bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Two ecotype-related long non-coding RNAs in the environmental control of root growth

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

[1]Institute of Plant Sciences Paris Saclay IPS2, CNRS, INRA, Université Paris-Sud, Université Evry, Université Paris-Saclay, Bâtiment 630, 91405 Orsay, France

[2]Institute of Plant Sciences Paris-Saclay IPS2, Paris Diderot, Sorbonne Paris-Cité, Bâtiment 630, 91405, Orsay, France.

[3]CEA, Institut de Biologie Environnementale et de Biotechnologie, Laboratoire de Biologie du Développement des Plantes, Saint-Paul-lez-Durance, F-13108, France.

[4]CNRS, Unité Mixte de Recherche 7265 Biologie Végétale & Microbiologie Environnementale, Saint-Paul-lez-Durance, F-13108, France.

[5]Aix-Marseille Université, Saint-Paul-lez-Durance, F-13108, France.

[6]UMR MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay, 75005, Paris, France

[7]I2BC, Institute for Integrative Biology of the Cell, CEA, CNRS, Université Paris Sud, 1 avenue de la terrasse, 91198 Gif sur Yvette, France.

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

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Abstract

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

Introduction

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

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

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

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

113 elongation in the elongation zone rapidly decreases and cell proliferation in the RAM
114 progressively ceases [37–40]. By contrast, in L*er* 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 L*er* root apexes
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 L*er* 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* (L*er*) 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

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

158 In total, we identified 5313 and 6408 novel putative genes respectively in Col and
159 L*er* 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 L*er* 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 L*er*
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 L*er* does not
177 result from a sequencing bias.

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

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

Evolutionary analysis of lncRNA genes expressed in root tips

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

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

Few lncRNA transcripts co-localize with small RNA generating loci

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

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

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

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

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

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

Differential accumulation of transcripts between ecotypes along a low phosphate kinetic

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

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

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

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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 L*er* and 106 lincRNAs and 187 NATs in L*er*
299 compared to Col (Additional file 2: Fig. S4c).

300 We used qRT-PCR in independent replicates of Col, Col^{er105} and L*er* to confirm the
301 differential expression of 14 intergenic lncRNA 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 lncRNAs (Fig. 4 e, f). Globally Col and Col^{er105} (used for the RNA sequencing data)
304 showed similar expression levels despite minor differences for some lncRNAs. To
305 investigate any dominant expression effect from one ecotype, we investigated the level
306 of expression of these lncRNAs in the F1 offsprings of Col and L*er* 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 lncRNAs may be expressed in the Col
312 and L*er* genomes in relation with known regulators of root growth responses triggered
313 by phosphate starvation. Indeed, we could identify 2 specific antisense lncRNAs 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

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

Differential accumulations of small RNAs between ecotypes

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

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

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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 L*er*, 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 L*er*; 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 L*er*. 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

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

365 expression was only detected in the Col^{er105} 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 L*er* genomes). No significant
373 modifications (except few SNPs) were detected between Col and L*er* for the *NPC34*,
374 *NPC43* and *NPC48* loci. By contrast, the *NPC15* locus contains an insertion of 2417 nt in
375 L*er* v8 and *NPC72* is completely missing in L*er* 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 L*er*.

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 L*er*, minor differences in root length were observed. However,
388 the OE of *NPC43* gene clearly increased the inhibition of root growth (Fig. 5c) when

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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 L*er*-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 µM 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).

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

Discussion

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

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

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

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

456 phenotypic difference of behavior observed among ecotypes [44]. In our experiments,
457 the expression polymorphism (differential gene expressions observed between Col and
458 L*er*) 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 L*er*. In the F1 offspring, the 15 confirmed

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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-

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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 (16 h photoperiod; intensity 90 µE;
534 21 °C). The growth medium contained 0.15 mM $MgSO_4$, 2.1 mM NH_4NO_3, 1.9 mM KNO_3,
535 0.34 mM $CaCl_2$, 0.5 µM KI, 10 µM $FeCl_2$, 10 µM H_3BO_3, 10 µM $MnSO_4$, 3 µM $ZnSO_4$, 0.1 µM
536 $CuSO_4$, 0.1 µM $CoCl_2$, 0.1 µM Na_2MoO_4, 0.5 $g.L^{-1}$ sucrose. The agar (0.8 $g.L^{-1}$) for plates
537 was from Sigma-Aldrich (A1296 #BCBL6182V). The -Pi and +Pi agar medium contained
538 10 and 500 µM 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

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

548 background. Col^{er105} 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^{er105} and L*er* 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 apexes 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 apexes grown and collected in the same condition
565 as for the RNA-Seq were used. Small RNAs were extracted using *mir*Vana™ 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.

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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 L*er* v7, were positioned
579 on the other genome, respectively L*er* v7 and TAIR10, and on L*er* 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 L*er*-predicted transcripts only those positioned on L*er* 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

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

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

Library saturation analysis

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

Small RNA analysis

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

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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.

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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.

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

Declarations

Availability of data and material

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

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

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

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

Additional file 2: supplemental figures:

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

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

Legends

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

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

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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 L*er* is the lncRNA 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

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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 L*er*, *Col*^{er105} and hybrids between Col and L*er* for Col upregulated lincRNAs (e) and L*er*
755 upregulated lincRNAs (f). Measures represent log 2 fold changes compared to Col (e) or
756 L*er* (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 ($p \leq 0.05$)

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

(a) Expression profile of selected intergenic lincRNAs in Col and L*er* early phosphate starvation kinetics (RNAseq data, average expression and standard deviation). Five selected lincRNAs showed differential expression at each time point between Col and L*er*. (b) Level of expression of selected intergenic lincRNAs in 11 days old root grown on in high phosphate condition in Col, L*er*, *Col*^{er105} and hybrids between Col and L*er*. Measures represent log 2 fold changes compared to Col measured by RT-qPCR. (c) Mean primary root length according to genotype and phosphate condition at the age of 11 days after sowing. (d) Expression level of *NPC48* and *NPC72* in 11 days old root grown on high phosphate condition of lines deregulated in NPC48 or NPC72 and mutants affected in phosphate related root arrest. Measures represent log 2 fold changes compared to Col. (b) - (d) Measure represent corrected means of the FC (b) and (d) or of primary root growth (c) computed according to generalized linear model fitted on several experiments (see the "Methods" section) and taking into account the following factors: genotypes, growth condition, interaction between genotype and growth condition. Error bars represent standard error. Results were analyzed by one-way analysis of variance (ANOVA) in (b) and (d) or two way ANOVA in (c) followed by Tukey's post-hoc test: groups with different letters are statistically different ($p \leq 0.05$) and groups with the same letters are statistically equal ($p \leq 0.05$)

Fig. S1 - Characteristic of identified transcripts

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

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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 L*er* v7 for L*er*). 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 L*er* (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.

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

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

Fig. S4 - The ecotype effect on gene expression

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

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

(a) Increased expression of *PHT1;2* in L*er* correlated with the specific expression of the two NATs *Ler_NEW_R_34181* and *Ler_NEW_R_34180* (b) Decreased expression of *SPX4* gene in Col correlated to the expression of the NAT *Col_NEW_RNA_R_29088* (c) Differential expression of *NIP3;1* and *AT1G1910* between Col and L*er* correlated with the expression of lincRNA *At1NC041650*.

Fig. S6 - The ecotype effect on siRNA accumulation

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

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

Fig. S7 - Genome homology at selected ncRNAs loci.

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table S1 - Mapping efficiency for each sequence sample

Table S2 - Genomic information of new transcripts compared to TAIR10

Table S3 - Differential gene expression analysis

For each comparison, the list of genes differentially expressed

Table S4 - Sequence of primers used in this study.

Table S5 - Number of samples used for each genotype and condition in the qPCR experiments

'n' values represent the number of independent samples in a repetition, i.e. the number of pools of root per genotype and condition. The number of independent experiments is denoted as "repetition".

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

'n' values represent the number of independent samples in a repetition, i.e. the number of root per genotype and condition. The number of independent experiments is denoted as "repetition".

References

1. Ariel F, Romero-Barrios N, Jégu T, Benhamed M, Crespi M. Battles and hijacks: Noncoding transcription in plants. Trends Plant Sci [Internet]. 2015;20:362–71. Available from: http://dx.doi.org/10.1016/j.tplants.2015.03.003

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

2. Borges F, Martienssen RA. The expanding world of small RNAs in plants. Nat. Rev. Mol. Cell Biol. 2015. p. 727–41.

3. Rinn JL, Chang HY. Genome Regulation by Long Noncoding RNAs. Annu Rev Biochem [Internet]. 2012;81:145–66. Available from: http://www.annualreviews.org/doi/10.1146/annurev-biochem-051410-092902

4. Chekanova JA. Long non-coding RNAs and their functions in plants. Curr. Opin. Plant Biol. 2015. p. 207–16.

5. Ma L, Bajic VB, Zhang Z. On the classification of long non-coding RNAs. RNA Biol. 2013;10:924–33.

6. Ulitsky I, Shkumatava A, Jan CH, Sive H, Bartel DP. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. Cell [Internet]. 2011;147:1537–50. Available from: http://dx.doi.org/10.1016/j.cell.2011.11.055

7. Hon CC, Ramilowski JA, Harshbarger J, Bertin N, Rackham OJL, Gough J, et al. An atlas of human long non-coding RNAs with accurate 5 ends. Nature. 2017;543:199–204.

8. Liu J, Jung C, Xu J, Wang H, Deng S, Bernad L, et al. Genome-Wide Analysis Uncovers Regulation of Long Intergenic Noncoding RNAs in Arabidopsis. Plant Cell [Internet]. 2012;24:4333–45. Available from: http://www.plantcell.org/cgi/doi/10.1105/tpc.112.102855

9. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure,

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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 lncRNAs 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: 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

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

JP, Igartua E, et al. Analysis of Plant Pan-Genomes and Transcriptomes with GET_HOMOLOGUES-EST, a Clustering Solution for Sequences of the Same Species. Front Plant Sci. 2017;8.

26. Vernikos G, Medini D, Riley DR, Tettelin H. Ten years of pan-genome analyses. Curr Opin Microbiol. 2015;23:148–54.

27. Clark RM, Schweikert G, Toomajian C, Ossowski S, Zeller G, Shinn P, et al. Common sequence polymorphisms shaping genetic diversity in Arabidopsis thaliana. Science (80-) [Internet]. 2007;317:338–42. Available from: http://www.sciencemag.org/cgi/doi/10.1126/science.1138632

28. Consortium T 1001 G. 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. Cell. 2016;166:481–91.

29. Zhao Q, Feng Q, Lu H, Li Y, Wang A, Tian Q, et al. Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. Nat Genet. 2018;50:278–84.

30. Kutter C, Watt S, Stefflova K, Wilson MD, Goncalves A, Ponting CP, et al. Rapid Turnover of Long Noncoding RNAs and the Evolution of Gene Expression. PLoS Genet. 2012;8:e1002841.

31. Necsulea A, Soumillon M, Warnefors M, Liechti A, Daish T, Zeller U, et al. The evolution of lncRNA repertoires and expression patterns in tetrapods. Nature. 2014;505:635–40.

32. Nelson ADL, Devisetty UK, Palos K, Haug-Baltzell AK, Lyons E, Beilstein MA. Evolinc: A tool for the identification and evolutionary comparison of long intergenic

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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, Svistoonoff 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. Svistoonoff 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/doifinder/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, Dartevelle 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

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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, Bezdan 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.

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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 lncRNA identification and characterization based on multiple features. Nucleic Acids
1010 Res. 2017;45:e2.

1011 52. Hupalo 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.

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

55. Rigaill G, Balzergue S, Brunaud V, Blondet E, Rau A, Rogier O, et al. Synthetic data sets for the identification of key ingredients for RNA-seq differential analysis. Brief Bioinform. 2018;19:65–76.

56. Park BS, Seo JS, Chua N-H. NITROGEN LIMITATION ADAPTATION Recruits PHOSPHATE2 to Target the Phosphate Transporter PT2 for Degradation during the Regulation of Arabidopsis Phosphate Homeostasis. Plant Cell. 2014;26:454–64.

57. Bayle V, Arrighi J-FJ-F, Creff A, Nespoulous C, Vialaret J, Rossignol M, et al. Arabidopsis thaliana High-Affinity Phosphate Transporters Exhibit Multiple Levels of Posttranslational Regulation. Plant Cell. 2011;23:1523–35.

58. Nussaume L, Kanno S, Javot H, Marin E, Pochon N, Ayadi A, et al. Phosphate import in plants: focus on the PHT1 transporters. Front Plant Sci. 2011;2.

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. Identification of Downstream Components of Ubiquitin-Conjugating Enzyme PHOSPHATE2 by Quantitative Membrane Proteomics in Arabidopsis Roots. Plant Cell. 2013;25:4044–60.

60. Ayadi A, David P, Arrighi J-F, Chiarenza S, Thibaud M-C, Nussaume L, et al. Reducing the Genetic Redundancy of Arabidopsis PHOSPHATE TRANSPORTER1 Transporters to Study Phosphate Uptake and Signaling. Plant Physiol. 2015;167:1511–26.

61. Shin H, Shin HS, Dewbre GR, Harrison MJ. Phosphate transport in Arabidopsis: Pht1;1 and Pht1;4 play a major role in phosphate acquisition from both low- and high-phosphate environments. Plant J. 2004;39:629–42.

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

62. Yuan J, Zhang Y, Dong J, Sun Y, Lim BL, Liu D, et al. Systematic characterization of novel lncRNAs responding to phosphate starvation in Arabidopsis thaliana. BMC Genomics. 2016;17.

63. Wang X-J, Gaasterland T, Chua N-H. Genome-wide prediction and identification of cis-natural antisense transcripts in Arabidopsis thaliana. Genome Biol [Internet]. 2005;6:R30. Available from: http://genomebiology.com/2005/6/4/R30

64. Petricka JJ, Winter CM, Benfey PN. Control of *Arabidopsis* Root Development. Annu Rev Plant Biol [Internet]. 2012;63:563–90. Available from: http://www.annualreviews.org/doi/10.1146/annurev-arplant-042811-105501

65. Pickrell JK. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. Am J Hum Genet. 2014;94:559–73.

66. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res. 2013;42:D1001–6.

67. Wang YM, Xu HB, Wang MS, Otecko NO, Ye LQ, Wu DD, et al. Annotating long intergenic non-coding RNAs under artificial selection during chicken domestication. BMC Evol Biol. 2017;17.

68. Ning S, Yue M, Wang P, Liu Y, Zhi H, Zhang Y, et al. LincSNP 2.0: An updated database for linking disease-associated SNPs to human long non-coding RNAs and their TFBSs. Nucleic Acids Res. 2017;45:D74–8.

69. Giral H, Landmesser U, Kratzer A. Into the Wild: GWAS Exploration of Non-coding RNAs. Front Cardiovasc Med. 2018;5.

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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.

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

79. Li L, Eichten SR, Shimizu R, Petsch K, Yeh CT, Wu W, et al. Genome-wide discovery and characterization of maize long non-coding RNAs. Genome Biol. 2014;15:R40.

80. Schnable PS, Jia Y, DeCook R, Swanson-Wagner RA, Borsuk LA, Nettleton D. All possible modes of gene action are observed in a global comparison of gene expression in a maize F1 hybrid and its inbred parents. Proc Natl Acad Sci. 2006;103:6805–10.

81. Lin SI, Chiang SF, Lin WY, Chen JW, Tseng CY, Wu PC, et al. Regulatory Network of MicroRNA399 and PHO2 by Systemic Signaling. Plant Physiol. 2008;147:732–46.

82. Lin SI, Chiou TJ. Long-distance movement and differential targeting of microRNA399s. Plant Signal Behav. 2008;3:730–2.

83. Shah S, Wittmann S, Kilchert C, Vasiljeva L. lncRNA recruits RNAi and the exosome to dynamically regulate pho1 expression in response to phosphate levels in fission yeast. Genes Dev. 2014;28:231–44.

84. Ard R, Tong P, Allshire RC. Long non-coding RNA-mediated transcriptional interference of a permease gene confers drug tolerance in fission yeast. Nat Commun. 2014;5.

85. Thieme CJ, Rojas-Triana M, Stecyk E, Schudoma C, Zhang W, Yang L, et al. Endogenous Arabidopsis messenger RNAs transported to distant tissues. Nat Plants. 2015;1:15025.

86. Torii KU. The Arabidopsis ERECTA Gene Encodes a Putative Receptor Protein Kinase with Extracellular Leucine-Rich Repeats. Plant Cell Online [Internet].

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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: 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

bioRxiv preprint doi: https://doi.org/10.1101/579656; this version posted March 16, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15.

96. Pound MP, Wells DM, Atkinson JA, Bennett MJ, French AP, Pridmore T. RootNav: Navigating Images of Complex Root Architectures. Plant Physiol. 2013;162:1802–14.

97. R Core Team. R: A language and environment for statistical computing. http://www.R-project.org/. R Found. Stat. Comput. Vienna, Austria. 2017.

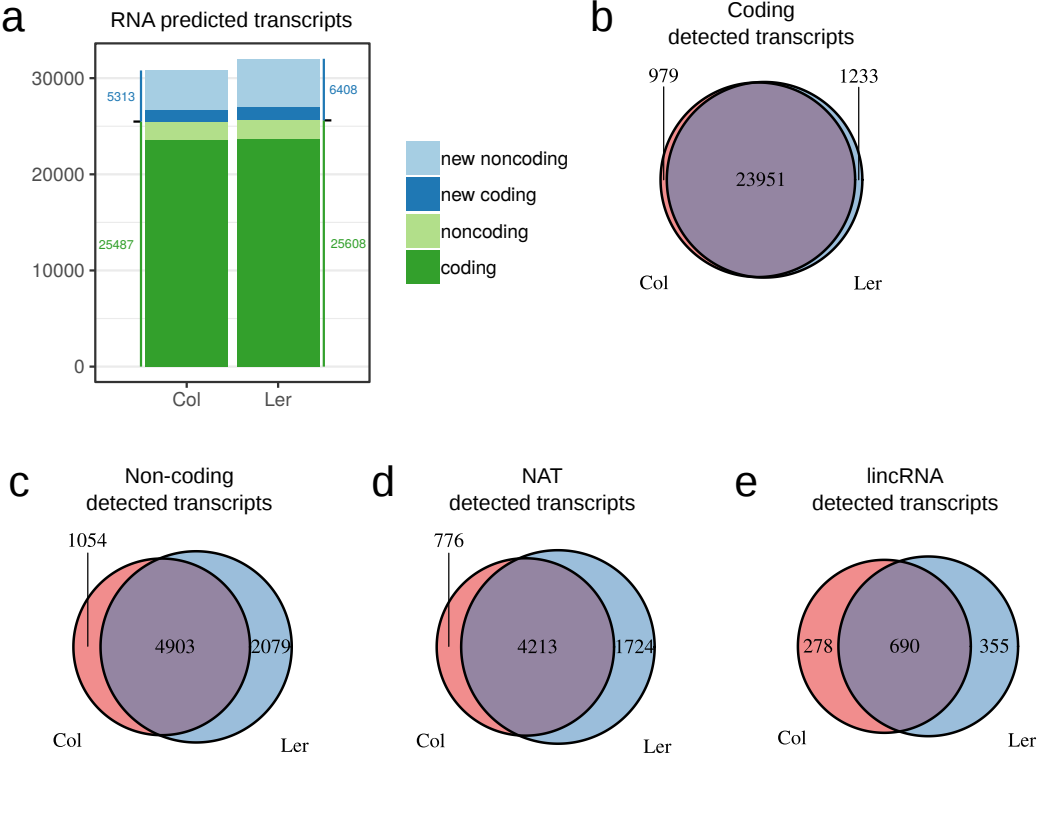
98. Wickham H. tidyverse: Easily Install and Load the "Tidyverse" [Internet]. 2017. Available from: https://cran.r-project.org/package=tidyverse

99. Lenth R. emmeans: Estimated Marginal Means, aka Least-Squares Means [Internet]. 2019. Available from: https://cran.r-project.org/package=emmeans

100. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002;30:207–10.

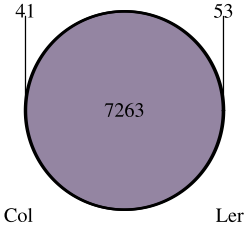
101. Balzergue C, Gabriel M, Taconnat L, Balzergue S, Blein T. Non-coding RNAs in the adaptation of root growth to phosphate starvation [Internet]. [cited 2019 Mar 15]. Available from: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE128250

102. Balzergue C, Gabriel M, Taconnat L, Balzergue S, Blein T. Non-coding RNAs in the adaptation of root growth to phosphate starvation [miRNA-Seq] [Internet]. [cited 2019 Mar 15]. Available from: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE128256


a
RNA predicted transcripts
30000
20000
10000
0
5313
6408
25487
25608
Col
Ler
new noncoding
new coding
noncoding
coding
b
Coding
detected transcripts
979
1233
23951
Col
Ler
c
Non-coding
detected transcripts
1054
4903
2079
Col
Ler
d
NAT
detected transcripts
776
4213
1724
Col
Ler
e
lincRNA
detected transcripts
278
690
355
Col
Ler

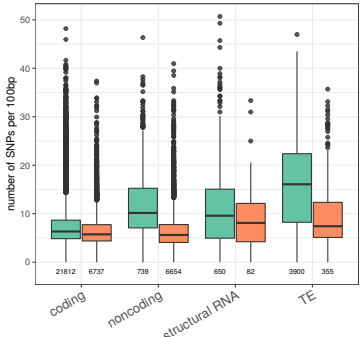
a

Presence of new predicted transcripts at DNA level



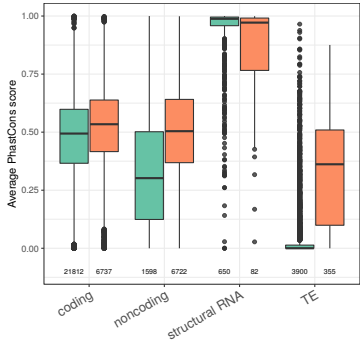
b

SNPs accumulation in all predicted transcripts

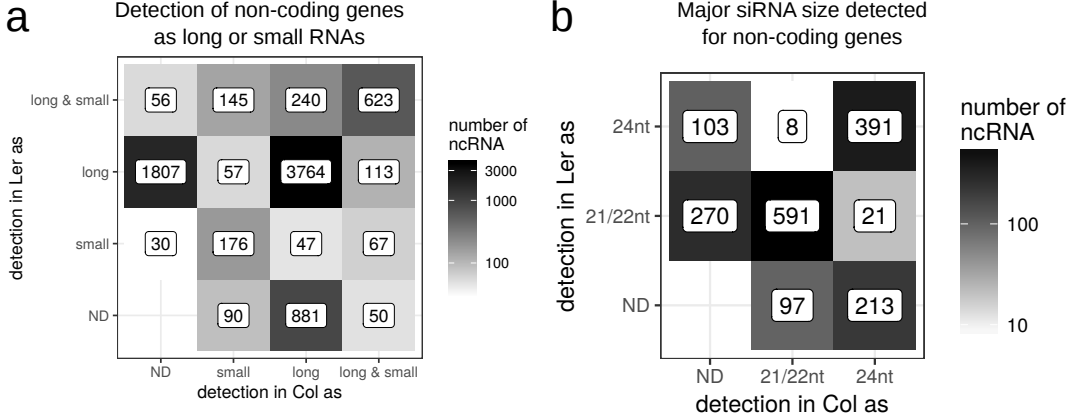


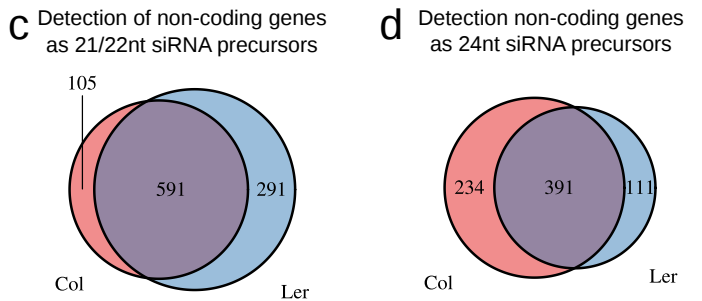
c

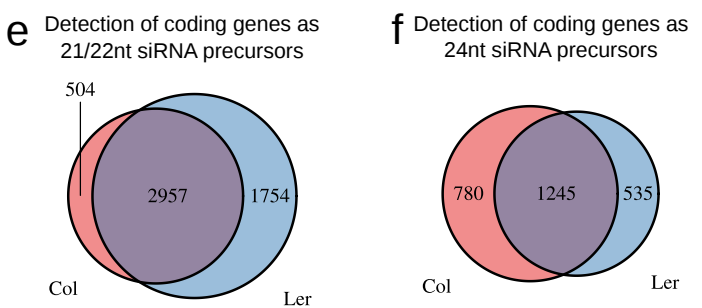
Conservation of all predicted transcripts

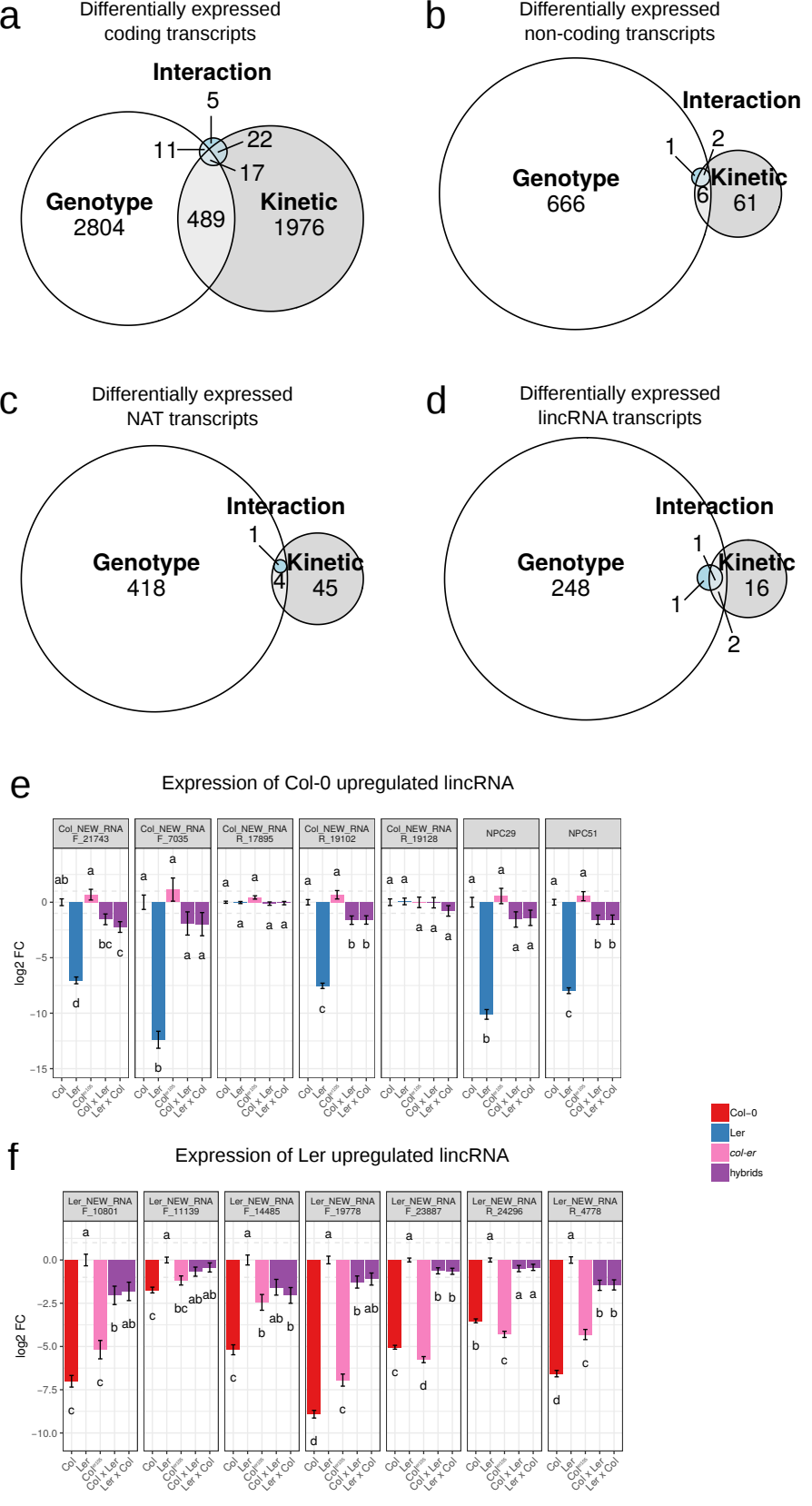


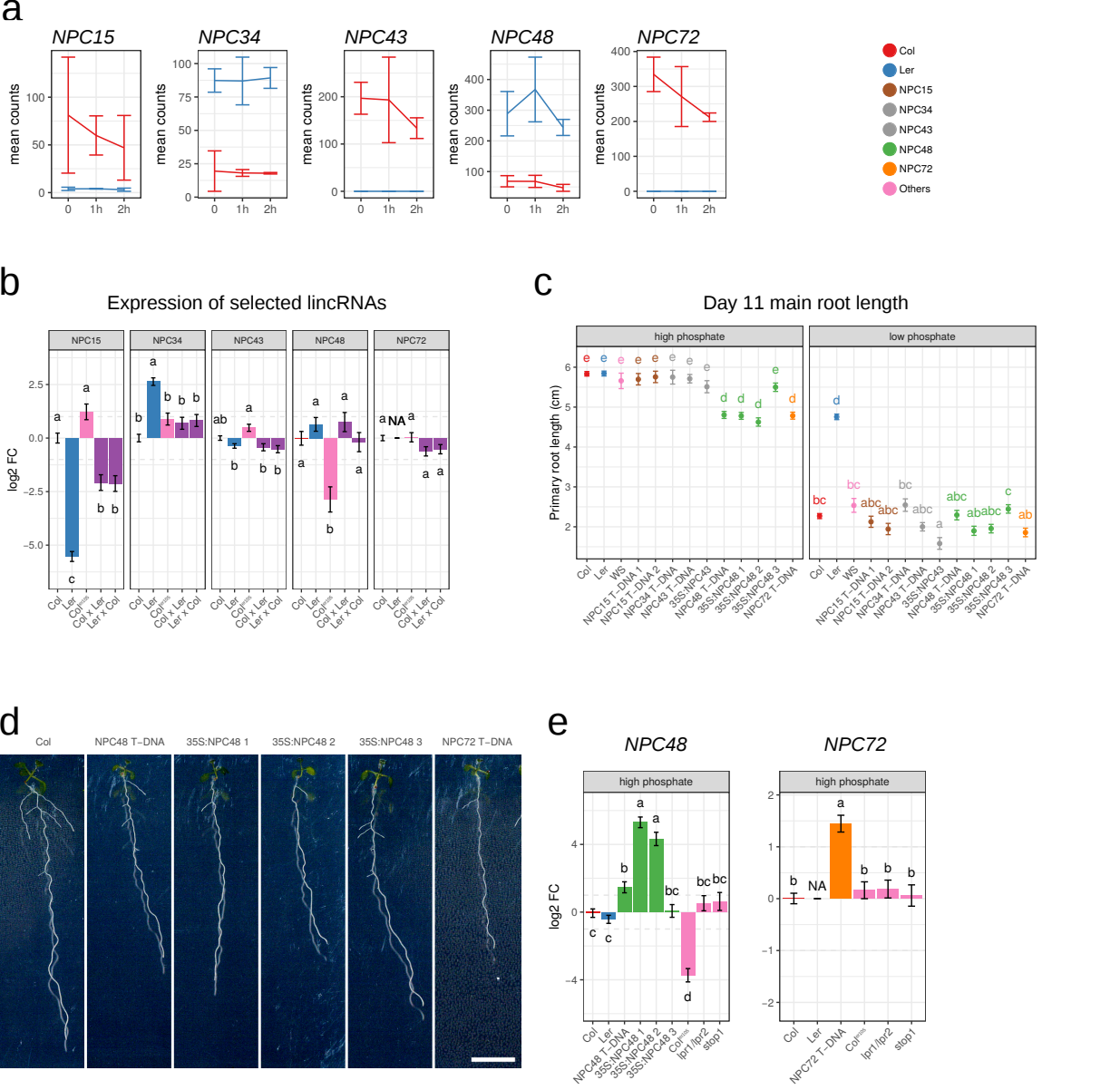
Non NAT NAT


a
Detection of non-coding genes as long or small RNAs
long & small
long
small
ND
56
145
240
623
1807
57
3764
113
30
176
47
67
90
881
50
detection in Ler as
ND
small
long
long & small
detection in Col as
number of ncRNA
3000
1000
100
b
Major siRNA size detected for non-coding genes
24nt
21/22nt
ND
103
8
391
270
591
21
97
213
detection in Ler as
ND
21/22nt
24nt
detection in Col as
number of ncRNA
100
10


c
Detection of non-coding genes as 21/22nt siRNA precursors
105
591
291
Col
Ler
d
Detection non-coding genes as 24nt siRNA precursors
234
391
111
Col
Ler


e
Detection of coding genes as 21/22nt siRNA precursors
504
2957
1754
Col
Ler
f
Detection of coding genes as 24nt siRNA precursors
780
1245
535
Col
Ler


a
Differentially expressed
coding transcripts
Interaction
5
11
22
17
Genotype
2804
489
Kinetic
1976
b
Differentially expressed
non-coding transcripts
Interaction
1
2
Genotype
666
6
Kinetic
61
c
Differentially expressed
NAT transcripts
Interaction
1
Genotype
418
4
Kinetic
45
d
Differentially expressed
lincRNA transcripts
Interaction
1
1
Genotype
248
Kinetic
16
2
e
Expression of Col-0 upregulated lincRNA
Col_NEW_RNA F_21743
Col_NEW_RNA F_7035
Col_NEW_RNA R_17895
Col_NEW_RNA R_19102
Col_NEW_RNA R_19128
NPC29
NPC51
log2 FC
Col
Ler
Col^res
Col x Ler
Ler x Col
Col−0
Ler
col-er
hybrids
f
Expression of Ler upregulated lincRNA
Ler_NEW_RNA F_10801
Ler_NEW_RNA F_11139
Ler_NEW_RNA F_14485
Ler_NEW_RNA F_19778
Ler_NEW_RNA F_23887
Ler_NEW_RNA R_24296
Ler_NEW_RNA R_4778
log2 FC


a
NPC15
NPC34
NPC43
NPC48
NPC72
mean counts
0
1h
2h
Col
Ler
NPC15
NPC34
NPC43
NPC48
NPC72
Others
b
Expression of selected lincRNAs
log2 FC
c
Day 11 main root length
high phosphate
low phosphate
Primary root length (cm)
d
Col
NPC48 T−DNA
35S:NPC48 1
35S:NPC48 2
35S:NPC48 3
NPC72 T−DNA
e
NPC48
NPC72
log2 FC