1 **The transcription and export complex THO/TREX contributes to**

2 **transcription termination in plants**

3

4

5

6 Ghazanfar Abbas Khan[1,2], Jules Deforges[1], Rodrigo S. Reis[1], Yi-Fang Hsieh[1], Jonatan

7 Montpetit[1], Wojciech Antosz[3], Luca Santuari[1], Christian S Hardtke[1], Klaus Grasser[3] and Yves

8 Poirier[1]

9

10 [1]Department of Plant Molecular Biology, University of Lausanne, Switzerland.

11 [2]School of Biosciences, University of Melbourne, VIC, Australia.

12 [3] Department of Cell Biology & Plant Biochemistry, Biochemistry Centre, University of

13 Regensburg, Universitätsstr. 31, D-93053 Regensburg, Germany

14

15 **Corresponding authors:**

16 Ghazanfar Abbas Khan, ghazanfar.khan@unimelb.edu.au

17 Yves Poirier; yves.poirier@unil.ch

18

19 **Short title:** The THO/TREX complex participates in transcription termination

1

20 **Abstract**

21   Transcription termination has important regulatory functions, impacting mRNA stability,

22   localization and translation potential. Failure to appropriately terminate transcription can also

23   lead to read-through transcription and the synthesis of antisense RNAs which can have

24   profound impact on gene expression. The Transcription-Export (THO/TREX) protein complex

25   plays an important role in coupling transcription with splicing and export of mRNA. However,

26   little is known about the role of the THO/TREX complex in the control of transcription

27   termination. In this work, we show that two proteins of the THO/TREX complex, namely

28   TREX COMPONENT 1 (TEX1 or THO3) and HYPER RECOMBINATION1 (HPR1 or

29   THO1) contribute to the correct transcription termination at several loci in *Arabidopsis*

30   *thaliana*. We first demonstrate this by showing defective termination in *tex1* and *hpr1* mutants

31   at the nopaline synthase (NOS) terminator present in a T-DNA inserted between exon 1 and 3

32   of the *PHO1* locus in the *pho1-7* mutant. Read-through transcription beyond the NOS

33   terminator and splicing-out of the T-DNA resulted in the generation of a near full-length *PHO1*

34   mRNA (minus exon 2) in the *tex1 pho1-7* and *hpr1 pho1-7* double mutants, with enhanced

35   production of a truncated PHO1 protein that retained phosphate export activity. Consequently,

36   the strong reduction of shoot growth associated with the severe phosphate deficiency of the

37   *pho1-7* mutant was alleviated in the *tex1 pho1-7* and *hpr1 pho1-7* double mutants. Additionally,

38   we show that RNA termination defects in *tex1* and *hpr1* mutants leads to 3'UTR extensions in

39   several plant genes. These results demonstrate that THO/TREX complex contributes to the

40   regulation of transcription termination.

41

42 **Author summary**

43 Production of messenger RNAs (mRNAs) involves numerous steps including initiation of

44 transcription, elongation, splicing, termination, as well as export out of the nucleus. All these

45 steps are highly coordinated and failure in any steps has a profound impact on the level and

46 identity of mRNAs produced. The THO/TREX protein complex is associated with nascent

47 RNAs and contributes to several mRNA biogenesis steps, including splicing and export.

48 However, the contribution of the THO/TREX complex to mRNA termination was poorly

49 defined. We have identified a role for two components of the THO/TREX complex, namely the

50 proteins TEX1 and HPR1, in the control of transcription termination in the plant *Arabidopsis*

51 *thaliana*. We show that the *tex1* and *hpr1* mutants have defects in terminating mRNA at the

52 nopaline synthase (NOS) terminator found in T-DNA insertion mutants leading to the

53 transcriptional read-through pass the NOS terminator. We also show that *tex1* and *hpr1* mutants

54 have defects in mRNA termination at several endogenous genes, leading to the production of

55 3'UTR extensions. Together, these results highlight a role for the THO/TREX complex in

56 mRNA termination.

57

58

Introduction

In eukaryotes, mRNAs are generated by several dynamic and coordinated processes including transcriptional initiation, elongation and termination, as well as splicing and nuclear export. Failure in any of these processes has a profound impact on the level and identity of transcripts [1-3]. These transcriptional steps are sequentially orchestrated by a multitude of RNA-binding protein complexes that co-transcriptionally couple with the nascent RNA [4]. For example, protein complex required for transcription termination cleaves pre-mRNA close to RNA polymerase II (RNAPII) and adds a poly (A) tail to the 3'end of nascent RNA [3]. Pre-mRNA cleavage exposes 5'end of nascent mRNA to 5'-3' exonucleases which degrades the RNA attached to RNAPII, leading to transcription termination [5].

Cleavage and polyadenylation define the transcription termination at a given locus and has a decisive role in regulating gene expression as it can influence stability and translation potential of the RNA via the inclusion of regulatory sequence elements [6]. Moreover, transcription termination avoid interference with the transcription of downstream genes and facilitates RNAPII recycling [7]. It also prevents synthesis of antisense RNAs which can have a severe effect on RNA production and overall gene expression [8]. Additional regulatory role of transcription termination is the synthesis of chimeric transcripts formed by tethering of two neighboring genes on the same chromosomal strand [8]. Considering its importance, molecular mechanisms which regulate RNA termination are relatively poorly understood.

After transcription termination, nascent RNA is assembled in a ribonucleoprotein complex is delivered for RNA export into the cytosol. Similar to other steps in RNA biogenesis which are closely coupled in a sequential manner, it is likely that transcription termination is associated with nuclear export of RNAs. The TREX (TRanscription-EXport) protein complex has emerged as an important component in coupling transcription with RNA processing and export [9]. In metazoans, TREX consists of THO core complex, which includes THO1/HPR1, THO2, THO3/TEX1, THO5, THO6 and THO7 [10]. The proteins associating with the THO core components and forming the TREX complex include the RNA helicase and splicing factor DDX39B (SUB2 in yeast) as well as the RNA export adaptor protein ALY (YRA1 in yeast) [10]. TREX is co-transcriptionally recruited to the nascent mRNA and regulates splicing, elongation and export [11]. Moreover, THO/TREX complex is required for the genetic stability as it is required for preventing DNA:RNA hybrids that lead to transcription impairment and are

92   responsible for genetic instability phenotypes observed in these mutants [12]. The genes
93   encoding the THO core components are conserved in plants, including in the model plant
94   *Arabidopsis thaliana* [13, 14]. *A. thaliana* mutants defective in THO components show a wide
95   range of phenotypes, from no obvious alteration and relatively mild phenotypes to lethal
96   phenotypes, suggesting overlapping but independent functions of these components [14-16]. *A.*
97   *thaliana* mutants in THO components, including HPR1, THO2 and TEX1, show defects in
98   small RNA biogenesis, mRNA elongation, splicing and export, but no defect in mRNA
99   termination has been reported [14-20].

100

101   In this work, we explored the regulatory function of two components of the THO/TREX
102   complex, namely TEX1 and HPR1, in transcription termination in *A. thaliana*. We show that
103   the *tex1* and *hpr1* mutants are defective in RNA termination at the Nopaline Synthase (NOS)
104   terminator present on a T-DNA inserted in the *PHO1* locus. Additionally, genome-wide
105   analysis of mRNAs revealed RNAPII termination defects in *tex1* and *hpr1* mutants at several
106   loci leading to the 3'UTR extensions.

107

108 **Results**

109 **A forward genetic screen using the T-DNA insertion mutant *pho1-7* for reversion of the**
110 **growth phenotype identified the *TEX1* gene**

111 The *PHO1* gene encodes an inorganic phosphate (Pi) exporter involved in loading Pi into the
112 root xylem for its transfer to the shoot [21-24]. Consequently, *pho1* mutants in both *A. thaliana*
113 and rice have reduced shoot Pi contents and shows all symptoms associated with Pi deficiency,
114 including highly reduced shoot growth and the expression of numerous genes associated with
115 Pi deficiency [23, 25, 26]. However, it has previously been shown that low shoot Pi content can
116 be dissociated from its major effects on growth and other responses normally associated with
117 Pi deficiency through the modulation of *PHO1* expression or activity [25, 27]. We thus used
118 the *pho1* mutant as a tool, in a forward genetic screen, to identify mutants which restore *pho1*
119 shoot growth to wild type (Col-0) level, while maintaining low shoot Pi contents. We performed
120 ethyl methane sulphonate (EMS) mutagenesis on seeds of the *pho1-7* mutant, derived from the
121 SALK line 119520 containing a single T-DNA inserted in the *PHO1* gene in between the first
122 and third exon of the *PHO1* gene leading to the deletion of the second exon from the genome
123 (S1 Text). Screening of 10'300 mutagenized *pho1-7* plants grown in soil for improved rosette
124 growth lead to the isolation of a suppressor mutant and had rosette growth similar to Col-0
125 while maintaining a low shoot Pi content similar to *pho1-7* (Fig 1AB) (see Material and
126 methods for further detail). Both *pho1-7* and the suppressor mutant maintained resistance to
127 kanamycin associated with the T-DNA insertion in *PHO1*.

128 Mapping-by-sequencing revealed that the mutation C116T is introduced into the *TEX1* gene in
129 the *pho1-7* suppressor mutant. This leads to a conversion of amino acid serine 39 to
130 phenylalanine in the TEX1 protein. Transformation of *TEX1* gene into *pho1-7* suppressor led
131 to *pho1*-like phenotype, confirming that mutation in *tex1* was the causal mutation for restoration
132 of *pho1-7* shoot growth (Fig 1AB). Furthermore, crossing of a T-DNA allele *tex1-4*
133 (SALK_100012) to *pho1-7* also resulted in the suppression of *pho1-7* shoot growth phenotype
134 (Fig 2A), further confirming that mutation in *TEX1* is responsible for the suppression of shoot
135 growth phenotype in *pho1-7 suppressor* mutant. Therefore, we named this new S39P mutant in
136 the *TEX1* gene as *tex1-6*. In agreement with previous reports, TEX1 protein was localized to
137 the nucleus [19] (S1A Fig). *TEX1* promoter fusion with GUS showed that *TEX1* is expressed
138 in root, cotyledon and rosette (S1B Fig).

139

6

140   The *pho1-7 tex1-6* mutant shows Col-0-like shoot growth while maintaining a low Pi content

141   that is only slightly higher to the parental *pho1-7* (Fig 1A, B). A key molecular response of

142   *pho1* mutants is the manifestation of gene expression and lipid profiles in the shoots that are

143   associated with strong Pi deficiency [25]. To determine if the *tex1-6* mutation can also suppress

144   the induction of Pi starvation responses (PSR) in the rosettes of *pho1-7 tex1-6*, we performed

145   quantitative RT-PCR (qRT-PCR) to see the expression of PSR genes. *pho1-7 tex1-6* shoots

146   showed an expression profile of PSR genes that was comparable to Pi-sufficient Col-0 plants

147   (S2A Fig). Additionally, lipid analysis in *pho1-7* mutants showed a decrease of phospholipids

148   and an increase in galactolipids expected for Pi-deficient plants, while *pho1-7 tex1-6* plants

149   showed lipids profiles similar to Col-0 plants (S2B Fig), confirming that the *tex1-6* mutation

150   suppressed morphological as well as molecular response to Pi deficiency displayed by the *pho1-*

151   *7* mutant.

152

153

154   **Mutation of *TEX1* in *pho1-7* resulted in the synthesis of a truncated PHO1 protein**

155   To determine if the *tex1* mutation also suppresses the growth phenotype associated with other

156   *pho1* alleles generated by EMS mutagenesis, a double mutant *pho1-4 tex1-4* was generated.

157   Surprisingly, *pho1-4 tex1-4* double mutant showed only minor improvement in shoot growth

158   and maintained low shoot Pi content (Fig 2A, B). In order to understand how the *tex1* mutation

159   can result in restoration of *pho1-7* shoot growth, we performed a detailed analysis of transcripts

160   produced at the *PHO1* locus in the *pho1-7 tex1-4* mutant. Interestingly, we identified a

161   truncated $PHO1^{\Delta249\text{-}342}$ transcript which only lacked the 2$^{\text{nd}}$ exon suggesting that the T-DNA is

162   spliced out from the mature mRNA (Fig 2C, S1 Text). The mRNA produced is in frame and

163   resulted in the production a truncated $PHO1^{\Delta84\text{-}114}$ protein (S1 Text). Western blot experiments

164   confirmed the presence of a $PHO1^{\Delta84\text{-}114}$ truncated protein in both *pho1-7* and *pho1-7 tex1-4*

165   roots with a strong increase of expression in the *pho1-7 tex1-4* double mutant as compared to

166   *pho1-7* (Fig 2D). This increase in protein quantity can be attributed to an increase in expression

167   of $PHO1^{\Delta249\text{-}342}$ RNA in the *pho1-7 tex1-4* double mutant (Fig 2E). We hypothesized that the

168     $PHO1^{\Delta84\text{-}114}$ protein variant was at least partially active as a Pi exporter and that its increased

169   expression in *pho1-7 tex1-4* partially restored PHO1 function, resulting in an improvement of

170   the shoot growth phenotype. We confirmed this hypothesis by expressing the $PHO1^{\Delta84\text{-}114}$

171   variant and the wild type *PHO1* fused to GFP using the *PHO1* promoter in the *pho1-4* null

172   mutant. As expected, *pho1-4* mutant which expressed the wild type PHO1 fully complemented

7

173   the growth and Pi content to Col-0 level (Fig 3A, B). However, plants expressing the PHO1$^{\Delta 84\text{-}}$

174   ^{114} variant only restored the shoot growth phenotype while maintaining low Pi contents

175   comparable to *pho1-7 tex1-4* plants (Fig 3A, B). Confocal analysis of roots showed that

176   PHO1$^{\Delta 84\text{-}114}$ variant protein was localized similarly to wild type PHO1 (Fig 3C), which was

177   previously shown to be primarily in the Golgi and *trans*-Golgi network (TGN) [21].

178   Furthermore, transient expression of the PHO1$^{\Delta 83\text{-}114}$-GFP fusion in *Nicotiana benthamiana*

179   leaves led to Pi export to the apoplastic space, demonstrating that the protein was competent in

180   Pi export (Fig 3D) [21]. Collectively, these results confirmed that restored expression of

181   *PHO1*$^{\Delta 249\text{-}342}$ RNA is responsible for the improved shoot growth in *pho1-7 tex1-4* double

182   mutants.

183

184   **Mutation in HPR1, another components of THO/TREX complex can also restore *pho1-7***

185   **shoot growth but not mutations affecting tasiRNA biogenesis**

186   To investigate if TEX1 exerts its function in the restoration of *PHO1* expression via the

187   THO/TREX complex, we crossed *pho1-7* to *hpr1-6* which is a mutant in another component of

188   THO/TREX core complex [14]. Double mutant *pho1-7 hpr1-6* partially restored shoot growth

189   while maintaining relatively low Pi contents comparable to the *pho1-7 tex1-6* but slightly higher

190   than *pho1-7* (Fig 1B and Fig 4 A-B).

191

192   THO/TREX complex has previously been demonstrated to participate in the biogenesis of trans

193   acting small interfering RNAs (tasiRNAs) and other small RNAs (siRNAs and miRNAs) that

194   can affect levels of transcription through DNA methylation and some unknown mechanisms.

195   We explored the possibility that changes in the biogenesis of tasiRNAs may be responsible for

196   the growth phenotype associated with the *pho1-7 tex1-4* and *pho1-7 hpr1-6* mutants. We

197   crossed *pho1-7* with two mutants in genes encoding core components required for the

198   biogenesis of tasiRNAs, namely *rdr6-11* and *sgs3-1* [28, 29]. Double mutants *pho1-7 rdr6-11*

199   and *pho1-7 sgs3-1* had shoot growth similar to the parental *pho1-7* (Fig 4C-D). Together, these

200   results indicate disruption in distinct genes of the THO/TREX complex, namely *TEX1* and

201   *HPR6*, can revert the growth phenotype of the *pho1-7* mutant and that biogenesis of tasiRNAs

202   is not implicated in these processes.

203

204

205

Impaired mRNA termination at the NOS terminator restores expression of truncated PHO1 in *pho1-7 tex1-4* mutant

To elucidate how mutations in *TEX1* and *HPR1* lead to changes in transcription at the *PHO1* locus, we performed paired-end next generation RNA sequencing of Col-0, *pho1-7*, *pho1-7 tex1-4* and *pho1-7 hpr1-6* mutants from roots. We first mapped the RNA reads of Col-0 and *pho1-7* against the Col-0 genome and confirmed the absence of the second exon of *PHO1* in the genome of *pho1-7* (S3A Fig). To understand the transcription dynamics at the *PHO1* locus in the various mutants derived from *pho1-7*, we mapped RNA sequencing reads to the *pho1-7* genomic configuration with the T-DNA insertion and exon 2 deletion. Detailed analysis of mRNAs from *PHO1* locus in *pho1-7* mutants indicated that transcription was initiated in the *PHO1* promoter and terminated at two different locations, namely at *NOS* terminator inside the T-DNA and at the endogenous *PHO1* transcription termination site (Fig 5A-C). Using RT-PCR and various primer combinations, we could detect four types of transcripts in the *pho1-7* mutant, namely one unspliced and two spliced mRNA ending at the *NOS* terminator, and one long transcript ending at the endogenous *PHO1* terminator and where *PHO1* exons 1 and 3 were appropriately spliced, removing the T-DNA and generating the $PHO1^{\Delta249\text{-}342}$ RNA variant (Figure 5A-D). While in the *pho1-7* mutant the four transcripts were expressed at similar low level, in *pho1-7 tex1-4* and *pho1-7 hpr1-6* double mutants the majority of transcripts was the $PHO1^{\Delta249\text{-}342}$ RNA variant (Fig 5C-D). Analysis by PacBio sequencing of full-length mRNAs produced at the *PHO1* locus in Col-0, *pho1-7*, *pho1-7 tex1-4* and *pho1-7 hpr1-6* supported to these conclusions and highlighted that essentially two classes of transcripts are produced in the various mutants, namely transcripts that include the 5'portion of the T-DNA and end at the NOS terminator and transcripts that end at the *PHO1* terminator and exclude the complete T-DNA (S3B Fig). While in the *pho1-7* mutant the majority of transcripts were of the first type, the *pho1-7 tex1-4* and *pho1-7 hpr1-6* mutants mostly expressed the second type. Such pattern of transcripts are not consistent with alternative splicing but rather indicate that transcription termination at the *NOS* terminator was suppressed in *pho1-7 tex1-4* and *pho1-7 hpr1-6* mutants, and this enabled the transcription machinery to reach the *PHO1* terminator and generate a transcript where exon 1 was spliced to exon 3, resulting in the removal of the T-DNA. Chromatin immunoprecipitation using an antibody against the elongating RNAPII (phosphorylated at S2 of the C-terminal domain) followed by qPCR showed that RNAPII occupation at the *PHO1* locus situated after the T-DNA insertion was significantly reduced in *pho1-7* mutants but increased in *pho1-7 tex1-4* mutant (Fig 5E), consistent with an increase in transcriptional read-through past the T-DNA in the *pho1-7 tex1-4* double mutant (Fig 5D).

9

240

THO/TREX complex is required for the correct termination of mRNAs in endogenous loci

243   To understand if TEX1 and HPR1 contribute to the termination of RNA transcription at
244 endogenous genes, RNA sequencing data generated from roots of Col-0, *pho1-7*, *pho1-7 tex1-*
245 *4* and *pho1-7 hpr1-6* grown in soil for 3 weeks were analyzed for the presence of 3' UTR
246 extensions. We observed significant changes in 3'UTR extensions in *pho1-7 tex1-4* and *pho1-*
247 *7 hpr1-6* mutants as compared to both Col-0 and *pho1-7*. Two examples of such 3'UTR
248 extensions are shown in Fig 6A. While 3'UTR extensions were observed in only 3 transcripts
249 in *pho1-7*, 72 and 51 transcripts showed 3'UTR extensions in the *pho1-7 tex1-4* and *pho1-7*
250 *hpr1-6* mutants, respectively, but not in the *pho1-7* parent, with a subset of 38 transcripts found
251 in common between *pho1-7 tex1-4* and *pho1-7 hpr1-6* but not *pho1-7* (Fig 6B). These results
252 indicate that while a large proportion of genes affected in their 3'UTR extension in the *tex1-4*
253 respond similarly in the *hpr1-6*, the effects of these two mutations on RNA termination are not
254 completely redundant.

255   We hypothesized that if regulation of RNA termination by THO/TREX complex is generic and
256 robust, changes in 3'UTR extensions should be relatively insensitive to growth conditions. To
257 assess the robustness of 3'UTR extensions, we analyzed an independent RNA sequencing
258 dataset generated from roots of Col-0 and *tex1-4* mutant grown *in vitro* for 7 days in MS
259 medium supplemented with sucrose. A total of 77 transcripts showed 3'UTR extensions in the
260 *tex1-4* mutant relative to Col-0, with 48 transcripts found also in the dataset of 3'UTR
261 extensions for *pho1-7 tex1-4* mutant grown for 3 weeks in soil (S4 Fig), indicating that a large
262 proportion of transcripts with 3'UTR extensions observed in the *tex1-4* genetic background are
263 insensitive to major changes in growth conditions.

264   Validation of 3'UTR extensions in a set of genes identified by RNA sequencing was first
265 performed by qPCR using (Fig 6C). A transient assay was also developed whereby the sequence
266 500 bp downstream of the stop codon (which includes the 3'UTR) of two genes, AT1G76560
267 and AT1G03160, was fused after the stop codon of the nano-luciferase gene. These constructs
268 were expressed in Arabidopsis mesophyll protoplasts and the ratio of transcripts with an
269 extended 3'end relative to the main transcription termination site was determined by qRT-PCR
270 16 hours after transformation. Analysis revealed an increase, for both constructs, in the ratio of
271 long-to-short transcripts by approximately 50-60% in the *hpr1-6* and *tex1-4* mutants compared

272   to Col-0 (Fig 6D), further supporting the implication of both TEX1 and HPR1 in mRNA

273   termination.

274   GO term enrichment analysis of transcripts with impaired RNA termination in the *tex1-4,*

275   *pho1-7 tex1-4* and *pho1-7 hpr1-6* showed that these transcripts did not belong to a particular

276   functional category (S5 Fig). Therefore, we looked at the sequences of RNA termination sites

277   for mechanistic clues of impaired RNA termination and 3'UTR extension in *tex1* and *hpr1*

278   mutants. RNA termination sites are defined characteristic motifs, including an A-rich region at

279   approximately -20 nucleotides (position -1 being defined as the last nucleotide before the polyA

280   tail), which includes the AAUAAA-like sequence. This is followed by a U-rich region at -7

281   nucleotide and a second peak of U-rich region at approximately +25 nucleotides [30-32].

282   Analysis of the distribution of nucleotides upstream and downstream of the 3' cleavage site did

283   not reveal a significant difference from this distribution for genes showing a 3'extension in the

284   *tex1-4* and *hpr1-6* mutant backgrounds (Fig 7A). Additionally, we analyzed the differences at

285   -20 polyadenylation signal for the genes with 3' extensions compared to all Arabidopsis genes.

286   Although not statistically significant (p=0.22), a lower representation of the canonical

287   AAUAAA sequence appeared associated with the group of genes with 3' extension compared

288   to all genes (Fig 7B). We used the transient expression to test the effect of changing the single

289   AAUGAA polyadenylation signal present in gene AT1G76560 to the canonical AAUAAA.

290   While the optimized polyadenylation signal led to a small decrease in 3'UTR extensions

291   relative to the wild type sequence in Col-0, there was still an approximately 50% increase in

292   the ratio of long-to-short transcripts in the *hpr1-6* and *tex1-4* mutant compared to Col-0 (Fig

293   7C). Altogether, these results indicate that while HPR1 and TEX4 contribute to mRNA

294   termination, they do not appear to act primarily via the nature of the -20 polyadenylation signal.

11

Discussion

The contribution of the THO/TREX complex to mRNA biogenesis has been particularly studied for splicing and export [9]. In contrast, the direct implication of the THO/TREX complex in mRNA termination is poorly defined and reported only in few studies in human [33, 34]. The THO5 was shown to interact with both CPSF100 and CFIm, two proteins involved in 3'end-processing and polyadenylation site choice, and differences in mRNA polyadenylation were observed in human cells depleted for THO5 [33, 34]. Recruitment of the cyclin-dependent kinase CDK11 by the THO/TREX complex was shown to be essential for the phosphorylation of RNAPII and the proper 3'end processing of the human immunodeficiency virus RNA, although it is unknown if this interaction is also mediated via THO5 [35].

Although considerably less is known about mRNA biogenesis in plants compared to yeast and metazoans, proteins forming the THO core complex are also found in plants, implicating a conservation in their mode of action [13]. Indeed, *A. thaliana* mutants in the *HPR1* and *TEX1* genes show defects in mRNA splicing and export [18, 19, 36]. However, most of our knowledge in plants on THO core complex relates to its implication in small RNA biogenesis. *A. thaliana* mutants in either *HPR1*, *TEX1*, *THO2* or *THO6* are defective in the synthesis of one or multiple forms of small RNAs, including miRNA, siRNA, and tasiRNAs, although the mode of action behind this defect is currently unknown [14-17]. Some of the phenotypes associated with the *tex1* and *hpr1* mutants in *A. thaliana*, such as repression of female germline specification or reduction in scopolin biosynthesis under abiotic stress, are likely caused by a reduction in tasiRNA or miRNA production [37, 38].

This current work highlights the contribution of both TEX1 and HPR1 to mRNA termination. Mutations in these genes in a *pho1-7* mutant background having a T-DNA insertion between the *PHO1* exons 1 and 3 led to the suppression of mRNA termination at the NOS terminator present in the T-DNA, followed by transcriptional read-through and splicing of the T-DNA, resulting in the generation of a near full-length *PHO1* mRNA (minus exon 2). The truncated PHO1 protein generated from this mRNA maintained some Pi export activity, resulting in a reversion of the growth phenotype associated with the severe Pi deficiency of the *pho1-7* mutant.

Beyond its effect on the NOS terminator, mutations in the *HPR1* and *TEX1* genes also affected mRNA 3' processing of endogenous loci leading to 3' UTR extensions. Interestingly, the

12

326   majority of loci with impaired transcription termination were shared between *tex1* and *hpr1*

327   mutants, suggesting that both proteins have overlapping functions in RNA termination.

328   Analysis of nucleotides surrounding of the 3' cleavage site did not reveal a significant

329   difference for genes showing a 3'extension in the *tex1-4* and *hpr1-6* mutants compared to the

330   Arabidopsis genome. The best defined DNA sequence involved in the control of mRNA

331   polyadenylation site is a A-rich region at approximately -20 nucleotides defined as the near

332   upstream element (NUE). Although the NUE canonical AAUAAA sequence is found very

333   frequently in animal genomes, the heterogeneity in NUE sequences is larger in plants [30, 39].

334   An apparent deviation (but not statistically significant) from the AAUAAA was observed in

335   genes showing a 3'UTR extension in the *hpr1-6* and *tex1-4* . Furthermore, optimization of the

336   polyadenylation site of the gene AT1G76560 from AAGAAA to AAUAAA did not affect the

337   extent of 3'UTR extension in the *tex1-4* and *hpr1-6* mutants compared to Col-0. It is likely that

338   the relatively low number of genes showing 3'UTR extensions significantly limits our ability

339   to identify nucleotide features that are important in 3'UTR extension in the *tex1-4* and *hpr1-6*

340   mutants through a bioinformatic approach. A more systematic analysis of the 3'UTR of the

341   genes affected in the *tex1-4* and *hpr1-6* mutants, such as AT1G76560 and AT1G03160, using

342   the transient assay described in this study may lead to the identification of the cis-elements

343   involved.

344

345   Analysis of the *A. tumefaciens* NOS transcript revealed two putative NUE sequences, namely

346   AAUAAA and AAUAAU, at position -135 and -50 nucleotides, which is much further

347   upstream than the usual -20 nucleotides [40]. It is thus likely that other sequences within the

348   NOS terminator play a determinant role in RNA transcription termination. Furthermore, while

349   the prominent dinucleotide located immediately before the cleavage site are typically CA or

350   UA, the dinucleotide GA is present in the NOS terminator [40]. While no detailed functional

351   analysis of the polyadenylation signal of the NOS gene has been reported, it is likely that the

352   its unusual structure is related to the bacterial origin of the gene. While the NOS terminator is

353   a common feature of many T-DNA vectors, several studies have shown that transgene

354   expression can be considerably enhanced either when the NOS terminator is combined with a

355   second terminator or when it is replaced by the terminator of plant endogenous genes [41-43].

356   For example, replacement of the NOS terminator with an extensin terminator was shown to

357   reduce read-through transcript and improve expression of transgenes [44]. Altogether, these

358   data suggest that mutations in the *HPR1* and *TEX1* genes may more prominently affect mRNA

13

359 3'processing for genes having unusual or weak polyadenylation signals, such as found in the
360 NOS terminator.

361 The 3'UTR of mRNAs have important regulatory functions, impacting mRNA stability,
362 localization and translation potential via interaction with numerous RNA binding proteins as
363 well as miRNAs [45]. Extension of the 3'UTRs of genes in the *hpr1-6* and *tex1-4* genetic
364 background may thus impact their expression in numerous ways. In some cases, extension of
365 3'UTR may also lead to disruption of the downstream gene by the formation of an antisense
366 RNA with potential to trigger siRNA-mediated gene silencing, or by transcriptional
367 interference via RNAPII collision [46]. The *tex1* and *hpr1* mutant are known to have multiple
368 phenotypes, ranging from defects in vegetative and reproductive development [18, 38],
369 responses to both biotic and abiotic stress [36, 37] and the expression of genes encoding acid
370 phosphatases [16] or ethylene signaling pathway repressor [20]. It would be of interest to
371 determine if the genes affected by 3'UTR extensions contribute to some extent to these
372 phenotypes.

373 Further work is necessary to gain detailed mechanistic insights as to how mutations in *tex1* and
374 *hpr1* lead to both suppression of termination at the NOS terminator and changes in the 3'UTR
375 of endogenous genes. Being part of the TREX complex associated with the mRNA transcription
376 machinery, TEX1 and HPR1 could affect mRNA transcription termination through interactions
377 with proteins more specifically involved in 3'end processing. This would be analogous to the
378 implication of the THOC5 protein, another component of the TREX complex, in mRNA 3'-end
379 processing in mammals via interactions with mRNA cleavage factors, including CPSF100 [33,
380 34]. Since TEX1 and HPR6 have both been implicated in the generation of small RNAs,
381 including tasiRNA, siRNA and miRNA, the contribution of these pathway to mRNA
382 termination should also be further examined. Reversion of the *pho1-7* growth phenotype could
383 not be reproduced by mutations in the *SGS3* and *RDR6* genes involved in small RNA
384 biogenesis, in particular of tasiRNAs, indicating that the effects of the *hpr1* and *tex1* mutations
385 in *pho1-7* could not be caused by changes in tasiRNAs generation [14, 15, 17, 18]. siRNA is
386 associated with DNA methylation, which in turn could impact RNA polymerase activity and
387 mRNA processing, including splicing and termination [47, 48]. Although the fact that T-DNA
388 present in the *pho1-7* mutant is both well transcribed and mediates resistance to kanamycin,
389 suggesting that it is unlikely to be strongly methylated, more subtle effects of siRNA-mediated
390 methylation on RNA transcription termination at the NOS terminator and endogenous loci
391 should be explored.

14

392 The N-terminal half of the PHO1 protein harbors a SPX domain which is involved in binding
393 inositol polyphosphate at high affinity via interactions with conserved tyrosine and lysine
394 residues [49, 50]. A PHO1 protein with mutations in these SPX conserved amino acids is unable
395 to complement the *pho1-2* null mutant, indicating that the binding of inositol polyphosphate is
396 important for PHO1 activity *in plantae* [50]. The protein PHO1$^{\Delta84-114}$ synthesized from the
397 *pho1-7* mutant does not affect the core of the SPX domain but only leads to a small deletion at
398 the N-terminal end of the second SPX subdomain (see S1 text). Heterologous expression of the
399 PHO1$^{\Delta84-114}$ protein in tobacco leaves led to specific Pi export activity, indicating that the 31
400 amino acid deletion does not completely inactivates the protein. Considering that *pho1-7 tex1-4*
401 and *pho1-7 hpr1-6* as well as the *pho1-4* null mutant expressing the PHO1$^{\Delta84-114}$ protein have
402 low shoot Pi, is thus likely that the PHO1$^{\Delta84-114}$ retains some Pi export activity in root xylem
403 parenchyma cells, but lower than the wild type protein. The high level of expression of the
404 PHO1$^{\Delta84-114}$ protein in the *pho1-7 tex1-4* mutant cannot be simply explained by a higher
405 expression of the truncated *PHO1*$^{\Delta249-342}$ mRNA in the double mutant background relative the
406 *pho1-7* parent, since the *PHO1*$^{\Delta249-342}$ mRNA remained lower than the full length *PHO1* mRNA
407 in Col-0 plants. PHO1 is known to be degraded by PHO2, a key protein involved in the Pi-
408 deficiency signaling pathway [51]. Whether the high level of PHO1$^{\Delta84-114}$ accumulation is a
409 reflection of greater stability of the truncated protein and/or increased translation efficiency of
410 the truncated mRNA remains to be determined.

411

412 Uncoupling low leaf Pi from its main effect on shoot growth has previously been reported for
413 plants under-expressing *PHO1* via silencing, indicating a role for high root Pi content and
414 PHO1 in modulating the response of the shoot to Pi deficiency [25]. The improved shoot growth
415 observed in the *pho1-7 tex1-4* and *pho1-7 hpr1-6* compared to the parent *pho1-7* is likely due
416 to the increased expression of the PHO1$^{\Delta84-114}$ hypomorphic protein and an increase in Pi export
417 activity. However, both *hpr1* and *tex1* mutants have low expression of the *RTE1* gene involved
418 in the repression of the ethylene signaling pathway which has been linked to an increase in root-
419 associated acid phosphatase activity and root hair elongation in these mutants, two
420 characteristics that can positively impact Pi acquisition [16, 20]. It is thus possible that a small
421 part of the growth improvement observed in the *tex1-4 pho1-7* and *hpr1-6 pho1-7* may also be
422 associated with the repression of the ethylene pathway.

15

Materials and methods

Plant material and growth conditions

All Arabidopsis plants used in this study, including mutants and transgenic plants, were in the Columbia (Col-0) background. For *in vitro* experiments, plants were grown in half-strength Murashige and Skoog (MS) salts (Duchefa M0255) containing 1% sucrose and 0.7% agar. For Pi-deficient medium MS salts without Pi (Caisson Labs, MSP11) and purified agar (Conda, 1806) was used. Pi buffer pH5.7 (93.5% KH_2PO_4 and 6.5% K_2HPO_4) was added to obtain different Pi concentrations. In the Pi-deficient media, Pi buffer was replaced by equimolar amounts of KCl. Plants were also grown either in soil or in a clay-based substrate (Seramis) supplemented with half-strength MS for the isolation of roots. Growth chamber conditions were 22°C, 60% humidity, and a 16h light/8h dark photoperiod with 100 μE/m2 per sec of white light for long days and 10h light/14h dark for short days. The *pho1-2*, *pho1-4*, *pho1-7* (SALK_119529) and *tex1-4* were previously described [15, 22, 52] and *hpr1-6* (SAIL_1209_F10) is a T-DNA mutant from SAIL collection. The *rdr6-11* is an EMS-derived mutant while *sgs3-1* is a T-DNA mutant from the SALK collection (SALK_001394) and have both been previously described [28].

pho1 suppressor screen

Ethyl methanesulfonate (EMS) mutagenesis was performed on approximately 20,000 seeds of homozygous *pho1-7*. Seeds were treated with 0.2% EMS in 100mM phosphate buffer for 8 hours and were rinsed with water 10 times afterwards. Approximately 10, 000 individual M1 plants were grown and seeds of their progeny were collected in bulk. Approximately 10'300 M2 plants were grown in soil for 4 weeks under an 18h/8h day/night light cycle. Plants showing an increased rosettes size relative to the *pho1-7* parent were identified and their seeds collected. In the next generation, plants retaining 100% kanamycin resistance mediated by the T-DNA in *PHO1* were sown in soil and the Pi content in 3-week-old rosettes was determined using the molybdate assay [53] as previously described [27]. Plants showing the combination of increased rosettes size with low shoot Pi similar to *pho1-7* were then genotyped by PCR to further confirm the presence of the T-DNA in the *PHO1* locus in an homozygous state.

Identification of mutant genes by next generation sequencing

pho1-7 suppressor mutant was back-crossed to the parent *pho1-7* line to test if the mutation was

16

456   dominant or recessive and to generate an isogenic mapping population. Approximately 40

457   plants with a *pho1-7* suppressor phenotype were selected from the segregating F2 population.

458   DNA was extracted from the pool of these 40 mutants and sequencing was performed using

459   Illumina Hiseq 2000 system. DNA sequencing yielded an average coverage of 80 to 100

460   nucleotides per nucleotide per sample. Sequencing reads were mapped with Burrows-Wheeler

461   Aligner software (BWA) version 0.5.9-r16 using default parameters to the TAIR10 release of

462   the *A. thaliana* genome. Using SAM (Sequence Alignment/Map) tool the alignments were

463   converted to BAM format. SNPs were called with the Unified Genotyper tool of the Genome

464   Analysis Toolkit (GATK) version v1.4-24-g6ec686b. SNPs present in the parental line *pho1-7*

465   were   filtered   out   using   BEDTools   utilities   version   v2.14.2.

466   TAIR10_GFF3_genes_transposons.gff file was used to filter out the SNPs present in the

467   transposons. The predicted effect of the remaining SNPs in coding regions was assessed with

468   SNPEff version 2.0.4 RC1. Unix command awk was used to extract the SNP frequencies (the

469   number of reads supporting a given SNP over the total number of reads covering the SNP

470   location) and were plotted with R 2.15.1.

471

472   **Cloning and transgenic lines**

473   For complementation of *pho1-7 suppressor* with *TEX1*, genomic sequence including 2 kb

474   promoter 5'UTR and 3'UTR was amplified using primers TEX1-gen-F -and TEX1-gen-R

475   (sequences of all primers used are listed in S1 Table). The amplicon was cloned into pENTR/D

476   TOPO vector (Invitrogen). The entry vector was then shuttled into the binary vector pMDC99

477   [54] using Gateway technology (Invitrogen). For pTEX1:TEX1:GFP fusion, *TEX1* promoter

478   and gene was amplified using primers TEX1-gen-F and TEX1-gen-R-w/o-stop. Reverse primer

479   was designed to remove the stop codon from *TEX1* gene. The amplicon was cloned into

480   pENTR/D TOPO vector (Invitrogen). The entry vector was then shuttled into the binary vector

481   pMDC107, which contains GFP at the C-terminal [54]. For TEX1 promoter GUS fusion,

482   promoter was amplified using the primers TEX1-Pro-infu-LP and TEX1-Pro-infu-RP. The

483   amplicon was cloned into the gateway entry vector pE2B using Infusion technology

484   (Clonetech). The entry vector was then shuttled into the binary vector pMDC63 [54] using

485   Gateway technology (Invitrogen). For cloning pPHO1:gPHO1$^{\Delta 83-114}$:GFP, promoter and first

486   exon was amplified using primers P2BJ pPHO1 1exon L and P2BJ pPHO1 1exon R (fragment

487   1), *PHO1* gene from third exon until before the stop codon was amplified using primers P2BJ

488   PHO1gene 3rd exon F and P2BJ PHO1gene 3rd exon R (fragment 2). The two fragments were

489   combined together in the Gateway entry vector pE2B using Infusion technology (Clonetech).

17

490   The entry vector was then shuttled into the binary vector pMDC107 [54] using Gateway

491   technology (Invitrogen). All the binary vectors were transformed into *Agrobacterium*

492   *tumefaciens* strain GV3101 and plants were transformed using flower dip method [55].

493   For analysis of transcript termination by transient expression in Arabidopsis protoplasts, the

494   500 nucleotides located immediately after the stop codon of the genes AT1G76560 and

495   AT1G03160 were fused after the stop codon of the nano-luciferase (nLUC) gene. To achieve

496   this, the 500 bp 3'sequences from AT1G76560 (wild-type and mutated) and AT1G03160

497   flanked by attR1 and attR2 sites were synthesized and inserted in the pUC57 plasmid by

498   Genscript . The DNA insert was then shuttled by Gateway cloning into the dual-luciferase

499   vector nLucFlucGW (Genbank MH552885) [56] modified to lack the original nLuc 3'UTR and

500   terminator sequences. The final constructs had the hybrid genes expressed under the control of

501   the ubiquitin promoter, as well as the firefly luciferase gene (Fluc) constitutively expressed,

502   used for loading control.

503

504   **Quantitative RT PCR and phosphate measurement and Pi export assay.**

505   Quantitative RT-PCR was performed as previously described [27]. For transient expression of

506   PHO1, *Nicotiana benthamiana* tobacco plants were infiltrated with *A. tumefaciens* as

507   previously described [21]. Pi measurements were performed using the molybdate-ascorbic acid

508   method [53].

509

510   **Confocal microscopy**

511   Seedlings were incubated for ten min in a solution of 15 mM Propidium Iodide (sigma, P4170)

512   and rinsed twice with water. Excitation and detection window for GFP was set at 488 nm for

513   excitation and 490–555 nm for detection. Propidium iodide was excited at 555 nm and detected

514   at 600-700nm. All experiments were performed using Zeiss LSM 700 confocal microscope.

515

516   **Western blot analysis**

517   Plants were grown on clay-based substrate (Seramis) supplemented with half-strength MS

518   liquid medium. Proteins were extracted from homogenized 25-day-old roots at 4 °C in

519   extraction buffer containing 10 mM phosphate buffer pH 7.4, 300 mM sucrose, 150 mM NaCl,

520   5 mM EDTA, 5 mM EGTA, 1 mM DTT, 20 mM NaF and 1× protease inhibitor (Roche EDTA

521   free complete mini tablet), and sonicated for 10 min in an ice-cold water bath. Fifty micrograms

18

522 of protein were separated on an SDS-PAGE and transferred to an Amersham Hybond-P PVDF

523 membrane (GE healthcare). The rabbit polyclonal antibody to PHO1 [51] and goat anti-rabbit

524 IgG-HRP (Santa Cruz Biotechnology, USA) was used along with the Western Bright Sirius

525 HRP substrate (Advansta, USA). Signal intensity was measured using a GE healthcare

526 ImageQuant RT ECL Imager.

527

Illumina RNA-sequencing data analysis

529 RNA was extracted from roots of plants grown for 3 weeks in pots containing clay-based

530 substrate (Seramis) or for 7 days on vertical agar plates containing half-strength MS media with

531 1% sucrose. Strand-specific libraries were prepared using the TruSeq Stranded Total RNA kit

532 (Illumina). PolyA$^+$ RNAs were selected according to manufacturer's instructions and the cDNA

533 libraries were sequenced on a HiSeq 2500 Illumina sequencer. The reads were mapped against

534 TAIR10.31 reference genome using Hisat2 [57] and the readcount for each gene was

535 determined using HTSeqcount [58]. Readcounts were normalized using DESeq2 [59]. Figures

536 showing read density from RNAseq data were generated using Integrative genomics viewer

537 (IGV) [60].

Analysis of full-length mRNA using PacBio sequencing

539 One µg of total RNA was used to generate cDNA with the SMARTer PCR cDNA Synthesis

540 kit (Clontech, Mountain View, CA, USA). Fifty µl of cDNA were amplified by 13 PCR cycles

541 with the Kapa HiFi PCR kit (Kapa Biosystems, Wilmington, MA, USA) followed by size

542 selection from 1.5kb to 3.5kb with a BluePippin system (Sage Science, Beverly, MA, USA).

543 Seventy ng of the size selected fragment were further amplified with Kapa HiFi PCR kit for 5

544 cycles and 2 minutes extension time and 750 ng was used to prepare a SMRTbell library with

545 the PacBio SMRTbell Template Prep Kit 1 (Pacific Biosciences, Menlo Park, CA, USA)

546 according to the manufacturer's recommendations. The resulting library was sequenced with

547 P4/C2 chemistry and MagBeads on a PacBio RSII system (Pacific Biosciences, Menlo Park,

548 CA, USA) at 240 min movie length using one SMRT cell v2. Bioinformatics analysis were

549 performed through SMRT® Analysis Server v2.3.0. using RS_IsoSeq.1 workflow and

550 TAIR10.31 as reference genome.

551

552   **Identification of 3'UTR extensions**

553   3' UTR extensions were identified following a procedure adapted from Sun et al. 2017 [8].

554   Briefly, reads obtained by single or paired-end polyA+ RNAseq were mapped with Hisat2 [57]

555   against the intergenic regions extracted from TAIR10.31 annotation. Each intergenic region

556   was divided into 10 nucleotide bins and the normalized readcount was determined for each bin

557   with HTSeq-count [58] and DESeq2 [59]. 3' extensions were then contiguously assembled

558   from the 5' end of intergenic intervals until a bin had a normalized readcount < 1. Only

559   extensions longer than 200 nucleotides were kept for further analyses.

560   The number of reads mapping each TAIR10.31 gene and newly identified 3' extensions was

561   determined with HTSeq-count [58]. Differential expression analysis was performed with

562   DESeq2 [59] to identify extensions significantly up or down-regulated independently of the

563   expression level of the TAIR10.31 annotated gene body, comparing different genotypes. An

564   extension was considered significantly differentially expressed if the adjusted pvalue corrected

565   for false discovery rate was < 0.1 and the fold change of the ratios normalized readcount 3'

566   extension / normalized readcount gene body between 2 genotypes was > 2.

567

568   To analyze the polyadenylation signal present in genes with and without 3'UTR extensions, the

569   frequency of each nucleotide at the polyadenylation consensus sequence AAUAAA was

570   calculated for each gene and a Chi square test was used to test for statistical significance.

571

572   **Chromatin immunoprecipitation analyzed by qPCR**

573   Leaves from 3-week-old *A. thaliana* seedlings from different genotypes were harvested and

574   immediately incubated in 37 ml of pre-chilled fixation buffer (1% formaldehyde in 0.4M

575   sucrose, 10 mM Tris pH 8, 1mM EDTA, 1 mM PMSF, 0.05% Triton   X-100) for 10 min under

576   vacuum. 2.5 ml of Glycine (2.5 M) was added and samples were incubated 5 additional min

577   under vacuum, rinsed 3 times with water and frozen in liquid nitrogen. Frozen samples were

578   ground and the powder resuspended in 30 ml of extraction buffer I (0.4 M sucrose, 10 mM

579   HEPES pH 8, 5 mM ß-mercaptoethanol, 0.1 g/ml 4-(2-aminoethyl) benzenesulfonyl fluoride

580   hydrochloride (AEBSF). After 20 min incubation at 4°C, the mixture was filtered through

20

581 Miracloth and centrifuged for 20 min at 3000 g at 4°C. The pellet was resuspended in 300 µl

582 of extraction buffer III (1.7 M sucrose, 10 mM HEPES pH 8, 0.15% Triton X-100, 2 mM

583 MgCl$_2$, 5 mM 5 mM ß-mercaptoethanol, 0.1 g/ml AEBSF), loaded on top of a layer of 300 µl

584 of extraction buffer III and centrifuge for 1h at 16000 g at 4°C. The pellet was resuspended in

585 300 µl of Nuclei Lysis Buffer (50 mM HEPES pH 8, 10 mM EDTA, 1% SDS, 0.1 g/ml AEBSF)

586 and incubated on ice for 30 min.

587 Chromatin solution was centrifuged twice for 10 min at 14000 g, 4°C and incubated over-night

588 with S2P-RNApolII specific antibodies. The mixture was then incubated with Protein A beads

589 for 3h at 4°C. After washing, immune complexes were eluted twice with 50 µl of Elution Buffer

590 (1 % SDS, 0.1 NaHCO3). To reverse crosslinking, 4 µl of a 5 M NaCl solution was added to

591 100 µl of eluate and the mixture incubated overnight at 65 °C. Two µl of 0.5 M EDTA, 1.5 µl

592 of 3 M Tris-HCl pH 6.8 and 20 µg of proteinase K were then added and the mixture incubated

593 for 3h at 45 °C. DNA was then extracted using the NucleoSpin kit from Macherey Nagel. DNA

594 samples were diluted 10 times and 2 µl were used for quantification by qPCR using Master Mix

595 SYBR Select (Applied Biosystems).

Transient expression in Arabidopsis protoplasts

597 Arabidopsis protoplasts were produced and transformed as previously described [61]. In brief,

598 wild type Col-0, as well as *hrp1-6* and *tex1-4* mutant plants were grown in long photoperiod

599 (16 h light and 8 h dark at 21 ^{0}C) for 4-5 weeks and leaves were cut with razor blades to produce

600 0.5-1 mm leaf strips. These were submerged in enzyme solution (1% cellulase, 0.25

601 % macerozyme, 0.4 M mannitol, 20 mM KCl, 20 mM MES and 10 mM CaCl2),

602 vacuum infiltrated and incubated at room temperature for 2 h. Protoplasts were harvested by

603 centrifugation at 100 xg for 3 min, washed with W5 solution (154 mM NaCl, 125 mM CaCl2,

604 5 mM KCl and 2 mM MES) and resuspended in MMG solution (4 mM MES, pH 5.7, 0.4 M

605 mannitol and 15 mM MgCl2) at 1x10^6 protoplast/ml. Protoplast transformation was performed

606 by combining ~1.5 x10^5 protoplasts, 8µg of plasmid, and PEG solution (40% PEG4000, 0.2 M

607 mannitol and 100 mM CaCl2). After replacing PEG solution with W5 solution by

608 consecutive washings, protoplasts were kept in the dark for approximately 16 hours at 21^{0}C.

609 Transformed protoplasts were harvested by centrifugation at 6000 xg for 1 min, and

610 resuspended in 1X Passive Lysis buffer (Promega, E1941). The lysate was cleared by

611 centrifugation and RNA was extracted using RNA purification kit as described by

612 the manufacture (Jena Bioscience, PP-210), followed by DNase I treatment. cDNA

613   was synthesized from 0.1 μg RNA using M-MLV Reverse Transcriptase (Promega,

614   M3681) and oligo d(T)15 as primer following the manufacturer's instructions. qPCR analysis

615   was performed using SYBR select Master Mix (Applied Biosystems, 4472908) with primer

616   pairs specific to transcripts of interest and firefly luciferase mRNA, used for

617   data normalization. Long/short transcript ratio was calculated with the following formula:

618    $\Delta CT(\text{long transcript}) = CT(\text{long transcript}) - CT(\text{Fluc})$

619    $\Delta CT(\text{short transcript}) = CT(\text{short transcript}) - CT(\text{Fluc})$

620    $\Delta\Delta CT(\text{long/short}) = \Delta CT(\text{long transcript}) / \Delta CT(\text{short transcript})$

621   Long/short transcript ratio$= 2^{\Delta\Delta CT(\text{long/short})}$.

622

623

Acknowledgment

627

628

References

629

630 1. Chen FX, Smith ER, Shilatifard A. Born to run: control of transcription elongation by
631 RNA polymerase II. Nat Rev Mol Cell Biol. 2018;19(7):464-478.

632 2. Herzel L, Ottoz DSM, Alpert T, Neugebauer KM. Splicing and transcription touch base:
633 co-transcriptional spliceosome assembly and function. Nat Rev Mol Cell Biol.
634 2017;18(10):637-650.

635 3. Richard P, Manley JL. Transcription termination by nuclear RNA polymerases. Genes
636 Dev. 2009;23(11):1247-1269.

637 4. Hsin J-P, Manley JL. The RNA polymerase II CTD coordinates transcription and RNA
638 processing. Genes Dev. 2012;26(19):2119-2137.

639 5. Nagarajan VK, Jones CI, Newbury SF, Green PJ. XRN 5 ' -> 3 ' exoribonucleases:
640 Structure, mechanisms and functions. Biochim Biophys Acta. 2013;1829(6-7):590-603.

641 6. Chen W, Jia Q, Song Y, Fu H, Wei G, Ni T. Alternative Polyadenylation: Methods,
642 Findings, and Impacts. Genom Proteom Bioinf. 2017;15(5):287-300.

643 7. Mapendano CK, Lykke-Andersen S, Kjems J, Bertrand E, Jensen TH. Crosstalk
644 between mRNA 3 ' End Processing and Transcription Initiation. Mol Cell. 2010;40(3):410-422.

645 8. Sun H-X, Li Y, Niu Q-W, Chua N-H. Dehydration stress extends mRNA 3 ' untranslated
646 regions with noncoding RNA functions in Arabidopsis. Genome Res. 2017;27(8):1427-1436.

647 9. Katahira J. mRNA export and the TREX complex. Biochim Biophys Acta.
648 2012;1819(6):507-513.

649 10. Heath CG, Viphakone N, Wilson SA. The role of TREX in gene expression and disease.
650 Biochem J. 2016;473:2911-2935.

651 11. Meinel DM, Burkert-Kautzsch C, Kieser A, O'Duibhir E, Siebert M, Mayer A, Cramer
652 P, Soding J, Holstege FCP, Straesser K. Recruitment of TREX to the Transcription Machinery
653 by Its Direct Binding to the Phospho-CTD of RNA Polymerase II. PLoS Genet. 2013;9(11).

654 12. Gomez-Gonzalez B, Garcia-Rubio M, Bermejo R, Gaillard H, Shirahige K, Marin A,
655 Foiani M, Aguilera A. Genome-wide function of THO/TREX in active genes prevents R-loop-
656 dependent replication obstacles. EMBO J. 2011;30(15):3106-3119.

657 13. Ehrnsberger HF, Grasser M, Grasser KD. Nucleocytosolic mRNA transport in plants:
658 export factors and their influence on growth and development. J Exp Bot.
659 2019;doi:10.1093/jxb/erz173.

660   14.   Yelina NE, Smith LM, Jones AME, Patel K, Kelly KA, Baulcombe DC. Putative
661   Arabidopsis THO/TREX mRNA export complex is involved in transgene and endogenous
662   siRNA biosynthesis. Proc Natl Acad Sci USA. 2010;107(31):13948-13953.

663   15.   Jauvion V, Elmayan T, Vaucheret H. The Conserved RNA Trafficking Proteins HPR1
664   and TEX1 Are Involved in the Production of Endogenous and Exogenous Small Interfering
665   RNA in Arabidopsis. Plant Cell. 2010;22(8):2697-2709.

666   16.   Tao S, Zhang Y, Wang X, Xu L, Fang X, Lu ZJ, Liu D. The THO/TREX Complex
667   Active in miRNA Biogenesis Negatively Regulates Root-Associated Acid Phosphatase
668   Activity Induced by Phosphate Starvation. Plant Physiol. 2016;171(4):2841-2853.

669   17.   Francisco-Mangilet AG, Karlsson P, Kim M-H, Eo HJ, Oh SA, Kim JH, Kulcheski FR,
670   Park SK, Andres Manavella P. THO2, a core member of the THO/TREX complex, is required
671   for microRNA production in Arabidopsis. Plant J. 2015;82(6):1018-1029.

672   18.   Furumizu C, Tsukaya H, Komeda Y. Characterization of EMU, the Arabidopsis
673   homolog of the yeast THO complex member HPR1. RNA. 2010;16(9):1809-1817.

674   19.   Sorensen BB, Ehrnsberger HF, Esposito S, Pfab A, Bruckmann A, Hauptmann J,
675   Meister G, Merkl R, Schubert T, Laengst G *et al*. The Arabidopsis THO/TREX component
676   TEX1 functionally interacts with MOS11 and modulates mRNA export and alternative splicing
677   events. Plant Mol Biol. 2017;93(3):283-298.

678   20.   Xu C, Zhou X, Wen C-K. HYPER RECOMBINATION1 of the THO/TREX Complex
679   Plays a Role in Controlling Transcription of the REVERSION-TO-ETHYLENE
680   SENSITIVITY1 Gene in Arabidopsis. PLoS Genet. 2015;11(2).

681   21.   Arpat AB, Magliano P, Wege S, Rouached H, Stefanovic A, Poirier Y. Functional
682   expression of PHO1 to the Golgi and *trans*-Golgi network and its role in export of inorganic
683   phosphate. Plant J. 2012;71:479-491.

684   22.   Hamburger D, Rezzonico E, MacDonald-Comber Petétot J, Somerville C, Poirier Y.
685   Identification and characterization of the *Arabidopsis PHO1* gene involved in phosphate
686   loading to the xylem. Plant Cell. 2002;14:889-902.

687   23.   Poirier Y, Thoma S, Somerville C, Schiefelbein J. A mutant of *Arabidopsis* deficient in
688   xylem loading of phosphate. Plant Physiol. 1991;97:1087-1093.

689   24.   Stefanovic A, Arpat AB, Bligny R, Gout E, Vidoudez C, Bensimon M, Poirier Y.
690   Overexpression of PHO1 in Arabidopsis leaves reveals its role in mediating phosphate efflux.
691   Plant J. 2011;66:689-699.

692   25.   Rouached H, Stefanovic A, Secco D, Arpat AB, Gout E, Bligny R, Poirier Y.
693   Uncoupling phosphate deficiency from its major effects on growth and transcriptome via PHO1
694   expression in Arabidopsis. Plant J. 2011;65:557-570.

695   26.   Secco D, Baumann A, Poirier Y. Charcaterization of the rice *PHO1* gene family reveals
696   a key role for *OsPHO1;2* in phosphate homeostasis and the evolution of a distinct clade in
697   dicotyledons. Plant Physiol. 2010;152:1693-1704.

698   27.   Wege S, Khan GA, Jung J-Y, Vogiatzaki E, Pradervand S, Aller I, Meyer AJ, Poirier
699   Y. The EXS domain of PHO1 participates in the response of shoots to phosphate deficiency via
700   a root-to-shoot signal. Plant Physiol. 2016;170(1):385-400.

701   28.   Peragine A, Yoshikawa M, Wu G, Albrecht HL, Poethig RS. SGS3 and
702   SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting
703   siRNAs in Arabidopsis. Genes Dev. 2004;18(19):2368-2379.

704   29.   Ronemus M, Vaughn MW, Martienssen RA. MicroRNA-targeted and small interfering
705   RNA-mediated mRNA degradation is regulated by Argonaute, Dicer, and RNA-dependent
706   RNA polymerase in Arabidopsis. Plant Cell. 2006;18(7):1559-1574.

707   30.   Loke JC, Stahlberg EA, Strenski DG, Haas BJ, Wood PC, Li QQ. Compilation of
708   mRNA polyadenylation signals in Arabidopsis revealed a new signal element and potential
709   secondary structures. Plant Physiol. 2005;138(3):1457-1468.

710   31.   Sherstnev A, Duc C, Cole C, Zacharaki V, Hornyik C, Ozsolak F, Milos PM, Barton
711   GJ, Simpson GG. Direct sequencing of Arabidopsis thaliana RNA reveals patterns of cleavage
712   and polyadenylation. Nat Struct Mol Biol. 2012;19(8):845-852.

713   32.   Wu XH, Liu M, Downie B, Liang C, Ji GL, Li QQ, Hunt AG. Genome-wide landscape
714   of polyadenylation in Arabidopsis provides evidence for extensive alternative polyadenylation.
715   Proc Natl Acad Sci USA. 2011;108(30):12533-12538.

716   33.   Doan Duy Hai T, Saran S, Williamson AJK, Pierce A, Dittrich-Breiholz O, Wiehlmann
717   L, Koch A, Whetton AD, Tamura T. THOC5 controls 3 ' end- processing of immediate early
718   genes via interaction with polyadenylation specific factor 100 (CPSF100). Nucleic Acids Res.
719   2014;42(19):12249-12260.

720   34.   Katahira J, Okuzaki D, Inoue H, Yoneda Y, Maehara K, Ohkawa Y. Human TREX
721   component Thoc5 affects alternative polyadenylation site choice by recruiting mammalian
722   cleavage factor I. Nucleic Acids Res. 2013;41(14):7060-7072.

723   35.   Pak V, Eifler TT, Jaeger S, Krogan NJ, Fujinaga K, Peterlin BM. CDK11 in
724   TREX/THOC Regulates HIV mRNA 3 ' End Processing. Cell Host Microbe. 2015;18(5):560-
725   570.

726   36.   Pan H, Liu S, Tang D. HPR1, a component of the THO/TREX complex, plays an

727   important role in disease resistance and senescence in Arabidopsis. Plant J. 2012;69(5):831-

728   843.

729   37.   Doll S, Kuhlmann M, Rutten T, Mette MF, Scharfenberg S, Petridis A, Berreth DC,

730   Mock HP. Accumulation of the coumarin scopolin under abiotic stress conditions is mediated

731   by the Arabidopsis thaliana THO/TREX complex. Plant J. 2018;93(3):431-444.

732   38.   Su Z, Zhao L, Zhao Y, Li S, Won S, Cai H, Wang L, Li Z, Chen P, Qin Y *et al*. The

733   THO Complex Non-Cell-Autonomously Represses Female Germline Specification through the

734   TAS3-ARF3 Module. Curr Biol. 2017;27(11):1597-+.

735   39.   Tian B, Hu J, Zhang HB, Lutz CS. A large-scale analysis of mRNA polyadenylation of

736   human and mouse genes. Nucleic Acids Res. 2005;33(1):201-212.

737   40.   Depicker A, Stachel S, Dhaese P, Zambryski PC, Goodman H. Nopaline synthase:

738   transcript mapping and DNA sequence. J Mol Appl Genet. 1982;6:561-573.

739   41.   Beyene G, Buenrostro-Nava MT, Damaj MB, Gao SJ, Molina J, Mirkov TE.

740   Unprecedented enhancement of transient gene expression from minimal cassettes using a

741   double terminator. Plant Cell Rep. 2011;30(1):13-25.

742   42.   Nagaya S, Kawamura K, Shinmyo A, Kato K. The HSP Terminator of Arabidopsis

743   thaliana Increases Gene Expression in Plant Cells. Plant Cell Physiol. 2010;51(2):328-332.

744   43.   Richter LJ, Thanavala Y, Arntzen CJ, Mason HS. Production of hepatitis B surface

745   antigen in transgenic plants for oral immunization. Nat Biotechnol. 2000;18(11):1167-1171.

746   44.   Rosenthal SH, Diamos AG, Mason HS. An intronless form of the tobacco extensin gene

747   terminator strongly enhances transient gene expression in plant leaves. Plant Mol Biol.

748   2018;96(4-5):429-443.

749   45.   Mayr C. Regulation by 3 '-untranslated regions. Ann Rev Genet. 2017;51:171-194.

750   46.   Kindgren P, Ard R, Ivanov M, Marquardt S. Transcriptional read-through of the long

751   non-coding RNA SVALKA governs plant cold acclimation. Nature Comm. 2018;9.

752   47.   Cuerda-Gil D, Slotkin RK. Non-canonical RNA-directed DNA methylation. Nature

753   Plants. 2016;2(11):e16163.

754   48.   Mathieu O, Bouche N. Interplay between chromatin and RNA processing. Curr Opin

755   Plant Biol. 2014;18:60-65.

756   49.   Jung J-Y, Ried MK, Hothorn M, Poirier Y. Control of plant phosphate homeostasis by

757   inositol pyrophosphates and the SPX domain. Curr Opin Biotechnol. 2018;49:156-162.

758 50.   Wild R, Gerasimaite R, Jung J-Y, Truffault V, Pavlovic I, Schmidt A, Saiardi A, Jessen
759 HJ, Poirier Y, Hothorn M *et al*. Control of eukaryotic phosphate homeostasis by inositol
760 polyphosphate sensor domains. Science. 2016;352(6288):986-990.

761 51.   Liu T-Y, Huang T-K, Tseng C-Y, Lai Y-S, Lin S-I, Lin W-Y, Chen J-W, Chiou T-J.
762 PHO2-dependent degradation of PHO1 modulates phosphate homeostasis in Arabidopsis. Plant
763 Cell. 2012;24(5):2168-2183.

764 52.   Khan GA, Vogiatzaki E, Glauser G, Poirier Y. Phosphate Deficiency Induces the
765 Jasmonate Pathway and Enhances Resistance to Insect Herbivory. Plant Physiol.
766 2016;171(1):632-644.

767 53.   Ames BN. Assay of inorganic phosphate, total phosphate and phosphatases. Methods
768 Enzymol. 1966;8:115-118.

769 54.   Curtis MD, Grossniklaus U. A gateway cloning vectors set for high-throughput
770 functional analysis of genes in planta. Plant Physiol. 2003;133:462-469.

771 55.   Clough SJ, Bent AF. Floral dip: a simplified method for *Agrobacterium*-mediated
772 transformation of *Arabidopsis thaliana*. Plant J. 1998;6:735-743.

773 56.   Deforges J, Reis RS, Jacquet P, Sheppard S, Geadekar VP, Hart-Smith G, Tanzer A,
774 Hofacker IL, Iseli C, Xenarios I *et al*. Control of cognate mRNA translation by *cis*-natural
775 antisense. Plant Physiol. 2019;180(1):305-322.

776 57.   Kim D, Landmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory
777 requirements. Nat Methods. 2015;12(4):357-360.

778 58.   Anders S, Pyl PT, Huber W. HTSeq-a Python framework to work with high-throughput
779 sequencing data. Bioinformatics. 2015;31(2):166-169.

780 59.   Love M, Huber W, Anders S. Moderated estimation of fold change and dispersion for
781 RNA-seq data with DESeq2. Genome Biol. 2014;15:550.

782 60.   Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov
783 JP. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24-26.

784 61.   Yoo S-D, Cho Y-H, Sheen J. Arabidopsis mesophyll protoplasts: a versatile cell system
785 for transient gene expression analysis. Nat Protoc. 2007;2(7):1565-1572.

786

787

788

789   **Figure legends**

790

791   **Fig 1. The *tex1* mutation suppresses the shoot growth phenotype of the *pho1-7* mutant.**

792   (A) Phenotype of Col-0, *pho1-7*, *pho1-7 suppressor*, and *pho1-7 suppressor* transformed with

793   the p*TEX1:gTEX1:GFP*. (B) Pi contents of 4-week-old rosettes. Data from a representative

794   experiment shows the mean Pi content from ten different plants grown in independent pots.

795   Errors bars represent standard deviation. Values marked with lowercase letters are statistically

796   significantly different from those for other groups marked with different letters ($P < 0.05$,

797   ANOVA with the Tukey-Kramer HSD test). Panel display a representative experiment

798

799

800   **Fig 2. *tex1* restores the expression of PHO1 in the *pho1-7* mutant.** (A) Phenotype of 4-

801   week-old Col-0, *tex1-4*, *pho1-7*, *pho1-4*, *pho1-7 tex1-4* and *pho1-4 tex1-4* mutants. (B) Shoot

802   Pi contents of 4-week-old plants. Data from a representative experiment shows the means of Pi

803   contents from six individual plants grown in independent pots. Error bars represent standard

804   deviation. (C) Structure of truncated *PHO1* mRNA produced at the *PHO1* locus in the *pho1-7*

805   mutant. For the *PHO1* locus, exons are indicated as black open boxes, except for the second

806   exon, indicated as a green doted box. Exon 2 is present in Col-0 but deleted in the *pho1-7*

807   mutant as a result of T-DNA insertion. The structure of the mRNA *PHO1$^{\Delta249-342}$* found in the

808   *pho1-7* mutant is shown below. The pairs of oligonucleotides P1 and P2 used for RT-PCR is

809   shown. (D) Western blot showing full length and truncated PHO1 protein in roots of Col-0,

810   *pho1-2* null mutant, *pho1-7* and *pho1-7 tex1-4*. Plants were grown for 4 weeks in a clay

811   substrate and total protein were extracted from roots. (E) Relative expression level of *PHO1*

812   gene in the roots of 4-week-old plants grown in a clay substrate. Data are means of three

813   samples from plants grown in independent pots and three technical replicates for each sample,

814   with each sample being a pool of three plants. Error bars represent standard deviation. For both

815   B and E, values marked with lowercase letters are statistically significantly different from those

816   for other groups marked with different letters. ($P < 0.05$, ANOVA with the Tukey-Kramer HSD

817   test).

818

819   **Fig 3. Truncated PHO1$^{\Delta84-114}$ is sufficient to restore the shoot growth phenotype of *pho1-4***

820   **null mutant.** (A) Phenotype of four-week-old plants. (B) Pi contents in the rosettes of 4-week-

821   old plants. Data from a representative experiment shows the means of Pi contents from six

822   individual plants grown in independent pots. Error bars represent standard deviation. (C)

823   Localization of full-length PHO1:GFP and PHO1$^{\Delta 83\text{-}114}$:GFP in the roots of 7-day-old

824   seedlings. Both *PHO1:GFP* and *PHO1$^{\Delta 83\text{-}114}$:GFP* were expressed under the control of native

825   *PHO1* promoter in the *pho1-4* null mutant. (D) Pi export mediated by PHO1:GFP and PHO1$^{\Delta 84\text{-}}$

826   ^{114}:GFP from transient expression in *N. benthamiana* leaf discs. As controls, Pi export was

827   measured in leaf discs expressing either free GFP (EV) or not infiltrated (none). Data are means

828   of 4 measurements taken from independent infiltrated leaves. For B and D, values marked with

829   lowercase letters are statistically significantly different from those for other groups marked with

830   different letters. ($P < 0.05$, ANOVA with the Tukey-Kramer HSD test).

831

832   **Fig 4. *hpr1* can restore *pho1-7* shoot growth but not mutations in tasiRNA biogenesis.**

833   (A, C, D) Phenotype of four-week-old plants. (B) Pi contents in the rosettes of four-week-old

834   plants. Data from a representative experiment shows the means of Pi contents from ten different

835   plants grown in independent pots. Errors bars represent standard deviation. Values marked with

836   lowercase letters are statistically significantly different from those for other groups marked with

837   different letters ($P < 0.05$, ANOVA with the Tukey-Kramer HSD test).

838

839   **Fig 5. *tex1* mutant impairs transcription termination at the NOS terminator.** (A) Structure

840   of the wild type *PHO1* gene (top) and T-DNA (red) integrated in the *pho1-7* mutant (bottom).

841   Exons are indicated as black open boxes, except for the second exon, indicated as a green doted

842   box. Exon 2 is present in Col-0 but deleted in the *pho1-7* mutant as a result of T-DNA insertion,

843   which is shown in red. Blue lines represent introns. pNOS, NOS promoter; NPTII, neomycin

844   phosphotransferase gene; tNOS, NOS terminator. (B) Structure of the four mRNA variants

845   produced at the *PHO1* locus in *pho1-7* mutants. The thick black and blue lines represent

846   sequences derives from *PHO1* exons and introns, respectively, while the thick red lines are

847   derived from the T-DNA. The sizes of the different transcripts are shown. Arrows indicate the

848   location of primers used for the qPCR shown in panel D. Forward primers (discontinued

849   arrows) in transcript 2 and 3 are at the junction of cryptic splicing sites to ensure specificity.

850   (C) Illumina RNA sequencing reads density graph showing mRNA expression at the *PHO1*

851   locus in various genotypes. The RNA sequencing reads are mapped against the *PHO1* locus in

852   *pho1-7,* represented in the lower section of panel A (D) Quantification via qPCR of four

853   different mRNA transcripts produced at the *PHO1* locus in the roots of 4-week-old plants. The

854   numbers associated with each transcript and primers used for qPCR are shown in panel B. (E)

855   CHIP-qPCR experiments showing the density of RNAPII-S2 at the 3'end of the *PHO1* gene.

856   For D and E, data are means of three samples from plants grown in independent pots and three

857   technical replicates for each sample. Error bars represent standard deviation. For D, statistical

858   analysis was performed comparing each *PHO1* transcript isoform in *pho1-7 tex1-4* and *pho1-7*

859   *hpr1-6* double mutants relative to the *pho1-7* parent, with asterisks denoting statistical

860   significance (*, $P < 0.05$; and ***, $P < 0.001$) according to Student's t test. For E, values marked

861   with lowercase letters are statistically significantly different from those for other groups marked

862   with different letters ($P < 0.05$, ANOVA with the Tukey-Kramer HSD test).

863

864   **Fig 6. 3′ UTR extensions in endogenous genes of the *pho1-7 tex1-4* and *pho1-7 hpr1-6***

865   **mutants .** (A) Illumina RNAseq reads density maps (blue) showing examples of 3' UTR

866   extensions for genes AT1G03160 (left) and AT3G11310 (right) in the *pho1-7 tex1-4* and *pho1-*

867   *7 hpr1-6* mutants. The 3′ UTR extensions are depicted by a red rectangle. The exons of the

868   genes are indicated as red boxes below. (B) Venn diagram showing the number and overlap in

869   genes with 3'UTR extensions. (C) Validation of 3' UTR extensions via qPCR. (D) Transient

870   expression of Luciferase gene fused after the stop codon to the 3'end of genes AT1G76560 and

871   AT1G03160. The constructs were expressed in Arabidopsis mesophyll protoplasts obtained

872   from Col-0, *hpr1-6* and *tex1-4*. The bar chart shows the relative ratios of long-to-short

873   transcripts in the various genotypes. For C, data are means of three samples with each sample

874   being a pool of seven seedlings and three technical replicates for each sample. For D, data are

875   means of four samples with each sample being an independent transfection of protoplasts. Error

876   bars represent standard deviation. Asterisks denote statistical significance ($P < 0.05$) from the

877   Col-0 control according to Student's t test (n.s.= not statistically different).

878

879   **Fig 7. Nucleotide composition near the 3' cleavage site of mRNAs.**

880   (A) Proportion of bases in a 200 nucleotide region 5' and 3' of the transcripts termination sites

881   for all genes (left) or genes showing 3'UTR extensions in the *hpr1-6* and *tex1-4* mutant

882   background (right). (B) Graphical representation of base enrichments found at the near

883   upstream element sequence of all genes (left) and genes with 3'UTR extensions in the *hpr1-6*

884   and *tex1-4* mutant background (right). Chi square test pvalue = 0.22 (C) Transient expression

885   of Luciferase gene fused after the stop codon to the 3'end of the gene AT1G76560 in which the

886   wild-type polyadenylation site AAUGAA was mutated to AAUAAA. The construct was

887   expressed in Arabidopsis mesophyll protoplasts obtained from Col-0, *hpr1-6* and *tex1-4*. The

888   bar chart shows the ratios of long-to-short transcripts in the various genotypes and values are

889   expressed relative to the wild-type construct in Col-0. Data are means of four samples with

890   each sample being an independent transfection of protoplasts. Error bars represent standard

891   deviation. Asterisks denote statistical significance (P < 0.05) from the Col-0 control according

892   to Student's t test.

893   **Supporting Information**

894   **Supporting Figure Legends**

895   **S1 Fig. TEX1 is broadly expressed and localized in the nucleus.** (A) TEX1 localization in

896   root tips of 5-day-old seedlings. Expression of *TEX1::GFP* fusion is under the control of the

897   endogenous *TEX1* promoter. (B) GUS expression from the *TEX1* promoter in roots and

898   cotyledons of 5-day-old seedlings (left) as well as rosette leaves of 4-week-old plants (right) .

899   Plants were transformed with a *TEX1 promoter:GUS* construct. Inset (lower left) shows

900   expression in a section of mature root. (C) Expression profile of the *TEX1* gene in Arabidopsis

901   as visualized with the eFP Browser 2.0 (https://bar.utoronto.ca/efp2/).

902

903   **S2 Fig. *pho1-7* suppressor mutants restores the expression of phosphate starvation**

904   **response genes and lipid dynamics comparable to Col-0 level.** (A) Relative expression of

905   phosphate starvation-induced genes in the shoots of 4-week-old plants. (B) Lipid quantification

906   in the shoots of 4-week-old plants. DGDG, digalactosyldiacylglycerol; PG,

907   phosphatidylglycerol; PE, phosphatidylethalonamine. Data in A and B are means of three

908   samples from plants grown in independent pots and three technical replicates. Error bars

909   represent standard deviation. Values marked with lowercase letters are statistically significantly

910   different from those for other groups marked with different letters (P < 0.05, ANOVA with the

911   Tukey-Kramer HSD test).

912   **S3 Fig. Mapping of *pho1-7* RNA onto the *PHO1* locus.** (A) Illumina RNA sequencing reads

913   density graph mapped against the wild-type *PHO1* locus. Note that exon 2 of *PHO1* is missing

914   in the *pho1-7* mutant because of T-DNA insertion while it is present in Col-0. (B) Pac-Bio read

915   density graph showing full length mRNA structure at the *PHO1* locus. The top red box shows

916   the location of the T-DNA in *pho1-7*. pNOS, NOS promoter; NPTII, neomycin

917   phosphotransferase gene; tNOS, NOS terminator. The black boxes below shows the positions

918   of the 15 *PHO1* exons, except for exon 2 which is shown in green. Exon 2 is present in Col-0

919   but deleted in the *pho1-7*, *pho1-7 tex1-4* and *pho1-7 hpr1-6* mutants as a result of T-DNA

920   insertion. For each genotype, the sequence of independent cDNAs are shown by individual lines

921   (red and blue lines) with the grey areas representing the sequence density in each region.

922

923   **S4 Fig. Overlap in genes showing 3'UTR extensions in *tex1-4* mutant grown under various**

924   **conditions.** RNA extracted from roots of *pho1-7 tex1-4* and *pho1-7* plants grown in pots for 4

925   weeks (green) or *tex1-4* and Col-0 plants grown in petri dishes for 7 days (blue) were used.

926   **S5 Fig. Gene ontology of genes showing 3'UTR extensions in the *hpr1-6* and *tex1-4***

927   **mutants.** The histograms show the fold enrichment of a given gene ontology term.

928

929

930

931   **Supporting Text S1**

932   Sequence of the *pho1-7* locus with the production of the truncated PHO1 protein.
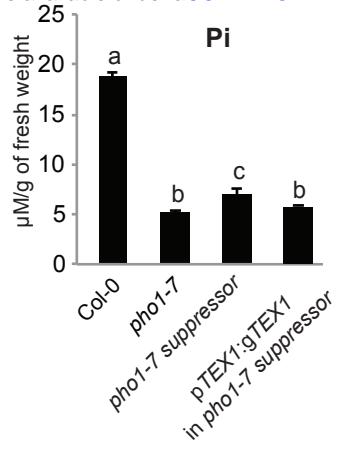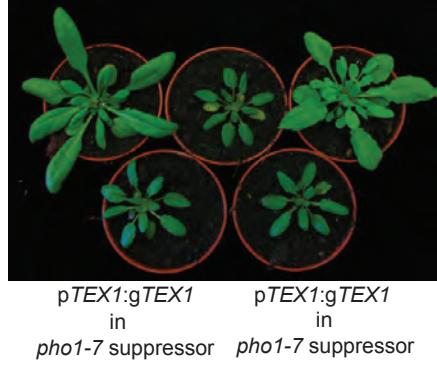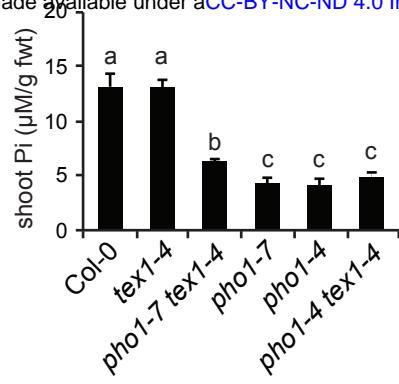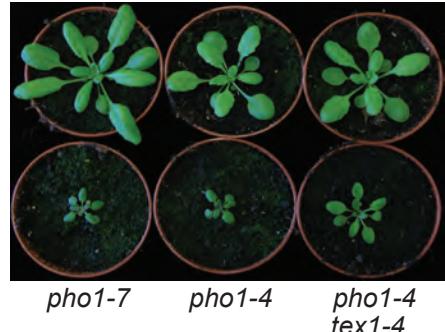
933

934

935   **Supporting Table S1**
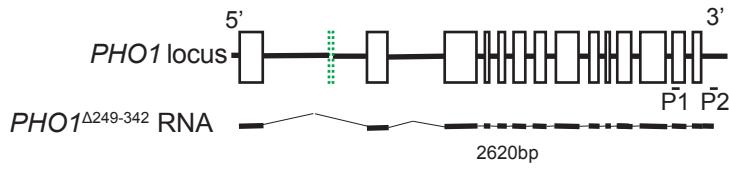
936   List of oligonucleotides used in this study.

937

A

Col-0   *pho1-7*   suppressor

p*TEX1*:g*TEX1*
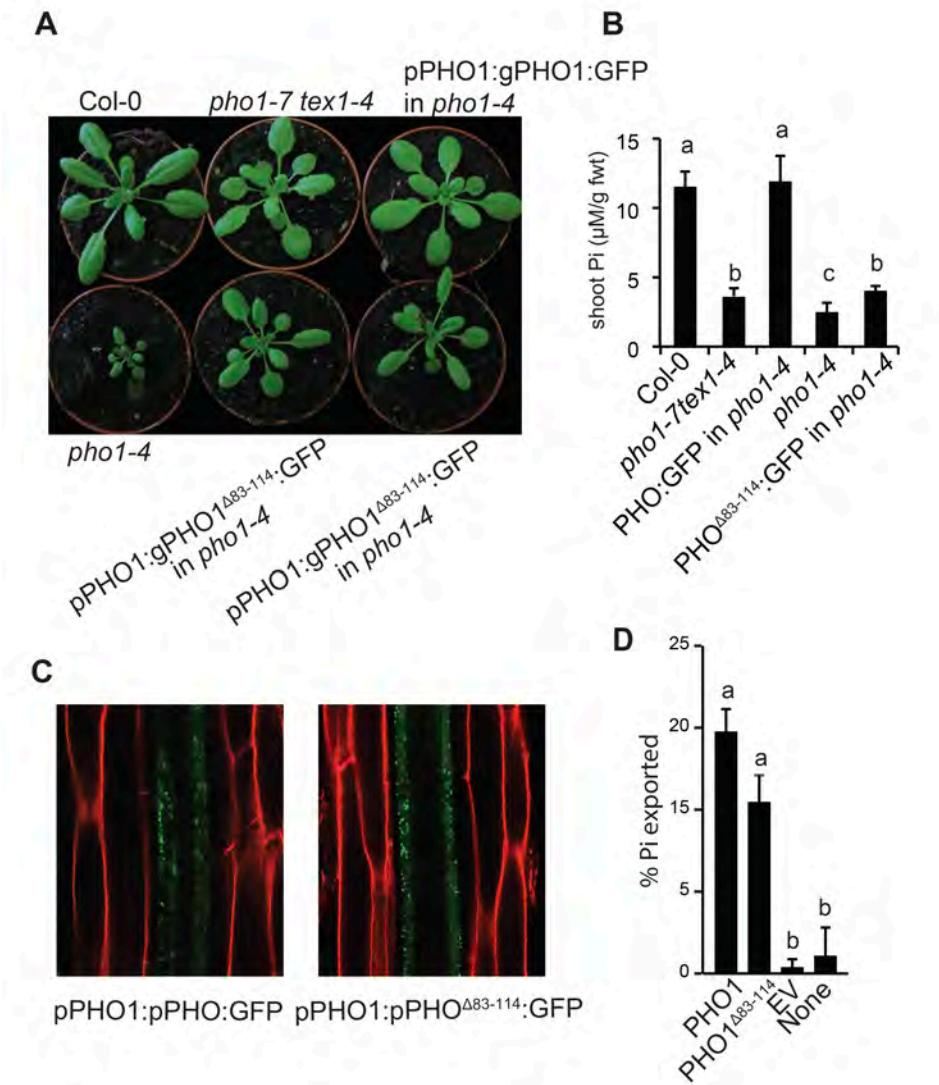in
pho1-7 suppressor

p*TEX1*:g*TEX1*
in
pho1-7 suppressor

B

Figure 1

A

Col-0 tex1-4 pho1-7 tex1-4

pho1-7 pho1-4 pho1-4 tex1-4

B

shoot Pi (μM/g fwt)

Col-0: a
tex1-4: a
pho1-7 tex1-4: b
pho1-7: c
pho1-4: c
pho1-4 tex1-4: c

C

PHO1 locus 5' 3'

P1 P2

PHO1$^{\Delta249-342}$ RNA

2620bp

D

PHO1
PHO1$^{\Delta83-114}$

Loading control

Col-0 pho1-2 pho1-7 pho1-7tex1-4

E *PHO1*

Relative expression

Col-0: a
tex1-4: a
pho1-7: b
pho1-7tex1-4: c

Figure 2

Figure 3

Figure 4

Figure 5

Figure 6

Figure 7

A

All genes

Genes with 3' extension

B

All genes

5'   3'

Genes with 3' extensions

5'   3'

C

AT1G76560 3'UTR mutated poly(A) site

Poly(A) site

AAUGAA (WT)

↓

AAUAAA (Mut.)

S1 Figure

Illumina sequencing

A.

Exon2

PHO1

Col-0

Read density

pho1-7

Read density

Pac-Bio sequencing

B.

pNOS NPTII tNOS

1   2   3   4 5 6 7 8 9 10 12 13 14 15

Col-0

Read density

pho1-7

Read density

pho1-7 tex1-4

Read density

pho1-7 hpr1-6

Read density
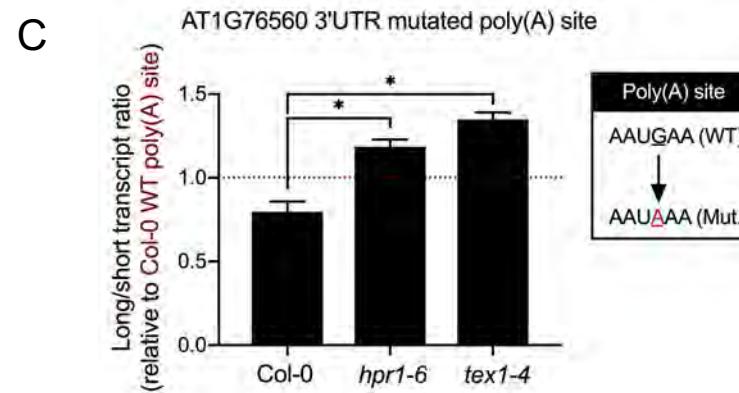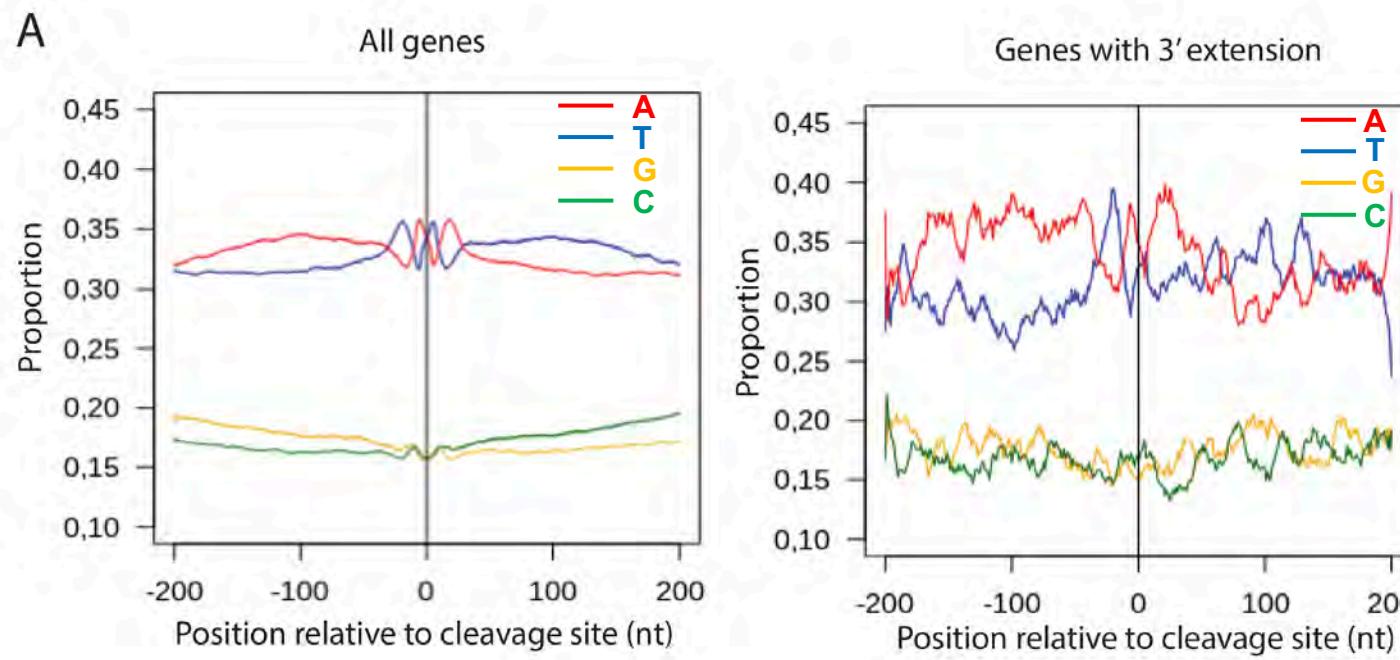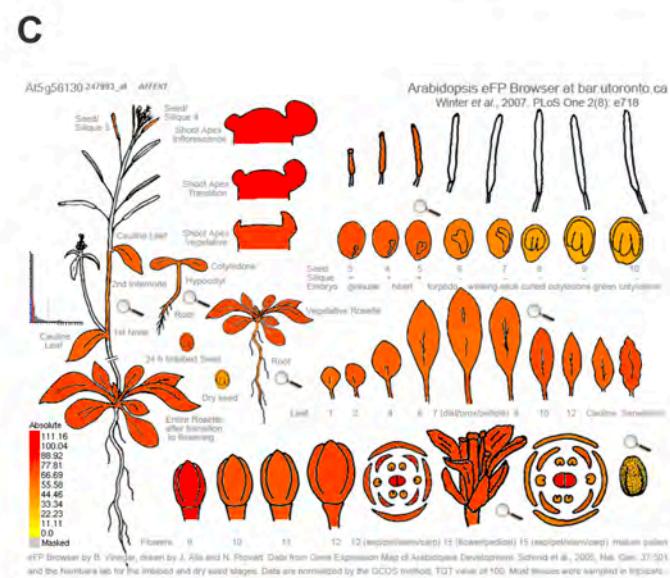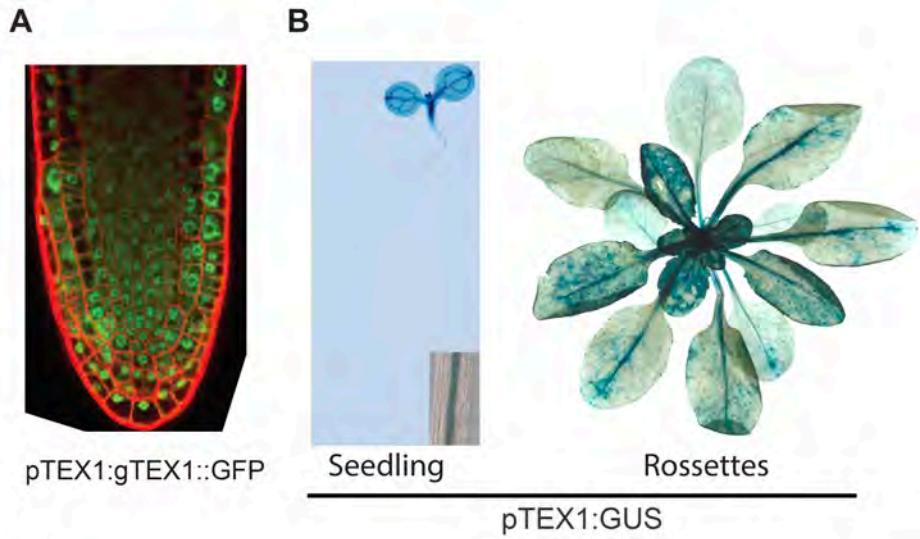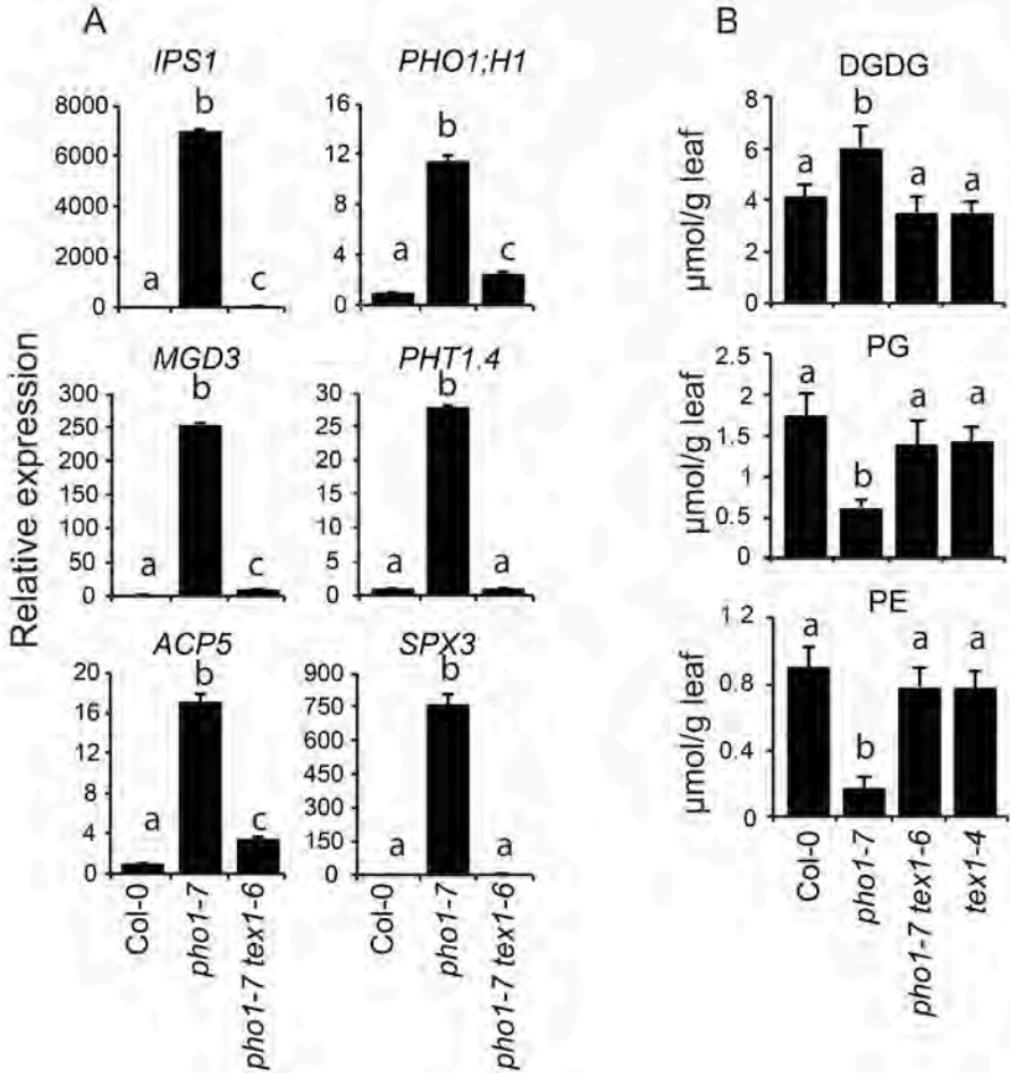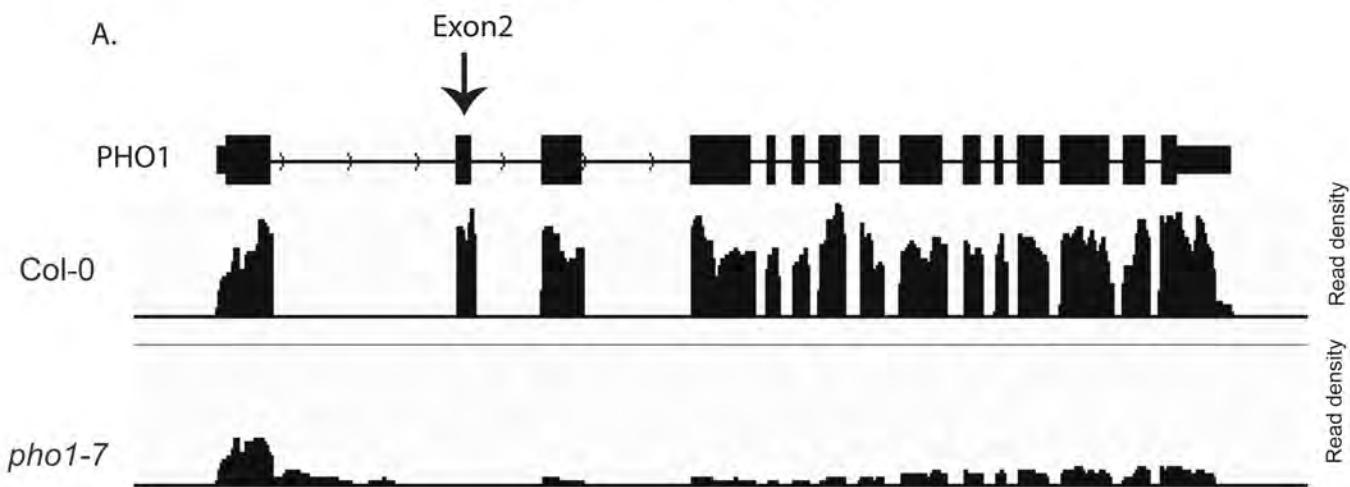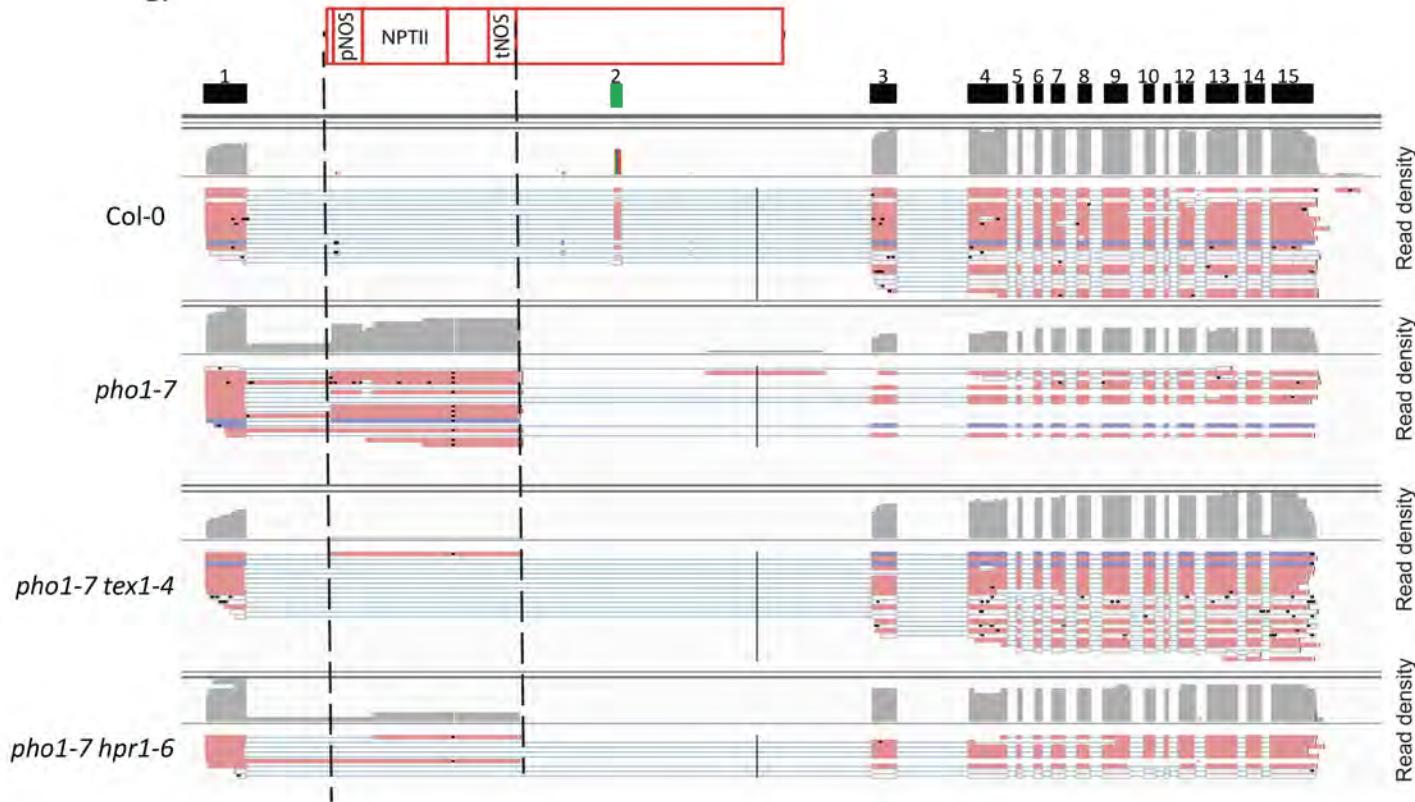
pho1-7tex1-4 vs pho1-7 in pot

24

29   48

tex1-4 vs Col0 in petris

Biological process
- Electron transport or energy pathways
- Developemntal processes
- Other biological processes
- Other metabolic processes
- Response to abiotic or biotic stimulus
- Transcription DNA-dependent
- Other cellular processes
- Response to stress
- Cell organization and biogenesis
- Protein metabolism
- Transport
- signal transduction
- Unknown biological processes

Molecular function
- Transcription factor activity
- Kinase activity
- Transferase activity
- Other enzyme activity
- DNA or RNA binding
- Nucleic acid binding
- Hydrolase activity
- Unknown molecular functions
- Protein binding
- Other binding

Cellular components
- ER
- Cell wall
- Other cellular components
- Chloroplast
- Plastid
- Other intracellular components
- Cytosl
- Ribosome
- Plasma membrane
- Unknown cellular components
- Other membranes
- Other cytoplasmic components
- Nucleus
- Extracellular
- Mitocondria
- No classification whatsoever

Enrichment level

0 1 2 3 4 5 6

Supplementary Text S1. Sequence of the *pho1-7* locus with the production of the truncated PHO1 protein.

T-DNA integration in *pho1-7* mutants

Capital letters show the exon.

Sequence highlighted in blue is T-DNA.

ATGGTGAAGTTCTCGAAGGAGCTAGAGGCACAACTTATACCGGAGTGGAAAGAGGCCTTTGT
TAACTATTGTTTACTAAAGAAACAAATCAAGAAAATCAAAACCTCTCGTAAACCAAAACCGG
CTTCTCATTACCCCATTGGTCATCACTCCGATTTTGGTCGATCTCTTTTCGACCCGGTTCGC
AAATTGGCCAGGACCTTCTCCGATAAACTATTTTCCAACTCAGAAAAACCAGAGATTCTCCA
Ggtaattaatcaactactttagttttgtcttaagaaaacatgcttgattcttttgtcggt
ttgaatgattagtctaaaattccgtgtaactttgtaacctagctctttatgtctaaatgcat
tttacggagtttgaacatgataccattagggactaaaagattacaattggtgagaccgtatg
tttttgagtttgtcggataaagaatttagattatttgtgataatgtagtgattttgttgtta
acttaatttaaattagcgtctttttggtcccatgattgttttgtattgagtttcgtgtatg
cttacgtctttgtaattgcttacgcgtcagaattaccagttatttctgctttgtctagtact
acatgagaaatctgcctttttcttgttttttctactttcttacaatattgatttgcttttcaa
aatatttattaacgagatatgtaaattttacatatttgacatgatttggtagagttctaata
tgttaattcttgtagagtaagttggtaattttgtgtgtgtattaaatatgtaatatataca
tatctaagtatttctgttcaggaaaacaatcattatacacatgatttagaccctatagctca
tatcttaacctagtaacctaccgctttccaaaattaagaaaaatcctataaattaactaacc
ctgaaatctgcaaaatatataattttgcggaaaataaatgctt**CCGGCTCTATCAAACACTG
ATAGTTTAAACTGAAGGCGGAAACCGAAATTCTAATAAGGGGGGAAAAATAAGGGGAGCCAC
TTTATCCCCCCCGCCGATGACGCGGGACAAGCCGTTTTACGTTTGGAACTGACAGAACCGCA
ACGTTGAAGGAGCCACTCAGCCGCGGGTTTCTGGAGTTTAATGAGCTAAGCACATACGTCAG
AAACCATTATTGCGCGTTCAAAAGTCGCCTAAGGTCACTATCAGCTAGCAAATATTTCTTGT
CAAAAATGCTCCACTGACGTTCCATAAATTCCCCTCGGTATCCAATTAGAGTCTCATATTCA
CTCTCAAtCCAAATAATCTGCACCGGATCTGGATCGTTTCGCATGATTGAACAAGATGGATT
GCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGA
CAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTT
GTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTG
GCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGG
ACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCC
GAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTG
CCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTC
TTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCC
AGGCTCAAGGCGCGCATGCCCGACGGCGATGATCTCGTCGTGACCCATGGCGATGCCTGCTT
GCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTG
TGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGC
GAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCgtagttCGATTCGCAG
CGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATG
ACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGA
AAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATC
TCATGCTGGAGTTCTTCGCCCACGGGATCTCTGCGGAACAGGCGGTCGAAGGTGCCGATATC
ATTACGACAGCAACGGCCGACAAGCACAACGCCACGATCCTGAGCGACAATATGATCGGGCC
CGGCGTCCACATCAACGGCGTCGGCGGCGACTGCCCAGGCAAGACCGAGATGCACCGCGATA
TCTTGCTGCGTTCGGATATTTTCGTGGAGTTCCCGCCACAGACCCGGATGATCCCCGATCGT
TCAAACATTTGGCAATAAAGTTTCTTAAGATTGAATCCTGTTGCCGGTCTTGCGATGATTAT
CATATAATTTCTGTTGAATTACGTTAAGCATGTAATAATTAACATGTAATGCATGACGTTAT
TTATGAGATGGGTTTTTATGATTAGAGTCCCGCAATTATACATTTAATACGCGATAGAAAAC
AAAATATAGCGCGCAAACTAGGATAAATTATCGCGCGCGGTGTCATCTATGTTACTAGATCG
GGCCTCCTGTCAATGCTGGCGGCGGCTCTGGTGGTGGTTCTGGTGGCGGCTCTGAGGGTGGT
GGCTCTGAGGGTGGCGGTTCTGAGGGTGGCGGCTCTGAGGGAGGCGGTTCCGGTGGTGGCTC
TGGTTCCGGTGATTTTGATTATGAAAAGATGGCAAACGCTAATAAGGGGGCTATGACCGAAA**

```
ATGCCGATGAAAACGCGCTACAGTCTGACGCTAAAGGCAAACTTGATTCTGTCGCTACTGAT
TACGGTGCTGCTATCGATGGTTTCATTGGTGACGTTTCCGGCCTTGCTAATGGTAATGGTGC
TACTGGTGATTTTGCTGGCTCTAATTCCCAAATGGCTCAAGTCGGTGACGGTGATAATTCAC
CTTTAATGAATAATTTCCGTCAATATTTACCTTCCCTCCCTCAATCGGTTGAATGTCGCCCT
TTTGTCTTTGGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAG
CTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTT
AGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGA
ATTGTGAGaagCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCT
TGCATGCCTGCAGGTCCCCAGATTAGCCTTTTCAATTTCAGAAGAATGCTAACCCACAGAT
GGTTAGAGAGGCTTACGCAGCAGGTCTCATCAAGACGATCTACCCGAGCAATAATCTCCAGG
AAATCAAATACCTTCCCAAGAAGGTTAAAGATGCAGTCAAAGATTCAGGACTAACTGCATC
AAGAACACAGAGAAAGATATATTTCTCAAGATCAGAAGTACTATTCCAGTATGGACGATTCA
AGGCTTGCTTCACAAACCAAGGCAAGTAATAGAGATTGGAGTCTCTAAAAAGGTAGTTCCCA
CTGAATCAAAGGCCATGGAGTCAAAGATTCAAATAGAGGACCTAACAGAACTCGCCGTAAAG
ACTGGCGAACAGTTCATACAGAGTCTCTTACGACTCAATGACAAGAAGAAAATCTTCGTCAA
CATGGTGGAGCACGACACACTTGTCTACTCCAAAAATATCAAAGATACAGTCTCAGAAGACC
AAAGGGCAATTGAGACTTTTCAACAAAGGGTAATATCCGGAAACCTCCTCGGATTCCATTGC
CCAGCTATCTGTCACTTTATTGTGAAGATAGTGGAAAAGGAAGGTGGCTCCTACAAATGCCA
TCATTGCGATAAAGGAAAGGCCATCGTTGAAGATGCCTCTGCCGACAGTGGTCCCAAAGATG
GACCCCCACCCACGAGGAGCATCGTGGAAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAA
GTGGATTGATGTGATATCTCCACTGACGTAAGGGATGACGCACAATCCCACTATCCTTCGCA
AGACCCTTCCTCTATATAAGGAAGTTCATTTCATTTGGAGAGAaagtACACGGGGGACTCTA
GAGGATCCCCGGGTACCGAGCTCGAATTCCCCGATCGTTCAAACATTTGGCAATAAAGTTT
CTTAAGATTGAATCCTGTTGCCGGTCTTGCGATGATTATCATATAATTTCTGTTGAATTACG
TTAAGCATGTAATAATTAACATGTAATGCATGACGTTATTTATGAGATGGGTTTTTATGATT
AGAGTCCCGCAATTATACATTTAATACGCGATAGAAAACAAAATATAGCGCGCAAACTAGGA
TAAATTATCGCGCGCGGTGTCATCTATGTTACTAGATaaggCGGGAATTCACTGGCCGTCGT
TTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATC
CCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTG
CGCAGCCTGAATGGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCG
GCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGG
CACCTCGACCCCAAAAAACTTGATTTGGGTGtATGGTTCACGTAGTGGGCCATCGCCCTGAT
AGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAA
ACTGGAACAACACTCAACCCTATCTCGGGCTATTCTTTTGATTTATAAGGGATTTTGCCGAT
TTCGGAACCACCATCAAACAGGATTTTCGCCTGCTGGGGCAAACCAGCGTGGACCGCTTGCT
GCAACTCTCTCAGGGCCAGGCGGTGAAGGGCAATCAGCTGTTGCCCGTCTCACTGGTGAAAA
GAAAAACCACCCCAGTACATTAAAAACGTCCGCAATGTGTTATTAAGTTGTCTAAGCGTCAA
TTTGTTTACACCACAATAtaattacatccctatattattttctataaaagctgaaacttaa
attttatttattattgggattctatataactagctatattaatactcttttgcttatactg
tataagcttgaaataaggacctattatttggtgcatttatttgactttattaatatcatcat
cactattactatattacgaattcctttttcttaagcttcaaaagcagaaaattaaataacat
agcacatagagatagtggtttattgctagcctaaattgcacaacacgaccacatcctaaat
cctaatgaagatggtcaaatttagactactttttcttccattaagagtttataaatgtttg
atctgtatttgaatgcttttgttattgttactttagGTTAAAGTGTTTTTCGCTAGATTGGA
TGAAGAACTAAACAAAGTGAATCAGTTTCACAAGCCCAAAGAAACAGAGTTTTTGGAAAGAG
GAGAGATTCTGAAGAAACAGTTGGAGACTCTTGCAGAACTCAAACAGATCTTAAGTGATCGG
AAGAAGAGAAACTTGTCTGGCTCAAATTCACATCGCTCCTTCTCATCTTCTGTTCGAAACTC
TGATTTCTCTGCAGgttagtcttcttcttattaatgcttttactacaagttaggtcagattt
agggtactttagtaaattctgcattttgacttaagagttaagactcttattaaaagctaacc
attagtatatttgagatatgacgcggcctgcacccattagaaatcacaaaagtgcatgtcct
agaggccgtttacgtagtgcggctagtaaaagaaaagtaggttatttaatattttatatc
attgcacaatttagcgttttt gaaataatttaagtaaaaattcaaatgttttgttataacac
```

```
aatagaaattctactactgaatgacaagatttgttaagcatacaaatcttttgcatgtggat
gcaaaactcgaagctgcatatatgtgaaagaaagaattttagagttttgagattttattatg
tagacatgcatatcttcaatggccgacaagaacacaaatcgattctactaacaaacgaacaa
aatgtttgtgagtaagaaaataaaaatggaccgacaatagatggctaggtcgacaaaaattt
ttattttcttcttctctttagacctgtgttaatagtcacaaaatatataattatttgtaatt
tatttttttcttattgttcagGGTCTCCAGGAGAACTAAGTGAGATACAAAGTGAAACATCAA
GAACAGATGAAATCATAGAGGCCTTAGAGAGGAACGGTGTGAGTTTCATAAACTCTGCAACG
AGGAGCAAAACAAAAGGAGGCAAACCAAAAATGTCTCTCCGCGTCGACATTCCCGACGCTGT
GGCCGGAGCCGAAGGTGGTATCGCAAGATCCATCGCCACCGCCATGTCTGTTCTTTGGGAAG
AGCTCGTTAACAACCCAAGATCAGATTTCACCAACTGGAAAAATATTCAAAGCGCCGAGAAG
AAGATACGAAGTGCCTTTGTTGAACTCTATAGAGGTCTTGGCTTGTTAAAGACTTACAGgta
tattcattaatgactcaatatcatttatttatttagtgatgatcttatcatttttaattctt
ttgttggctttattatggcagCTCGTTGAATATGATAGCTTTCACAAAAATAATGAAGAAAT
TCGATAAGgtaaattggttatagattgtactttcggtgataaaccaatgaaaaagataatca
tagtggttaatgatgaatctttgatatatatatgaatcatcagGTTGCTGGTCAGAATGCAT
CATCAACGTATCTCAAAGTCGTAAAGAGATCGCAATTTATCAGCTCTGATAAGgtaaaataa
aaggaggtcttatgctatgaataatttattgaaagattgtttgaaaatgattgtttgattaa
aattaaaaaagGTGGTAAGACTTATGGACGAAGTGGAGTCCATATTCACAAAGCACTTCGCC
AACAATGACAGGAAAAAGGCCATGAAATTCTTGAAACCCCACCAGACCAAAGATTCTCACAT
GGTCACTTTCTTTGTTGgtacttatttcattttctctatctttataccttttatataccattc
aaaaaaacagttcatcagagttttaacaaattgagattgtgtatctatgcagGGTTATTTAC
GGGTTGCTTCATCTCATTGTTTGTTATTTACATAATACTAGCCCATCTTTCTGGAATCTTCA
CTTCTAGTGATCAAGTCTCTTATCTGGAGACTGTTTATCCTGTTTTCAGgtaatgaataatt
atacgaattaatgatcaattcaacaaaactgtcaccatccaatgagacttaaccatttatcg
cttacattttgatgattttttttaaaaacgcagCGTTTTTGCGTTGCTGAGTCTACACATGT
TCATGTATGGATGCAATCTATACATGTGGAAGAACACGAGGATAAACTACACCTTTATTTTT
GAGTTTGCACCAAACACAGCGTTGCGTTACCGAGACGCGTTTCTGATGGGAACCACGTTCAT
GACCTCAGTTGTGGCAGCTATGGTCATCCACCTCATCCTCCGAGCCTCCGGTTTCTCAGCTA
GTCAAGTAGACACCATTCCAGGCATCCTCCTCCTGgtaaatcaaattacttagttcattaat
tatcatatggcgcgtttcaatcgcaatcgctatcacaatcacaatttgaaaccgctaatttc
tttttcggtgtgcatgctacagATCTTCATATGCGTCTTGATATGCCCATTTAACACATTCT
ACCGTCCAACAAGGTTCTGCTTCATCCGCATCTTGCGGAAGATTGTTTGCTCACCGTTCTAC
AAGgtaacaattggagttatttggttactttcagcacaagaaatagcagaaacatgattttt
ttttcttgtacggtaaatttagGTTTTGATGGTTGATTTCTTCATGGGCGATCAACTTACTA
GCCAGgtaaaactaagttatgcaacttcaatagatggtgacgacatctaatttagtcgataa
ttaacccattaatcgtgttctattcagATTCCATTGCTGAGACACCTTGAGACAACCGGGTG
TTACTTCTTGGCTCAAAGCTTCAAAACTCACGAATACAATACCTGCAAAAACGGAAGATACT
ATAGAGAATTTGCTTACTTGATTTCTTTCTTACCCTACTTCTGGCGTGCCATGCAAgtaagc
tcacttagggtttctctgtttttttttttttttgtcaagtcccttaaaacctttcttctaa
gacactatgaacattaatttacagTGTGTAAGGAGATGGTGGGACGAATCAAACCCTGATCA
CCTAATCAACATGGGAAAATACGTGTCAGCGATGGTTGCAGCCGGAGTCCGCATAACCTACG
CGAGAGAAACAACGACTTGTGGTTAACAATGGTGCTCGTAAGCTCCGTTGTGGCAACTATT
TACCAATTATACTGGGACTTTGTCAAGGATTGGGGTCTTCTAAACCCTAAATCGAAAAATCC
ATGGCTAAGAGACAATTTGGTTCTCCGGAACAAGAACTTCTACTACCTCTCCATTgtaagcc
aattacataactaactatagcgtgtttcacaatttatgatcttcgactaaatgttgagttgt
tcagGCGTTGAATTTGGTGTTGCGAGTTGCTTGGATCGAGACAATTATGAGATTCAGGGTCA
GTCCTGTTCAGTCTCATTTGCTAGATTTCTTCTTGGCGTCACTTGAAGTCATTCGTCGAGGC
CACTGGAACTTTTACAGgtaaataaaaaaacttcacctaggtttattaaaacttgattttgg
atgttattgaacatgaatctttctttcggtatttacagAGTGGAGAATGAGCACTTAAACAA
TGTCGGCCAATTTAGGGCAGTGAAGACCGTACCGTTACCGTTCCTTGACAGGGACTCAGACG
GTTAA
```

PHO1 coding cDNA

Sequence highlighted in green is deleted in *pho1-7* mutants (exon 2)

ATGGTGAAGTTCTCGAAGGAGCTAGAGGCACAACTTATACCGGAGTGGAAAGAGGCCTTTGTTAAC
TATTGTTTACTAAAGAAACAAATCAAGAAAATCAAAACCTCTCGTAAACCAAAACCGGCTTCTCATTA
CCCCATTGGTCATCACTCCGATTTTGGTCGATCTCTTTTCGACCCGGTTCGCAAATTGGCCAGGACCT
TCTCCGATAAACTATTTTCCAACTCAGAAAAACCAGAGATTCTCCAGGTAAGGAGAAGAAGAGGTA
GCTCAGAAACTGGGGATGACGTCGATGAGATTTACCAAACTGAACTTGTTCAGTTGTTTTCCGAAGA
AGACGAGGTTAAAGTGTTTTTCGCTAGATTGGATGAAGAACTAAACAAAGTGAATCAGTTTCACAA
GCCCAAAGAAACAGAGTTTTTGGAAAGAGGAGAGATTCTGAAGAAACAGTTGGAGACTCTTGCAG
AACTCAAACAGATCTTAAGTGATCGGAAGAAGAGAAACTTGTCTGGCTCAAATTCACATCGCTCCTT
CTCATCTTCTGTTCGAAACTCTGATTTCTCTGCAGGGTCTCCAGGAGAACTAAGTGAGATACAAAGT
GAAACATCAAGAACAGATGAAATCATAGAGGCCTTAGAGAGGAACGGTGTGAGTTTCATAAACTCT
GCAACGAGGAGCAAAACAAAAGGAGGCAAACCAAAAATGTCTCTCCGCGTCGACATTCCCGACGCT
GTGGCCGGAGCCGAAGGTGGTATCGCAAGATCCATCGCCACCGCCATGTCTGTTCTTTGGGAAGAG
CTCGTTAACAACCCAAGATCAGATTTCACCAACTGGAAAAATATTCAAAGCGCCGAGAAGAAGATAC
GAAGTGCCTTTGTTGAACTCTATAGAGGTCTTGGCTTGTTAAAGACTTACAGCTCGTTGAATATGATA
GCTTTCACAAAAATAATGAAGAAATTCGATAAGGTTGCTGGTCAGAATGCATCATCAACGTATCTCA
AAGTCGTAAAGAGATCGCAATTTATCAGCTCTGATAAGGTGGTAAGACTTATGGACGAAGTGGAGT
CCATATTCACAAAGCACTTCGCCAACAATGACAGGAAAAAGGCCATGAAATTCTTGAAACCCCACCA
GACCAAAGATTCTCACATGGTCACTTTCTTTGTTGGGTTATTTACGGGTTGCTTCATCTCATTGTTTGT
TATTTACATAATACTAGCCCATCTTTCTGGAATCTTCACTTCTAGTGATCAAGTCTCTTATCTGGAGAC
TGTTTATCCTGTTTTCAGCGTTTTTGCGTTGCTGAGTCTACACATGTTCATGTATGGATGCAATCTATA
CATGTGGAAGAACACGAGGATAAACTACACCTTTATTTTTGAGTTTGCACCAAACACAGCGTTGCGT
TACCGAGACGCGTTTCTGATGGGAACCACGTTCATGACCTCAGTTGTGGCAGCTATGGTCATCCACC
TCATCCTCCGAGCCTCCGGTTTCTCAGCTAGTCAAGTAGACACCATTCCAGGCATCCTCCTCCTGATC
TTCATATGCGTCTTGATATGCCCATTTAACACATTCTACCGTCCAACAAGGTTCTGCTTCATCCGCATC
TTGCGGAAGATTGTTTGCTCACCGTTCTACAAGGTTTTGATGGTTGATTTCTTCATGGGCGATCAACT
TACTAGCCAGATTCCATTGCTGAGACACCTTGAGACAACCGGGTGTTACTTCTTGGCTCAAAGCTTCA
AAACTCACGAATACAATACCTGCAAAAACGGAAGATACTATAGAGAATTTGCTTACTTGATTTCTTTC
TTACCCTACTTCTGGCGTGCCATGCAATGTGTAAGGAGATGGTGGGACGAATCAAACCCTGATCACC
TAATCAACATGGGAAAATACGTGTCAGCGATGGTTGCAGCCGGAGTCCGCATAACCTACGCGAGAG
AAAACAACGACTTGTGGTTAACAATGGTGCTCGTAAGCTCCGTTGTGGCAACTATTTACCAATTATAC
TGGGACTTTGTCAAGGATTGGGGTCTTCTAAACCCTAAATCGAAAAATCCATGGCTAAGAGACAATT
TGGTTCTCCGGAACAAGAACTTCTACTACCTCTCCATTGCGTTGAATTTGGTGTTGCGAGTTGCTTGG
ATCGAGACAATTATGAGATTCAGGGTCAGTCCTGTTCAGTCTCATTTGCTAGATTTCTTCTTGGCGTC
ACTTGAAGTCATTCGTCGAGGCCACTGGAACTTTTACAGAGTGGAGAATGAGCACTTAAACAATGTC
GGCCAATTTAGGGCAGTGAAGACCGTACCGTTACCGTTCCTTGACAGGGACTCAGACGGTTAA

PHO1 protein sequence

Sequence highlighted in red is deleted in *pho1-7* mutants
Sequence underlined form the 3 SPX subdomains, with amino acids in green involved in inositol polyphosphate binding

MVKFSKELEAQLIPEWKEAFVNYCLLKKQIKKIKTSRKPKPASHYPIGHHSDFGRSLFDPVRKLARTFSDKL
FSNSEKPEILQVRRRRGSSETGDDVDEIYQTELVQLFSEEDEVKVFFARLDEELNKVNQFHKPKETEFLERG
EILKKQLETLAELKQILSDRKKRNLSGSNSHRSFSSSVRNSDFSAGSPGELSEIQSETSRTDEIIEALERNGVS
FINSATRSKTKGGKPKMSLRVDIPDAVAGAEGGIARSIATAMSVLWEELVNNPRSDFTNWKNIQSAEKKI
RSAFVELYRGLGLLKTYSSLNMIAFTKIMKKFDKVAGQNASSTYLKVVKRSQFISSDKVVRLMDEVESIFTK
HFANNDRKKAMKFLKPHQTKDSHMVTFFVGLFTGCFISLFVIYIILAHLSGIFTSSDQVSYLETVYPVFSVFA
LLSLHMFMYGCNLYMWKNTRINYTFIFEFAPNTALRYRDAFLMGTTFMTSVVAAMVIHLILRASGFSAS
QVDTIPGILLLIFICVLICPFNTFYRPTRFCFIRILRKIVCSPFYKVLMVDFFMGDQLTSQIPLLRHLETTGCYFL
AQSFKTHEYNTCKNGRYYREFAYLISFLPYFWRAMQCVRRWWDESNPDHLINMGKYVSAMVAAGVRI
TYARENNDLWLTMVLVSSVVATIYQLYWDFVKDWGLLNPKSKNPWLRDNLVLRNKNFYYLSIALNLVLR
VAWIETIMRFRVSPVQSHLLDFFLASLEVIRRGHWNFYRVENEHLNNVGQFRAVKTVPLPFLDRDSDG

Primers Table

Primer Name	Primer sequence	Description
SAIL_1209_F10 LP	GTTTAGGCACCAGGGAAGAAC	Genotyping hpr1-6 mutants
SAIL_1209_F10 RP	TTCTATGCCAGCTGATGGATC	
TEX1-gen F	CACCGAGTGGTCACTGTATCAAGGCATT	Cloning of TEX1 Genomic region
TEX1-gen R	TGTGCTCTTCTTGTTCTTTAATCCA	
TEX1-gen F	CACCGAGTGGTCACTGTATCAAGGCATT	Cloning of TEX1 promoter and gene without stop for GFP fusion
TEX1-gen R w/o stop	CGAGCTCTCAAAACCAAATATCCGGA	
TEX1 Pro infu LP	CCGGTACCGAATTCGGAAGCACAGGCAGGAATGGAGTAA	Cloning of TEX1 promoter for GUS fusion
TEX1 Pro infu RP	TAGATATCTCGAGTGTTTTTAATTCTCAACTCTTCTTC	
qRT-IPSI F	GCGTTTTAAGATATGGAGCAATG	qPCR
qRT-IPSI R	CGAAGCTTGCCAAAGGATAG	
qRT-MGD3 F	GGTACGATTGCGGAAGCACTG	qPCR
qRT-MGD3 R	GTCGAACACGGCTTCAGGTTG	
qRT PHO1;H1 LP	AACCGGTAGCTATGCAACACAGGA	qPCR
qRT PHO1;H1 RP	TTCCAGCGGCTAACATAGCTGAAA	
PHT1.4-qRT-F	CCA CGA TTC CTC AAG CTG AT	qPCR
PHT1.4-qRT-R	CAA CCA AAG CCG TGT ACC TT	
ACP5 qRT-F	TGC ACA AAT GGG TCA CTT CT	qPCR
ACP5 qRT-R	GAA GCC TAC CTA GCC TGC AA	
qRT-SPX3 LP	CGCCGGTGGAATCTATTTTCG	qPCR
qRT-SPX3 RP	CAGAACCATTTCGCCATGGAA	
qRT- ACT2 F	AGTGGTCGTACAACCGGTATTGT	qPCR
qRT- ACT2 R	GATGGCATGGAGGAAGAGAGAAAC	
qRTPHO17-1LP	ACCTAGTAACCTACCGCTTTCCAA	qPCR 1st RNA isoform at the PHO1 locus in pho1-7 mutants
qRTPHO17-1RP	TCCTTCAACGTTGCGGTTCTGTCA	
qRTPHO17-2LP	TCTCTTTTCGACCCGGTTCGCAAA	qPCR 2nd RNA isoform at the PHO1 locus in pho1-7 mutants
qRTPHO17-2RP	TCCTTCAACGTTGCGGTTCTGTCA	
qRTPHO17-3LP	AGAGATTCTCCAGGTTCTCCGGCCGC	qPCR 3rd RNA isoform at the PHO1 locus in pho1-7 mutants
qRTPHO17-3RP	CCTCGTCCTGCAGTTCATTCAGGG	
qRTPHO1 LP	GACAATTTGGTTCTCCGGAACAAG	qPCR PHO1 full length
qRTPHO1 RP	GAACGGTAACGGTACGGTCTTCAC	
AT1G76560-UTRext-F	ACCGTTGTTCATCGTGTTCACT	Expression of 3'UTR extensions by qPCR
AT1G76560 -UTRext-R	CGGACTCGGCGGATCCAGATTA	
AT1G03160-UTRext-F	CAGCTCGAAGACGCTATGCAA	Expression of 3'UTR extensions by qPCR
AT1G03160-UTRext-R	GAAGGGAAAACCCGACATTGTG	
AT3G13110-UTRext-F	TGGAGTTTCCTGTGGTTGGTTTC	Expression of 3'UTR extensions by qPCR
AT3G13110-UTRext-R	GCAAGCAATCCTGGATCCTGAC	
AT2G41200-UTRext-F	TGCGATCGCCGTCTTCATTA	
AT2G41200-UTRext-R	GCTTCTCCGTGTATCAAACACAT	
TEX1-m F	GCGAAGGAAGGTTGAAGAAGAAG	Confirm tex1-6 mutation by PCR and sequencing
TEX1-m R	TCCCAACATAGCTGATCCACACT	
P2BJ pPHO1 1exon L	CCGGTACCGAATTCGTCTTGATGATTCTTTGTCCAAGTA	Cloning PHO1Δ83-114: Amplification of PHO1 promoter and first exon.
P2BJ pPHO1 1exon R	TCTAAGATGTAGAGTAATAGAGTGTAGTAA	
P2BJ PHO1gene 3rd exon F	ACTCTACATCTTAGACCCTATATTATTTTTCTATAAAAG	Cloning PHO1Δ83-114: Amplification of PHO1 from third exon. Two amplicons were joned togther in intron deleting the second exon
P2BJ PHO1gene 3rd exon R	ATATCTCGAGTGCGGACCGTCTGAGTCCCTGTCAAGGAA	
Pho1_ex1_LP	CGAAGGAGCTAGAGGCACAACTTATACC	CHIP-qPCR: Primer binding within exon1 of pho1
Pho1_ex1_RP	GAAGCCGGTTTTGGTTTACGAGAGG	
Pho1_int1-3'_4_LP	CATAGCACATAGAGATAGTGGTTTTATTGCTAGCCT	CHIP-qPCR: Primers binding within 3' part of intron 1 of pho1.
Pho1_int1-3'_4_RP	GACCATCTTCATTAGGATTTAGGATGTGGTCGTG	
nLuc_qPCR-F	GACCCTGTGGAACGGCAACAA	Expression of luciferase fused to 3"UTRs of At1G76560 in protoplasts
At1g76560_3UTR_qPCR-R1	CAGTCTCTGTTCTAAGTCCTCTGT	Reverse primer for short RNA-Primer is placed before the predicted polyA signal
At1g76560_3UTR_qPCR-R2	CGGACTCGGCGGATCCAGATTA	Reverse primer for long RNA-Primer is placed after the predicted polyA signal
Fluc_qPCR-F	TGCGCGGAGGAGTTGTGTTTGTG	Primers for qPCR of luciferase as an internal control for normalisation of expression.
Fluc_qPCR-R	ACGGCGATCTTTCCGCCCTTCT	