

Battles and hijacks: noncoding transcription in plants

Federico Ariel¹, Natali Romero-Barrios¹, Teddy Jégu¹, Moussa Benhamed^{1,2}, and Martin Crespi¹

¹Institut of Plant Sciences Paris-Saclay (IPS2), UMR 9213/UMR1403, CNRS, INRA, Université Paris-Sud, Université d'Evry, Université Paris-Diderot, Sorbonne Paris-Cité, Bâtiment 630, 91405 Orsay, France

²Division of Biological and Environmental Sciences and Engineering, King Abdullah University of Science and Technology, Thuwal 23955-6900, Kingdom of Saudi Arabia

Noncoding RNAs have emerged as major components of the eukaryotic transcriptome. Genome-wide analyses revealed the existence of thousands of long noncoding RNAs (lncRNAs) in several plant species. Plant lncRNAs are transcribed by the plant-specific RNA polymerases Pol IV and Pol V, leading to transcriptional gene silencing, as well as by Pol II. They are involved in a wide range of regulatory mechanisms impacting on gene expression, including chromatin remodeling, modulation of alternative splicing, fine-tuning of miRNA activity, and the control of mRNA translation or accumulation. Recently, dual noncoding transcription by alternative RNA polymerases was implicated in epigenetic and chromatin conformation dynamics. This review integrates the current knowledge on the regulatory mechanisms acting through plant noncoding transcription.

The coding versus the noncoding genome

Long before the era of next-generation sequencing technologies, the poor correlation between DNA amount and organism complexity, the so-called 'C-value paradox' [1], had already drawn the attention of the scientific community. Since the initial genome sequence analyses in higher organisms we have learned that only a tiny percentage of the genome is occupied by protein-coding genes [2]. Even in the case of the model plant *Arabidopsis thaliana* (*Arabidopsis*), less than 50% of its relatively small and compact genome actually encodes proteins [3]. In recent years, in-depth transcriptome analyses revealed that up to 90% of eukaryotic genomes are transcribed into RNAs, although only a small part corresponds to protein-coding mRNAs. Although this finding cannot completely disprove the concept of 'junk DNA' [4], the emergence of the noncoding RNAs (ncRNAs, see Glossary) as novel riboregulators of gene expression shed light on the so-called 'dark matter' of the genome.

Noncoding RNAs comprise a diverse group of transcripts, including housekeeping ncRNAs (ribosomal RNAs, transfer RNAs, small nuclear RNAs, and small nucleolar

RNAs) and many regulatory ncRNAs [5]. Small ncRNAs, such as miRNAs and siRNAs, play important roles in post-transcriptional and transcriptional regulation of eukaryotic gene expression [6]. Although small RNAs are produced by cleavage of double-stranded RNA (dsRNA) derived from lncRNA precursors, they act as small molecules of less than 50 nt. By contrast, recent studies have identified and characterized a growing number of lncRNAs that are able to regulate gene expression without being processed. In recent years, several examples from the plant

Glossary

AGO4: ARGONAUTE 4, an essential component of the TGS machinery. It can load 24 nt siRNAs to recognize Pol V lncRNAs and trigger DNA methylation.

AS: alternative splicing, the regulated processing of mRNA that leads to the splicing of specific introns from a precursor mRNA and results in several mRNAs coding for multiple proteins, by combinations of alternative exons, from a single gene.

ceRNA: competing endogenous RNAs regulate other transcripts by being recognized by a common pool of miRNAs.

dsRNA: double-stranded RNA.

DCL3: DICER-like 3, an enzyme capable of cleaving RDR2-dependent dsRNAs into 24 nt siRNAs for TGS.

IR: inverted repeat, a genomic locus that is transcribed into a perfect or near-perfect dsRNA molecule.

lincRNA: long intergenic noncoding RNA (Figure 1).

lncNAT: long noncoding natural antisense transcript (Figure 1).

lncRNA: long noncoding RNA; ncRNAs longer than 200 nt.

miRNA: microRNA, a DICER-like 1(DCL1)-dependent 21 nt small RNA that, in complex with an AGO protein, induces cleavage or translational regulation of target mRNAs.

ncRNA: noncoding RNA, an RNA that does not encode a protein, but has other cellular functions.

Pol: DNA-dependent RNA polymerase.

PRC: polycomb repressive complex (1 or 2), a complex formed by polycomb group (PcG) proteins, with methyltransferase activity that catalyzes and maintains histone post-translational modifications.

PSMS: photoperiod-sensitive male sterility, fertility regulated by day length.

RdDM: RNA-dependent DNA methylation, a plant-specific DNA methylation pathway involving lncRNAs and 24 nt siRNAs.

RDR2: RNA-dependent RNA polymerase 2; it can interact with Pol IV, transforming a nascent lncRNA into a dsRNA for TGS.

siRNA: small interfering RNA, the product of DICER-like cleavage of dsRNA. DCL1-, DCL2- and DCL4-dependent siRNAs (21 or 22 nt) trigger post-TGS, whereas DCL3-dependent siRNAs (24 nt) induce TGS.

SNP: single-nucleotide polymorphism, a single-nucleotide DNA sequence variation in the genome between members from the same species.

SWI/SNF: an ATP-dependent nucleosome remodeling complex.

TF: transcription factor, a protein or complex that binds to specific DNA sequences to control the rate of transcription of a target gene.

TGS: transcriptional gene silencing, a mechanism depending on noncoding transcription and dsRNA production that controls the activity of transposable elements and expression of protein-coding genes through 24 nt siRNAs.

Corresponding author: Crespi, M. (crespi@isv.cnrs-gif.fr).

Keywords: lncRNAs; lncNATs; RdDM; dual noncoding transcription.

1360-1385/

© 2015 Elsevier Ltd. All rights reserved. <http://dx.doi.org/10.1016/j.tplants.2015.03.003>

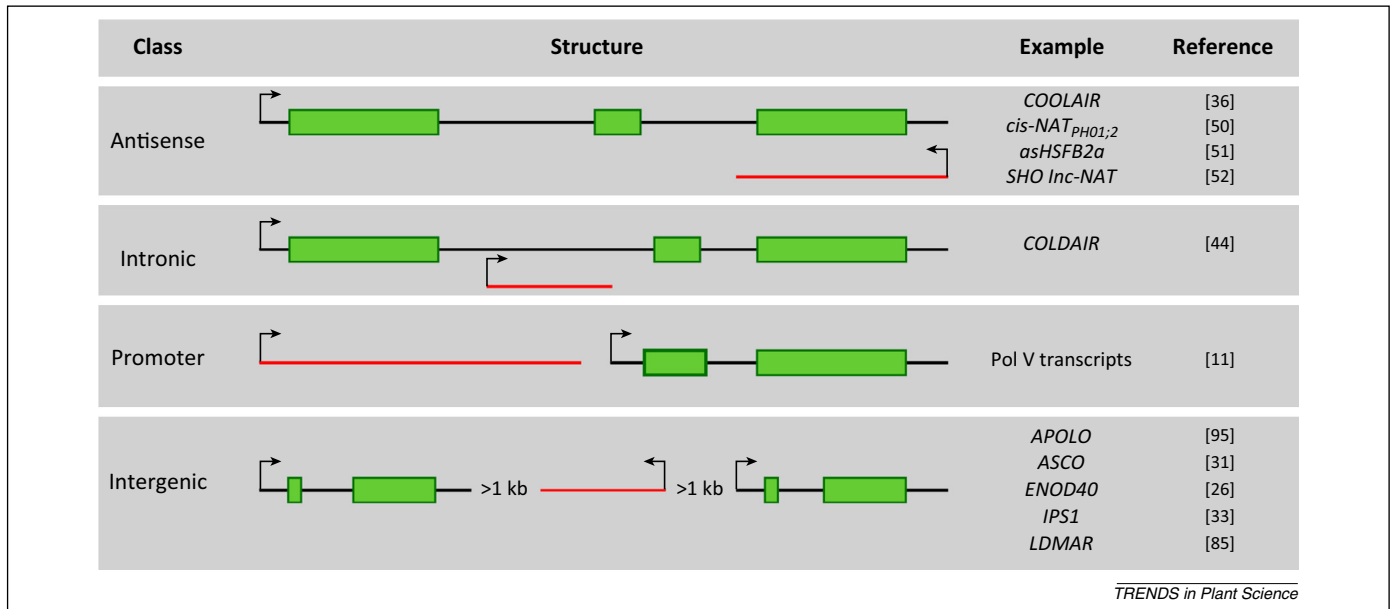


Figure 1. Classification of long noncoding (lnc)RNAs. Antisense transcripts (lncNATs) initiate inside or 3' to a protein-coding gene, are transcribed in its opposite direction, and overlap with at least one coding exon. Intronic lncRNAs initiate inside an intron in either direction and terminate without overlapping with an exon. Promoter lncRNAs are transcripts of the promoter region of a protein-coding gene. Long intergenic ncRNAs (lincRNAs) are separate transcriptional units at a distance of at least 1 kb from protein-coding genes. Protein-coding genes are represented with green exons and black introns, 5' and 3' untranslated regions. Red lines represent noncoding genes, which are shown without introns for clarity. Examples of each class are given with the corresponding reference. Based on the classification by [10,12].

kingdom have shed light on the role of lncRNAs in a wide range of molecular mechanisms, including the hijack of proteins and miRNAs, the modulation of mRNA stability and translation, as well as the control of chromatin modifications. Furthermore, noncoding transcription of a single locus by alternative RNA polymerases has been linked to chromatin conformation dynamics and genome topology.

Classification of plant lncRNAs

A single multisubunit RNA polymerase is in charge of transcribing genomic DNA in prokaryotes, whereas distinct roles are performed by three essential nuclear DNA-dependent RNA polymerases in eukaryotic organisms. RNA polymerase I (Pol I) transcribes ribosomal RNA, Pol II mainly transcribes protein-coding and miRNA genes, and Pol III transcribes short (<400 nt) structural RNAs such as tRNAs and 5S rRNA [7]. In plants, two additional RNA polymerases (closely related to Pol II) [8], Pol IV and Pol V, have likely evolved to specifically mediate transcriptional gene silencing (TGS) [9].

Long noncoding transcripts can be defined by their location relative to nearby protein-coding genes [10]. Long noncoding natural antisense transcripts (lncNATs) initiate inside or 3' to a protein-coding gene, are transcribed in its opposite direction, and overlap with at least one coding exon. Intronic lncRNAs initiate inside an intron in either direction and terminate without overlapping with an exon. Promoter lncRNAs are transcripts of the promoter region of a protein-coding gene and, in animals, these lncRNAs are usually initiated in a divergent fashion by the promoter of the protein-coding gene (divergent lncRNAs) [10]. However, we propose here to establish a wider lncRNA classification to include Pol V-dependent lncRNAs transcribed in gene promoters in any of both directions [11]. Finally, long

intergenic ncRNAs (lincRNAs) are independent transcriptional units between protein-coding genes, separated for at least 1 kb [12]. Examples of known plant lncRNAs of each type are given in Figure 1.

Hunting lncRNAs in plant genomes

Initial genome-wide analyses for lncRNA identification were based on the bioinformatic search for RNAs with poor coding capacity in cDNA databases. This type of approach generated a list of 76 lncRNAs in *Arabidopsis* [13,14], among which there are miRNA precursors, lncNATs, and lincRNAs. Some genomic regions producing lincRNAs also aligned with loci producing 24 nt siRNA clusters, hinting at TGS activity. Abiotic stress altered the accumulation of 22 of these lncRNAs, whereas overexpression analyses in *Arabidopsis* identified two lncRNAs as regulators of root growth during salt stress and leaf morphology, respectively [14]. Similarly, in the model legume *Medicago truncatula*, an *in silico* approach allowed the identification of a set of 503 mRNA-like transcripts that appeared not to encode proteins. Many of these potential lncRNAs shared structural similarities to known lncRNAs, and they clearly differed from protein-coding genes and non-transcribed regions in their sequence composition as well as in potential secondary structure [15]. Based on expressed sequence tag (EST) databases of wheat infected by *Puccinia striiformis* (Pst), three lincRNAs and one lncNAT cDNAs were isolated and shown to be differentially expressed in three near-isogenic lines after Pst inoculation, suggesting their participation in pathogen-defense responses [16].

High-resolution analyses of plant transcriptomes by RNA-sequencing (RNA-seq; [17]) allowed a more comprehensive view of the noncoding genome of several species.

Download English Version:

<https://daneshyari.com/en/article/2825928>

Download Persian Version:

<https://daneshyari.com/article/2825928>

[Daneshyari.com](https://daneshyari.com)