



**HAL**  
open science

## Landscape of the Noncoding Transcriptome Response of Two Arabidopsis Ecotypes to Phosphate Starvation

Thomas Blein, Coline Balzergue, Thomas Roulé, Marc Gabriel, Laetitia Scalisi, Tracy François, Céline Sorin, Aurélie Christ, Christian Godon, Etienne Delannoy, et al.

### ► To cite this version:

Thomas Blein, Coline Balzergue, Thomas Roulé, Marc Gabriel, Laetitia Scalisi, et al.. Landscape of the Noncoding Transcriptome Response of Two Arabidopsis Ecotypes to Phosphate Starvation. *Plant Physiology*, 2020, 183 (3), pp.1058-1072. 10.1104/pp.20.00446 . cea-02962421

**HAL Id: cea-02962421**

**<https://cea.hal.science/cea-02962421v1>**

Submitted on 19 Nov 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.

# 1                    **Two ecotype-related long non-coding RNAs in the** 2                    **environmental control of root growth**

3    Thomas Blein<sup>1,2,\*</sup>, Coline Balzergue<sup>3,4,5</sup>, Thomas Roulé <sup>1,2</sup>, Marc Gabriel<sup>7</sup>, Laetitia Scalisi<sup>1,2</sup>,  
4    Céline Sorin<sup>1,2</sup>, Aurélie Christ<sup>1,2</sup>, Etienne Delannoy<sup>1,2</sup>, Marie-Laure Martin-Magniette<sup>1,2,6</sup>,  
5    Laurent Nussaume<sup>3,4,5</sup>, Caroline Hartmann<sup>1,2</sup>, Daniel Gautheret<sup>7</sup>, Thierry Desnos<sup>3,4,5</sup> ,  
6    Martin Crespi<sup>1,2,\*</sup>

7    <sup>1</sup>Institute of Plant Sciences Paris Saclay IPS2, CNRS, INRA, Université Paris-Sud,  
8    Université Evry, Université Paris-Saclay, Bâtiment 630, 91405 Orsay, France

9    <sup>2</sup>Institute of Plant Sciences Paris-Saclay IPS2, Paris Diderot, Sorbonne Paris-Cité,  
10    Bâtiment 630, 91405, Orsay, France.

11    <sup>3</sup>CEA, Institut de Biologie Environnementale et de Biotechnologie, Laboratoire de  
12    Biologie du Développement des Plantes, Saint-Paul-lez-Durance, F-13108, France.

13    <sup>4</sup>CNRS, Unité Mixte de Recherche 7265 Biologie Végétale & Microbiologie  
14    Environnementale, Saint-Paul-lez-Durance, F-13108, France.

15    <sup>5</sup>Aix-Marseille Université, Saint-Paul-lez-Durance, F-13108, France.

16    <sup>6</sup>UMR MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay, 75005, Paris, France

17    <sup>7</sup>I2BC, Institute for Integrative Biology of the Cell, CEA, CNRS, Université Paris Sud, 1  
18    avenue de la terrasse, 91198 Gif sur Yvette, France.

19    \*Corresponding author. Correspondence and requests for materials should be addressed  
20    to T. Blein ([thomas.blein@ips2.universite-paris-saclay.fr](mailto:thomas.blein@ips2.universite-paris-saclay.fr)) and M. Crespi  
21    ([martin.crespi@ips2.universite-paris-saclay.fr](mailto:martin.crespi@ips2.universite-paris-saclay.fr))

## 22           **Abstract**

23   **Background:** Root architecture varies widely between species and even between  
24 ecotypes of the same species despite the strong conservation of the protein-coding  
25 portion of their genomes. In contrast, non-coding RNAs evolved rapidly between  
26 ecotypes and may control their differential responses to the environment as several long  
27 non-coding RNAs (lncRNAs) can quantitatively regulate gene expression.

28   **Results:** Roots from Columbia (Col) and Landsberg *erecta* (Ler) ecotypes respond  
29 differently to phosphate starvation. We compared complete transcriptomes (mRNAs,  
30 lncRNAs and small RNAs) of root tips from these two ecotypes during early phosphate  
31 starvation. We identified thousands of new lncRNAs categorized as intergenic or  
32 antisense RNAs that were largely conserved at DNA level in these ecotypes. In contrast  
33 to coding genes, many lncRNAs were specifically transcribed in one ecotype and/or  
34 differentially expressed between ecotypes independently of the phosphate condition.  
35 These ecotype-related lncRNAs were characterized by analyzing their sequence  
36 variability among plants and their link with siRNAs. Our analysis identified 675 lncRNAs  
37 differentially expressed between the two ecotypes including specific antisense RNAs  
38 targeting key regulators of root growth responses. Mis-regulation of several intergenic  
39 lncRNAs showed that at least two ecotype-related lncRNAs regulate primary root  
40 growth in Col.

41   **Conclusions:** The in depth exploration of the non-coding transcriptome of two ecotypes  
42 identified thousands of new lncRNAs showing specific expression in root apices. De-  
43 regulation of two ecotype-related lncRNAs revealed a new pathway involved in the

44 regulation of primary root growth. The non-coding genome may reveal novel  
45 mechanisms involved in ecotype adaptation of roots to different soil environments.

## 46 **Introduction**

47 Over the last decade, genome-wide transcriptomics studies revealed that a large  
48 part of eukaryotic genomes, intergenic to protein-coding genes, was transcribed. These  
49 molecules, globally known as non-coding RNAs [1] may regulate genome expression at  
50 transcriptional, post-transcriptional and epigenetic levels and are generally divided in  
51 small (21 to 24 nt) and long (> 200nt to 100 kb) non-coding RNAs. Plant small RNAs are  
52 processed from longer non-coding transcripts that generally contain a hairpin structure  
53 or lead to double stranded RNA formation. Plant small RNAs include miRNAs,  
54 endogenous siRNAs (both generally 21-22 nt long) and the most abundant  
55 heterochromatin siRNA (hc-siRNAs with a length of 24 nt [2]). On the other hand, long  
56 non-coding RNAs (lncRNAs) are a heterogeneous group of RNA molecules with a coding  
57 capacity shorter than 50 amino acids [3,4]. LncRNA transcripts are generally  
58 polyadenylated (polyA) and can be intergenic (lincRNAs), intronic (incRNAs) or natural  
59 antisense (NATs) with respect of protein coding genes [5,6]. When compared to mRNAs,  
60 lncRNAs are expressed at low levels in a tissue-specific manner or in response to  
61 environmental stresses [7,8] and are more frequently accumulated in the nucleus  
62 relative to the cytoplasm [9]. Recently, some RNA motifs that promote nuclear retention  
63 have been identified in lncRNAs (reviewed in [10]). In human cells, Mukherjee et al [11]  
64 observed that lncRNA showed lower synthesis and higher degradation rates than  
65 protein-coding mRNAs. [11].

66 LncRNAs utilize both *cis* and *trans* modalities of action to regulate gene  
67 expression through interaction with ribonucleoproteins and can form scaffolds and/or  
68 sequester proteins or RNA molecules as decoys or sponges. However, molecular  
69 function has only been identified for a very small proportion of lncRNAs both in animals  
70 and plants. As lncRNA genes lack regions with high primary sequence constraints [9], it  
71 is difficult to use sequence conservation to identify potential functions as it has generally  
72 been done for protein coding genes or housekeeping RNA genes [12,13].

73 Recently, the utilization of new methods (Structure-seq [14] SHAPE-MaP [15]  
74 and icSHAPE [16]) allows the prediction of lncRNA structures *in vivo* (named RNA  
75 structuromes, [17]). It has also been shown that these RNA structures can be modified  
76 under different environmental conditions such as temperature or metabolic changes.

77 It is noteworthy that in 16 vertebrate species, Herzoni et al. [18] have shown that  
78 thousands of lncRNAs, without sequence conservation, appear in syntenic positions and  
79 some present conserved promoters [7,18–20]. In rice and maize, the characterization of  
80 lncRNAs also revealed a higher positional conservation (around 7 times more) than  
81 sequence conservation [21]. In the same way, in Brassicaceae, the fraction of lincRNAs,  
82 for which sequence conservations with *Arabidopsis* could be detected, decreased as the  
83 phylogenetic distance increased [22]. In fact, emerging mechanisms involved lncRNAs in  
84 the regulation of genome expression, however, we cannot exclude that part of lncRNAs  
85 could be simply transcriptional by-products and/or that the sole act of their  
86 transcription rather than their sequence, is the source of the regulation activity [23,24].

87 This last decade, whole-genome variations and evolution in many model species  
88 had been determined through resequencing approaches (from yeast and humans to

89 Arabidopsis, rice and certain crops). These studies provided the characterization of pan-  
90 genomes composed of “core” genomes (present in all accessions) and “dispensable”  
91 genomes (specific to two or more accessions or cultivars or even unique sequences  
92 specific to only one accession). Core genes are frequently highly expressed while  
93 dispensable genes are variably expressed, and generally in a tissue specific manner [25].  
94 The dispensable genomes may play important roles in the capacity of individual  
95 organisms to cope with environmental conditions [26]. Indeed, identification of natural  
96 variations in large worldwide populations (accessions) of *Arabidopsis* ([27] and more  
97 recently in [28]) showed an average of one SNP per 10 bp more frequently located in  
98 intergenic regions than in coding mRNAs [28]; this has also been observed recently in  
99 rice for which only 3,5 % of SNP and 2,5 % of small InDels were located in the coding  
100 region [29]. This latter observation would explain why lncRNAs differ between closely  
101 related species (e.g. rat and mouse, [30], human and chimp, [31] or different plant  
102 species, [32]). Recently, lncRNAs were found to vary, at the individual level, much more  
103 than the protein coding genes [33]. These modifications may drive new regulatory  
104 patterns of gene expression as lncRNA evolved quickly and many of them emerged  
105 recently [31]. Indeed, regulations derived from novel transcription units due to  
106 transposon rearrangements were suggested to be highly variable until they are selected  
107 for a strict regulatory pattern [34].

108         The Col and Ler accessions display different primary root growth and  
109 architecture in their response to phosphate starvation [35]. The identification of LPR1, a  
110 major QTL, has been done in RIL lines obtained by the crosses of accessions presenting  
111 this opposite root response to low Pi [35,36]. Pi deficiency is perceived at the root apex  
112 [36]. When the primary root tip of a Col seedling encounters a low Pi medium, cell

113 elongation in the elongation zone rapidly decreases and cell proliferation in the RAM  
114 progressively ceases [37–40]. By contrast, in *Ler* seedlings elongation and proliferation  
115 of root cells continue, thereby sustaining root growth. In addition, the root tip  
116 concentrate a high proportion of Pi transporters which provide an important  
117 contribution to Pi nutrition and Pi systemic signaling [41,42]. Hence, we decided to  
118 identify and characterize the non-coding transcriptomes of Col and *Ler* root apices  
119 during early phosphate starvation responses to search for lncRNAs potentially linked to  
120 this differential growth response. Interestingly, we identified thousands of new  
121 Arabidopsis lncRNAs, notably in the *Ler* accession, with only a minor fraction being  
122 linked to small RNA production. Several “ecotype-specific” or “-enriched” variants,  
123 highly conserved at DNA level, showed expression variation correlated with changes in  
124 the expression of key regulators of phosphate starvation response. Functional analysis  
125 of 6 lncRNAs in Col revealed two new regulators of primary root growth allowing us to  
126 hypothesize that novel lncRNA expression patterns contribute to the modulation of  
127 environmental responses in different ecotypes.

## 128 **Results**

### 129 ***Columbia and Landsberg root tip transcriptome assemblies***

130 We first reconstructed the transcriptome of root tip of Columbia (Col) and  
131 Landsberg *erecta* (*Ler*) ecotypes. These two ecotypes present contrasting root  
132 phenotypes in response to Pi deficiency [35]. The root growth arrest of the Col ecotype  
133 occurred in the first hours of low phosphate sensing by the root tip [38]. Therefore, in  
134 these two ecotypes, we performed comparative whole genome transcriptomic analyses

135 using paired end sequencing of three biological replicates of root tips during a short  
136 kinetics (0 h, 1 h and 2 h) of low phosphate treatment (10 $\mu$ M, see the “Methods”  
137 section). To discard possible differences related to the *erecta* mutation known to be  
138 present in *Ler* ecotype, we used Col<sup>er105</sup> mutant carrying this mutation in Col. We  
139 obtained between 47M and 65M of reads (Additional file 1: Table S1). For each ecotype,  
140 the reads were independently mapped to their reference genome (Additional file 2: Fig.  
141 S1a): TAIR10 for Col [43] and *Ler* v7 for *Ler* [44]. We selected these two genome  
142 versions since they shared the same gene annotation (TAIR10). For these two genomes,  
143 we predicted new transcripts by comparing our data to those available in TAIR10  
144 annotations. The homology of newly predicted transcripts between the Col and *Ler*  
145 genomes was determined by mapping them on the other genome (TAIR10 and *Ler* v7).  
146 When the newly transcripts overlapped with pre-existing annotations, fusions between  
147 the two transcripts were done to generate the new transcripts. We retained as new  
148 transcripts only RNA molecules of at least 200 nt. The *Ler* v8 version of the *Ler* genome  
149 lacks the official TAIR10 annotation[45]. Therefore, for the transcripts mapping on *Ler*  
150 genome, we retained only the ones that also mapped on this version. Newly identified  
151 transcripts by this pipeline (Additional file 2: Fig. S1a) were then compared with the  
152 ones already described in different Arabidopsis databases: Araport 11 [46], RepTas [8],  
153 CANTATAdb [47], miRBase v21 [48] and with the results of two previous studies  
154 concerning lncRNAs [49,50]. Finally, we used the COME software [51] to determine the  
155 potential coding capacity of new identified transcripts (Additional file 2: Fig. S1a). On the  
156 basis of both database information and COME predictions we classified the  
157 corresponding genes as coding or non-coding.



158           In total, we identified 5313 and 6408 novel putative genes respectively in Col and  
159 *Ler* ecotypes (Fig. 1a; Additional file 1: Table S2; Additional file 3). In root apexes, newly  
160 discovered genes were predominantly non-coding RNAs: 76% for Col and 77% for *Ler* of  
161 total number of new genes (Fig. 1a; Additional file 2: Fig. S1b, c). This suggests that the  
162 coding capacity of the Arabidopsis genome is now well documented in TAIR10 and  
163 Araport11 databases (Additional file 2: Fig. S1b). As expected, non-coding genes were  
164 globally less expressed than coding genes (Additional file 2: Fig. S2a, b). Genes  
165 specifically detected only in one ecotype belong much more to the non-coding class  
166 (>40%) than to the coding one (<8%; Fig. 1b, c). Moreover, the proportion of ecotype-  
167 specific expressed lncRNAs was higher for intergenic lncRNA genes (52%) than for NATs  
168 (34%; Fig. 1d,e). Overall, these results show that at the ecotype level the expression of  
169 non-coding genes is more ecotype specific than for coding genes.

170           We detected a greater number of new genes (essentially non-coding) in the *Ler*  
171 ecotype (Fig. 1a, c). We wondered whether this difference might be due to a technical  
172 bias. Hence, we examined the sequencing saturation of the different libraries (see the  
173 “Methods” section; Additional file 2: Fig. S2c). We noted that in the last 2% of sequencing  
174 reads, less than 10 additional genes were newly detected. We considered then that  
175 sequencing was deep enough to detect the very large majority of expressed genes in  
176 both ecotypes. Hence, the difference of gene detection between Col and *Ler* does not  
177 result from a sequencing bias.

178           We next sought whether these newly detected expressed *Ler* genes (coding or  
179 non-coding) corresponded to specific parts of *Ler* genome that were missed or  
180 rearranged in Col genome and can easily explain expression differences between

181 ecotypes. Out of the 7357 newly identified genes, only 41 and 53 genes, respectively in  
182 Col and *Ler*, coincided with missing DNA sequences in the other ecotype (Fig. 2a)  
183 showing that the DNA sequence of the different new genes is largely conserved. Thus,  
184 the large majority of the differences in transcript accumulation between ecotypes came  
185 from a shift in transcription in one of the two ecotypes of an almost identical DNA region  
186 (except for few SNPs). This change in expression could be due to the deregulation of  
187 some master gene regulator, or to the accumulation of small sequence differences in  
188 gene promoter regions with functional consequences at transcription initiation level or  
189 the consequence of specific differences in epigenetic status in the lncRNA-producing  
190 region due to TE insertions or other rearrangements at large distance from the loci.

#### 191 ***Evolutionary analysis of lncRNA genes expressed in root tips***

192 We then characterized the evolution of Arabidopsis genes expressed in root tips  
193 profiting from the extensive sequence information in Arabidopsis accessions [28].  
194 According to current annotations, genes were classified as non-NAT (no gene on the  
195 other strand) or NAT (presence of a gene on the other strand). First at the level of the  
196 *Arabidopsis thaliana* species, we calculated the rate of SNP accumulated in each gene  
197 among all the different ecotypes (Fig. 2b). As expected, transposable elements (TE)  
198 accumulated much more SNPs than coding genes, whereas non-NAT lncRNAs and  
199 structural RNA genes showed an intermediate level of SNPs between TEs and coding  
200 genes. By contrast, the level of SNPs was generally similar for NAT lncRNA and coding  
201 genes as the major part of NAT lncRNA genes are antisense of coding genes which are  
202 under clear selection pressure.

203           In a second step, to investigate sequence evolution at a larger scale, we used the  
204 PhastCons score that represents an inter-species level of nucleotide conservation  
205 (normalized between 0 and 1) according to the alignment of 20 Angiosperm genomes  
206 (Fig. 2c, [52]). As expected, structural RNA genes were strongly conserved (median  
207 PhastCons score of 1), while transposable elements were not (median PhastCons score  
208 of 0). Coding genes present a score between these two extremes (around 0.5).  
209 Interestingly, non-NAT lncRNA genes showed an intermediate score between coding and  
210 transposable genes (median PhastCons score around 0.3) whereas NAT lncRNA genes  
211 had again the same degree of conservation than coding genes. These observations  
212 further suggest that the sequence of the NAT lncRNA genes are strongly constrained,  
213 likely due to the strong selection pressure on their overlapping coding genes, whereas  
214 intergenic non-coding genes allow more variability even though they are more  
215 constrained than TEs.

### 216           ***Few lncRNA transcripts co-localize with small RNA generating loci***

217           In animals and plants, some lncRNA loci co-localize with genomic regions  
218 producing small RNA molecules [53,54]. Therefore, we asked whether lncRNAs loci  
219 expressed in root tips could generate small RNAs. Small RNA sequencing was done on  
220 similar samples used for the long RNA sequencing and the identification of small RNAs  
221 was conducted independently on each genome (TAIR10 and Ler v7; see the “Methods”  
222 section). Then, small RNAs were mapped on genes. The majority of the lncRNAs  
223 containing small RNAs generated non-phased molecules of 21/22nt or 24nt (Additional  
224 file 2: Fig. S3a, b). A minor proportion of lncRNAs overlapped with phasiRNAs or were  
225 likely miRNA precursors (Additional file 2: Fig. S3a, b).

226 We then analyzed the potential link between small RNAs and lncRNAs in each  
227 ecotype. The majority of lncRNAs did not accumulate siRNAs (6452 genes out of 7850  
228 detected lncRNAs) that were specifically observed in one ecotype. Many of them also did  
229 not generate any siRNA on either ecotype (2688 genes out of 3110 ecotype specific  
230 detected genes; Fig. 3a, long in Col and ND in *Ler* or *vice versa*). Thus, the differential  
231 presence of lncRNA between ecotypes could not be linked to a change in siRNA  
232 production from the encoding loci.

233 We then wondered whether different small RNA processing from lncRNAs could  
234 occur between the two ecotypes. We first looked at lncRNA that accumulated small RNA  
235 only in one accession: (i) lncRNAs that could generate siRNAs in only one ecotype while  
236 detected as lncRNAs in both ecotypes (long in *Ler* and long plus small in Col and *vice*  
237 *versa*, respectively 113 and 240 genes) and (ii) loci that produced only siRNAs in one  
238 ecotype and only detected as lncRNAs in the other one (long in *Ler* and small in Col or  
239 *vice versa*, respectively 57 and 47 genes). It is known that 21/22nt siRNAs act on gene  
240 transcripts whereas 24nt siRNAs mediate chromatin modifications [53,54]. Thus,  
241 alterations in the size of specific siRNAs, in one ecotype, could indicate a modification of  
242 the type of gene regulation: post-transcriptional (21/22nt) or epigenetic (24nt) in the  
243 other ecotype. Therefore, we analyzed the major small RNA size that accumulated on  
244 lncRNA genes. Among the lncRNA genes accumulating siRNA, a large portion  
245 accumulated the same size in both ecotypes (591 for 21/22nt siRNA and 391 for 24nt  
246 siRNA) or produced siRNA in one ecotype (367 for 21/22nt and 316 for 24nt, Fig. 3b).  
247 Only 29 lncRNA genes accumulated a different size of siRNAs between the two ecotypes  
248 among the 1694 lncRNA genes accumulating siRNAs. Therefore, no major change of

249 reciprocal posttranscriptional or transcriptional regulation of lncRNA by small RNAs  
250 could be established between ecotypes.

251 We finally investigated the specificity of detection of siRNAs between the two  
252 ecotypes. First, we studied lncRNA genes predicted to produce phased 21/22nt siRNAs.  
253 Among the 7 predicted lncRNAs, only 2 were specific to *Ler* (Additional file 2: Fig. S3d).  
254 Second, we searched for miRNA loci and 23 and 12 miRNAs of the 191 detected miRNAs  
255 were specifically detected respectively in *Col* and *Ler* (Additional file 2: Fig. S3c), a  
256 proportion related to the variation detected for protein-coding genes. Third, we  
257 analyzed the proportion of specific expression for the vast majority of 21/22nt and 24nt  
258 siRNAs (Fig. 3c, d) located in the different genome annotations (coding or non-coding).  
259 Altogether, the *Ler* ecotype produces a larger number of 21-22nt siRNAs specifically  
260 linked to this ecotype whereas *Col* is more enriched for ecotype-specific 24nt siRNAs,  
261 potentially suggesting links with differential post-transcriptional and epigenetic  
262 regulations among ecotypes.

263 Nevertheless, the major difference in the non-coding transcriptome of the two  
264 ecotypes was linked to lncRNAs and not associated to small RNAs, even though in  
265 certain cases small RNAs may be involved in ecotype-specific regulations.

266 ***Differential accumulation of transcripts between ecotypes along a low***  
267 ***phosphate kinetic***

268 The root growth arrest of *Col* ecotype occurred in the first hours of low  
269 phosphate sensing by the root tip [38] whereas *Ler* continues its growth. To determine  
270 the effect of a short kinetics of phosphate deficiency, we examined gene expression  
271 patterns in the two ecotypes in response to this stress. Principal component analysis

272 (PCA) showed a data dispersion that allowed a clear distinction between the effect of  
273 ecotype (first axis, Additional file 2: Fig. S4a) and the effect of the kinetics (second axis,  
274 Additional file 2: Fig. S4a). Thus, we used a multifactor analysis that takes into account  
275 the ecotype, the kinetics, their interaction and the replicate to investigate differential  
276 gene expression. Coding and non-coding genes had comparable dispersion in our  
277 experiments. Therefore we were able to use both types of gene in the same analysis. For  
278 each comparison, we confirmed the distribution of p-value as a criterion of robustness of  
279 the test [55]. After processing the differential analyses, we interpreted the results by  
280 separating the genes as “coding” or “non-coding” using the categories defined  
281 previously.

282 For coding genes, we observed that 3321 genes were differentially expressed  
283 between the two ecotypes on average over the 3 time points of the kinetic and 2504  
284 were differentially expressed between at least two points of the kinetic on average over  
285 the 2 ecotypes (Fig. 4a; Additional file 1: Table S3). In our experiment, the number of  
286 differentially expressed coding genes between ecotypes or along the stress kinetics was  
287 of the same order of magnitude. However the response to phosphate starvation of only  
288 55 genes was significantly impacted by the ecotype (“interaction” of both factors, Fig.  
289 4a). Respectively 1566 and 1749 coding genes were up-regulated in Col and in Ler on  
290 average over the 3 time points of the kinetic (Additional file 2: Fig. S4b). Interestingly, a  
291 clear bias of expression between ecotypes could be observed for non-coding genes (Fig.  
292 4b). Indeed, 675 (666 + 6 + 2 + 1) non-coding genes were differentially expressed  
293 between the two ecotypes on average over the 3 time points of the kinetic while only 70  
294 (61 + 6 + 2) were differentially expressed along at least one point of the kinetic. This  
295 bias could be correlated with the number of non-coding genes specifically detected in

296 the two ecotypes (Fig. 1a). Comparable biases were observed for both classes of non-  
297 coding genes, lincRNAs and NATs (Fig. 4c, d). Globally, 146 lincRNAs and 236 NATs were  
298 significantly up-regulated in Col compared to *Ler* and 106 lincRNAs and 187 NATs in *Ler*  
299 compared to Col (Additional file 2: Fig. S4c).

300 We used qRT-PCR in independent replicates of Col, Col<sup>er105</sup> and *Ler* to confirm the  
301 differential expression of 14 intergenic lincRNA genes (7 in Col and 7 in *Ler*) previously  
302 identified in our statistical analysis. We were able to confirm the differential expression  
303 of 12 lincRNAs (Fig. 4 e, f). Globally Col and Col<sup>er105</sup> (used for the RNA sequencing data)  
304 showed similar expression levels despite minor differences for some lincRNAs. To  
305 investigate any dominant expression effect from one ecotype, we investigated the level  
306 of expression of these lincRNAs in the F1 offsprings of Col and *Ler* crosses. Among the 12  
307 differentially expressed genes, an intermediate level of repression has always been  
308 detected (8 were statistically significant; Fig. 4e, f). This suggests independent  
309 regulation of lincRNAs between the parental genomes and discards major dominant  
310 “trans” regulatory effects of lincRNA expression between genomes.

311 One interesting possibility is that specific lincRNAs may be expressed in the Col  
312 and *Ler* genomes in relation with known regulators of root growth responses triggered  
313 by phosphate starvation. Indeed, we could identify 2 specific antisense lincRNAs in *Ler*  
314 complementary to the phosphate transporter *AT5G43370/PHT1.2* which itself is  
315 differentially expressed between ecotypes (Additional file 2: Fig. S5a). Furthermore, a  
316 Col expressed NAT RNA is complementary to *SPX4*, a critical regulator of phosphate  
317 responses, that shows reduced expression in Col compared to *Ler* (Additional file 2: Fig.  
318 S5b). In other cases, we observed that two consecutive coding transcripts showing

319 differential levels of expression among ecotypes flank an intergenic lncRNA with an  
320 ecotype-specific expression pattern (Additional file 2: Fig. S5c), suggesting that various  
321 cis-effects may be involved in these differential ecotype-linked expression patterns.

### 322 ***Differential accumulations of small RNAs between ecotypes***

323 The differential accumulations of small RNAs of 21/22nt and 24nt have been also  
324 examined in each ecotype and during the kinetics response. PCA of these sequencing  
325 data separated again clearly small RNA abundance between ecotypes but not at the level  
326 of the kinetics response (Additional file 1: Table S3; Additional file 2: Fig. S6a, b). We  
327 identified 416 coding and 211 non-coding genes that accumulated 21/22 nt siRNAs  
328 differentially between ecotypes on average over the 3 time points of kinetic with  
329 generally more siRNAs in *Ler* (298 coding genes, 83 lincRNAs and 40 NATs) than in *Col*  
330 (118 coding genes, 49 lincRNAs and 39 NATs; Additional file 1: Table S3; Additional file  
331 2: Fig. S6d). A greater number of genes differentially accumulated 24 nt siRNAs co-  
332 localizing with 1149 coding-genes and 429 non-coding genes between ecotypes on  
333 average over the 3 time points of kinetic, again with a greater number of differentially  
334 up-regulated siRNAs in *Ler* (758 coding genes, 189 lincRNAs and 67 NATs) compared to  
335 *Col* (391 coding genes, 104 lincRNAs and 69 NATs; Additional file 2: Fig. S6e).

336 Concerning miRNA, the only difference that could be analyzed according to the  
337 PCA was also between ecotypes (Additional file 2: Fig. S6c). Indeed, among the 240  
338 miRNAs indexed in miRBase, 38 were differentially expressed between the two ecotypes  
339 (15 and 23 for *Col* and *Ler* respectively, Additional file 2: Fig. S6f, Additional file 1: Table  
340 S3). Interestingly, the families of miR399 and miR397 were specifically accumulated in  
341 the *Ler* ecotype. These miRNAs target the PHOSPHATE 2 (PHO2) and NITROGEN



342 LIMITATION ADAPTATION (NLA) transcripts. The PHO2 and NLA proteins are known to  
343 act together to allow the degradation of the phosphate transporter PT2 (Pht1;4) [56]. In  
344 *Ler*, the higher amount of miR399 and miR397 would lead to a lower level of PHO2 and  
345 NLA and therefore a higher level of PT2 protein and could have consequently increased  
346 Pi uptake, even if the numerous post translational regulations affecting Pi transporters  
347 could limit such impact [57–60]. However, in our experiments, no difference in the  
348 accumulation of these three gene transcripts has been detected between ecotypes (as  
349 previously reported [60,61]). This suggests that the promoter activity of *PHO2* and *NLA*  
350 could enhance the transcription of these genes to compensate the increased  
351 accumulation of these miRNAs in *Ler*.

### 352 ***Mis-regulation of lncRNA expression affects primary root growth in Col***

353 The different patterns of gene expression of lncRNAs between ecotypes may be  
354 relevant for the generation of novel regulatory patterns linked to root growth responses.  
355 As the interactions of lncRNAs with different ribonucleoproteins may lead to changes in  
356 gene expression potentially linked to the differential growth responses between *Col* and  
357 *Ler*; we selected 5 lncRNA genes: *NPC15*, *NPC34*, *NPC43*, *NPC48* and *NPC72* showing  
358 differential expression among ecotypes in order to study the impact of their expression  
359 on *Col* primary root growth. In our RNA-Seq data, three of these lncRNA genes were  
360 more expressed in *Col* (*NPC15*, *NPC43* and *NPC72*) and two in *Ler* (*NPC34* and *NPC48*)  
361 (Fig. 5a). To evaluate the potential dominant expression patterns of these lincRNA  
362 genes, we investigated their expression in corresponding two F1 of reciprocal crosses  
363 *Col* x *Ler*. Expression analysis by qRT-PCR confirmed the previous RNA-Seq expression  
364 levels for *NPC15*, *NPC34* and *NPC72* genes (Fig. 5b) whereas for *NPC48*, the differential

365 expression was only detected in the Col<sup>er105</sup> mutant (the line used for the RNA-Seq  
366 experiment), suggesting that the *erecta* mutation directly affects the expression of this  
367 gene. No differential expression could be detected for the *NPC43* gene. This latter result  
368 might be due to the accumulation of antisense transcripts (*NPC504*) that could generate  
369 siRNAs against *NPC43* transcripts. One explanation for the differential expression of  
370 lincRNA genes in these two ecotypes could be linked to genetic changes at the transcript  
371 locus. Hence, we examined the DNA sequences at these specific loci (Additional file 2:  
372 Fig. S7; profiting from the well characterized Col and *Ler* genomes). No significant  
373 modifications (except few SNPs) were detected between Col and *Ler* for the *NPC34*,  
374 *NPC43* and *NPC48* loci. By contrast, the *NPC15* locus contains an insertion of 2417 nt in  
375 *Ler* v8 and *NPC72* is completely missing in *Ler* v7 and v8 genomes (Additional file 2: Fig.  
376 S7a-e). Therefore, genome modifications could explain the specific expression pattern of  
377 *NPC72* and *NPC15* genes in *Ler*.

378 To support the biological relevance of lincRNAs differentially regulated between  
379 ecotypes at phenotypic level, we used T-DNA insertion and over-expressing (OE) lines to  
380 monitor the effect of lincRNA genes on Col root growth in control and low phosphate  
381 conditions. In control conditions, only *NPC48* and *NPC72* overexpression lead to a  
382 significant root growth reduction when compared to Col (Fig. 5c). Indeed, T-DNA  
383 insertions have been mapped to the 5' region of the *NPC48* and *NPC72* loci respectively  
384 for *npc48* T-DNA and *npc72* T-DNA and lead to an overexpression of these lincRNAs in  
385 the respective lines (Fig. 5d). For *npc48* T-DNA, it strongly supported the phenotype  
386 observed with the 35S::*NPC48* lines. In low phosphate conditions, known to inhibit root  
387 growth in Col but not in *Ler*, minor differences in root length were observed. However,  
388 the OE of *NPC43* gene clearly increased the inhibition of root growth (Fig. 5c) when

389 compared to Col plants. *A priori*, the ratio of root growth in control and low phosphate  
390 conditions should highlight potential phenotypic differences in phosphate sensitivity of  
391 transgenic lines. This ratio is significantly increased for *NPC48* and *NPC72* lines  
392 compared to Col, partly mimicking a *Ler*-type response. On the other hand, this ratio is  
393 significantly reduced for *NPC43*- and *NPC15*-deregulated lines (Fig. 5c).

394 *NPC48* and *NPC72* de-regulated lines present a significant decrease of root length  
395 in control, but not in low phosphate conditions. The 500  $\mu$ M of Pi used in our control  
396 conditions could be perceived as a low Pi concentration in these lines. Hence, we asked  
397 whether these phenotypes could be linked to a root growth arrest due to oversensitive  
398 perception of phosphate starvation under control conditions or an alteration of Pi  
399 systemic sensing (which would affect Pi uptake). This does not seem the case as i)  
400 known phosphate-starvation markers were not induced in root in control conditions  
401 and ii) these markers were induced in these lines at the same level as in Col (Additional  
402 file 2: Fig. S8a, b). Then we investigated the local Pi signaling response which control  
403 primary root growth. As the genes *LOW PHOSPHATE ROOT 1* and *2* (*LPRs*) and the  
404 transcription factor *STOP1* are known to be involved in primary root growth arrest  
405 under low Pi [38] we analyzed the expression of *NPC48* and *NPC72* genes in *lpr1/lpr2*  
406 and *stop1* mutant lines. Fig. 5e shows that we cannot observe any significant variation of  
407 these lncRNA expression patterns in these mutant lines. Reciprocally, no significant  
408 expression variation was detected of the *LPR1/LPR2* pathway in *NPC48* or *NPC72* lines  
409 (*LPR1*, *LPR2*, *STOP1*, *ALMT1* and *MATE* genes; Additional file 2: Fig. S8c-g).

410           Altogether, the identification of ecotype-related lncRNAs allowed us to  
411 characterize new regulators of primary root growth acting through a distinct pathway  
412 from that involving *LPR1/LPR2* and *STOP1*.

## 413           **Discussion**

414           Until recently, transcriptome studies were mainly focused on protein coding gene  
415 transcripts and discarded lncRNAs. Variation in the nucleotide sequences or expression  
416 patterns of the non-coding genome can have less pleiotropic effects than changes in the  
417 protein sequence of critical regulators. Therefore, in addition to promoters, introns and  
418 transposons, regions encompassing non-coding RNAs emerge as actors of plant  
419 adaptation to environmental constraints. It is now clear that certain lncRNAs play  
420 important roles in development and response to environmental conditions [1], even if  
421 the majority of lncRNAs awaits a functional characterization. In addition, lncRNAs  
422 exhibit a high cell specificity which could explain specific functions in a particular cell-  
423 type in plants and animals. In this study, by using strand specific RNA analysis on root  
424 tips, we succeeded in identifying thousands of different forms of lncRNAs: lincRNAs,  
425 NATs or antisense RNAs expressed from two *Arabidopsis* accessions, *Col* and *Ler*. As  
426 previously shown in numerous studies, in comparison with mRNAs, lncRNAs observed  
427 here were expressed at low levels [62,63]. We opted to focus on root growth as it is a  
428 complex trait resulting of the expression of a large number of loci spread across the  
429 plant genome and is highly susceptible to the soil environment [64]. Interestingly, in our  
430 study, we observed many lncRNA genes differentially or specifically expressed in *Col* or  
431 *Ler*, in contrast to protein-coding genes for which the expression were more stable in

432 the two ecotypes. We think reasonable to assume that certain of these lncRNAs could  
433 contribute to *Col* and *Ler* root growth specificities. Indeed, in animals, certain complex  
434 traits are mainly driven by non-coding variants [65,66]. In chicken, domestication traits  
435 governing body morphology and behavior are under selection and correspond to  
436 lincRNA genes [67]. Moreover, recently, the human LincSNP 2.0 database, which  
437 contains 809451 unique disease associated SNPs, 11,6 million of linkage disequilibrium  
438 SNPs and 244 545 lncRNAs, identified approximately 45% of disease-associated human  
439 SNPs that mapped to non-coding regions of the genome [68]. Some lncRNAs containing  
440 SNPs have been recently associated to cardiometabolic traits [69]. Similarly, in plants,  
441 the comparison of SNPs observed in fruit transcripts of two tomato cultivars also  
442 corresponded to non-coding genomic regions or lncRNA genes [70]. The SNPs could act  
443 directly at the level of lncRNA expression or affect the expression of lncRNA-neighboring  
444 genes [24,71]. lncRNAs are thus elements to be considered in genetic association  
445 studies.

446         Root apices play an important role in sensing external stimuli. We examined the  
447 gene expression profile soon after the stress application (1 h and 2 h) in two ecotypes  
448 that present different root growth phenotypes in response to phosphate starvation. The  
449 number of coding genes differentially expressed during the phosphate kinetic and  
450 between the two ecotypes was of the same order of magnitude. A clear bias of specific  
451 expression of lncRNA genes was identified between the two ecotypes. This ecotype bias  
452 was also observed for the accumulation of siRNA. The analysis of pan-genome  
453 (restricted to coding genes) using genomic and RNAseq of 19 *Arabidopsis* ecotypes  
454 showed that at least 70 accessory genes could be identified in each ecotype [25]. In  
455 response to stress, it is known that accessory genes can explain, at least, part of the

456 phenotypic difference of behavior observed among ecotypes [44]. In our experiments,  
457 the expression polymorphism (differential gene expressions observed between Col and  
458 *Ler*) corresponded mainly to regions where SNPs accumulate or neighboring DNA  
459 rearrangements (e.g. transposon insertions) were detected. Very few ecotype-specific  
460 lncRNAs coincided with the absence of specific DNA sequences in the particular ecotype.  
461 Hence, we propose that the lncRNA difference of expression of a relatively similar DNA  
462 molecule would result in shifts in transcription or stability of lncRNAs that could be  
463 connected to SNP or InDels polymorphisms in promoters and/or lncRNA gene  
464 sequences in the two ecotypes. As lncRNAs can repress or activate the transcription of  
465 other genes the expression polymorphisms observed between the two ecotypes could  
466 also result in a cascade of *cis*-local or *trans*-distal action on target genes [72,73]. It is  
467 noteworthy that the majority of ecotype-specific lncRNAs identified did not co-localize  
468 with siRNAs and thus could not reflect only gene silencing differences between ecotypes  
469 (either transcriptional or post-transcriptional processes, [53,54]). This points to the  
470 lncRNA itself or its transcription *per se* to be linked to the quantitative regulation of  
471 target gene expression [73].

472 We were able to confirm, by qRT-PCR, the expression of 12 lincRNAs among the  
473 14 chosen for validation, supporting the expression polymorphisms identified. Allele-  
474 specific expression is known to affect specific diseases in human and productivity in  
475 plant and animal agriculture [74,75]. Moreover, in Arabidopsis, heterosis has been  
476 reported for different traits such as flowering time [76] leaf area and plant biomass [77]  
477 but also for phosphate acquisition [78]. As lncRNAs are able to modify chromatin and  
478 thus alter gene expression, we added, in our expression analysis, F1 resulting from  
479 reciprocal crosses between Col and *Ler*. In the F1 offspring, the 15 confirmed

480 differentially expressed lncRNA genes chosen for validation exhibited globally an  
481 additive expression pattern when compared to their parents. This is consistent with  
482 results obtained in maize F1 hybrids, where additivity are frequently observed for  
483 lncRNAs [79] and 78% of coding genes [80]. In the case of non-additive expression  
484 patterns in the heterozygote, a *trans*-regulation could affect the two alleles and products  
485 a specific accumulation of lncRNA transcripts.

486         The general adaptation of root architecture in response to low Pi comprises an  
487 arrest of PR growth and the activation of LR development after the perception of Pi  
488 limitation. In Arabidopsis, studies concerning plant phosphate homeostasis, during Pi  
489 deficiency, characterized the IPS1/AT4 lncRNA controlling the distribution of Pi from  
490 root to shoot. It acts as a target mimic for miR399, which regulates *PHO2* mRNAs  
491 [81,82]. Moreover, Yuan et al [62] identified lncRNAs differentially expressed in root and  
492 shoot of plants grown in the presence or absence of Pi for 10 days. The authors  
493 suggested that a co-expression between lncRNAs and adjacent coding genes may be  
494 linked to a *cis*-regulation by lncRNAs of target genes involved in Pi starvation processes.  
495 Interestingly, in fission yeast *S. pombe*, two of the three genes of the phosphate regulon  
496 are repressed, in Pi rich medium, by the transcription of lncRNA genes [83,84] that are  
497 present in the 5' region (*cis*-regulation). The molecular mechanisms that govern root  
498 growth modification by Pi have been mostly elucidated in Col plants. For the local  
499 impact of Pi (restricted to root architecture), Pi deficiency is sensed by the root tips and  
500 a primary root growth inhibition is induced by both the reduction of cell elongation and  
501 the progressive arrest of meristem division, notably linked to the presence of iron in the  
502 medium. The pathway that governs this meristem arrest involves the STOP1  
503 transcription factor that controls the expression of the transporter ALMT1 involved in

504 malate secretion [38]. In the apoplasm, malate leads in turn to the production, by  
505 LPR1/LPR2, of ROS that induces plasmodesmata (PD) closure by callose deposition [40].  
506 The interruption of trafficking through PD progressively blocks meristem division. In  
507 parallel, in low Pi, it has been shown recently that the small peptide CLE14, which acts  
508 after LPR1/LPR2 module, is able to trigger the differentiation of root meristem in  
509 absence of callose deposition [37]. We selected several ecotype-specific lncRNAs for  
510 functional root growth analysis and misregulation of two lncRNAs affected primary root  
511 growth. Expression analysis in response to Pi deficiency in our mutant lines *NPC48* and  
512 *NPC72*, allowed us to suggest the existence of a potential new pathway that do not  
513 overlap with the known LPR1/LPR2 and STOP1 pathways. In Arabidopsis, by using  
514 grafts between ecotypes presenting a high frequency of SNPs, Thieme et al [85] have  
515 shown that about 2000 mRNAs, among 9300 containing SNPs, could move in plants that  
516 were subjected to Pi deficiency for two weeks. These mRNAs were transported from  
517 root-to-shoot or shoot-to-root. The authors suggested that these mobile mRNAs might  
518 function widely as specific signaling molecules coordinating growth, cell differentiation  
519 and stress adaptation of distant body parts. As the lncRNAs described here have 3' polyA  
520 tails and are probably 5' capped, it is tempting to assume that at least some of them can  
521 be transported through the xylem and/or the phloem and may contribute to systemic  
522 signaling responses.

523 Globally, the in depth exploration of the non-coding transcriptome of two ecotypes  
524 identified thousands of new lncRNAs with ecotype-specific expression. Statistical  
525 analysis among ecotypes identified several co-regulated events between coding and  
526 non-coding genes (including small RNAs) potentially linked to the evolution of different  
527 regulatory mechanisms among ecotypes grown in diverse soil environments. De-



528 regulation of two ecotype-related lncRNAs revealed a new pathway involved in the  
529 regulation of primary root growth.

## 530 **Methods**

### 531 ***Plant growth.***

532 Seeds were surface-sterilized seeds were sown on a horizontal line in plates that  
533 were vertically disposed in a growing chamber (16h photoperiod; intensity 90  $\mu$ E;  
534 21 °C). The growth medium contained 0.15 mM MgSO<sub>4</sub>, 2.1 mM NH<sub>4</sub>NO<sub>3</sub>, 1.9 mM KNO<sub>3</sub>,  
535 0.34 mM CaCl<sub>2</sub>, 0.5  $\mu$ M KI, 10  $\mu$ M FeCl<sub>2</sub>, 10  $\mu$ M H<sub>3</sub>BO<sub>3</sub>, 10  $\mu$ M MnSO<sub>4</sub>, 3  $\mu$ M ZnSO<sub>4</sub>, 0.1  $\mu$ M  
536 CuSO<sub>4</sub>, 0.1  $\mu$ M CoCl<sub>2</sub>, 0.1  $\mu$ M Na<sub>2</sub>MoO<sub>4</sub>, 0.5 g·L<sup>-1</sup> sucrose. The agar (0.8 g·L<sup>-1</sup>) for plates  
537 was from Sigma-Aldrich (A1296 #BCBL6182V). The -Pi and +Pi agar medium contained  
538 10 and 500  $\mu$ M Pi, respectively; the media were buffered at pH 5.6-5.8 with 3.4 mM 2-  
539 (N-morpholino)ethane sulfonic acid.

### 540 ***Arabidopsis lines***

541 The *stop1* (SALK\_114108, NASC reference N666684), *lpr1;lpr2* [36], npc15 T-DNA  
542 1 (SALK\_027817, NASC reference N527817) npc15 T-DNA 2 (SALK\_090867, NASC  
543 reference N590867), npc43 T-DNA (SALK\_007967, NASC reference N507967), npc48 T-  
544 DNA (SAIL\_1165\_H01, NASC reference N843057) and npc72 T-DNA (SAIL\_571\_C12,  
545 NASC reference N824316) are in the Col-0 (Col) background. Surexpressor lines  
546 35S:NPC43, 35S:NPC48 1, 35S:NPC48 2, 35S:NPC48 3 were retrieved from [49] and are  
547 in Col-0 background. npc34 T-DNA (FLAG\_223D08 or FLAG\_228A07) is in WS

548 background. Col<sup>er105</sup> is in a Columbia background (Col-0) with the null allele *erecta-105*  
549 [86]

550 ***Libraries construction and sequencing.***

551 Three biological replicates of Col<sup>er105</sup> and *Ler* were sown vertically on 1 cm high  
552 bands of nylon membrane (Nitex 100 µm). After one week on +Pi agar medium, while  
553 the roots were out of the membrane, the membranes were transferred on -Pi agar  
554 medium. Plants were then sampled at time point 0, 1 h and 2 h after transfer. Each  
555 biological replicate is a pool of more than 100 root apices cut at 0.5cm from root  
556 extremity. Total RNA were extracted using RNeasy micro kit (Qiagen 74004, n°lot  
557 136257409), tissues protocol. One microgram of total RNA from root tips of each sample  
558 was used for mRNA library preparation using the Illumina TruSeq Stranded mRNA  
559 library preparation kit according to the manufacturer instruction. Libraries were  
560 sequenced on HiSeq™ 2000 Sequencing System (Illumina) using 100 nt paired-end  
561 reads. Samples were multiplexed by 6 and sequenced on 4 sequencing lines. All reads  
562 were quality trimmed using Trimmomatic and remaining ribosomal sequences were  
563 removed using sortMeRNA [87].

564 For small RNAs libraries, root apices grown and collected in the same condition  
565 as for the RNA-Seq were used. Small RNAs were extracted using *mirVana*™ miRNA  
566 Isolation Kit (Ambion AM 1560, part IV/A. Isolation of Small RNAs from Total RNA  
567 Samples). Small RNA libraries were construct using Ion Total RNA-Seq Kit v2 (Ion  
568 Torrent, Life Technologies) according to manufacturer instruction. Libraries were then  
569 sequenced using IonProton and the adapters removed.

## 570 ***New transcript identification***

571 According to their ecotype of origin, mRNA cleaned reads were aligned on  
572 TAIR10 [43] or *Ler* v7 [44] genome using tophat2 (version 2.0.13, [88]) with the  
573 following arguments: `--max-multihits 1 --num-threads 8 -i 20 --min-segment-`  
574 `intron 20 --min-coverage-intron 20 --library-type fr-firststrand --`  
575 `microexon-search -I 1000 --max-segment-intron 1000 --max-coverage-intron`  
576 `1000 --b2-very-sensitive`. Independently of ecotype, new transcripts were predicted  
577 using GFFprof included in RNAprof [89]

578 The transcripts predicted on each genome, TAIR10 and *Ler* v7, were positioned  
579 on the other genome, respectively *Ler* v7 and TAIR10, and on *Ler* v8 [45] using blastn  
580 (from BLAST suite 2.2.29+) using a maximum e-value of  $10^{-4}$ . For each transcript, the  
581 different blast hits fragments were fused together if the distance between two fragments  
582 was less than 5000 nucleotides and placed on the same strand of the chromosome. Only  
583 hits with at least 90 % of sequence identity and where the length was conserved (at least  
584 90 % and less than 110 % of length outside of insertion) were kept. For each transcript  
585 only the best hit was conserved according first to the conservation of the sequence  
586 length and then identity. In case of hits of the same strength, a higher priority was given  
587 when the chromosome and then the strand were conserved. Each transcript was  
588 therefore placed on each of the three genomes.

589 For *Ler*-predicted transcripts only those positioned on *Ler* v8 were kept. On each  
590 genome independently, transcripts coming from the same ecotype (GFFprof prediction)  
591 or the other one (blastn positioning) were fused using cuffmerge (version 1.0.0) with

592 default parameters. Only transcripts longer than 200nt in either ecotype were kept for  
593 further processing.

594 Based on the position of the transcripts on the TAIR10 genome, new transcripts  
595 were annotated according to already known transcripts in the following databases:  
596 Araport 11 [46], RepTas [8], CANTATADB [47], miRBase v21 [48] lncRNAs predicted  
597 from Ben Amor et al. [49] and root predicted lncRNAs from Li et al. [50]. GffCompare  
598 (version 0.10.4, <https://ccb.jhu.edu/software/stringtie/gffcompare.shtml>) was used for  
599 the comparison. In case of overlap with a known transcript (=, c, k, j, e, o codes of  
600 GffCompare), the closest transcript was used to determine the identification and the  
601 coding potential of the transcript. For previously non-discovered transcripts we used  
602 the COME software [90] to predict their coding potential.

### 603 ***Library saturation analysis***

604 The saturation of libraries was computed using the `RPKM_saturation.py` script  
605 included in RSeQC v3.0.0 [91,92]. Raw counts were quantified at gene level from  
606 resampled bam for each additional 2% of the reads from 2% to 100% of the reads (`-l 2`  
607 `-u 100 -s 2`). The saturation was computed 100 times. For each library and each  
608 sampling the number of detected gene (known or new and coding or non-coding) was  
609 defined as the number of genes having at least one read.

### 610 ***Small RNA analysis***

611 The cleaned small RNA reads were aligned on TAIR10 or Ler v7 genome using  
612 ShortStack (version 3.8.5, [93]) without mismatch (`--mismatches 0`), keeping all

613 primary multi-mapping (`--bowtie_m all`) and correcting for multi-mapped reads  
614 according to the uniquely mapped reads (`--mmap u`).

615 For each annotation in Araport11 (mRNA coding and non-coding and TE) and  
616 each new annotation predicted in this study according to mRNA sequencing the  
617 accumulation of small RNA was analysed using ShortStack with default parameters. The  
618 counts for 21nt and 22nt were summed for each sample. Using these new counts, the  
619 DicerCall, defined as the size of the majority of the reads of a cluster, was recomputed.  
620 The other description and counting are according to ShortStack prediction.

### 621 ***Expression analysis***

622 For each annotation, coding and non-coding, mRNA reads were counted with  
623 `htseq-count` [94] using strand specific and intersection strict mode (`--`  
624 `stranded=reverse -t gene --mode=intersection-strict`). These counts were used  
625 for differential gene expression analysis with DESeq2 (v1.16.1[95]) using a linear model  
626 and as factors the ecotype (two levels), the kinetic time (3 levels), the interaction  
627 between the two and the replicate (3 levels). Low counts were discarded using the  
628 default DESeq2 threshold and raw p-values were adjusted with the Bonferroni method.  
629 Differentially expressed genes were defined as having an adjusted p-value lower than  
630 0.01.

631 Differential siRNA accumulation was computed using DESeq2 with a model  
632 taking into account only the genotype (two levels) as factor, and using the counts of  
633 ShortStack. Differential accumulation was computed independently for the 21/22nt on  
634 one side and the 24nt on the other side and limited to coding and non-coding genes.

635 Bonferroni correction of the p-value was used and differential siRNAs were defined as  
636 having an adjusted p-value inferior to 0.01.

### 637 ***Measurement of the primary root length***

638 Pictures of the plates were taken with a flat scanner and root lengths measured  
639 using RootNav software [96]. Each measure corresponds to a different plant.

### 640 ***Quantitative RT-PCR***

641 Total RNA was extracted from whole roots using the Quick-RNA MiniPrep™ kit  
642 (Zymo Research, USA) and treated with the included DNase treatment according to the  
643 manufacturer's instructions. Reverse transcription was performed on 500 ng total RNA  
644 using the Maxima Reverse Transcriptase (Thermo Scientific™). Quantitative PCR (qRT-  
645 PCR) was performed on a 480 LightCycler thermocycler (Roche) using the  
646 manufacturer's instructions with Light cycler 480sybr green I master (Roche) and with  
647 primers listed in Additional file 1: Table S4. We used PP2A subunit PD (AT1G13320) as a  
648 reference gene for normalization.

### 649 ***Statistics and reproducibility of experiments***

650 Statistical analyses were performed using R (v3.4.2 [97]) with the help of the  
651 tidyverse (v1.2.1 [98]) and emmeans packages [99]. For each measure (root length,  
652 qPCR expression level, experimental repetition), the least-squares means were  
653 computed taking into account all the factors (genotype and condition) in a linear model.  
654 This allows correcting for inter-repetition variation. Data are presented as least-squares  
655 means±SEM.

656 The results for statistical significance tests are included in the legend of each  
657 figure. 'n' values represent the number of independent samples in a repetition, i.e. the  
658 number of roots or pools of root per condition. The number of independent experiments  
659 is denoted as "repetition". For each analysis, the detail number of repetition and number  
660 of sample per repetition are available in Additional file 1: Table S5 and Table S6.

## 661 **Declarations**

### 662 *Availability of data and material*

663 Sequence files generated during this study have been deposited into the NCBI  
664 GEO database [100] under the accessions GSE128250[101] and GSE128256[102].

### 665 *Ethics approval and consent to participate*

666 Not applicable.

### 667 *Consent for publication*

668 Not applicable.

### 669 *Competing interests*

670 The authors declare that they have no competing interests.

### 671 *Funding*

672 This work was supported by grants from Agence Nationale pour la Recherche  
673 (ANR) RNAdapt (grant no. ANR-12-ADAP-0019), SPLISIL (grant no. ANR-16-CE12-0032)  
674 and grants of The King Abdulla University of Science and Technology (KAUST)

675 International Program OICRF-2014-CRG4 and 'Laboratoire d'Excellence (LABEX)' Saclay  
676 Plant Sciences (SPS; ANR-10-LABX-40).

677 ***Authors' contributions***

678 TB, CB, TR, LS performed gene expression analysis; TB, MG, DG, ED and MLM  
679 performed statistical analysis and bioinformatics; CB, CS and AC were involved in  
680 sample preparation and processing; TB, CH, TD, LN and MC directed experimental work;  
681 TB, CH, TD and MC designed experiments and wrote the manuscript. All authors  
682 provided comments to the manuscript and approved it.

683 ***Acknowledgement***

684 We thank Ambre Miassod and Janina Lüders (IPS2, CNRS, INRA, Université Paris-  
685 Sud, Université Evry, Université Paris-Saclay, France) for technical assistance with gene  
686 expression experiments.

687 ***Additional files***

688 **Additional file 1:** supplemental tables.

- 689
- 690 • Table S1 - Mapping efficiency for each sequence sample
  - 691 • Table S2 - Genomic information of new transcripts compared to TAIR10
  - 692 • Table S3 - Differential gene expression analysis
  - 693 • For each comparison, the list of genes differentially expressed
  - 694 • Table S4 - Sequence of primers used in this study.
  - 695 • Table S5 - Number of samples used for each genotype and condition in the  
qPCR experiments



- 696           • Table S6 - Number of samples used for each genotype and condition in  
697           root length measurements

698           **Additional file 2:** supplemental figures:

- 699           • Fig. S1 - Characteristic of identified transcripts  
700           • Fig. S2 - Expression level and detection of coding and non-coding genes  
701           • Fig. S3 - Ecotype specific classification of lncRNAs as siRNA precursors  
702           • Fig. S4 - The ecotype effect on gene expression  
703           • Fig. S5 - Genome organization and correlation of expression at selected  
704           loci  
705           • Fig. S6 - The ecotype effect on siRNA accumulation  
706           • Fig. S7 - Genome homology at selected ncRNAs loci.  
707           • Fig. S8 - Deregulation of NPC48 and NPC72 do not change the expression  
708           of phosphate starvation related genes.

709           **Additional file 3:** new transcript localization on Col and Ler v7 genome as GFF  
710 files.

## 711           Legends

712           ***Fig. 1 - Identification of the transcripts and their repartition between the two***  
713 ***ecotypes.***

714           (a) Number of predicted coding and non-coding transcripts in the two ecotypes  
715 classed by type. New transcripts were not discovered in previously published studies.  
716 Then, detection at RNA levels of all transcripts predicted in each ecotype: classified as

717 coding (b) or as non-coding (c). For the latter two classes are defined if they are antisens  
718 of another annotation (NAT, d) or are intergenic (lincRNA, e). In contrast to coding  
719 genes, many ncRNAs, notably those intergenic, were detected only in one ecotype  
720 despite the DNA sequence similarity in both ecotypes.

721 ***Fig. 2 - Characterization of transcript at the DNA levels***

722 (a) Detection of the DNA sequence of the newly predicted transcripts in the two  
723 ecotypes (minimum of 90% of sequence identity and minimum of 90% of RNA length).  
724 The large majority of RNAs come from common DNA regions from both ecotypes. (b)  
725 SNPs accumulation per 100 bp of transcript length for each type of transcript according  
726 to data of 1001 genomes project. (c) Conservation among plant species (average  
727 PhasCons score) of each type of transcript according to their genomic position in  
728 relation to other annotations. In (b) and (c), NAT are natural antisense RNAs  
729 (overlapping with another mRNA, see Materials and Methods) and Non-NAT are  
730 transcript without any transcript overlapping on the other DNA strand (intergenic).

731 ***Fig. 3 - Long ncRNAs as small RNA precursors***

732 (a) Detection specificity of non-coding transcript as siRNAs (at the level of  
733 1 RPM) or long RNA (at the level of 1 TPM) between the two accessions. ND = not  
734 detected. The major difference between Col and Ler is the lincRNA component of the  
735 transcriptome. (b) Repartition of the major siRNA size for non-coding transcripts  
736 detected as long and short (level of 1 RPM) in the two ecotypes. ND = not detected.  
737 There is no major change of siRNA size between the two accessions. (c) Detection  
738 specificity of non-coding RNA as 21 nt and 22 nt siRNAs precursors at the level of 1 RPM  
739 in each ecotype. (d) Detection specificity of non-coding RNA as 24 nt siRNAs precursors

740 at the level of 1 RPM in each ecotype. (e) Detection specificity of coding RNA as 21 nt and  
741 22 nt siRNAs precursors at the level of 1 RPM in each ecotype. (f) Detection specificity of  
742 coding RNA as 24 nt siRNAs precursors at the level of 1 RPM in each ecotype.

743 ***Fig. 4 - Differentially expressed gene according to ecotype and kinetic effects.***

744 Statistical analysis revealed differentially expressed genes between ecotypes and  
745 kinetics treatments for coding and non-coding genes. The differentially expressed genes  
746 can be grouped according to their significant link with genotype effect (different level  
747 between the two ecotypes), kinetic effect (differential between any couple of the points  
748 of the phosphate starvation kinetic) and their interaction (showing differential  
749 expression in response to phosphate according to the genotype). This global distribution  
750 was partitioned between coding (a) and non-coding (b) transcripts. Among non-coding  
751 genes we separated non-coding transcript antisense to another annotation (c) or  
752 intergenic ones (d). The level of expression of strongly upregulated lincRNAs was  
753 investigated by RT-qPCR in 11 days old root grown on high phosphate condition of Col,  
754 *Ler*, *Col<sup>ler105</sup>* and hybrids between Col and *Ler* for Col upregulated lincRNAs (e) and *Ler*  
755 upregulated lincRNAs (f). Measures represent log 2 fold changes compared to Col (e) or  
756 *Ler* (f). Error bars represent standard error. Results were analyzed by one-way analysis  
757 of variance (ANOVA) followed by Tukey's post-hoc test: groups with different letters are  
758 statistically different ( $p \leq 0.05$ ) and groups with the same letters are statistically equal  
759 ( $p \leq 0.05$ )

760 **Fig. 5 - Overexpression of the lincRNAs NPC48 and NPC72 affects primary root**  
761 **growth.**

762 (a) Expression profile of selected intergenic lincRNAs in Col and Ler early  
763 phosphate starvation kinetics (RNAseq data, average expression and standard  
764 deviation). Five selected lincRNAs showed differential expression at each time point  
765 between Col and Ler. (b) Level of expression of selected intergenic lincRNAs in 11 days  
766 old root grown on in high phosphate condition in Col, Ler, Col<sup>ler105</sup> and hybrids between  
767 Col and Ler. Measures represent log 2 fold changes compared to Col measured by RT-  
768 qPCR. (c) Mean primary root length according to genotype and phosphate condition at  
769 the age of 11 days after sowing. (d) Expression level of NPC48 and NPC72 in 11 days old  
770 root grown on high phosphate condition of lines deregulated in NPC48 or NPC72 and  
771 mutants affected in phosphate related root arrest. Measures represent log 2 fold changes  
772 compared to Col. (b) - (d) Measure represent corrected means of the FC (b) and (d) or of  
773 primary root growth (c) computed according to generalized linear model fitted on  
774 several experiments (see the "Methods" section) and taking into account the following  
775 factors: genotypes, growth condition, interaction between genotype and growth  
776 condition. Error bars represent standard error. Results were analyzed by one-way  
777 analysis of variance (ANOVA) in (b) and (d) or two way ANOVA in (c) followed by  
778 Tukey's post-hoc test: groups with different letters are statistically different ( $p \leq 0.05$ )  
779 and groups with the same letters are statistically equal ( $p \leq 0.05$ )

780 **Fig. S1 - Characteristic of identified transcripts**

781 (a) Flowchart of identification of lincRNAs responsive to Pi starvation in two  
782 Arabidopsis ecotypes. Plants of the two ecotypes were grown on control condition for

783 seven day before transfer on low phosphate condition. Root tip were then sampled at  
784 time point 0 h, 1 h and 2 h after transfer. After RNA extraction, PolyA transcripts were  
785 sequenced. The retrieve reads were then mapped independently for the two ecotypes on  
786 their respective genomes (TAIR10 for Col and *Ler* v7 for *Ler*). Based on this mapping we  
787 predicted new transcriptional units on each genome compared to TAIR10 annotation  
788 available on both genomes. We then aligned the resulting transcriptional on the opposite  
789 genome to compute a homology and fused the overlapping transcripts. Only the  
790 transcript with a length superior to 200 nt were kept for further analysis. Each  
791 transcript was then categorized in one of the 4 following classes: coding, structural RNA  
792 (rRNA, tRNA, snRNA, snoRNA, ...), transposable element and non-coding RNA. The  
793 classification was estimated based first the overlap with already annotated transcripts  
794 (by order of importance in Araport11, RepTAS database, CANTATA database, BenAmor  
795 et al. 2009, Li et al. 2016 and miRBase v21). For the transcripts that were not found in  
796 any of these databases, their coding potential was predicted using COME. Repartition of  
797 the different detected genes in the different Arabidopsis databases for genes predicted  
798 as coding (b) or as non-coding (c).

799 ***Fig. S2 - Expression level and detection of coding and non-coding genes***

800 Distribution of coding genes, non-coding genes, structural RNA and transposable  
801 element according to their level of expression in Col (a) or *Ler* (b). (c) Number of newly  
802 detected genes per percent of additional sequencing reads of the library. Line  
803 correspond to the median of 100 bootstraps, grey shadow correspond to the standard  
804 deviation.

805 ***Fig. S3 - Ecotype specific classification of lncRNAs as siRNA precursors***

806 ShortStack classification of all non-coding genes as siRNA precursor expressed at  
807 more than 1 RPM of major siRNA in Col (a) or *Ler* (b). (c) Detection specificity of non-  
808 coding RNA as phased 21 nt and 22 nt siRNAs precursors at the level of 1 RPM in each  
809 ecotype. (d) Detection specificity of miRBase miRNAs at the level of 1 RPM in each  
810 ecotype. The majority of the miRNAs are detected in both ecotypes.

811 ***Fig. S4 - The ecotype effect on gene expression***

812 (a) PCA analysis showing the effect of ecotype and phosphate kinetic on the  
813 variance of gene expression. For the 3321 coding genes (b) and 675 non-coding genes  
814 (c) that are ecotype differentially expressed, the number of genes that accumulated  
815 more in Col or *Ler*.

816 ***Fig. S5 - Genome organization and correlation of expression at selected loci***

817 (a) Increased expression of *PHT1;2* in *Ler* correlated with the specific expression  
818 of the two NATs *Ler\_NEW\_R\_34181* and *Ler\_NEW\_R\_34180* (b) Decreased expression of  
819 *SPX4* gene in Col correlated to the expression of the NAT *Col\_NEW\_RNA\_R\_29088* (c)  
820 Differential expression of *NIP3;1* and *AT1G1910* between Col and *Ler* correlated with the  
821 expression of lincRNA *At1NC041650*.

822 ***Fig. S6 - The ecotype effect on siRNA accumulation***

823 PCA analysis showing the effect of genotype and phosphate kinetic on the  
824 variance between samples for 21/22nt (a), 24nt (b) and miRNAs (c). The samples can be  
825 well separated according to genotype (Col or *Ler*) but not according to the phosphate  
826 kinetic point. For ecotype differentially accumulated siRNAs, the number of differentially

827 accumulated siRNAs in either ecotype for 21/22nt siRNA precursors (d), 24nt siRNA  
828 precursors (e) and miRNAs (f).

829 ***Fig. S7 - Genome homology at selected ncRNAs loci.***

830 Genome alignment around *NPC15* (a), *NPC34* (b), *NPC43* (c), *NPC48* (d) and  
831 *NPC72* (e) between the reference sequence of Arabidopsis Col and two sequences from  
832 *Ler* (*Ler* v7 and *Ler* v8). The identity line is color coded according to conservation: green  
833 full conservation, yellow mismatches. For each sequence, dash represent missing  
834 sequences, plain grey identity sequence and plain black specific insertion or mismatches.

835 ***Fig. S8 - Deregulation of NPC48 and NPC72 do not change the expression of***  
836 ***phosphate starvation related genes.***

837 (a) - (g) Level of expression in 11 days old root grown on in high phosphate  
838 condition or low phosphate condition of genes involved phosphate sensing, *IPS1* (a) and  
839 *SPX3* (b), or in the response to phosphate related growth arrest, *LPR1* (c), *LPR2* (d),  
840 *STOP1* (e), *AMLT1* (f) and *MATE* (g). Measures represent log 2 fold changes compared to  
841 Col measured by RT-qPCR. Measure represent corrected means of the FC computed  
842 according to generalized linear model fitted on several experiments (see the "Methods"  
843 section) and taking into account the following factors: genotypes, growth condition,  
844 interaction between genotype and growth condition. Error bars represent standard  
845 error. Results were analyzed by two-way analysis of variance (ANOVA) followed by  
846 Tukey's post-hoc test: groups with different letters are statistically different ( $p \leq 0.05$ )  
847 and groups with the same letters are statistically equal ( $p \leq 0.05$ )

848 ***Table S1 - Mapping efficiency for each sequence sample***

849 ***Table S2 - Genomic information of new transcripts compared to TAIR10***

850 ***Table S3 - Differential gene expression analysis***

851 For each comparison, the list of genes differentially expressed

852 ***Table S4 - Sequence of primers used in this study.***

853 ***Table S5 - Number of samples used for each genotype and condition in the***

854 ***qPCR experiments***

855 'n' values represent the number of independent samples in a repetition, i.e. the

856 number of pools of root per genotype and condition. The number of independent

857 experiments is denoted as "repetition".

858 ***Table S6 - Number of samples used for each genotype and condition in root***

859 ***length measurements***

860 'n' values represent the number of independent samples in a repetition, i.e. the

861 number of root per genotype and condition. The number of independent experiments is

862 denoted as "repetition".

## 863 **References**

864 1. Ariel F, Romero-Barrios N, Jégu T, Benhamed M, Crespi M. Battles and hijacks:

865 Noncoding transcription in plants. Trends Plant Sci [Internet]. 2015;20:362–71.

866 Available from: <http://dx.doi.org/10.1016/j.tplants.2015.03.003>



- 867           2. Borges F, Martienssen RA. The expanding world of small RNAs in plants. *Nat.*  
868 *Rev. Mol. Cell Biol.* 2015. p. 727–41.
- 869           3. Rinn JL, Chang HY. Genome Regulation by Long Noncoding RNAs. *Annu Rev*  
870 *Biochem* [Internet]. 2012;81:145–66. Available from:  
871 <http://www.annualreviews.org/doi/10.1146/annurev-biochem-051410-092902>
- 872           4. Chekanova JA. Long non-coding RNAs and their functions in plants. *Curr. Opin.*  
873 *Plant Biol.* 2015. p. 207–16.
- 874           5. Ma L, Bajic VB, Zhang Z. On the classification of long non-coding RNAs. *RNA*  
875 *Biol.* 2013;10:924–33.
- 876           6. Ulitsky I, Shkumatava A, Jan CH, Sive H, Bartel DP. Conserved function of  
877 lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell*  
878 [Internet]. 2011;147:1537–50. Available from:  
879 <http://dx.doi.org/10.1016/j.cell.2011.11.055>
- 880           7. Hon CC, Ramilowski JA, Harshbarger J, Bertin N, Rackham OJL, Gough J, et al. An  
881 atlas of human long non-coding RNAs with accurate 5' ends. *Nature.* 2017;543:199–  
882 204.
- 883           8. Liu J, Jung C, Xu J, Wang H, Deng S, Bernad L, et al. Genome-Wide Analysis  
884 Uncovers Regulation of Long Intergenic Noncoding RNAs in Arabidopsis. *Plant Cell*  
885 [Internet]. 2012;24:4333–45. Available from:  
886 <http://www.plantcell.org/cgi/doi/10.1105/tpc.112.102855>
- 887           9. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The  
888 GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure,

- 889 evolution, and expression. *Genome Res.* 2012;22:1775–89.
- 890 10. Gandhi M, Caudron-Herger M, Diederichs S. RNA motifs and combinatorial  
891 prediction of interactions, stability and localization of noncoding RNAs. *Nat Struct Mol*  
892 *Biol.* 2018;25:1070–6.
- 893 11. Mukherjee N, Calviello L, Hirsekorn A, De Pretis S, Pelizzola M, Ohler U.  
894 Integrative classification of human coding and noncoding genes through RNA  
895 metabolism profiles. *Nat Struct Mol Biol.* 2017;24:86–96.
- 896 12. Quinn JJ, Ilik IA, Qu K, Georgiev P, Chu C, Akhtar A, et al. Domain ChIRP reveals  
897 the modularity of long noncoding RNA architecture, chromatin interactions, and  
898 function. *Nat Biotechnol.* 2014;32:933–40.
- 899 13. Tsuiji H, Yoshimoto R, Hasegawa Y, Furuno M, Yoshida M, Nakagawa S.  
900 Competition between a noncoding exon and introns: Gomafu contains tandem UACUAAC  
901 repeats and associates with splicing factor-1. *Genes to Cells.* 2011;16:479–90.
- 902 14. Ding Y, Tang Y, Kwok CK, Zhang Y, Bevilacqua PC, Assmann SM. In vivo  
903 genome-wide profiling of RNA secondary structure reveals novel regulatory features.  
904 *Nature.* 2013;505:696–700.
- 905 15. Siegfried NA, Busan S, Rice GM, Nelson JAE, Weeks KM. RNA motif discovery  
906 by SHAPE and mutational profiling (SHAPE-MaP). *Nat Methods.* 2014;11:959–65.
- 907 16. Spitale RC, Flynn RA, Zhang QC, Crisalli P, Lee B, Jung JW, et al. Structural  
908 imprints in vivo decode RNA regulatory mechanisms. *Nature.* 2015;519:486–90.
- 909 17. Qian X, Zhao J, Yeung PY, Zhang QC, Kwok CK. Revealing lncRNA Structures

910 and Interactions by Sequencing-Based Approaches. *Trends Biochem Sci.* 2018;44:33–52.

911 18. Hezroni H, Koppstein D, Schwartz MG, Avrutin A, Bartel DP, Ulitsky I.

912 Principles of Long Noncoding RNA Evolution Derived from Direct Comparison of

913 Transcriptomes in 17 Species. *Cell Rep.* 2015;11:1110–22.

914 19. Guttman M, Rinn JL. Modular regulatory principles of large non-coding RNAs.

915 *Nature.* 2012;482:339–46.

916 20. Melia T, Hao P, Yilmaz F, Waxman DJ. Hepatic lincRNAs: high promoter

917 conservation and dynamic, sex-dependent transcriptional regulation by growth

918 hormone. *Mol Cell Biol.* 2015;MCB.00861-15.

919 21. Wang H, Niu QW, Wu HW, Liu J, Ye J, Yu N, et al. Analysis of non-coding

920 transcriptome in rice and maize uncovers roles of conserved lincRNAs associated with

921 agriculture traits. *Plant J.* 2015;84:404–16.

922 22. Nelson ADL, Forsythe ES, Frank MR, Clausen DS, Haug-Batzell AK, Beilstein

923 MA, et al. A Genomic Analysis of Factors Driving lincRNA Diversification: Lessons from

924 Plants. *G3&#58; Genes|Genomes|Genetics.* 2016;6:2881–91.

925 23. Groff AF, Sanchez-Gomez DB, Soruco MML, Gerhardinger C, Barutcu AR, Li E,

926 et al. In Vivo Characterization of Linc-p21 Reveals Functional cis -Regulatory DNA

927 Elements. *Cell Rep.* 2016;16:2178–86.

928 24. Kopp F, Mendell JT. Functional Classification and Experimental Dissection of

929 Long Noncoding RNAs. *Cell.* 2018;172:393–407.

930 25. Contreras-Moreira B, Cantalapiedra CP, García-Pereira MJ, Gordon SP, Vogel

- 931 JP, Igartua E, et al. Analysis of Plant Pan-Genomes and Transcriptomes with  
932 GET\_HOMOLOGUES-EST, a Clustering Solution for Sequences of the Same Species. Front  
933 Plant Sci. 2017;8.
- 934 26. Vernikos G, Medini D, Riley DR, Tettelin H. Ten years of pan-genome analyses.  
935 Curr Opin Microbiol. 2015;23:148–54.
- 936 27. Clark RM, Schweikert G, Toomajian C, Ossowski S, Zeller G, Shinn P, et al.  
937 Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*.  
938 Science (80- ) [Internet]. 2007;317:338–42. Available from:  
939 <http://www.sciencemag.org/cgi/doi/10.1126/science.1138632>
- 940 28. Consortium T 1001 G. 1,135 Genomes Reveal the Global Pattern of  
941 Polymorphism in *Arabidopsis thaliana*. Cell. 2016;166:481–91.
- 942 29. Zhao Q, Feng Q, Lu H, Li Y, Wang A, Tian Q, et al. Pan-genome analysis  
943 highlights the extent of genomic variation in cultivated and wild rice. Nat Genet.  
944 2018;50:278–84.
- 945 30. Kutter C, Watt S, Stefflova K, Wilson MD, Goncalves A, Ponting CP, et al. Rapid  
946 Turnover of Long Noncoding RNAs and the Evolution of Gene Expression. PLoS Genet.  
947 2012;8:e1002841.
- 948 31. Necsulea A, Soumillon M, Warnefors M, Liechti A, Daish T, Zeller U, et al. The  
949 evolution of lncRNA repertoires and expression patterns in tetrapods. Nature.  
950 2014;505:635–40.
- 951 32. Nelson ADL, Devisetty UK, Palos K, Haug-Baltzell AK, Lyons E, Beilstein MA.  
952 Evolinc: A tool for the identification and evolutionary comparison of long intergenic

- 953 non-coding RNAs. *Front Genet.* 2017;8.
- 954 33. Kornienko AE, Dotter CP, Guenzl PM, Gisslinger H, Gisslinger B, Cleary C, et al.  
955 Long non-coding RNAs display higher natural expression variation than protein-coding  
956 genes in healthy humans. *Genome Biol.* 2016;17.
- 957 34. Quadrana L, Colot V. Plant Transgenerational Epigenetics. *Annu Rev Genet.*  
958 2016;50:467-91.
- 959 35. Reymond M, Svistonoff S, Loudet O, Nussaume L, Desnos T. Identification of  
960 QTL controlling root growth response to phosphate starvation in *Arabidopsis thaliana*.  
961 *Plant, Cell Environ.* 2006;29:115-25.
- 962 36. Svistonoff S, Creff A, Reymond M, Sigoillot-Claude C, Ricaud L, Blanchet A, et  
963 al. Root tip contact with low-phosphate media reprograms plant root architecture. *Nat*  
964 *Genet* [Internet]. 2007;39:792-6. Available from:  
965 <http://www.nature.com/doi/10.1038/ng2041>
- 966 37. Gutiérrez-Alanís D, Yong-Villalobos L, Jiménez-Sandoval P, Alatorre-Cobos F,  
967 Oropeza-Aburto A, Mora-Macías J, et al. Phosphate Starvation-Dependent Iron  
968 Mobilization Induces CLE14 Expression to Trigger Root Meristem Differentiation  
969 through CLV2/PEPR2 Signaling. *Dev Cell.* 2017;41:555-570.e3.
- 970 38. Balzergue C, Darteville T, Godon C, Laugier E, Meisrimler C, Teulon J-M, et al.  
971 Low phosphate activates STOP1-ALMT1 to rapidly inhibit root cell elongation. *Nat*  
972 *Commun.* 2017;8:15300.
- 973 39. Sánchez-Calderón L, López-Bucio J, Chacón-López A, Cruz-Ramírez A, Nieto-  
974 Jacobo F, Dubrovsky JG, et al. Phosphate starvation induces a determinate

- 975 developmental program in the roots of *Arabidopsis thaliana*. *Plant Cell Physiol*  
976 [Internet]. 2005;46:174–84. Available from: <http://dx.doi.org/10.1093/pcp/pci011>
- 977 40. Müller J, Toev T, Heisters M, Teller J, Moore KL, Hause G, et al. Iron-Dependent  
978 Callose Deposition Adjusts Root Meristem Maintenance to Phosphate Availability. *Dev*  
979 *Cell*. 2015;33:216–30.
- 980 41. Thibaud MC, Arrighi JF, Bayle V, Chiarenza S, Creff A, Bustos R, et al.  
981 Dissection of local and systemic transcriptional responses to phosphate starvation in  
982 *Arabidopsis*. *Plant J*. 2010;64:775–89.
- 983 42. Kanno S, Arrighi JF, Chiarenza S, Bayle V, Berthomé R, Péret B, et al. A novel  
984 role for the root cap in phosphate uptake and homeostasis. *Elife*. 2016;5.
- 985 43. Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, et al. The  
986 *Arabidopsis* Information Resource (TAIR): improved gene annotation and new tools.  
987 *Nucleic Acids Res*. 2012;40:D1202–10.
- 988 44. Gan X, Stegle O, Behr J, Steffen JG, Drewe P, Hildebrand KL, et al. Multiple  
989 reference genomes and transcriptomes for *Arabidopsis thaliana*. *Nature*. 2011;477:419–  
990 23.
- 991 45. Zapata L, Ding J, Willing E-M, Hartwig B, Bezdán D, Jiao W-B, et al.  
992 Chromosome-level assembly of *Arabidopsis thaliana* L er reveals the extent of  
993 translocation and inversion polymorphisms. *Proc Natl Acad Sci*. 2016;113:E4052–60.
- 994 46. Cheng CY, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S, Town CD.  
995 Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant*  
996 *J*. 2017;89:789–804.

- 997           47. Szcześniak MW, Rosikiewicz W, Makałowska I. CANTATAdb: A collection of  
998 plant long non-coding RNAs. *Plant Cell Physiol.* 2016;57:e8.
- 999           48. Kozomara A, Griffiths-Jones S. MiRBase: Annotating high confidence  
1000 microRNAs using deep sequencing data. *Nucleic Acids Res.* 2014;42.
- 1001           49. Amor B Ben, Wirth S, Merchan F, Laporte P, D'Aubenton-Carafa Y, Hirsch J, et  
1002 al. Novel long non-protein coding RNAs involved in Arabidopsis differentiation and  
1003 stress responses. *Genome Res [Internet].* 2009;19:57–69. Available from:  
1004 <http://dx.doi.org/10.1101/gr.080275.108>
- 1005           50. Li S, Yamada M, Han X, Ohler U, Benfey PN. High-Resolution Expression Map  
1006 of the Arabidopsis Root Reveals Alternative Splicing and lincRNA Regulation. *Dev Cell.*  
1007 2016;39:508–22.
- 1008           51. Hu L, Xu Z, Hu B, Lu ZJ. COME: A robust coding potential calculation tool for  
1009 lincRNA identification and characterization based on multiple features. *Nucleic Acids*  
1010 *Res.* 2017;45:e2.
- 1011           52. Hupaló D, Kern AD. Conservation and functional element discovery in 20  
1012 angiosperm plant genomes. *Mol Biol Evol [Internet].* 2013;30:1729–44. Available from:  
1013 <http://dx.doi.org/10.1093/molbev/mst082>
- 1014           53. Fei Q, Xia R, Meyers BC. Phased, Secondary, Small Interfering RNAs in  
1015 Posttranscriptional Regulatory Networks. *Plant Cell.* 2013;25:2400–15.
- 1016           54. Matzke MA, Mosher RA. RNA-directed DNA methylation: An epigenetic  
1017 pathway of increasing complexity. *Nat. Rev. Genet.* 2014. p. 394–408.

- 1018            55. Rigaille G, Balzergue S, Brunaud V, Blondet E, Rau A, Rogier O, et al. Synthetic  
1019 data sets for the identification of key ingredients for RNA-seq differential analysis. *Brief*  
1020 *Bioinform.* 2018;19:65–76.
- 1021            56. Park BS, Seo JS, Chua N-H. NITROGEN LIMITATION ADAPTATION Recruits  
1022 PHOSPHATE2 to Target the Phosphate Transporter PT2 for Degradation during the  
1023 Regulation of Arabidopsis Phosphate Homeostasis. *Plant Cell.* 2014;26:454–64.
- 1024            57. Bayle V, Arrighi J-F, Creff A, Nespoulous C, Vialaret J, Rossignol M, et al.  
1025 Arabidopsis thaliana High-Affinity Phosphate Transporters Exhibit Multiple Levels of  
1026 Posttranslational Regulation. *Plant Cell.* 2011;23:1523–35.
- 1027            58. Nussaume L, Kanno S, Javot H, Marin E, Pochon N, Ayadi A, et al. Phosphate  
1028 import in plants: focus on the PHT1 transporters. *Front Plant Sci.* 2011;2.
- 1029            59. Huang T-K, Han C-L, Lin S-I, Chen Y-RY-JY-S, Tsai Y-C, Chen Y-RY-JY-S, et al.  
1030 Identification of Downstream Components of Ubiquitin-Conjugating Enzyme  
1031 PHOSPHATE2 by Quantitative Membrane Proteomics in Arabidopsis Roots. *Plant Cell.*  
1032 2013;25:4044–60.
- 1033            60. Ayadi A, David P, Arrighi J-F, Chiarenza S, Thibaud M-C, Nussaume L, et al.  
1034 Reducing the Genetic Redundancy of Arabidopsis PHOSPHATE TRANSPORTER1  
1035 Transporters to Study Phosphate Uptake and Signaling. *Plant Physiol.* 2015;167:1511–  
1036 26.
- 1037            61. Shin H, Shin HS, Dewbre GR, Harrison MJ. Phosphate transport in Arabidopsis:  
1038 Pht1;1 and Pht1;4 play a major role in phosphate acquisition from both low- and high-  
1039 phosphate environments. *Plant J.* 2004;39:629–42.



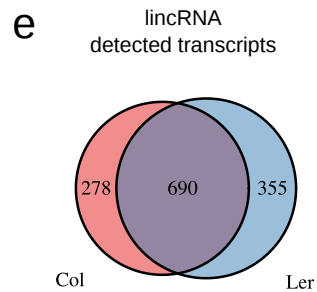
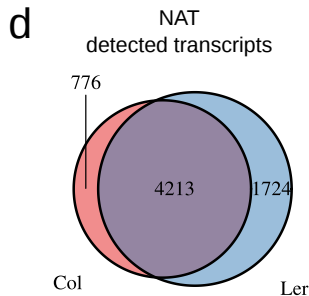
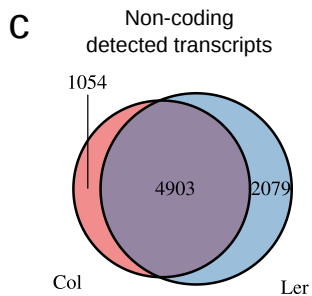
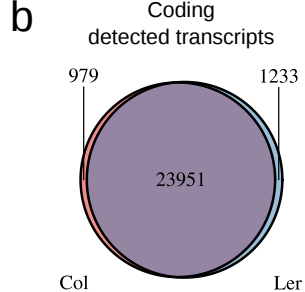
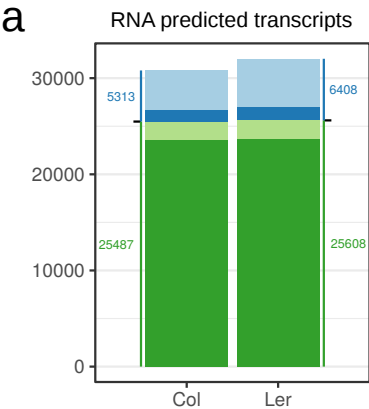
- 1040           62. Yuan J, Zhang Y, Dong J, Sun Y, Lim BL, Liu D, et al. Systematic characterization  
1041 of novel lncRNAs responding to phosphate starvation in *Arabidopsis thaliana*. *BMC*  
1042 *Genomics*. 2016;17.
- 1043           63. Wang X-J, Gaasterland T, Chua N-H. Genome-wide prediction and  
1044 identification of cis-natural antisense transcripts in *Arabidopsis thaliana*. *Genome Biol*  
1045 [Internet]. 2005;6:R30. Available from: <http://genomebiology.com/2005/6/4/R30>
- 1046           64. Petricka JJ, Winter CM, Benfey PN. Control of *Arabidopsis* Root Development.  
1047 *Annu Rev Plant Biol* [Internet]. 2012;63:563–90. Available from:  
1048 <http://www.annualreviews.org/doi/10.1146/annurev-arplant-042811-105501>
- 1049           65. Pickrell JK. Joint analysis of functional genomic data and genome-wide  
1050 association studies of 18 human traits. *Am J Hum Genet*. 2014;94:559–73.
- 1051           66. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI  
1052 GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res*.  
1053 2013;42:D1001–6.
- 1054           67. Wang YM, Xu HB, Wang MS, Otecko NO, Ye LQ, Wu DD, et al. Annotating long  
1055 intergenic non-coding RNAs under artificial selection during chicken domestication.  
1056 *BMC Evol Biol*. 2017;17.
- 1057           68. Ning S, Yue M, Wang P, Liu Y, Zhi H, Zhang Y, et al. LincSNP 2.0: An updated  
1058 database for linking disease-associated SNPs to human long non-coding RNAs and their  
1059 TFBSs. *Nucleic Acids Res*. 2017;45:D74–8.
- 1060           69. Giral H, Landmesser U, Kratzer A. Into the Wild: GWAS Exploration of Non-  
1061 coding RNAs. *Front Cardiovasc Med*. 2018;5.

- 1062           70. Scarano D, Rao R, Corrado G. In Silico identification and annotation of  
1063 noncoding RNAs by RNA-seq and de Novo assembly of the transcriptome of Tomato  
1064 Fruits. PLoS One. 2017;12.
- 1065           71. Goff LA, Rinn JL. Linking RNA biology to lncRNAs. Genome Res.  
1066 2015;25:1456–65.
- 1067           72. Wu H, Yang L, Chen LL. The Diversity of Long Noncoding RNAs and Their  
1068 Generation. Trends Genet. 2017. p. 540–52.
- 1069           73. Marchese FP, Raimondi I, Huarte M. The multidimensional mechanisms of  
1070 long noncoding RNA function. Genome Biol. 2017;18.
- 1071           74. Springer NM, Stupar RM. Allele-Specific Expression Patterns Reveal Biases  
1072 and Embryo-Specific Parent-of-Origin Effects in Hybrid Maize. Plant Cell Online.  
1073 2007;19:2391–402.
- 1074           75. De La Chapelle A. Genetic predisposition to human disease: Allele-specific  
1075 expression and low-penetrance regulatory loci. Oncogene. 2009;28:3345–8.
- 1076           76. Seymour DK, Chae E, Grimm DG, Pizarro CM, Habring-Müller A, Vasseur F, et  
1077 al. Genetic architecture of nonadditive inheritance in Arabidopsis thaliana hybrids. Proc  
1078 Natl Acad Sci. 2016;113:E7317–26.
- 1079           77. Yang M, Wang X, Ren D, Huang H, Xu M, He G, et al. Genomic architecture of  
1080 biomass heterosis in Arabidopsis. Proc Natl Acad Sci. 2017;114:8101–6.
- 1081           78. Narang RA, Altmann T. Phosphate acquisition heterosis in Arabidopsis  
1082 thaliana: A morphological and physiological analysis. Plant Soil. 2001;234:91–7.

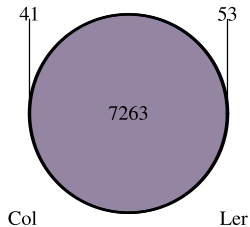
- 1083           79. Li L, Eichten SR, Shimizu R, Petsch K, Yeh CT, Wu W, et al. Genome-wide  
1084 discovery and characterization of maize long non-coding RNAs. *Genome Biol.*  
1085 2014;15:R40.
- 1086           80. Schnable PS, Jia Y, DeCook R, Swanson-Wagner RA, Borsuk LA, Nettleton D. All  
1087 possible modes of gene action are observed in a global comparison of gene expression in  
1088 a maize F1 hybrid and its inbred parents. *Proc Natl Acad Sci.* 2006;103:6805–10.
- 1089           81. Lin SI, Chiang SF, Lin WY, Chen JW, Tseng CY, Wu PC, et al. Regulatory  
1090 Network of MicroRNA399 and PHO2 by Systemic Signaling. *Plant Physiol.*  
1091 2008;147:732–46.
- 1092           82. Lin SI, Chiou TJ. Long-distance movement and differential targeting of  
1093 microRNA399s. *Plant Signal Behav.* 2008;3:730–2.
- 1094           83. Shah S, Wittmann S, Kilchert C, Vasiljeva L. lncRNA recruits RNAi and the  
1095 exosome to dynamically regulate *pho1* expression in response to phosphate levels in  
1096 fission yeast. *Genes Dev.* 2014;28:231–44.
- 1097           84. Ard R, Tong P, Allshire RC. Long non-coding RNA-mediated transcriptional  
1098 interference of a permease gene confers drug tolerance in fission yeast. *Nat Commun.*  
1099 2014;5.
- 1100           85. Thieme CJ, Rojas-Triana M, Stecyk E, Schudoma C, Zhang W, Yang L, et al.  
1101 Endogenous Arabidopsis messenger RNAs transported to distant tissues. *Nat Plants.*  
1102 2015;1:15025.
- 1103           86. Torii KU. The Arabidopsis ERECTA Gene Encodes a Putative Receptor Protein  
1104 Kinase with Extracellular Leucine-Rich Repeats. *Plant Cell Online [Internet].*

- 1105 1996;8:735–46. Available from: <http://www.plantcell.org/cgi/doi/10.1105/tpc.8.4.735>
- 1106 87. Kopylova E, Noé L, Touzet H. SortMeRNA: Fast and accurate filtering of  
1107 ribosomal RNAs in metatranscriptomic data. *Bioinformatics*. 2012;28:3211–7.
- 1108 88. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2:  
1109 accurate alignment of transcriptomes in the presence of insertions, deletions and gene  
1110 fusions. *Genome Biol*. 2013;14:R36.
- 1111 89. Tran VDT, Souiai O, Romero-Barrios N, Crespi M, Gautheret D. Detection of  
1112 generic differential RNA processing events from RNA-seq data. *RNA Biol*. 2016;13:59–  
1113 67.
- 1114 90. Hu L, Xu Z, Hu B, Lu ZJ. COME: A robust coding potential calculation tool for  
1115 lncRNA identification and characterization based on multiple features. *Nucleic Acids*  
1116 *Res*. 2017;45.
- 1117 91. Wang L, Wang S, Li W. RSeQC: quality control of RNA-seq experiments.  
1118 *Bioinforma Oxford Engl*. 2012;28:2184–5.
- 1119 92. Wang L, Nie J, Sicotte H, Li Y, Eckel-Passow JE, Dasari S, et al. Measure  
1120 transcript integrity using RNA-seq data. *BMC Bioinformatics*. 2016;17:1–16.
- 1121 93. Johnson NR, Yeoh JM, Coruh C, Axtell MJ. Improved Placement of Multi-  
1122 mapping Small RNAs. *G3&#58; Genes|Genomes|Genetics*. 2016;6:2103–11.
- 1123 94. Anders S, Pyl PT, Huber W. HTSeq-A Python framework to work with high-  
1124 throughput sequencing data. *Bioinformatics*. 2015;31:166–9.
- 1125 95. Love MI, Huber W, Anders S. Moderated estimation of fold change and

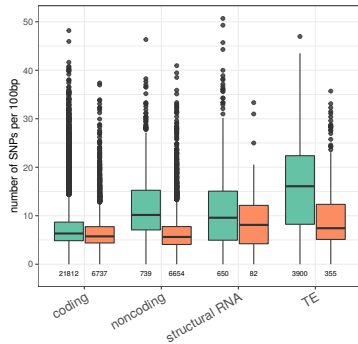
- 1126 dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15.
- 1127 96. Pound MP, Wells DM, Atkinson JA, Bennett MJ, French AP, Pridmore T.
- 1128 RootNav: Navigating Images of Complex Root Architectures. *Plant Physiol.*
- 1129 2013;162:1802–14.
- 1130 97. R Core Team. R: A language and environment for statistical computing.
- 1131 <http://www.R-project.org/>. R Found. Stat. Comput. Vienna, Austria. 2017.
- 1132 98. Wickham H. tidyverse: Easily Install and Load the “Tidyverse” [Internet].
- 1133 2017. Available from: <https://cran.r-project.org/package=tidyverse>
- 1134 99. Lenth R. emmeans: Estimated Marginal Means, aka Least-Squares Means
- 1135 [Internet]. 2019. Available from: <https://cran.r-project.org/package=emmeans>
- 1136 100. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene
- 1137 expression and hybridization array data repository. *Nucleic Acids Res.* 2002;30:207–10.
- 1138 101. Balzergue C, Gabriel M, Taconnat L, Balzergue S, Blein T. Non-coding RNAs in
- 1139 the adaptation of root growth to phosphate starvation [Internet]. [cited 2019 Mar 15].
- 1140 Available from: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE128250>
- 1141 102. Balzergue C, Gabriel M, Taconnat L, Balzergue S, Blein T. Non-coding RNAs in
- 1142 the adaptation of root growth to phosphate starvation [miRNA-Seq] [Internet]. [cited
- 1143 2019 Mar 15]. Available from:
- 1144 <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE128256>
- 1145



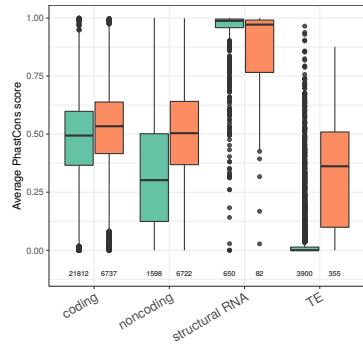
**a** Presence of new predicted transcripts at DNA level



**b** SNPs accumulation in all predicted transcripts

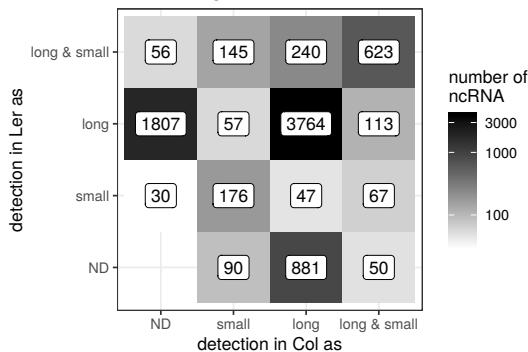
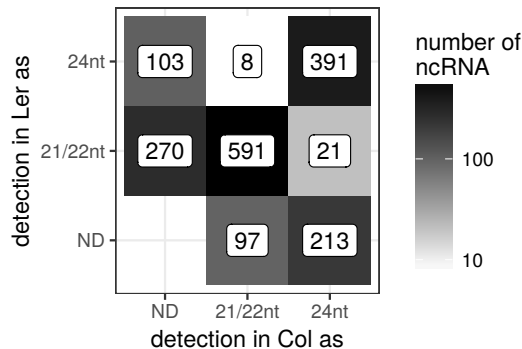
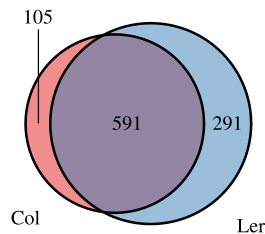
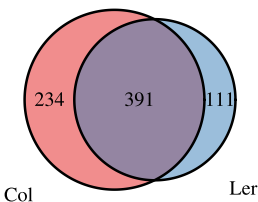
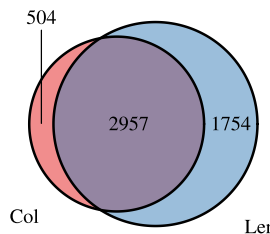


**c** Conservation of all predicted transcripts



Non NAT

NAT

**a** Detection of non-coding genes as long or small RNAs**b** Major siRNA size detected for non-coding genes**c** Detection of non-coding genes as 21/22nt siRNA precursors**d** Detection non-coding genes as 24nt siRNA precursors**e** Detection of coding genes as 21/22nt siRNA precursors**f** Detection of coding genes as 24nt siRNA precursors