



HAL
open science

Novel long non-protein coding RNAs involved in Arabidopsis differentiation and stress responses.

Besma Ben Amor, Sonia Wirth, Francisco Merchan, Philippe Laporte, Yves d'Aubenton-Carafa, Judith Hirsch, Alexis Maizel, Allison Mallory, Antoine Lucas, Jean Marc Deragon, et al.

► **To cite this version:**

Besma Ben Amor, Sonia Wirth, Francisco Merchan, Philippe Laporte, Yves d'Aubenton-Carafa, et al.. Novel long non-protein coding RNAs involved in Arabidopsis differentiation and stress responses.. Genome Research, 2009, 19 (1), pp.57-69. 10.1101/gr.080275.108 . hal-00736193

HAL Id: hal-00736193

<https://hal.science/hal-00736193>

Submitted on 31 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Novel long non-protein coding RNAs involved in *Arabidopsis* differentiation and stress responses

Besma Ben Amor, Sonia Wirth, Francisco Merchan, et al.

Genome Res. 2009 19: 57-69 originally published online November 7, 2008

Access the most recent version at doi:[10.1101/gr.080275.108](https://doi.org/10.1101/gr.080275.108)

**Supplemental
Material**

<http://genome.cshlp.org/content/suppl/2008/12/05/gr.080275.108.DC1.html>

References

This article cites 67 articles, 28 of which can be accessed free at:
<http://genome.cshlp.org/content/19/1/57.full.html#ref-list-1>

**Email alerting
service**

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#)

To subscribe to *Genome Research* go to:
<http://genome.cshlp.org/subscriptions>

Novel long non-protein coding RNAs involved in *Arabidopsis* differentiation and stress responses

Besma Ben Amor,^{1,6} Sonia Wirth,^{1,6,7} Francisco Merchan,¹ Philippe Laporte,¹ Yves d'Aubenton-Carafa,² Judith Hirsch,^{3,8} Alexis Maizel,¹ Allison Mallory,⁵ Antoine Lucas,^{2,9} Jean Marc Deragon,⁴ Herve Vaucheret,⁵ Claude Thermes,² and Martin Crespi^{1,10}

¹Institut des Sciences du Végétal (ISV), CNRS, 91198 Gif-sur-Yvette, France; ²Centre de Génétique Moléculaire (CGM), CNRS, 91198 Gif-sur-Yvette, France; ³Biomedal S.L., 41092 Sevilla, Spain; ⁴Université de Perpignan Via Domitia, CNRS UMR5096 LGDP, 66860 Perpignan Cedex, France; ⁵INRA Laboratoire de Biologie Cellulaire, 78026 Versailles Cedex, France

Long non-protein coding RNAs (npcRNA) represent an emerging class of riboregulators, which either act directly in this long form or are processed to shorter miRNA and siRNA. Genome-wide bioinformatic analysis of full-length cDNA databases identified 76 *Arabidopsis* npcRNAs. Fourteen npcRNAs were antisense to protein-coding mRNAs, suggesting *cis*-regulatory roles. Numerous 24-nt siRNA matched to five different npcRNAs, suggesting that these npcRNAs are precursors of this type of siRNA. Expression analyses of the 76 npcRNAs identified a novel npcRNA that accumulates in a *dcl1* mutant but does not appear to produce *trans*-acting siRNA or miRNA. Additionally, another npcRNA was the precursor of miR869 and shown to be up-regulated in *dcl4* but not in *dcl1* mutants, indicative of a young miRNA gene. Abiotic stress altered the accumulation of 22 npcRNAs among the 76, a fraction significantly higher than that observed for the RNA binding protein-coding fraction of the transcriptome. Overexpression analyses in *Arabidopsis* identified two npcRNAs as regulators of root growth during salt stress and leaf morphology, respectively. Hence, together with small RNAs, long npcRNAs encompass a sensitive component of the transcriptome that have diverse roles during growth and differentiation.

[Supplemental material is available online at www.genome.org.]

Non-protein coding RNAs (npcRNAs) are a class of RNAs that do not encode proteins, but instead their function lies on the RNA molecule. They are a heterogeneous group and have been divided into different classes according to their length and function. With respect to length, npcRNAs can range from 20 to 27 nucleotides (nt) for the families of microRNAs (miRNAs) and small interfering RNAs (siRNAs), 20–300 nt for small RNAs commonly found as transcriptional and translational regulators, or up to and beyond 10,000 nt for medium and large RNAs involved in other processes, including splicing, gene inactivation, and translation (Costa 2007). We use the term non-protein-coding RNAs instead of noncoding RNAs as every sequence has the potential to be coding, and certain large npcRNAs might encode small oligopeptides, which could be translated under specific conditions as shown for a pentapeptide located inside rRNA, a canonical RNA in *Escherichia coli* (Tenson et al. 1996). In recent years, numerous novel npcRNA candidates have been identified in a variety of organisms from *E. coli* to *Homo sapiens* (Argaman et al. 2001; Storz et al. 2004; Washietl et al. 2005).

⁶These authors contributed equally to this work.

Present addresses: ⁷Laboratorio de Agrobiotecnología, Piso 2, Pabellón 2, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina; ⁸INRA, Campus International de Baillarguet, UMR BGPI, TA A-54/K 34398, Montpellier, France; ⁹Alten Ouest, 12i Rue du Patis Tatelin, ZAC Saint-Sulpice, 35000 Rennes, France.

¹⁰Corresponding author.

E-mail crespi@isv.cnrs-gif.fr; fax 33-1-69-82-36-95.

Article published online before print. Article and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.080275.108>.

Several strategies have been employed to detect and discover novel npcRNAs, including both experimental and computational screenings (Huttenhofer et al. 2002). Genomic approaches, such as tiling arrays and systematic sequencing of full-length cDNA libraries, in model organisms have recently revealed that much larger portions of eukaryote transcriptomes represent non-protein-coding transcripts than previously believed (Okazaki et al. 2002; Numata et al. 2003; Rinn et al. 2003; Ota et al. 2004; Chekanova et al. 2007). Diverse npcRNAs, including a surprising number of antisense RNA transcripts, pseudogenes, and truncated transcripts, have been described (Prasanth and Spector 2007). Certain npcRNAs, referred to as riboregulators, control the stability or translation of specific mRNAs and, in this way, regulate developmental events or stress responses in eukaryotic cells (Erdmann et al. 2001). As very few riboregulators involved in development were previously revealed by classical genetic approaches, it has been proposed that riboregulators may fine-tune mRNA levels in the cell and play a more critical role in the adaptation of developmental processes rather than in differentiation per se.

The most well-studied npcRNA species are single-stranded, 20- to 27-nt small RNAs belonging to two classes, miRNAs and siRNAs, both known to have essential roles in the four eukaryote kingdoms (protists, fungi, plants, and animals). In plants, miRNAs and siRNAs differ in their biogenesis, but both function by guiding target mRNA cleavage after integration into a ribonucleoprotein complex: the RISC (RNA-induced silencing complex) invariably containing a member of the AGO protein family (Vaucheret 2006). In contrast, most animal miRNAs appear to

repress translation (Chapman and Carrington 2007). The miRNAs are single-stranded, 21-nt RNA molecules deriving from partially-complementary RNA precursors, which are mainly transcribed by RNA polymerase II from intergenic regions, although few miRNA genes are located in introns of protein-coding genes. It has been estimated that miRNA genes could represent more than 1% of the expressed genome in worms and humans where it has been proposed that a single miRNA could regulate at least 100 mRNA targets (Lim et al. 2005), underlining the relevance of this post-transcriptional regulatory mechanism. In *Arabidopsis*, 118 putative miRNA loci have been identified (Jones-Rhoades et al. 2006), which represent 42 families, 20 families being conserved in other species such as rice (Bonnet et al. 2004; Sunkar et al. 2005) or poplar (Lu et al. 2005). DICER-LIKE 1 (DCL1) is the main enzyme responsible for mature miRNA production, although DCL4 was shown to process at least two young miRNA genes (Rajagopalan et al. 2006).

In plants, in addition to miRNAs, there exists a great diversity of siRNA: the heterochromatic siRNAs (hc-siRNAs), the trans-acting siRNAs (tasiRNAs), and the natural antisense siRNAs (nat-siRNAs). The common trait during the biogenesis of the first two siRNA species is the generation of a fully complementary double-stranded RNA (dsRNA) by the action of an RNA-dependent RNA polymerase (RDR) (Vaucheret 2006). The highly diverse 24-nt hc-siRNAs are produced through the action of *DCL3* and *RDR2* and are linked to the formation of heterochromatin. The tasiRNAs derive from large npcRNAs that are targets of miRNAs. This miRNA-dependent processing generates shorter npcRNA molecules that are targeted by RDR6 to produce dsRNA that are processed by DCL4 to generate 21-nt tasiRNAs that integrate into RISC complexes. In this way, novel tasiRNAs are generated from the action of nonhomologous miRNAs, amplifying the diversity of small RNA regulated targets (Vaucheret 2006). The nat-siRNAs are generated from a pair of genes transcribed in antisense orientation generating natural double-strand transcripts. They have been identified in *Arabidopsis* plants growing under high NaCl conditions where one overlapping gene is constitutively expressed and the other induced by salt stress. The resulting dsRNA molecules are cleaved by DCL2 and DCL1 enzymes to generate the nat-siRNA, which cleaves the mRNA of the constitutive gene and leads to salt tolerance (Borsani et al. 2005).

In contrast to small RNAs, much less is known about the large and diverse population of long npcRNAs. This heterogeneous class of transcripts generally does not contain any long open reading frame (ORF) (e.g., no ORF coding more than 70 amino acids). Even though some of these genes may encode oligopeptides, it has been shown that their RNA moiety plays a critical role in their function. In order to achieve their function, they interact with proteins to regulate transcription, translation or mRNA stability (Yamashita et al. 1998; Campalans et al. 2004; Prasanth and Spector 2007; Filipowicz et al. 2008; Sasidharan and Gerstein 2008). Furthermore, several of these npcRNAs are precursors of miRNAs and tasiRNAs (Reinhart et al. 2002; Hirsch et al. 2006). Like some miRNAs, certain npcRNAs are induced in various developmental processes as well as during abiotic stress responses in plants and animals (Jones-Rhoades et al. 2006; Mendes Soares and Valcarcel 2006; Prasanth and Spector 2007; Sunkar et al. 2007). In *Caenorhabditis elegans*, 25 npcRNAs are regulated in seven developmental stages and two stimulated conditions (He et al. 2006). We have identified 15 plant npcRNAs displaying diverse tissue-specific expression patterns and/or regulation by environmental stimuli (Hirsch et al. 2006).

In this work, we performed a genome-wide in silico screen of a full-length *Arabidopsis* cDNA library (Castelli et al. 2004) and identified 33 npcRNAs, including 13 that are antisense to protein-coding transcripts. Analyses of these new npcRNAs together with the 43 npcRNAs that we reported previously revealed that 34 potentially give rise to small RNAs, and consistently, we identified several novel mi/siRNA precursors from this group. A new npcRNA accumulated in *dcl1* mutants without forming known mi/siRNAs was identified, while abiotic stresses regulated the expression of 22 npcRNAs. Based on their expression profiles, stable secondary structures, and/or sequence homologies, we selected 12 npcRNAs for functional studies in *Arabidopsis* plants. Overexpression of two of these npcRNAs revealed roles in leaf differentiation and salt stress responses, respectively, suggesting that at least a subset of these newly identified npcRNAs has roles in developmental or stress adaptation programs.

Results

Identification of new *Arabidopsis* npcRNAs

A number of npcRNAs detected in *Arabidopsis* databases could be truncated mRNAs, and as such, further analyses are required to identify bona fide npcRNAs. Our previous (Hirsch et al. 2006) bioinformatic analyses, which filtered 172,495 expressed sequence tags (ESTs) and 24,985 mRNAs retrieved from NCBI databases, identified 43 npcRNA genes that shared certain functional elements, such as a high GC content, an atypical compositional skew (over-representation of T versus A) and the presence, in certain cases, of significantly stable RNA secondary structures. Using a similar approach, we analyzed a collection of 18,000 full-length cDNAs (Castelli et al. 2004) and identified 22 additional candidate npcRNA genes, as well as another 11 npcRNAs that are antisense to annotated genes with an overlapping region of at least 50 bp (for the 33 new npcRNAs and their names, see Supplemental Table 1). Our 43 previously identified npcRNAs also were identified by our new analysis, although in certain cases, we introduced small changes in their annotations; notably three were antisense to protein-coding mRNAs. The size of these 33 new npcRNAs ranges from 265–1879 nt with a mean of 1013 nt, and they harbor one or several introns in one third of the cases. We looked for potential ORFs in the npcRNA, i.e., segments of RNA that, between a start and a stop codon, are multiples of 3 nt. In 22 of them, the longest ORF (mean, 100 nt) is preceded by two to 30 ATG codons, rendering ribosome initiation at this ORF unlikely (Supplemental Table 1). The majority of the selected npcRNAs displays a high GC content (mean GC%, 38%) and a significant abundance of T over A (skew, S already observed for the previous set of npcRNAs, calculated as in Hirsch et al. 2006). Nineteen out of the 22 npcRNAs without an annotated antisense partner display a positive T/A skew, further reinforcing their strand-specific transcription (Touchon et al. 2004). Highly significant Z-scores (Z -score > 5) (Hirsch et al. 2006) suggest the presence of stable RNA secondary structures for five of the new putative npcRNAs (Supplemental Fig. 1). BLASTN analysis revealed that three npcRNAs show significant nucleotide similarities independent of their encoded ORFs, whereas three contain conserved small ORFs (sORFs) (Supplemental Table 1). When the sORF encoded in an npcRNA is much smaller than the similar protein, these genes may correspond to pseudogenes (three genes, indicated as “pp” in Supplemental Table 1). Apart from antisense genes, we could not find any evidence of preferential

silent codon nucleotide substitutions. Finally, npc511, npc513, npc520, npc530, and npc541 contain at least one 100-nt box (conserved greater than 90%) in another region of the *Arabidopsis thaliana* genome.

Nine npcRNAs correspond to miRNA, tasiRNA, and 24-nt siRNA precursors

By mapping small RNA present in databases (Rajagopalan et al. 2006) to the complete collection of 76 npcRNAs (Supplemental Table 1; Hirsch et al. 2006), we revealed that 34 npcRNA loci give rise to at least one small RNA (Table 1 and Supplemental Fig. 2, npcRNA spanning small RNA loci). Some were already characterized small RNA precursors, such as the tasiRNA precursor *TAS3* (npc41) or the *MIR162A* precursor (npc78) (Hirsch et al. 2006). The npc83 and npc521 corresponded to the miRNA precursors *MIR869A* and *MIR160C*, respectively. In addition, we identified new small RNA precursor candidates (npc34, npc351, npc375, npc520, and npc523). The majority of siRNAs deriving from these five npcRNAs are 24 nt long and map to both DNA strands of the npcRNA region, suggesting that these npcRNAs correspond to 24-nt-small RNA precursors. Expression analysis in different organs (Fig. 1) revealed that npc351 is enriched in stems and npc34 in aerial parts, whereas npc375 and npc523 show maximal levels in flowers. Specific oligonucleotides for real-time PCR studies could not be obtained for npc520. Whereas the npc375 and npc523 levels appeared unchanged in the *rdr2 rdr6* double mutant, the npc34 and npc351 were up-regulated in both seedlings and flowers of *rdr2 rdr6* (Fig. 1, right panel), suggesting that RDR2 or RDR6 produces a dsRNA from these npcRNAs, consistent with small RNA production from both DNA strands of these loci.

Transcriptomic analysis identifies a novel DCL4-processed young miRNA gene and a novel npcRNA overaccumulating in *dcl1* mutants

In order to characterize further this set of npcRNAs, we constructed a “dedicated noncoding RNA biology” microarray (the “RIBOCHIP”) containing 274 different oligonucleotide probes (50–70 nt in length) corresponding to all 76 npcRNAs, the sense/antisense couples for those npcRNAs encoding antisense transcripts, and a large collection of npcRNA-related RNA-binding proteins (for a complete list of genes, their identification, and spotted oligonucleotides with their main characteristics, see Supplemental Table 2). At first, the RIBOCHIP was hybridized with RNA from wild-type and *dcl1-9* inflorescences (Fig. 2). Among seven known miRNA precursors spotted on the RIBOCHIP, six (*MIR160C*, *MIR162A*, *MIR164B*, *MIR166A*, *MIR166B*, and *MIR168A*) were up-regulated in *dcl1* inflorescences, consistent with DCL1-processed conserved miRNAs. In contrast, the nonconserved *MIR869A* precursor (npc83) did not overaccumulate in *dcl1*. This miRNA precursor has an unusually stable secondary structure (Z -score = 16.0) (Hirsch et al. 2006) and other small RNAs are produced from the long stem-loop of this precursor (Fig. 3). RT-PCR analysis performed on other *dcl* mutants revealed that the *MIR869A* precursor overaccumulates in *dcl4* (Fig. 3C). These results strongly suggest that, like *MIR822* and *MIR839*, *MIR869A* is a young miRNA gene and that its transcript is processed by DCL4 because it adopts a secondary structure closer to that of perfect dsRNA than that of a classic miRNA precursor processed by DCL1 (Rajagopalan et al. 2006).

In addition to npcRNAs corresponding to conserved

miRNAs, the npc531, which was not previously linked to RNA silencing mechanisms or known to encode a miRNA, also was up-regulated in a *dcl1-9* mutant, a difference confirmed using quantitative RT-PCR (Fig. 4A, left panel). No known miRNAs map to this gene, suggesting that, alternatively, it could be the target of a miRNA. We identified a putative miR319 target site in npc531 (Fig. 4A, right panel), but 5' RACE-PCR analyses did not reveal any specific cleavage of this transcript in this site (data not shown).

Several npcRNAs are regulated by abiotic stresses

A deeper analysis of npcRNA accumulation was performed by hybridizing the RIBOCHIP array with RNA from wild-type leaves and roots, as well as roots of plants grown under stress conditions (phosphate starvation, salt stress, water stress) versus nonstressed control plants. Of the 274 genes spotted on the RIBOCHIP, which include the 76 npcRNAs (Supplemental Table 1; Hirsch et al. 2006), 42 were differentially expressed in at least one of the conditions assayed. From these 42, 26 corresponded to npcRNAs and only six to putative or known RNA binding proteins (RBPs) (Table 2; Fig. 2). The remaining regulated genes corresponded to *cis*-antisense coding transcripts or positive controls. The observed bias toward npcRNAs (62% of the regulated genes were npcRNAs) points to a dynamic regulation of these functional RNAs over npcRNA-related protein-coding genes, which can be additionally controlled at translational and post-translational levels. Moreover, several of these genes were regulated in more than one condition suggesting pleiotropic roles (Table 2).

Independent validations of the results observed in the microarray experiments were obtained for several npcRNAs using Northern analysis (Fig. 4B; that also confirmed the predicted size for npc43, npc72, npc60, and npc536) and quantitative RT-PCR (Fig. 5). For the phosphate starvation assays, the up-regulation of npc43 and npc536 and the down-regulation of npc33 were confirmed (Fig. 5A). We also observed a slight up- and down-regulation of npc60 and npc311, respectively, which also were detected in the microarray experiments (albeit with low statistical significance). Salt stress resulted in a dramatic 100-fold increase in npc60 accumulation. For npc60, npc82, and npc536, the change in expression observed after 3 h of NaCl treatment was maintained after 24 h (Fig. 5B), whereas for npc72 the induction was transient (Fig. 5B). Differential expression of *TAS3*, npc43, and npc311 in roots and of npc15 and npc156 in leaves was confirmed, as we previously reported using semi-quantitative RT-PCR (Hirsch et al. 2006). Additionally, the *MIR160C* precursor was specifically enriched in roots, as already shown (Wang et al. 2005a). Taken together, the combination of microarray and quantitative RT-PCR allowed us to identify eight npcRNAs differentially expressed between roots and leaves and 15 npcRNAs regulated by diverse abiotic stresses.

Overexpression of two npcRNAs affects growth and differentiation in *Arabidopsis*

To explore the biological function of npcRNAs, we selected 12 candidates for further functional analyses. These candidates were chosen considering several criteria, which included conservation at the nucleotide level independent of encoded sORFs, the presence of statistically significant RNA secondary structures, tissue-specific expression patterns, and stress-regulation (Table 3). As a first step to elucidate the function of these 12 npcRNAs, we overexpressed the complete cDNAs under the control of the strong constitutive cauliflower mosaic virus 35S promoter in *Arabidop-*

Table 1. The *Arabidopsis* npcRNA collection and their matching small RNAs

npcRNA	Sense gene	Antisense gene	Gene start	Gene length	Chromosome	Strand ^a	No. of exons	RNA structure	Sense sRNA	Antisense sRNA
2	AT1G03106		748102	933	1	C	1	5.7		
4	AT1G06135		1862196	546	1	W	1			
14			17298005	735	1	W	1			1
15			18141841	833	1	W	1			
17	AT1G52855		19686614	553	1	C	1			
21			26179316	3700	1	C	3			
26			2330796	698	2	W	1		1	
29			6651362	1253	2	C	1			
30			8129244	912	2	W	1			
33	AT2G22122		9410344	545	2	W	1			
34	AT2G34655		14603710	910	2	W	1		14	42
40	AT3G15395		5200646	876	3	W	2			
41	AT3G17185 (<i>TAS3</i>)		5861497	947	3	W	1	6.4	120	116
43		npcRNA 504	6956996	692	3	W	1		1	1
48			11511953	983	3	W	1	5.6		
51	AT3G47965		17718903	992	3	C	2			
52	AT3G48115		17782155	1757	3	C	2			
58	AT4G04315	AT3G66658	2095257	1260	4	W	2			
60	AT4G12735		7493713	519	4	C	1			
62			9244083	589	4	W	1		1	
72			16617625	887	4	W	1			
75		AT4G38930	18151750	804	4	C	1	5.3	3	
78	AT5G08185 (<i>MIR162A</i>)		2634105	1333	5	C	4	5.2	14	1
79	AT5G10745		3396476	1589	5	W	3		1	
82	AT5G24105		8151927	944	5	W	2	9		
83	<i>MIR869A</i>		15908505	706	5	C	1	16.7	22	1
86	AT5G56795		22989135	603	5	C	2			
111			17255407	936	1	C	1		1	1
113			20126020	520	1	C	1		1	
131	AT2G31585		13453079	728	2	W	1			
149	AT3G24927		9103094	903	3	W	1			
150			9666799	788	3	W	1	5.6		
155			22001256	509	3	W	1		1	
156	AT4G04692		2375600	763	4	C	1			
157	AT4G08035		4866551	1790	4	W	1			
311			3544985	488	1	C	1	6.4		
325			11453507	415	1	W	2			
326	AT1G34418		12582100	600	1	W	2			
351			25061142	782	1	W	1	7.2	5	7
370		AT2G07783	3299858	522	2	C	1			1
375			3437581	832	2	C	1	5.1	4	22
415	AT3G09162		2807672	602	3	C	1		4	1
431			11155388	340	3	W	1		2	
502	AT1G16515		5647542	265	1	C	1			
504		npcRNA 43	6957092	591	3	C	1		1	1
507	AT3G21755	AT3G21760	7666815	1758	3	C	2			2
508			17460953	591	3	W	1		1	1
509	AT4G31985		15469901	665	4	W	2			
510	AT1G29357	AT1G29355	10274088	1617	1	C	2			
511	AT1G11592		3895179	1756	1	C	5	5.3	1	1
512			28940575	1500	1	C	5		7	5
513			11110207	1420	4	C	1		6	
514	AT3G27865 (snoRNA)		10329663	1360	3	W	1		8	
515		AT5G21020	7139131	542	5	C	1			
516		AT1G76500	28710088	1356	1	C	1			
518		AT1G58235	4360634	718	5	C	2			
519		AT4G39675	18413768	449	4	C	1			
520			16888083	1195	1	C	1	5	30	18
521	AT5G46845 (<i>MIR160C</i>)		19025187	1362	5	C	1	7.3	21	
522			8459600	705	4	W	1		11	
523	AT2G45245		18666309	1427	2	C	3		66	36
524			9832273	722	3	W	1	9.4	3	1
525	AT2G13665	AT2G13660	5699230	869	2	W	1			
526		AT1G12060	4076214	704	1	C	1		1	
527			11472056	740	1	W	1			
528		AT2G11240	4490324	1080	2	W	1			
529	AT1G67365	AT1G67370	25242073	826	1	W	1			
530	AT1G52347		19500175	1198	1	W	2			
531			17480129	900	1	W	1			
532	AT2G03972		1255245	1198	2	C	4			

(continued)

Table 1. *Continued*

npcRNA	Sense gene	Antisense gene	Gene start	Gene length	Chromosome	Strand ^a	No. of exons	RNA structure	Sense sRNA	Antisense sRNA
534	AT1G67265 (<i>DVL3</i>)		25179221	645	1	C	1			
535	AT3G05152		1443973	1121	3	W	2			
536	AT1G67920	AT1G67930	25477207	597	1	W	1		1	
539	AT3G59765		22087438	1879	3	C	3		1	4
540			20581088	850	1	C	1		1	
541			28298269	753	1	W	1			

Table shows npcRNA genes, annotation (name), antisense transcripts in this locus, gene position, length, chromosome, strand (Watson or Crick), number of exons, Z-score, as well as the number of small RNA hits matching to either the sense or antisense npcRNA molecule (relative to the transcription direction). npcRNAs from npc2 to npc431 are derived from Hirsch et al. (2006), whereas npcRNAs from npc502 to 541 are from this study. ^aW, Watson; C, Crick.

sis. Overexpression of 11 out of these 12 npcRNAs did not lead to obvious visible differences compared with wild-type plants in nonstressed growth conditions (for a list of the overexpressed npcRNAs and the number of independent lines analyzed, see Table 3). In contrast, T1 plants overexpressing the npc48 showed drastic developmental anomalies, including an increase in the rosette diameter, leaf serration, and a delay in the flowering time compared with wild-type plants (Fig. 6A). Quantitative RT-PCR analysis of npc48 accumulation revealed that all transgenic lines exhibiting this phenotype overaccumulated npc48 (Supplemental Fig. 3). The phenotype of 35S::npc48 plants resembled that of 35S::*MIR168* plants, which have reduced levels of the miR168 *AGO1* target, and of *AGO1*-sensitive miRNAs such as miR166 (Vaucheret et al. 2006). 35S::npc48 plants also showed a decrease in miR166 accumulation, but accumulated *AGO1* mRNA and miR168 at wild-type levels (Fig. 6B). In addition, 35S::npc48 plants, but not 35S::*MIR168* plants, exhibited reduced accumulation of miR164 (Fig. 6B), suggesting that npc48 could regulate a subset of miRNAs. T-DNA insertions within npc48 are not available.

As the expression of several npcRNAs was affected by environmental stress, we explored the response to salt stress and phosphate starvation in lines overexpressing npc43, npc60, npc311, and npc536. 35S::npc536 transformants that actually overaccumulate npc536 (Supplemental Fig. 4) displayed heightened root growth under salt stress conditions (Fig. 7) compared with wild-type plants. This increase is due to both primary root growth and secondary root length under salt stress (100 and 125 mM), whereas no differences were observed in the absence of salt treatment. No phenotype could be observed in mutants that carry T-DNA insertions within npc536 (data not shown). The npc536

transcript is antisense to the AT1G67930 mRNA (encoding a Golgi-transport complex related protein) (Fig. 7C). The npc536 has a large dynamic range of expression across a wide range of tissue and hormonal, biotic, or abiotic treatment, whereas AT1G67930 is much less variable (AtGenExpress atlas) (Schmid et

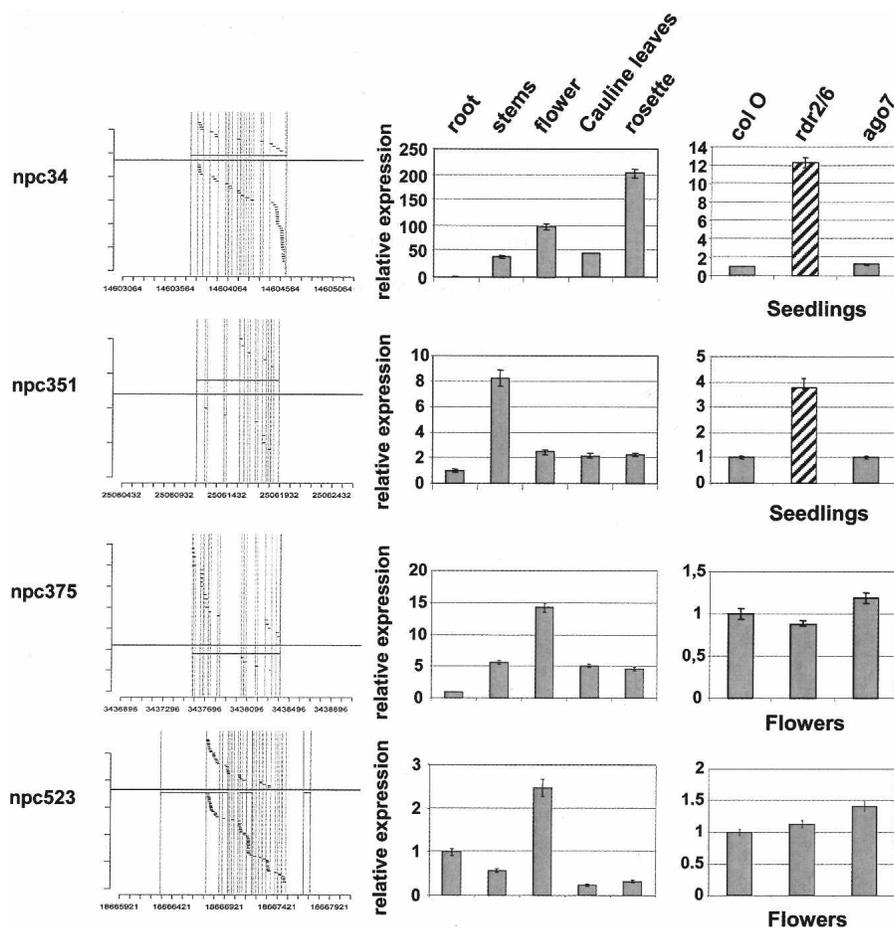


Figure 1. Expression patterns of four npcRNAs corresponding to new 24-nt siRNA precursors. (Left panels) The gray boxed regions (above or below are Watson and Crick strands, respectively) indicate the npcRNA gene and transcript (black line) on the *Arabidopsis* genomic DNA coordinates. Small RNAs identified in public databases are indicated by small bands on both strands. The size of each siRNA is indicated in Supplemental Figure 2; npcRNAs spanning small RNA loci (following a color code). (Middle panels) Expression of these npcRNAs in different tissues (roots, stems, flowers, cauline leaves, and rosette leaves) determined by qRT-PCR. Relative expression levels were normalized with ACT2 (AT3G18780), and values for roots were arbitrarily fixed to 1; SDs are shown. (Right panels) Expression in flowers or seedlings of Col-0 and a *rdr2/6* or *ago7* mutant. Values for Col-0 were fixed to 1.

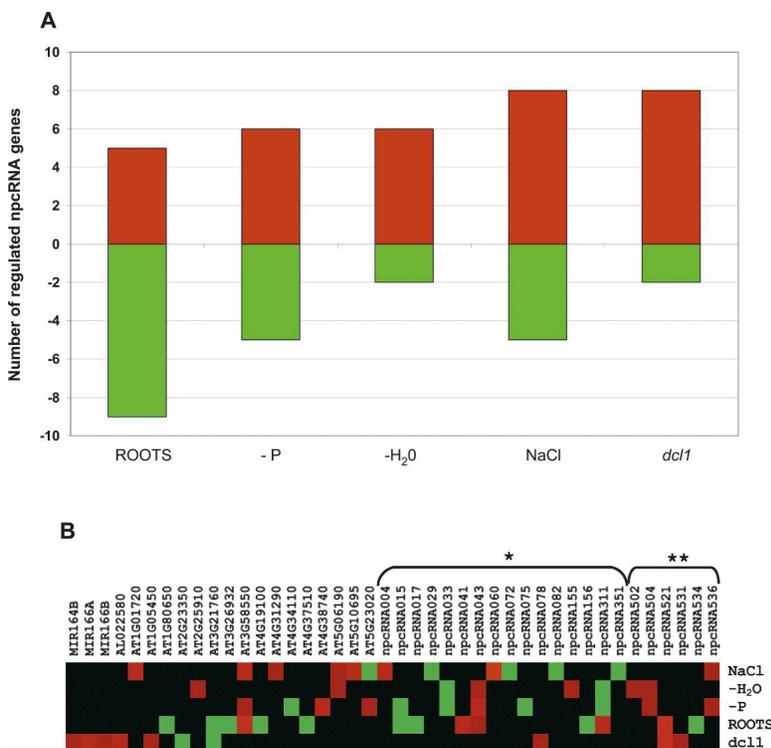


Figure 2. Regulation of genes in the RIBOCHIP. (A) Total number of genes that were either induced (red bars) or repressed (green bars) in each analyzed condition. Conditions include expression in roots versus leaves (roots), roots of phosphate starved plants (-P), plants under water stress (-H₂O), roots of plants treated with 150 mM NaCl (NaCl), and inflorescences of *dcl1-9* mutants (*dcl1*). (B) Heat map showing expression level of each individual gene summarized in A. (*) npcRNAs identified by Hirsch et al. (2006); (**) npcRNAs identified in this study. Genes without asterisks correspond to additional regulated genes from the RIBOCHIP (annotation is indicated: coding sense counterparts of npcRNAs, RNA-binding proteins, miRNA precursors, and control genes).

al. 2005). In response to certain abiotic stresses such as drought or cold treatment, the root expression of these two genes tends to be anti-correlated (Pearson correlation coefficient of -0.636 and -0.983 , respectively). However, accumulation of the AT1G67930 transcript was not significantly modified in 35S::npc536 plants or in npc536 mutants (Supplemental Fig.4). It remains possible that npc536 regulates the translation of AT1G67930 mRNA. Altogether, overexpression of two of 12 npcRNAs affected *Arabidopsis* differentiation and growth responses to abiotic stresses.

Discussion

Through this work we revealed a variety of *Arabidopsis* non-protein-coding transcripts, some of which likely regulate development and responses to abiotic stresses. Whole-genome mapping in the model plant *Arabidopsis*, based on the use of tiling arrays, revealed that >50% of observed transcription was intergenic and that numerous antisense RNA transcripts exist (Yamada et al. 2003). In *Drosophila*, related experiments based on tiling-arrays, performed at six developmental stages, detected RNA expression for 41% of the probes in intronic and intergenic regions (Stolc et al. 2004). In agreement, analysis of EST and full-length cDNA libraries identified many transcripts not previously assigned to genomic loci (Okazaki et al. 2002; Numata et al.

2003; Ota et al. 2004; Riano-Pachon et al. 2005; Hirsch et al. 2006). At least 13% and 26% of the unique full-length cDNAs in mice and humans, respectively, are thought to be poly(A) tail-containing mRNA-like npcRNAs. Since the functions of most of these npcRNAs are unknown, much work needs to be done to identify regulatory RNAs among these genes.

Using genome-wide mapping of full-length cDNAs, we describe 33 new npcRNAs that, when added to the 43 previously identified (Hirsch et al. 2006), make a set of 76 npcRNAs dispersed throughout the *Arabidopsis* genome. These npcRNAs belong to at least five classes:

1. Three npcRNAs correspond to miRNA precursors such as the MIR869A precursor that overaccumulates in *dcl4* mutants, suggesting that MIR869A is a young miRNA gene, similar to MIR822 and MIR839 (Rajagopalan et al. 2006).
2. Five npcRNAs give rise to 24-nt siRNAs that map to both DNA strands, suggesting that these npcRNAs define 24-siRNA precursors. Among these five npcRNAs, two (npc34 and npc351) overaccumulate in *rdp2/6* mutants, suggesting that these npcRNAs are likely RDR substrates or, alternatively, targets of RDR-dependent small RNAs such as the tasiRNAs.
3. Five npcRNAs correspond to small RNA classes other than 21, 22, or 24 nt reminiscent of the class of 30- to 40-nt long-siRNAs (lsiRNAs) (Katiyar-Agarwal et al. 2007) recently shown to be involved in gene regulation in *Arabidopsis*.
4. Fourteen npcRNAs corresponded to *cis* natural antisense transcripts (*cis*-NATs), based on partial complementarity to other endogenous RNAs. The *cis*-NATs transcripts are derived from the same genomic loci as their sense counterparts, but from opposite strands. Genome-wide computational and experimental studies have shown that ~5%–10% of gene transcripts in mammals and plants have *cis*-NATs (Wang et al. 2005b; Henz et al. 2007). cDNA sequence cluster analyses revealed that 7600 annotated genes in *Arabidopsis* (30%) had significant antisense expression (Yamada et al. 2003; Chekanova et al. 2007). Furthermore, characterization of 32,000 full-length rice cDNAs identified 600 antisense transcript pairs, half of which have no ORF in one member of the pair (Osato et al. 2003). Analysis of their regulation in comparison with their complementary transcripts indicates that there is a trend toward anticorrelated expression of *cis*-NAT pairs in *Arabidopsis*. However, currently available data do not produce a strong signature of small RNA mediated silencing for this process (Henz et al. 2007). The *Arabidopsis* transcriptome also contains a fairly large number of *trans*-NATs, ~1320 putative *trans*-NAT pairs. Among those with available expression data, more than 85% were found in the same tissue as their sense partners

Table 2. Analysis of gene regulation using the RIBOCHIP

Gene name	Roots	-P	-H ₂ O	NaCl	<i>dcl1</i>	Description
<i>MIR164B</i>					3	miRNA 164B
<i>MIR166A</i>					5	miRNA 166A
<i>MIR166B</i>					3	miRNA 166B
AL022580					5	T5K18 miRNA168a
AT1G01720				5		No apical meristem (NAM) family protein putative TF
AT1G05450					2	Protease inhibitor/seed storage/lipid transfer protein (LTP)
AT1G80650	-2					Double-stranded RNA-binding domain (DsRBD)-containing protein
AT2G23350					-3	Polyadenylate-binding protein, putative
AT2G25910			2			3-5 Exonuclease domain-containing protein
AT3G21760	-3				-2	UDP-glucuronosyl/UDP-glucosyl transferase family protein
AT3G26932	-3					DsRBD-containing protein
AT3G58550	8	3		4		Protease inhibitor/seed storage/lipid transfer protein (LTP) family
AT4G19100	-6					Expressed protein
AT4G31290				2		ChaC-like family protein
AT4G34110		-2				Polyadenylate-binding protein 2 (PABP2)
AT4G37510	-3					Ribonuclease III family protein
AT4G38740		4				Peptidyl-prolyl <i>cis-trans</i> isomerase/cyclophilin/rotamase
AT5G06190			2	3		Expressed protein
AT5G10695				4		Expressed protein
AT5G23020		2		-4		2-Isopropylmalate synthase 2 (IMS2), identical to 2-isopropylmalate
npcRNA 004				4		npcRNA (AT1G06135)
npcRNA 015	-3	-2				npcRNA
npcRNA 017	-2					npcRNA AT1G52855
npcRNA 029				-2		npcRNA
npcRNA 033		-4	-2			npcRNA AT2G22122
npcRNA 041	4					TAS3
npcRNA 043	3	6	3			npcRNA
npcRNA 060				97		npcRNA AT4G12735
npcRNA 072				-2		npcRNA
npcRNA 075		-2				npcRNA
npcRNA 078					3	<i>MIR162A</i>
npcRNA 082				-3		npcRNA AT5G24105
npcRNA 155			2			npcRNA
npcRNA 156	-5					npcRNA AT4G04692
npcRNA 311	4	-2	-2			npcRNA
npcRNA 351				-2		npcRNA
npcRNA 502			3			npcRNA AT1G16515
npcRNA 504		3	2			npcRNA (AS 43)
npcRNA 521	14				8	<i>MIR160C</i>
npcRNA 531					3	npcRNA
npcRNA 534	-7					npcRNA
npcRNA 536		3		2		npcRNA (AT1G67920)

Table shows genes spotted in the RIBOCHIP that were regulated (induced, positive numbers or repressed, negative numbers) in each analyzed condition: roots versus leaves (Roots), roots of phosphate starved plants (-P), plants under water stress (-H₂O), roots of plants treated with 150 mM NaCl (NaCl), and inflorescences of *dcl1-9* mutants (*dcl1*). Numbers indicate fold-change induction (positive) or repression (negative) respective to several constitutive controls.

and further purified by passage through RNAeasy columns (Qiagen) and residual genomic DNA removed by on-column DNase I treatment (Qiagen).

Bioinformatic analysis of npcRNAs

To identify candidate npcRNAs from the "full length" sequenced cDNA databank (Castelli et al. 2004), RNAs were first mapped with Sim4 (Florea et al. 1998) on the genome and selected by the length of the longest ORF (<210 nt). About 500 candidates were further tested for lack of significant BLAST hits against NCBI nr protein databank. Mapping of other ESTs and cDNAs on these genomic regions did not bridge the gap with already annotated neighboring genes. Mapping of short RNAs (sRNAs) on the total set of 76 npcRNA was performed using a large data set of massively sequenced *A. thaliana* sRNAs (Rajagopalan et al. 2006). Potential miRNAs regulating npc531 were searched in miRBASE database. Sequence comparisons were performed using BLASTN for nucleotide sequences and BLASTX for encoded sORFs sur TAIR, NCBI nonredundant and chromosome databases. Similarities

between nucleotide (BLASTN) and sORF-encoded peptide (BLASTX) were searched up to an expected value $E = 1 \times 10^{-5}$. Boxes of at least 21 nt conserved between number of species N independent of the encoded ORFs are indicated as N(+). Conserved sORFs are indicated as N(XX), where N is the number of species and XX is the size of the sORF in *A. thaliana*. Those cases where the sORF show similarity with a larger ORF (>100 amino acids) in other species are indicated as putative pseudogenes (pp). Exceptionally stable RNA secondary structures were predicted as by Hirsch et al. (2006). The sequences were scanned using sliding windows of increasing lengths. Briefly, for each window, the free-energy of folding of the most stable structure was computed using rnaFold from the Vienna Package (<http://www.tbi.univie.ac.at/~ivo/RNA/>) and compared with the distribution of free-energy of folding of random sequences similar in size and nucleotide composition. The Z-score, defined as the difference between these free-energies divided by the standard deviation, was calculated, and sequence regions with a Z-score > 5 were displayed in Supplemental Figure 1.

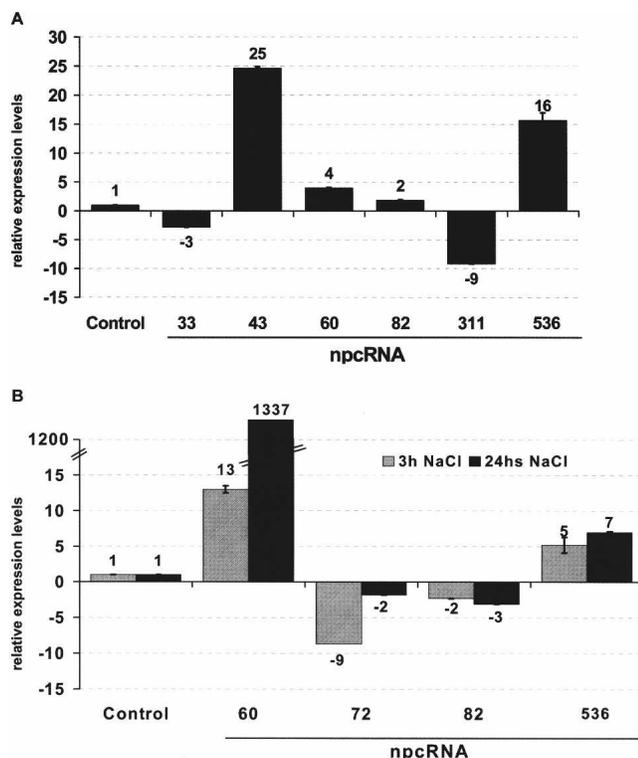


Figure 5. Regulation of specific npcRNAs in different stress conditions. Real-time RT-PCR expression analysis of indicated npcRNAs in roots of plants grown in phosphate starvation (A) and after 3h and 24h in 150 mM NaCl (B). In every case data was normalized with *ACT2* (AT3G18780) and values for non-treated or wild-type controls were arbitrary fixed to 1. For each cDNA synthesis, quantifications were made in triplicate and two biological replicates were analyzed. Values are means \pm SD.

RIBOCHIP array design

Oligonucleotide primers were designed using OligoWiz 2.0 (Wernersson and Nielsen 2005). Each 45–60mer oligonucleotide probe was printed as triplicates on Quantifoil QMT epoxy slides in 16 grids in a 4 metacolumns \times 4 metarows configuration with a 9 columns \times 7 rows spot pattern for each subarray by the Institute of Genome Research, Center of Biotechnology at Bielefeld

University, Germany. Additionally, the array contains probes specifically designed as labeling and hybridization controls (positive) corresponding to nine *A. thaliana* genes: AT5G44200 (nuclear cap-binding protein), AT1G07920 (EF1-alpha), AT5G03240 (*UBQ3*), AT4G05320 (*UBQ10*), AT3G13920 (*EIF4A1*), AT1G19910 (V-type H⁺ ATPase 16 kD subunit), M21415 (*TUB4*), AT1G49240 (ACTIN 8 protein), and AT3G18780 (ACTIN 2 protein). At least two independent biological replicates were performed for each condition analyzed in the RIBOCHIP.

Target preparation for RIBOCHIP array

Two micrograms of total RNA from each biological sample (treatments and controls) were amplified using the Amino Allyl Message Amp II aRNA Amplification Kit (Ambion), following the protocol provided by the manufacturer. Ten micrograms of amplified RNA (aRNA) was chemically labeled with Cy5 (Q15108, Amersham), monofunctional reactive dye or Cy3 (Q13108, Amersham), monofunctional reactive dye in coupling buffer (Ambion). Labeled aRNA was purified from noncoupled dyes using the columns supplied by the Amino Allyl Message Amp II aRNA Amplification Kit.

Array hybridization

Before use, the printed slides were processed according to the manufacturer's recommendations (Institute of Genome Research, Center of Biotechnology, Bielefeld University) and then hybridized with 200 pmol each of Cy5- and Cy3-labeled targets (treatment and control) in $5 \times$ SSC (0.75 M NaCl, 75 mM sodium citrate at pH 7), 25% formamide, 0.1% SDS for 60 h at 42°C. For each slide, 20 μ g of salmon sperm DNA and 4 μ g of yeast tRNA (Sigma) were added as blocking agents. After hybridization slides were washed once at 42°C in $1 \times$ SSC, 0.2% SDS for 4 min followed by a wash at room temperature in $1 \times$ SSC, 0.2% SDS for 4 min and twice at 4 min each in $0.1 \times$ SSC. A quick final wash at room temperature in $0.05 \times$ SSC for 1 min was done to remove any salt residue and particles bound to the slides.

Slides were processed using an Axon 4000B microarray scanner (Axon Instruments) with 635-nm and 532-nm lasers and a scan resolution of 10 μ m to generate TIFF images on each channel. PMTs for lasers were automatically set by software for each slide (GenePix 6.0, Axon Instruments).

Table 3. Characteristics of selected candidate genes for functional analysis

npcRNA	No. of independent lines	Gene annotation	Sequence similarities	Z-score > 5	Tissue specificity	Small RNA	Regulation
34	19	AT2G34655		—	Leaves	Yes	
TAS3	14	AT3G17185	2 (+)	6.4	Roots	Yes	
43	21			—		Yes	P, H ₂ O
48	236			5.6		No	
60	9	AT4G12735		—		No	NaCl
72	27			—		No	NaCl
82	9	AT5G24105	9 (63)	9	Flowers	Yes	NaCl
MIR869A	14			16.7	Stems	Yes	
311	19		Bn (+) ^a	6.4	Roots	No	P, NaCl
375	27		15 (+)	5.1	Flowers	Yes	
531	51			—		No	Dcl1
536	8		2 (67)	—		No	P, NaCl

Columns contain the name of candidate genes, sequence similarities (+, nucleotide similarities in N species independent of encoded sORFsl N(XX), number of species containing a conserved sORF of XX amino-acids in *Arabidopsis*; see Methods), the values of the Z-scores (statistically significant stable RNA secondary structure were selected, Z-scores > 5), tissue specificity (+, preferentially expressed), regulation in response to stress condition (+, up-regulated; P, phosphate starvation; H₂O, water stress; NaCl, salt stress; Dcl1, *dicer-like1* mutant), and their capacity to encode small RNAs (yes or no).

^a*Brassica napus* (a plant close to *Arabidopsis*).

Microarray data processing and normalization

Quantification of images on 635-nm and 532-nm channels was performed with GenePix 6.0 software (Axon Instruments). Normalization was performed by the Loess and median by block method using MANGO software developed at the Gif-Orsay DNA microarray platform (Center de Génétique Moléculaire, CNRS). Replicate spots outliers were detected by Grubb's test. Statistical significance of regulated probes was analyzed by the single slide methods of Newton et al. (2001), Sapir and Churchill (2000), and Chen et al. (1997).

Gene expression analysis

For real-time RT-PCR, cDNA was synthesized by reverse transcription performed on 1.5 µg of total RNA using the SuperScript II Reverse Transcriptase (Invitrogen) and (T)16 A/G/C oligonucleotides. Real-time RT-PCR was performed on the Roche Light Cycler instrument using SYBR Green I dye (LightCycler FastStart DNA MasterPLUS SYBR Green I, Roche) using the primer pairs listed in Supplemental Table 3. Technical triplicates were done for each assay, and at least two independent biological replicates were assayed.

5' RACE-PCR assays to detect npcRNA 531 cleavage products and Northern analysis of small RNAs were performed as described by Hirsch et al. (2006).

Construction of *A. thaliana* npcRNA overexpression lines

The coding sequences for each npcRNA were amplified by PCR from genomic *A. thaliana* Col-0 DNA using the primers showed in Supplemental Table 4. The full-length DNAs were cloned into

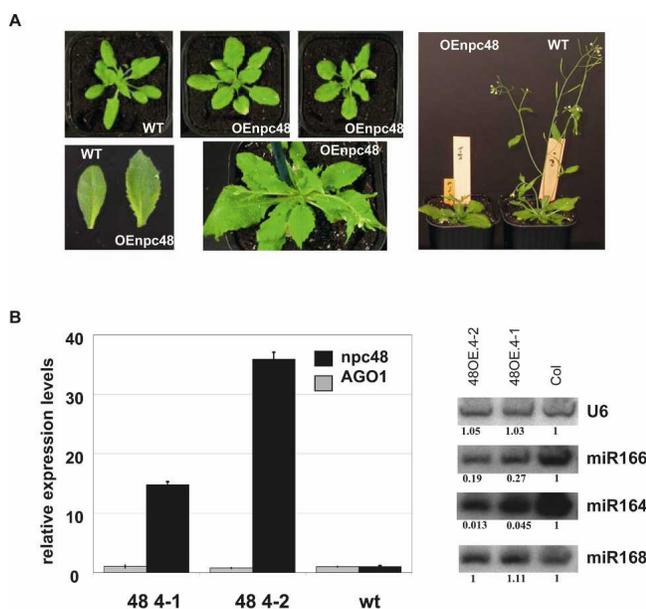


Figure 6. Phenotypic and molecular analysis of transgenic plants overexpressing npc48. (A) Phenotype of *Arabidopsis* plants transformed with 35S::npc48 construct. A wild-type (WT) plant (left) and two representative T1 35S::npc48 plants (OEnpc48) from independent lines displaying a characteristic serrated leaf phenotype at different growth stages (central photos) are shown. Rosette leaves from WT and OEnpc 48 plants (lower left panel). Image of whole plants (right panel) illustrates the delayed flowering of OEnpc48 plants. (B) Expression analysis of npc48 and AGO1 mRNA from leaves of two npc48 overexpressing lines and control WT using real-time RT-PCR (left panel). Northern blot analyses of miR164, miR166, miR168, and U6 RNAs (right panel). Quantification is indicated setting the Col-0 value arbitrarily to 1 and normalizing to U6 values.

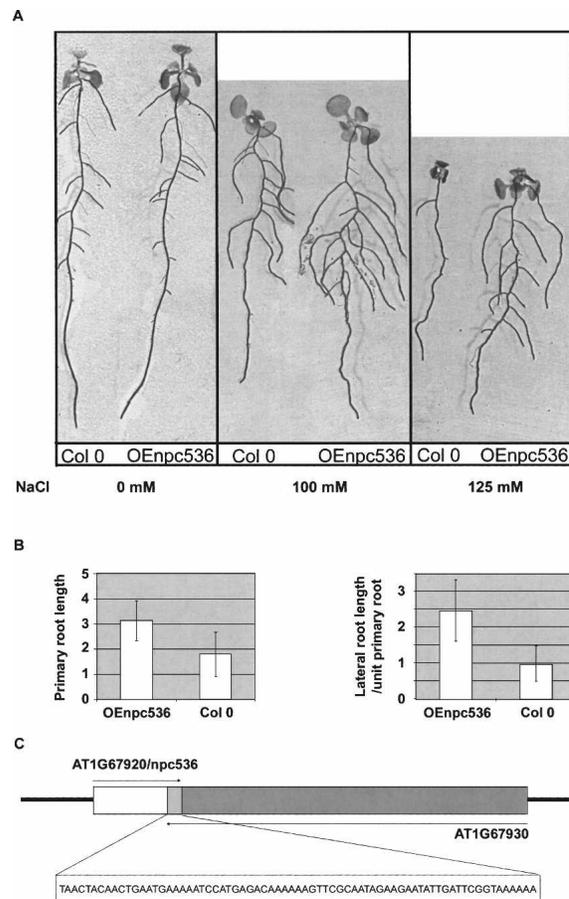


Figure 7. Phenotypic and molecular analysis of transgenic plants overexpressing npcRNA536. Phenotype of *Arabidopsis* plants transformed with 35S::npc536 constructs (Oenpc536). (A) Control and a representative OEnpc536 transgenic plants grown under the indicated salt stress conditions (0 mM, 100 mM, and 125 mM). (B) Quantification of primary root length and lateral root/primary root lengths in two independent transgenic OEnpc536 and control lines grown in 100 mM salt ($n < 45$ per experiment). SDs are indicated. These differences are statistically significant (t -test, 7×10^{-7}), whereas no differences could be detected under normal growth conditions. (C) Schematic of the genomic positions and transcripts deriving from the npc536 (AT1G67920, white box) and AT1G67930 (dark gray box) loci. The nucleotide sequence of the overlapping region (light gray box) between the two transcripts is indicated.

pENTR/D-TOPO (Invitrogen) and then in the pB7WG2 binary plasmid (Karimi et al. 2002), under the control of CaMV 35S promoter, using the gateway technology (Invitrogen) to be introduced into *Agrobacterium tumefaciens* (agl-0). These constructions were used to transform *A. thaliana* ecotype "Columbia" plants, by floral dip (Bechtold and Pelletier 1998). Transgenic plants were selected by spraying seedlings from T1 generation 7, 9, and 11 d after germination with a solution of 0.01% Basta (200 g/L glufosinate ammonium). Basta-resistant plantlets (T1) were then grown in soil under long-day conditions (16-h light). Expression levels of the transgene in the T1 and T2 generations were determined by qRT-PCR.

Acknowledgments

We thank Raquel Chan (Argentina) for water stress RNA samples and Javier Paz Ares (Spain) for helpful discussions. P.L. is the recipient of a fellowship from the Ministère de l'Éducation Na-

tionale et de la Recherche (MENR; France); S.W. was a fellow from the Fundacion Foro de Biotecnologia, Argentina. F.M. was supported by the Ministerio de Educacion y Ciencia, Spain. This work was also supported by the European FP6 RIBOREG project no. LSHG503 022 and partly by the French ANR-Genoplante RIBOROOT project.

References

- Adenot, X., Elmayan, T., Lauresseguies, D., Boutet, S., Bouche, N., Gascioli, V., and Vaucheret, H. 2006. DRB4-dependent TAS3 *trans*-acting siRNAs control leaf morphology through AGO7. *Curr. Biol.* **16**: 927–932.
- Argaman, L., Hershberg, R., Vogel, J., Bejerano, G., Wagner, E.G., Margalit, H., and Altuvia, S. 2001. Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Curr. Biol.* **11**: 941–950.
- Bechtold, N. and Pelletier, G. 1998. In planta *Agrobacterium*-mediated transformation of adult *Arabidopsis thaliana* plants by vacuum infiltration. *Methods Mol. Biol.* **82**: 259–266.
- Bonnet, E., Wuyts, J., Rouze, P., and Van de Peer, Y. 2004. Detection of 91 potential conserved plant microRNAs in *Arabidopsis thaliana* and *Oryza sativa* identifies important target genes. *Proc. Natl. Acad. Sci.* **101**: 11511–11516.
- Borsani, O., Zhu, J., Verslues, P.E., Sunkar, R., and Zhu, J.K. 2005. Endogenous siRNAs derived from a pair of natural *cis*-antisense transcripts regulate salt tolerance in *Arabidopsis*. *Cell* **123**: 1279–1291.
- Bouche, N., Lauresseguies, D., Gascioli, V., and Vaucheret, H. 2006. An antagonistic function for *Arabidopsis* DCL2 in development and a new function for DCL4 in generating viral siRNAs. *EMBO J.* **25**: 3347–3356.
- Brosius, J. 2003. Gene duplication and other evolutionary strategies: From the RNA world to the future. *J. Struct. Funct. Genomics* **3**: 1–17.
- Campalans, A., Kondoroski, A., and Crespi, M. 2004. *Enod40*, a short open reading frame-containing mRNA, induces cytoplasmic localization of a nuclear RNA binding protein in *Medicago truncatula*. *Plant Cell* **16**: 1047–1059.
- Castelli, V., Aury, J.M., Jaillon, O., Wincker, P., Clepet, C., Menard, M., Cruaud, C., Quetier, F., Scarpelli, C., Schachter, V., et al. 2004. Whole genome sequence comparisons and “full-length” cDNA sequences: A combined approach to evaluate and improve *Arabidopsis* genome annotation. *Genome Res.* **14**: 406–413.
- Chapman, E.J. and Carrington, J.C. 2007. Specialization and evolution of endogenous small RNA pathways. *Nat. Rev. Genet.* **8**: 884–896.
- Chekanova, J.A., Gregory, B.D., Reverdatto, S.V., Chen, H., Kumar, R., Hooker, T., Yazaki, J., Li, P., Skiba, N., Peng, Q., et al. 2007. Genome-wide high-resolution mapping of exosome substrates reveals hidden features in the *Arabidopsis* transcriptome. *Cell* **131**: 1340–1353.
- Chen, Y., Dougherty, E.R., and Bittner, M.L. 1997. Ratio-based decisions and the quantitative analysis of cDNA microarray images. *J. Biomed. Opt.* **2**: 364–374.
- Costa, F.F. 2007. Non-coding RNAs: Lost in translation? *Gene* **386**: 1–10.
- Deak, K.I. and Malamy, J. 2005. Osmotic regulation of root system architecture. *Plant J.* **43**: 17–28.
- Erdmann, V.A., Barciszewska, M.Z., Szymanski, M., Hochberg, A., de Groot, N., and Barciszewski, J. 2001. The non-coding RNAs as riboregulators. *Nucleic Acids Res.* **29**: 189–193.
- Filipowicz, W., Bhattacharyya, S.N., and Sonenberg, N. 2008. Mechanisms of post-transcriptional regulation by microRNAs: Are the answers in sight? *Nat. Rev. Genet.* **9**: 102–114.
- Florea, L., Hartzell, G., Zhang, Z., Rubin, G.M., and Miller, W. 1998. A computer program for aligning a cDNA sequence with a genomic DNA sequence. *Genome Res.* **8**: 967–974.
- Franco-Zorrilla, J.M., Martin, A.C., Solano, R., Rubio, V., Leyva, A., and Paz-Ares, J. 2002. Mutations at CRE1 impair cytokinin-induced repression of phosphate starvation responses in *Arabidopsis*. *Plant J.* **32**: 353–360.
- Franco-Zorrilla, J.M., Valli, A., Todesco, M., Mateos, I., Puga, M.I., Rubio-Somoza, I., Leyva, A., Weigel, D., Garcia, J.A., and Paz-Ares, J. 2007. Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat. Genet.* **39**: 1033–1037.
- Gascioli, V., Mallory, A.C., Bartel, D.P., and Vaucheret, H. 2005. Partially redundant functions of *Arabidopsis* DICER-like enzymes and a role for DCL4 in producing *trans*-acting siRNAs. *Curr. Biol.* **15**: 1494–1500.
- He, H., Cai, L., Skogerbo, G., Deng, W., Liu, T., Zhu, X., Wang, Y., Jia, D., Zhang, Z., Tao, Y., et al. 2006. Profiling *Caenorhabditis elegans* non-coding RNA expression with a combined microarray. *Nucleic Acids Res.* **34**: 2976–2983.
- Henz, S.R., Cumbie, J.S., Kasschau, K.D., Lohmann, J.U., Carrington, J.C., Weigel, D., and Schmid, M. 2007. Distinct expression patterns of natural antisense transcripts in *Arabidopsis*. *Plant Physiol.* **144**: 1247–1255.
- Hirsch, J., Lefort, V., Vankerschaver, M., Boualem, A., Lucas, A., Thermes, C., d’Aubenton-Carafa, Y., and Crespi, M. 2006. Characterization of 43 non-protein-coding mRNA genes in *Arabidopsis*, including the MIR162a-derived transcripts. *Plant Physiol.* **140**: 1192–1204.
- Huttenhofer, A., Brosius, J., and Bachelier, J.P. 2002. RNomics: Identification and function of small, non-messenger RNAs. *Curr. Opin. Chem. Biol.* **6**: 835–843.
- Jones-Rhoades, M.W., Bartel, D.P., and Bartel, B. 2006. MicroRNAs and their regulatory roles in plants. *Annu. Rev. Plant Biol.* **57**: 19–53.
- Karimi, M., Inze, D., and Depicker, A. 2002. GATEWAY vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci.* **7**: 193–195.
- Kastenmayer, J.P., Ni, L., Chu, A., Kitchen, L.E., Au, W.C., Yang, H., Carter, C.D., Wheeler, D., Davis, R.W., Boeke, J.D., et al. 2006. Functional genomics of genes with small open reading frames (sORFs) in *S. cerevisiae*. *Genome Res.* **16**: 365–373.
- Katiyar-Agarwal, S., Morgan, R., Dahlbeck, D., Borsani, O., Villegas Jr., A., Zhu, J.K., Staskawicz, B.J., and Jin, H. 2006. A pathogen-inducible endogenous siRNA in plant immunity. *Proc. Natl. Acad. Sci.* **103**: 18002–18007.
- Katiyar-Agarwal, S., Gao, S., Vivian-Smith, A., and Jin, H. 2007. A novel class of bacteria-induced small RNAs in *Arabidopsis*. *Genes & Dev.* **21**: 3123–3134.
- Kondo, T., Hashimoto, Y., Kato, K., Inagaki, S., Hayashi, S., and Kageyama, Y. 2007. Small peptide regulators of actin-based cell morphogenesis encoded by a polycistronic mRNA. *Nat. Cell Biol.* **9**: 660–665.
- Lim, L.P., Lau, N.C., Garrett-Engle, P., Grimson, A., Schelter, J.M., Castle, J., Bartel, D.P., Linsley, P.S., and Johnson, J.M. 2005. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**: 769–773.
- Lobbes, D., Rallapalli, G., Schmidt, D.D., Martin, C., and Clarke, J. 2006. SERRATE: A new player on the plant microRNA scene. *EMBO Rep.* **7**: 1052–1058.
- Lu, S., Sun, Y.H., Shi, R., Clark, C., Li, L., and Chiang, V.L. 2005. Novel and mechanical stress-responsive MicroRNAs in *Populus trichocarpa* that are absent from *Arabidopsis*. *Plant Cell* **17**: 2186–2203.
- Manavella, P.A., Arce, A.L., Dezar, C.A., Bitton, F., Renou, J.P., Crespi, M., and Chan, R.L. 2006. Cross-talk between ethylene and drought signalling pathways is mediated by the sunflower Habb-4 transcription factor. *Plant J.* **48**: 125–137.
- March-Diaz, R., Garcia-Dominguez, M., Florencio, F.J., and Reyes, J.C. 2007. SEF, a new protein required for flowering repression in *Arabidopsis*, interacts with PIE1 and ARP6. *Plant Physiol.* **143**: 893–901.
- Mendes Soares, L.M. and Valcarcel, J. 2006. The expanding transcriptome: The genome as the “Book of Sand.” *EMBO J.* **25**: 923–931.
- Newton, M.A., Kendziorski, C.M., Richmond, C.S., Blattner, F.R., and Tsui, K.W. 2001. On differential variability of expression ratios: Improving statistical inference about gene expression changes from microarray data. *J. Comput. Biol.* **8**: 37–52.
- Nikovics, K., Blein, T., Peaucelle, A., Ishida, T., Morin, H., Aida, M., and Laufs, P. 2006. The balance between the *MIR164A* and *CUC2* genes controls leaf margin serration in *Arabidopsis*. *Plant Cell* **18**: 2929–2945.
- Numata, K., Kanai, A., Saito, R., Kondo, S., Adachi, J., Wilming, L.G., Hume, D.A., Hayashizaki, Y., and Tomita, M. 2003. Identification of putative noncoding RNAs among the RIKEN mouse full-length cDNA collection. *Genome Res.* **13**: 1301–1306.
- Okazaki, Y., M. Furuno, T., Kasukawa, J., Adachi, H., Bono, S., Kondo, I., Nikaido, N., Osato, R., Saito, H., Suzuki, T., et al. 2002. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **420**: 563–573.
- Osato, N., Yamada, H., Satoh, K., Ooka, H., Yamamoto, M., Suzuki, K., Kawai, J., Carninci, P., Ohtomo, Y., Murakami, K., et al. 2003. Antisense transcripts with rice full-length cDNAs. *Genome Biol.* **5**: R5.
- Ota, T.Y., Suzuki, T., Nishikawa, T., Otsuki, T., Sugiyama, R., Irie, A., Wakamatsu, K., Hayashi, H., Sato, K., Nagai, K., et al. 2004. Complete sequencing and characterization of 21,243 full-length human cDNAs. *Nat. Genet.* **36**: 40–45.
- Prasanth, K.V. and Spector, D.L. 2007. Eukaryotic regulatory RNAs: An answer to the “genome complexity” conundrum. *Genes & Dev.* **21**: 11–42.

- Rajagopalan, R., Vaucheret, H., Trejo, J., and Bartel, D.P. 2006. A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. *Genes & Dev.* **20**: 3407–3425.
- Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B., and Bartel, D.P. 2002. MicroRNAs in plants. *Genes & Dev.* **16**: 1616–1626.
- Riano-Pachon, D.M., Dreyer, I., and Mueller-Roeber, B. 2005. Orphan transcripts in *Arabidopsis thaliana*: Identification of several hundred previously unrecognized genes. *Plant J.* **43**: 205–212.
- Rinn, J.L., Euskirchen, G., Bertone, P., Martone, R., Luscombe, N.M., Hartman, S., Harrison, P.M., Nelson, F.K., Miller, P., Gerstein, M., et al. 2003. The transcriptional activity of human chromosome 22. *Genes & Dev.* **17**: 529–540.
- Sapir, M. and Churchill, G.A. 2000. Estimating the posterior probability of differential gene expression from microarray data. The Jackson Laboratory. <http://research.jax.org/faculty/churchill/pubs/marina.pdf>.
- Sasidharan, R. and Gerstein, M. 2008. Genomics: Protein fossils live on as RNA. *Nature* **453**: 729–731.
- Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Scholkopf, B., Weigel, D., and Lohmann, J.U. 2005. A gene expression map of *Arabidopsis thaliana* development. *Nat. Genet.* **37**: 501–506.
- Stolc, V., Gauhar, Z., Mason, C., Halasz, G., van Batenburg, M.F., Rifkin, S.A., Hua, S., Herreman, T., Tongprasit, W., Barbano, P.E., et al. 2004. A gene expression map for the euchromatic genome of *Drosophila melanogaster*. *Science* **306**: 655–660.
- Storz, G., Opdyke, J.A., and Zhang, A. 2004. Controlling mRNA stability and translation with small, noncoding RNAs. *Curr. Opin. Microbiol.* **7**: 140–144.
- Sunkar, R., Girke, T., Jain, P.K., and Zhu, J.K. 2005. Cloning and characterization of microRNAs from rice. *Plant Cell* **17**: 1397–1411.
- Sunkar, R., Chinnusamy, V., Zhu, J., and Zhu, J.K. 2007. Small RNAs as big players in plant abiotic stress responses and nutrient deprivation. *Trends Plant Sci.* **12**: 301–309.
- Swarup, K., Benkova, E., Swarup, R., Casimiro, I., Peret, B., Yang, Y., Parry, G., Nielsen, E., De Smet, I., Vanneste, S., et al. 2008. The auxin influx carrier LAX3 promotes lateral root emergence. *Nat. Cell Biol.* **10**: 946–954.
- Tenson, T., DeBlasio, A., and Mankin, A. 1996. A functional peptide encoded in the *Escherichia coli* 23S rRNA. *Proc. Natl. Acad. Sci.* **93**: 5641–5646.
- Touchon, M., Arneodo, A., d'Aubenton-Carafa, Y., and Thermes, C. 2004. Transcription-coupled and splicing-coupled strand asymmetries in eukaryotic genomes. *Nucleic Acids Res.* **32**: 4969–4978.
- Vaucheret, H. 2006. Post-transcriptional small RNA pathways in plants: Mechanisms and regulations. *Genes & Dev.* **20**: 759–771.
- Vaucheret, H., Mallory, A.C., and Bartel, D.P. 2006. AGO1 homeostasis entails coexpression of *MIR168* and *AGO1* and preferential stabilization of miR168 by AGO1. *Mol. Cell* **22**: 129–136.
- Wang, J.W., Wang, L.J., Mao, Y.B., Cai, W.J., Xue, H.W., and Chen, X.Y. 2005a. Control of root cap formation by microRNA-targeted auxin response factors in *Arabidopsis*. *Plant Cell* **17**: 2204–2216.
- Wang, X.J., Gaasterland, T., and Chua, N.H. 2005b. Genome-wide prediction and identification of cis-natural antisense transcripts in *Arabidopsis thaliana*. *Genome Biol.* **6**: R30.
- Wang, H., Chua, N.H., and Wang, X.J. 2006. Prediction of trans-antisense transcripts in *Arabidopsis thaliana*. *Genome Biol.* **7**: R92.
- Wang, X., Arai, S., Song, X., Reichart, D., Du, K., Pascual, G., Tempst, P., Rosenfeld, M.G., Glass, C.K., and Kurokawa, R. 2008. Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature* **454**: 126–130.
- Washietl, S., Hofacker, I.L., Lukasser, M., Huttenhofer, A., and Stadler, P.F. 2005. Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome. *Nat. Biotechnol.* **23**: 1383–1390.
- Wernersson, R. and Nielsen, H.B. 2005. OligoWiz 2.0—Integrating sequence feature annotation into the design of microarray probes. *Nucleic Acids Res.* **33**: W611–W615.
- Xiong, L., Wang, R.G., Mao, G., and Koczan, J.M. 2006. Identification of drought tolerance determinants by genetic analysis of root response to drought stress and abscisic acid. *Plant Physiol.* **142**: 1065–1074.
- Yamada, K., Lim, J., Dale, J.M., Chen, H., Shinn, P., Palm, C.J., Southwick, A.M., Wu, H.C., Kim, C., Nguyen, M., et al. 2003. Empirical analysis of transcriptional activity in the *Arabidopsis* genome. *Science* **302**: 842–846.
- Yamashita, A., Watanabe, Y., Nukina, N., and Yamamoto, M. 1998. RNA-assisted nuclear transport of the meiotic regulator Mei2p in fission yeast. *Cell* **95**: 115–123.

Received April 30, 2008; accepted in revised form October 7, 2008.