in maize // J. Integr. Plant Biol. 2019. Vol. 61, No. 4, pp. 394–405. DOI: 10.1111/jipb.12708
87. *Subramanian S.*, *Kumar S.* Gene expression intensity shapes evolutionary rates of the proteins encoded by the vertebrate genome // Genetics. Oxford University Press. 2004. Vol. 168, No. 1, pp. 373–381.
88. *Yanai I., Benjamin H., Shmoish M. et al.* Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification // Bioinformatics. Oxford University Press. 2005. Vol. 21, No. 5, pp. 650–659.
89. *Ceriani L.*, *Verme P.* The origins of the Gini index: extracts from Variabilità e Mutabilità (1912) by Corrado Gini // J. Econ. Inequal. Springer. 2012. Vol. 10, pp. 421–443.
90. *Julien P., Brawand D., Soumillon M. et al.* Mechanisms and evolutionary patterns of mammalian and avian dosage compensation // PLoS Biol. Public Library of Science San Francisco, USA. 2012. Vol. 10, No. 5, p. e1001328.
91. *Xiao S.-J., Zhang C., Zou Q., Ji Z.L.* TiSGeD: a database for tissue-specific genes // Bioinformatics. Oxford University Press. 2010. Vol. 26, No. 9, pp. 1273–1275.
92. *Yu X., Lin J., Zack D.J., Qian J.* Computational analysis of tissue-specific combinatorial gene regulation: predicting interaction between transcription factors in human tissues // Nucleic Acids Res. Oxford University Press. 2006. Vol. 34, No. 17, pp. 4925–4936.
93. *Huang X.*, *Li S.Z.*, *Wang Y.* Jensen-shannon boosting learning for object recognition // 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). IEEE. 2005. Vol. 2, pp. 144–149.
94. *Marquardt S., Raitskin O., Wu Z. et al.* Functional consequences of splicing of the antisense transcript COOLAIR on FLC transcription // Mol. Cell. Elsevier. 2014. Vol. 54, No. 1, pp. 156–165.
95. *Liu X., Hao L., Li D. et al.* Long non-coding RNAs and their biological roles in plants // Genomics Proteomics Bioinformatics. Oxford University Press. 2015. Vol. 13, No. 3, pp. 137–147.
96. *Ding J., Lu Q., Ouyang Y. et al.* A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice // Proc. Natl. Acad. Sci. 2012. Vol. 109, No. 7, pp. 2654–2659. DOI: 10.1073/pnas.1121374109.
97. *Song J.-H.*, *Cao J.-S.*, *Wang C.-G.* BcMF11, a novel non-coding RNA gene from Brassica campestris, is required for pollen development and male fertility // Plant Cell Rep. Springer. 2013. Vol. 32, pp. 21–30.
98. *Kim J., Yi H., Choi G. et al.* Functional characterization of phytochrome interacting factor 3 in phytochrome-mediated light signal transduction // Plant Cell. American Society of Plant Biologists. 2003. Vol. 15, No. 10, pp. 2399–2407.
99. *Wang J., Meng X., Dobrovolskaya O.B. et al.* Non-coding RNAs and Their Roles in Stress Response in Plants: 5 // Genomics Proteomics Bioinformatics. 2017. Vol. 15, No. 5, pp. 301–312. DOI: 10.1016/j.gpb.2017.01.007.
100. *Wang A., Hu J., Gao C. et al.* Genome-wide analysis of long non-coding RNAs unveils the regulatory roles in the heat tolerance of Chinese cabbage (Brassica rapa ssp. chinensis) // Sci. Rep. Nature Publishing Group UK London. 2019. Vol. 9, No. 1, pp. 5002.

101. *Wang P., Dai L., Ai J. et al.* Identification and functional prediction of cold-related long non-coding RNA (lncRNA) in grapevine // Sci. Rep. Nature Publishing Group UK London. 2019. Vol. 9, No. 1, p. 6638.
102. *Chung P.J., Jung H., Jeong D.H. et al.* Transcriptome profiling of drought responsive noncoding RNAs and their target genes in rice // BMC Genomics. 2016. Vol. 17, No. 1, p. 563. DOI: 10.1186/s12864-016-2997-3.
103. *Calixto C.P., Tzioutziou N.A., James A.B. et al.* Cold-dependent expression and alternative splicing of Arabidopsis long non-coding RNAs // Front. Plant Sci. Frontiers Media SA. 2019. Vol. 10, p. 235.
104. *Jha U.C., Nayyar H., Jha R. et al.* Long non-coding RNAs: emerging players regulating plant abiotic stress response and adaptation // BMC Plant Biol. 2020. Vol. 20, No. 1, p. 466. DOI: 10.1186/s12870-020-02595-x
105. *Zhang X., Dong J., Deng F. et al.* The long non-coding RNA lncRNA973 is involved in cotton response to salt stress // BMC Plant Biol. 2019. Vol. 19, No. 1, p. 459. DOI: 10.1186/s12870-019-2088-0.
106. *Wang X., Fan H., Wang B., Yuan F.* Research progress on the roles of lncRNAs in plant development and stress responses // Front. Plant Sci. Frontiers Media SA. 2023. Vol. 14, p. 1138901.
107. *Chen J.*, *Zhong Y.*, *Qi X.* LncRNA TCONS_00021861 is functionally associated with drought tolerance in rice (Oryza sativa L.) via competing endogenous RNA regulation // BMC Plant Biol. 2021. Vol. 21, No. 1, p. 410. DOI: 10.1186/s12870-021-03195-z.
108. *Kazemzadeh M.*, *Safaralizadeh R.*, *Orang A.V.* LncRNAs: emerging players in gene regulation and disease pathogenesis // J. Genet. Springer. 2015. Vol. 94, pp. 771–784.
109. *Zhu Y., Chen L., Hong X. et al.* Revealing the novel complexity of plant long non-coding RNA by strand-specific and whole transcriptome sequencing for evolutionarily representative plant species // BMC Genomics. 2022. Vol. 23, No. S4, p. 381. DOI: 10.1186/s12864-022-08602-9.
110. *Ulitsky I.* Evolution to the rescue: using comparative genomics to understand long non-coding RNAs // Nat. Rev. Genet. Nature Publishing Group UK London. 2016. Vol. 17, No. 10, pp. 601–614.
111. *Wang H., Niu Q.W., Wu H.W. et al.* Analysis of non-coding transcriptome in rice and maize uncovers roles of conserved lnc RNA s associated with agriculture traits // Plant J. 2015. Vol. 84, No. 2, pp. 404–416. DOI: 10.1111/tpj.13018.
112. *Nitsche A.*, *Stadler P.F.* Evolutionary clues in lncRNAs // WIREs RNA. 2017. Vol. 8, No. 1, p. e1376. DOI: 10.1002/wrna.1376.
113. *Sang S., Chen W., Zhang D. et al.* Data integration and evolutionary analysis of long non-coding RNAs in 25 flowering plants: 3 // BMC Genomics. 2021. Vol. 22, No. 3, p. 739. DOI: 10.1186/s12864-021-08047-6.
114. *Zhang Y.-C., Liao J.Y., Li Z.Y. et al.* Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice // Genome Biol. 2014. Vol. 15, No. 12, p. 512. DOI: 10.1186/s13059-014-0512-1.
115. *Pronozin A.Y.*, *Bragina M.K.*, *Salina E.A.* Crop pangenomes // Vavilov J. Genet. Breed. Institute of Cytology and Genetics, Siberian Branch of the Russian Academy .... 2021. Vol. 25, No. 1, p. 57.
116. *Vernikos G.,Medini D., Riley D.R., Tettelin H. et al.* Ten years of pan-genome analyses // Curr. Opin. Microbiol. Elsevier. 2015. Vol. 23, pp. 148–154.
117. *Lapierre P.*, *Gogarten J.P.* Estimating the size of the bacterial pan-genome // Trends Genet. Elsevier. 2009. Vol. 25, No. 3, pp. 107–110.
118. *Chekanova J.A., Gregory B.D., Reverdatto S.V. et al.* Genome-wide high-resolution mapping of exosome substrates reveals hidden features in the Arabidopsis transcriptome // Cell. Elsevier. 2007. Vol. 131, No. 7, pp. 1340–1353.
119. *Medina C.A.*, *Samac D.A.*, *Yu L.-X.* Pan-transcriptome identifying master genes and regulation network in response to drought and salt stresses in Alfalfa (Medicago sativa L.) // Sci. Rep. Nature Publishing Group UK London. 2021. Vol. 11, No. 1, p. 17203.
120. *Jin M., Liu H., He C. et al.* Maize pan-transcriptome provides novel insights into genome complexity and quantitative trait variation // Sci. Rep. Nature Publishing Group UK London. 2016. Vol. 6, No. 1, p. 18936.
121. *Chowdhary A.*, *Satagopam V.*, *Schneider R.* Long Non-coding RNAs: Mechanisms, Experimental, and Computational Approaches in Identification, Characterization, and Their Biomarker Potential in Cancer // Front. Genet. 2021. Vol. 12, p. 649619. DOI: 10.3389/fgene.2021.649619
122. *Svergun D.I.*, *Koch M.H.* Small-angle scattering studies of biological macromolecules in solution // Rep. Prog. Phys. IOP Publishing. 2003. Vol. 66, No. 10, p. 1735.

123. *Schön P.* Atomic force microscopy of RNA: State of the art and recent advancements // Seminars in Cell & Developmental Biology. Elsevier. 2018. Vol. 73, pp. 209–219.
124. *Chillón I., Marcia M., Legiewicz M. et al.* Native purification and analysis of long RNAs // Methods in Enzymology. Elsevier. 2015. Vol. 558, pp. 3–37.
125. *Cheung F., Haas B.J., Goldberg S.M.D. et al.* Sequencing Medicago truncatula expressed sequenced tags using 454 Life Sciences technology // BMC Genomics. 2006. Vol. 7, No. 1. DOI: 10.1186/1471-2164-7-272
126. *Au P.C.K.*, *Zhu Q.-H.* Identification of lncRNAs Using Computational and Experimental Approaches // Regulatory RNAs / eds Mallick B., Ghosh Z. Berlin, Heidelberg: Springer Berlin Heidelberg. 2012. pp. 319–340.
127. *Babak T.*, *Blencowe B.J.*, *Hughes T.R.* A systematic search for new mammalian noncoding RNAs indicates little conserved intergenic transcription // BMC Genomics. 2005. Vol. 6, No. 1, p. 104. DOI: 10.1186/1471-2164-6-104.
128. *Shiraki T., Kondo S., Katayama S. et al.* Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage // Proc. Natl. Acad. Sci. 2003. Vol. 100, No. 26, pp. 15776–15781. DOI: 10.1073/pnas.2136655100.
129. *Merino E.J., Wilkinson K.A., Coughlan J.L., Weeks K.M.* RNA Structure Analysis at Single Nucleotide Resolution by Selective 2‘-Hydroxyl Acylation and Primer Extension (SHAPE) // J. Am. Chem. Soc. 2005. Vol. 127, No. 12, pp. 4223–4231. DOI: 10.1021/ja043822v.
130. *Kertesz M., Wan Y., Mazor E. et al.* Genome-wide measurement of RNA secondary structure in yeast // Nature. Nature Publishing Group UK London. 2010. Vol. 467, No. 7311, pp. 103–107.
131. *Hawkes E.J., Hennelly S.P., Novikova I.V. et al.* COOLAIR antisense RNAs form evolutionarily conserved elaborate secondary structures // Cell Rep. Elsevier. 2016. Vol. 16, No. 12, pp. 3087–3096.
132. *Kim D.N., Thiel B.C., Mrozowich T. et al.* Zinc-finger protein CNBP alters the 3-D structure of lncRNA Braveheart in solution // Nat. Commun. Nature Publishing Group UK London. 2020. Vol. 11, No. 1, p. 148.
133. *Li A.*, *Zhang J.*, *Zhou Z.* PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme // BMC Bioinformatics. 2014. Vol. 15, No 1, p. 311. DOI: 10.1186/1471-2105-15-311.
134. *Kang Y.-J., Yang D.C., Kong L. et al.* CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features // Nucleic Acids Res. Oxford University Press. 2017. Vol. 45, No. W1, pp. W12–W16.
135. *Wang L., Park H.J., Dasari S. et al.* CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model // Nucleic Acids Res. Oxford University Press. 2013. Vol. 41, No. 6, pp. e74–e74.
136. *Da Costa Negri T.*, *Paschoal A.R.*, *Alves W.A.L.* Comparison tools for lncRNA identification: analysis among plants and humans // 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB). IEEE. 2020. pp. 1–8.
137. *Pronozin A.Y.*, *Afonnikov D.A.* ICAnnoLncRNA: A Snakemake Pipeline for a Long Non-Coding-RNA Search and Annotation in Transcriptomic Sequences // Genes. MDPI. 2023. Vol. 14, No. 7, p. 1331.
138. *Gong Y., Huang H.T., Liang Y. et al.* lncRNA-screen: an interactive platform for computationally screening long non-coding RNAs in large genomics datasets // BMC Genomics. 2017. Vol. 18, No. 1, p. 434. DOI: 10.1186/s12864-017-3817-0.
139. *Kit O.I., Kirichenko E.Yu., Kirichenko Yu.G. et al.* Long non-coding RNAs associated with carcinogenesis: biological significance and prospects for use in diagnostics // Clinical Laboratory Diagnostics. OJSC "Izdatelstvo" Medicine ". 2016. Vol. 61, No. 1, pp. 13–16.
140. *Gao L., Gonda I., Sun H. et al.* The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor // Nat. Genet. Nature Publishing Group US New York. 2019. Vol. 51, No. 6, pp. 1044–1051.

Table S1. Comparison of lncRNA bioinformatic analysis pipelines

	LncPipe [1]	LncEvo [2]	CALINCA [3]	lncRNA Detector [4]	lncRNA-screen [5]	ICAnnoLncRNA [6]
Quality control and assembly of the transcriptome	+	+	+	–	+	–
Prediction model training for a non-model organism	–	–	–	–	+	+
Possibility of using data for plants	–	–	–	+	+	+
Prediction of lncRNA using specialized programs	+	+	–	+	+	+
Classification of lncRNAs according to their location relative to protein-coding genes	+	+	+	–	+	+
Removal of transcriptional noise and assembly redundancy	–	–	–	–	–	+
Deletion of sequences that are potential mobile elements	–	–	–	–	–	+
Differential expression analysis	+	–	+	–	+	-
Analysis of lncRNA conservation between different organisms	–	+	+	–	+	+
Analysis of lncRNA distribution in tissues based on its expression	+	–	+	–	+	+

Table S2. Databases for lncRNAs in plants

Title	Description	Data source	Reference
PLncDBv2.0	Plant database with more than 1,246,372 lncRNAs predicted for 80 organisms from chlorophytes to embryophytes. Presents information on expression, tissue specificity, mutations, and developmental stages of lncRNA transcripts in 13834 datasets from different organisms. Allows the user to study the relationship between lincRNAs and epigenetic markers.	Methods for lncRNA prediction	[7]
PNRD	Includes from 166 plant species, over 25,739 lncRNAs of 16 types. Offers several analytical tools and includes a proprietary genomic search engine, protein-coding sequence prediction methods and a microRNA prediction method.	Literature and databases	[8]
NONCODE v6	The database contains 16 animal species and 549,813 lncRNAs belonging to them, and 23 plant species and 94,697 lncRNA transcripts.	Literature and databases	[9]
GREENC v2	Includes more than 495,000 lncRNA transcripts for 94 plant and algal species	Methods for lncRNA prediction	[10]
CANTATAdb v2	The database contains more than 239,631 lncRNAs from 36 plant species and 3 algal species	Methods for lncRNA prediction	[11]
EVLncRNAs v2	Includes 4010 lncRNAs for 124 species of organisms. Provides the user with a network of lncRNA interactions with microRNAs, proteins, genes, and other functional elements. Also contains several lncRNA analysis and prediction programs	Literature	[12]

REFERENCES

1. *Zhao Q. et al.* LncPipe: A Nextflow-based pipeline for identification and analysis of long non-coding RNAs from RNA-Seq data // J Genet Genomics. 2018. Vol. 45, No. 7, pp. 399–401.
2. *Bryzghalov O., Makałowska I., Szcześniak M.W.* lncEvo: automated identification and conservation study of long noncoding RNAs // BMC Bioinformatics. 2021. Vol. 22, No. 1, p. 59.
3. *Talyan S. et al.* CALINCA—A Novel Pipeline for the Identification of lncRNAs in Podocyte Disease // Cells. MDPI, 2021. Vol. 10, No. 3, p. 692.
4. *Shukla B. et al.* lncRNADetector: a bioinformatics pipeline for long non-coding RNA identification and MAPslnc: a repository of medicinal and aromatic plant lncRNAs // RNA Biol. 2021. Vol. 18, No. 12, pp. 2290–2295.
5. *Gong Y. et al.* lncRNA-screen: an interactive platform for computationally screening long non-coding RNAs in large genomics datasets // BMC Genomics. 2017. Vol. 18, No. 1, p. 434.
6. *Pronozin A.Y., Afonnikov D.A.* ICAnnoLncRNA: A Snakemake Pipeline for a Long Non-Coding-RNA Search and Annotation in Transcriptomic Sequences // Genes. MDPI, 2023. Vol. 14, No. 7, p. 1331.
7. *Jin J. et al.* PLncDB V2. 0: a comprehensive encyclopedia of plant long noncoding RNAs // Nucleic Acids Res. Oxford University Press, 2021. Vol. 49, No. D1, pp. D1489–D1495.
8. *Yi X. et al.* PNRD: a plant non-coding RNA database // Nucleic Acids Res. Oxford University Press, 2015. Vol. 43, No. D1, pp. D982–D989.
9. *Zhao Y. et al.* NONCODE 2016: an informative and valuable data source of long non-coding RNAs // Nucleic Acids Res. Oxford University Press, 2016. Vol. 44, No. D1, pp. D203–D208.
10. *Gallart A.P. et al.* GREENC: a Wiki-based database of plant lncRNAs // Nucleic Acids Res. Oxford University Press, 2016. Vol. 44, No. Database issue, p. D1161.
11. *Szcześniak M.W. et al.* CANTATAdb 2.0: Expanding the Collection of Plant Long Noncoding RNAs // Plant Long Non-Coding RNAs / ed. Chekanova J.A., Wang H.-L.V. New York, NY: Springer New York, 2019. Vol. 1933. pp. 415–429.
12. *Zhou B. et al.* EVLncRNAs 2.0: an updated database of manually curated functional long non-coding RNAs validated by low-throughput experiments // Nucleic Acids Res. Oxford University Press, 2021. Vol. 49, No. D1, pp. D86–D91.