Evidence-based gene expression modulation correlates with transposable element knock-down

José Hernandes-Lopes[1*], Danielle M. Quintanilha[1*], Erika M. de Jesus[1], Fabrício M. Lopes[2], Raphael B. Parmigiani[3,4], Bruno Karolski[1,5], Henrique M. Dias[1], Thomas B. Jacobs[6], Anamaria A. Camargo[3,7], João P. Kitajima[8] and Marie-Anne Van Sluys[1**].

[1]Departamento de Botânica, Universidade de São Paulo, São Paulo, Brazil. [2]Universidade Federal Tecnológica do Paraná, Cornélio Procópio, Brazil. [3]Ludwig Institute for Cancer Research, São Paulo, Brazil. [4]Idengene Medicina Diagnóstica, São Paulo, Brazil. [5]Laboratório de Simulação e Controle de Processos (LSCP-POLI-USP), São Paulo, Brazil. [6]Center for Plant Systems Biology, VIB, Ghent, Belgium. [7]Centro de Oncologia Molecular, Hospital Sírio-Libanês, São Paulo, Brazil. [8]Mendelics Análise Genômica, São Paulo, Brazil.

[*] Both authors contributed equally to this work

**Correspondence: mavsluys@usp.br

ABSTRACT

Background: Transposable elements (TEs) are major components of plant genomes. Despite being regarded as "junk DNA" at first, TEs play important roles for the organisms they are found in. The most obvious and easily recognizable effects caused by TEs result from their mobility, which can disrupt coding sequences or promoter regions. However, with the recent advances in transcriptomics, it is becoming increasingly evident that TEs can act as an additional layer of gene expression regulation through a number of processes, which can involve production of non-coding RNAs. Here, we describe how Tnt1, a stress-responsive LTR-retrotransposon, interferes with gene expression and modulate a number of developmental aspects in tobacco.

Results: Through an RNAi approach, we generated tobacco (HP) lines knocked-down for Tnt1 expression. Quantitative RT-PCR experiments confirm that Tnt1 is downregulated in HP lines after ethylene exposure. A RNA-seq experiment was performed and through two independent bioinformatic approaches (with different stringencies) we found 932 and 97 differentially expressed genes in HP lines. A number of phenotypes were observed in such lines, namely lesion mimicry in leaves, underdevelopment of the root system, overproduction of root hairs and early loss of seed viability. Folding prediction of part of the Tnt1 mRNA reveals putative stem-loop secondary structures containing transcriptional regulation sequences, suggesting it could be a source of small RNAs. We also propose a model to explain the Tnt1 expression in both homeostatic and stress conditions, and how it could interact with stress-responsive genes.

Conclusions: Our results are consistent that interferences with Tnt1 transcript levels correlate with transcriptomic and phenotypic changes, suggesting a functional role for this element during plant development and stress response.

Keywords: ethylene, transcriptome, retrotransposon, Tnt1, tobacco, RNAi

BACKGROUND

Most genomes harbor a particular type of genetic elements collectively known as Transposable Elements (TE). TEs encode proteins that enable their own mobilization within the genome. Discovered by Nobel laureate Barbara McClintock in the 1940's, TEs were first described as "controlling elements" since mobilization of specific TEs (*Ac* and *Ds*) in maize caused a variegated phenotype in kernels due to an insertion into the C locus that is responsible for anthocyanin synthesis [1]. Despite this, TEs in eukaryotes were at first widely regarded as "junk DNA" or "selfish genes" due to their self-replicating nature, mutagenic potential and the lack of an obvious function for the host genome [2].

The idea of TEs as simply being genomic parasites was gradually abandoned with the ever-increasing understanding of eukaryotic genome structure. TEs are known to be found in virtually all domains of life [3]. Whereas TEs make up about 45% of the human genome [4,5],

the TE-derived content can be much higher in domesticated plants such as sorghum (~62%), tomato (~63%), wheat (~80%) and maize (~85%) [6, 7]. Nowadays it is well accepted that TEs are able to generate genome structural and functional variation as a result of their mobile nature and predisposition to recruit epigenetic silencing mechanisms. Through a number of processes, they can deeply affect epigenetic variation, alter or create new gene regulatory networks, as well as the formation of new proteins through the spreading of functional motifs [8, 9, 10].

Given the potential deleterious effects of transposition, expression and mobility of TEs are usually tightly controlled in eukaryotes. Modifications of histone tails, DNA methylation and alterations in chromatin packing and condensation are amongst the most well-known mechanisms involved in TEs silencing, but there are examples of post-transcriptional silencing of TEs by RNA interference (RNAi) (see 11 for a review). Interestingly, in vertebrates silencing can be achieved by DNA editing through APOBEC enzymes, which selectively edit the promoter region of LTR type retroelements [12, 13]. Thus, while deleterious effects of new insertions are negatively selected, advantageous changes can be incorporated (for reviews, see reference 9 and 14), playing an important role in evolution. Such is the case of maize cultivars, which are living examples of genome evolution driven by transposable elements [15].

Besides the classically-accepted role as drivers of genetic diversity, stress-related expression of TEs is also demonstrated to participate in different regulatory pathways, such as the human *Alu* element, which seems to regulate protein translation after exposure to stress [16]. While there are many examples of stress-induced expression of TEs, recent studies also recognize TEs as important components for the maintenance of biological processes. Also, TEs have been shown to be a significant source of noncoding RNAs and to interfere in the small RNA (sRNA) machinery, which are key regulators of gene expression in plants [17, 18]. An example of such interference is the expression of the TE MIKKI during rice root development. MIKKI transcripts act as decoys for miR171, which usually targets and silences *OsSCL21*. By mimicking *OsSCL21*, MIKKI sequesters miR171 molecules, culminating in *OsSCL21* upregulation [19].

The Tnt1 retrotransposon is an interesting transcriptionally-active TE in somatic tissues of *Nicotiana tabacum* plants growing under normal conditions [20, 21]. Tnt1 (Genbank: X13777) is a multicopy Long Terminal Repeat (LTR) retrotransposon that was first discovered after its insertion in the nitrate reductase coding sequence [22]. It is estimated to have more than 600 insertions in the tobacco genome [23] and it also has homologs in other Solanaceae under different names (*e.g.* Retrolyc1/TLC1 in tomato and Retrosol in potato) [24–26].

Being an LTR element, Tnt1 is composed of central open-reading frames (ORFs) flanked by 5' and 3' LTRs, which can be further divided in U3, R and U5 regions [27]. In tobacco, Tnt1 insertions are classified in three subfamilies (Tnt1A, Tnt1B and Tnt1C), which have highly conserved sequences except for their U3 regions [28]. Regulatory sequences present in each of the U3 are responsive to different hormone induction [20, 21, 29]. Interestingly, the U3 region of Tnt1A (U3A) shows sequence repeats highly similar to the GCC core, present in the promoter region of ethylene-responsive genes [30, 31]. These results suggest that expression of this subfamily could also be induced by ethylene, but they remain to be experimentally validated. Expression of Tnt1 is also known to occur in homeostatic conditions. During normal development, expression of Tnt1 was reported in roots, leaves and petals [20]. Although interesting, more data on this basal expression is lacking in the literature, and whether it has a role for tobacco development or represent a residual escape from endogenous silencing mechanisms is still an open question.

Given the ever-increasing number of proposed mechanisms through which TEs can exert important functions, either at genome or organism levels, Tnt1 is an interesting target for functional studies. Thus, this work aims to understand the Tnt1 pattern of expression and its potential role during plant development. Through an RNAi approach we observed phenotypic changes caused by Tnt1 downregulation, such as increased root hair production, underdevelopment of the root system and decreased seed viability. Transcriptome profiling of downregulated Tnt1 plants revealed the dynamics of Tnt1 expression and a close association with ethylene biosynthesis and responsive genes in tobacco. Taken together, these results reveal the importance of the Tnt1 retroelement for normal tobacco development.

RESULTS

Tnt1 expression knockdown reveals a connection with ethylene biosynthesis and responsive genes

Given the presence of GCC-like motifs in the U3A region of Tnt1A insertions, we first tested if expression of Tnt1 is upregulated upon ethylene stimulus. In parallel, to understand if perturbations in the level of Tnt1 transcripts would have any detectable effect, we developed transgenic tobacco lines expressing an RNAi (hairpin, HP) construction under the Cauliflower mosaic virus 35S promoter, targeting the Tnt1 reverse transcriptase (RT) domain (Additional file: Figure S1). These lines are herein called HP(x), where (x) corresponds to an independent transformation event, followed by T(n), where (n) correspond to the generation of the transgenic line (T0 = the plant regenerated from the callus).

We treated two WT and two HP1 (T2) plants with ethylene, while two other individuals of each genotype did not receive the treatment (control group). After 24 hours of treatment we quantified the expression (quantitative RT-PCR) of three different Tnt1 coding domains in all individuals. Among individuals that were not exposed to ethylene treatment, HP1 plants had higher expression level of all three Tnt1 domains than the WT (Figure 1A). However, upon ethylene treatment both WT and HP1 plants overexpressed the three Tnt1 domains when compared to the control group, demonstrating that this gaseous hormone indeed induces Tnt1 expression. Interestingly, while WT plants increased Tnt1 expression by a factor of 35 to 50, HP1 plants had an increase of around 20-fold. These qRT-PCR results are consistent with decrease of Tnt1 expression in HP lines treated with ethylene, confirming the knock-down of Tnt1 by the RNAi construct. Since ethylene induced Tnt1 expression and it was upregulated in untreated HP1 plants, we asked whether HP plants produced more ethylene under normal conditions. Thus, we tested ethylene emissions in 90-day-old WT, HP1 (T1), HP1 (T2) and HP8 (T1) plants using gas chromatography. All HP samples showed a significant increase in ethylene production when compared to WT (Figure 1B).

Because ethylene biosynthesis is known to follow a circadian cycle, we also investigated the expression dynamics of Tnt1 and some selected ethylene biosynthesis and responsive genes in WT plants throughout a period of 48 hours. For these, we used 15-day-old plants grown under a 12-hours-light / 12-hours-dark regime. Expression of two genes related to the circadian clock in *Nicotiana* species confirmed synchronization of the samples (*NtCP-23* and *NtTOC1*, reaching highest levels of expression at 12PM and 6PM respectively; Additional file 1: Figure S2A).

As expected, the ethylene biosynthetic genes *ACO1* and *ACO2* were transcribed in a circadian fashion in WT plants, peaking at the beginning of the light period (6h; Figure 2A). Accordingly, the ethylene responsive genes *ER24*, *JERF1* and *TEIL* also presented a circadian cycle: while *ER24* is consistently more expressed at noon, both *JERF1* and *TEIL* reached highest expression values at 6h, much like both *ACO1* and *ACO2* (Figure 2B). Expression of Tnt1 in the WT (evaluated by the RT domain) however, showed no signs of a circadian rhythm, presenting great variation in expression between biological replicates (Figure 2C). We then checked specific expression of the U3A region (expected to be responsive to ethylene). When biological replicates are plotted separately, it becomes clear that expression of Tnt1A is indeed decoupled from the circadian cycle as each sample showed different levels of U3A expression (Figure 2C).

To further analyze the connection between Tnt1 expression and ethylene biosynthetic and responsive genes, WT and HP plants (the same used for measurement of ethylene in gas chromatography) were also used for qRT-PCR experiments. Most genes assayed related to ethylene synthesis and response, were downregulated in HP plants when compared to the WT (Figure 3). Two copies of the ACC oxidase (*ACO1* and *ACO2*) gene, which participates in ethylene biosynthesis, were slightly downregulated (with relative expression ranging from 0.6 to 0.8 when compared to WT, except for *ACO2* in HP8 with expression 1.4 times higher than the WT line (Figure 3). The ethylene responsive genes *JERF1*, *ER24*, *SAR8.2b*, *TEIL* and *CHN48* were also downregulated in HP lines, except for *TEIL* in HP8, which had expression comparable to the WT (Figure 3). Finally, we also assayed three different Tnt1 coding domains

(Gag, Integrase and Reverse Transcriptase), and they were downregulated in all HP plants (Figure 3).

HP lines transcriptome profile differs from wild-type tobacco

Given the perturbations found in the ethylene biosynthesis and response pathways of HP lines, we asked whether these alterations could influence global gene expression. Thus, a whole transcriptome RNA-seq was performed on leaves of 45-day-old WT, empty-vector control plants (transformed with a vector with the same backbone but lacking the hairpin construction) and HP lines [HP1 (T1), HP1 (T2) and HP8 (T1)]. Because the most recent published tobacco genome is not yet complete, but is partly assembled into chromosomes by optical mapping, we expect a number of undefined nucleotides between scaffolds and is likely missing a number of gene annotations. Thus, we followed two approaches to process our RNA-seq data: (1) mapping reads against the tobacco whole genome in Solgenomics database and (2) mapping reads against unigenes from the tobacco database in Genbank.

Of the 35,519 annotated gene models in the tobacco genome, 16,331 had detectable expression levels [counts per million (CPM) > 1 in at least eight sequenced samples (including replicates)]. From these, we identified 932 differentially expressed genes (DEGs) in HP lines compared to the WT and control lines (FDR < 0.001, fold change $\geq$ 2). Hierarchical clustering of expression values for the 932 DEGs grouped samples in two classes, segregating HP lines from WT and control lines (Figure 4).

Enrichment of Gene Ontology (GO) terms (p-value $\leq$ 0.01) identified a number of modulated biological processes in the HP lines (Figure 5). The most represented terms among the upregulated genes were: organophosphate metabolic process (GO:0019637), carbohydrate catabolic process (GO:0016052) and anion transport (GO:0006820); while metabolic process (GO:0008152), cellular process (GO:0009987) and cellular metabolic process (GO:0044237) were the most frequent terms of downregulated genes. Interestingly, we observed enrichment for stress-related processes in both sets of genes [upregulated: reactive oxygen species metabolic process (GO:0072593), defense response (GO:0006952), response to ethylene

(GO:0009723) and reactive nitrogen species metabolic process (GO:2001057); downregulated: homeostatic process (GO:0042592), cellular homeostasis (GO:0019725) and cell redox homeostasis (GO:0045454)].

A second, more stringent approach was performed, in which each transgenic line (Control and HPs) were compared against WT. We then considered only DEGs consistently found in all HP samples and filtering out DEGs identified between WT and Control lines. Ninety seven DEGs were found (p-value $\leq$ 0.001, fold change $\geq$ 2), reflecting significant changes in gene expression. GO categorization of these 97 DEGs corroborates the results found in the previous analysis, with terms like “defense response”, “response to biotic stimulus” and “response to ethylene” appearing in the upregulated gene set (Additional file 1: Figure S2B). We used these 97 DEGs as seeds to start the inference of the gene regulatory networks using entropy based approach from gene expression patterns. The inference process was performed by selecting the predictors for each seed gene on each step. Thirty-five networks modules were identified revealing gene circuits in HP plants that were not identified in WT. From these, the most representative network module and which had the highest number of nodes connecting 44 genes is presented in Additional file 1: Figure S3.

Phenotypic changes in HP lines

HP plants displayed a varied range of phenotypes. Lesion mimicry was readily observed in leaves of all four HP lines selected for this study (Figure 6A). These necrotic spots were present in T1 plants growing under normal conditions, but also in subsequent generations (up to T5) when exposed to stress, such as transfer from *in vitro* cultivation to soil (Figure 6A).

To explore the consequences of interfering with Tnt1 expression we compared the organization of the shoot and root, as well as the germination rate of WT and HP lines (HP8 and HP13). Two measures were considered at two-leaf stage plants (15-days old): total foliar area and the maximum distance across the longer longitudinal axis of the two leaves. Under these parameters, the HP13 line showed significantly increased growth of the shoot system; the total

220 foliar area was 30% greater than the wild type and 15% longer in the longitudinal axis (Figure
221 6B, Table 1).

222 Several morphological phenotypes were observed in roots. HP lines growing *in vitro*
223 developed longer root hairs close to the root tip when compared to WT (Figure 7A, Additional
224 file 1: Figure S4A). Root growth was also altered in HP lines: primary root length, surface and
225 volume were significantly smaller in 15-days-old plants (Figure 7B; Table 2). The same result
226 was observed when comparing the whole root system (Additional file 1: Figure S4B; Table 2).
227 HP lines also tended to produce fewer lateral roots (Figure 7B; Additional file: Figure S4B;
228 Table 2).

229 We also compared germination of 6-year-old and fresh seeds of WT and HP lines
230 considering two parameters: root and cotyledon emergence. Germination was delayed in HP
231 seeds when compared to the WT (Figure 8; Additional file 1: Figure S5). For example, at nine
232 days after sowing most of the fresh WT seeds had both root and cotyledons emerged, with only
233 1.33% of the seeds partially germinated with emerged roots but no cotyledon. On the other
234 hand, in HP8 and HP13 lines 21.33% and 10.67% seedlings were partially germinated,
235 respectively, for the same period (Figure 8). HP13 line has the most impacted germination rate
236 (as measured 15 days after sowing), 58.92% compared to 94.66% in WT plants (Table 3). Not
237 only the germination rate was altered, but a decrease in seed vigor was indicated by both mean
238 germination time (MGT) and germination speed index (GSI) (Additional file 1: Figure S5;
239 Table 3). Although HP lines had a slight reduction in germination rate compared to the WT
240 (89.33% in HP13 versus 98.67% in WT; Additional file 1: Figure S5; Table 3), only the MGT
241 was significantly different in HP lines (Additional file 1: Figure S5; Table 3).

242

243 **Tnt1 genomic insertions**

244 We asked how interfering with Tnt1 expression could result in the phenotypes and
245 transcriptome profile changes observed in HP lines. The presence of Tnt1 insertions in the
246 vicinity of genes and / or within introns could lead to an indirect reduction in mRNA of those
247 genes in HP lines via the RNAi mechanism. To avoid this bias, we first identified genomic

248 copies of Tnt1 by searching and annotating its distinct domains (i.e. U3A, U3B, U3C, U5, the
249 whole LTR, as well as the coding domains GAG, PROT, INT, RT and RNH) .

250 A total of 276 U3A, 155 U3B and 35 U3C regions were found, each of them being part
251 of a complete Tnt1 insertion, an incomplete insertion or even a solo LTR. Not surprisingly,
252 many Tnt1 insertions are found in close proximity to scaffold borders, making it impossible to
253 determine the completeness of most of the insertions. Next we identified the first gene present
254 within a 5kb distance (in both upstream and downstream direction) of each U3 region. From the
255 216 genes found near Tnt1 insertions, only 11 were differentially expressed, being 8
256 upregulated and 3 downregulated. Interestingly, 127 of these 216 genes had no detectable
257 expression levels in our RNA-seq experiment, suggesting that Tnt1 insertions may play a role in
258 silencing mechanisms of neighbor genes.

259 Finally, given that there is not yet a complete assembly of the tobacco genome, it is
260 important to note that these results are possibly an underrepresentation of the total number of
261 genomic Tnt1 insertions. As described above, because of their repetitive nature, Tnt1 sequences
262 are prone to appear next to scaffold borders, and a number of unknown insertions are expected
263 to exist in the gaps between scaffolds.

264 **Tnt1 LTR as a putative source of sRNAs harboring GCC-motifs**

265 Because TEs are known to be a source of noncoding RNAs, we checked if Tnt1
266 transcripts could form secondary structures. Thus, we performed a folding prediction of part of
267 the Tnt1A mRNA, which includes three 3' U3 GCC-like motifs (Figure 9). The prediction
268 reveals stem-loop secondary structures, and GCC-like motifs would be located in the arms of
269 the putative hairpin loops (Figure 9), with the folded RNA free energy of dG = -24 kcal/mol.

270 Next, we used PsRNATarget (see "Methods" section for details) to search for potential
271 targets of three putative sRNAs derived from the stem-loop secondary structures predicted in
272 the previous step, each of them containing one GCC-like motif. To account for both
273 transcriptional and post-transcriptional possible targets of sRNAs, searches were performed in
274 two datasets: (1) a PsRNATarget built-in library of 25,398 *Nicotiana tabacum* SGN unigenes, in

which only the coding sequences are considered, and; (2) sequences of 3kb upstream of the 932 DEGs identified in this work, aiming to understand if the promoter region of such genes could be potential targets of the sRNAs. Because GCC motifs are regulatory sequences, one would expect them to be found in the promoter region of genes, particularly among the DEGs found in the present study.

As results, we found 224 putative targets within the coding sequences of the SGN unigenes library, from which 14 were present in our set of 932 DEGs. When targeting the upstream 3kb region, a total of 106 out of the 932 DEGs showed potential targets for the putative sRNAs, of which 67 were upregulated and 39 downregulated in HP lines. Based on these results, our hypothesis is that Tnt1 exerts transcriptional regulation of some genes via sRNAs.

DISCUSSION

The dynamics of Tnt1 expression in WT tobacco

Expression of the Tnt1 retroelement is known to be induced by both biotic and abiotic factors. Protoplast preparation by fungal extracts is the most studied way to induce Tnt1 expression, and was used to demonstrate mobilization of the element [32]. Likewise, expression of Tnt1 was reported to be induced upon viral infection and seems to be linked to plant defense responses, although its role in such cases is still unclear [33]. Interestingly, expression of Tnt1 subfamilies (Tnt1A, Tnt1B and Tnt1C) is induced by different stress-associated molecules, with Tnt1A being particularly upregulated by cryptogein and methyl jasmonate [29]. Activation of different subfamilies is accredited to the presence of specific promoter motifs in each of their U3 regions (U3A, U3B and U3C). Although Tnt1A responsiveness to ethylene has been previously hypothesized due the presence of repetitive GCC-like motifs in its U3A region [31], no experimental data was available to support this idea. Our results demonstrate that after exposure to the gaseous-hormone ethylene, expression of Tnt1 is upregulated (30 to 40-fold increase).

Expression of ethylene biosynthetic and responsive genes follow a basal circadian cycle, however that does not mean senescence, stress-responses, and other mechanisms dictated by ethylene also oscillates in a circadian fashion. Results from an experiment designed to address if Tnt1 followed either of the two ethylene expression patterns (circadian associated or not) is consistent with the non-circadian response and support basal and continuous Tnt1 expression. The hypothesis raised by the results presented is that Tnt1 basal expression is not controlled by ethylene, but rather induced by the hormone only under specific conditions, probably requiring additional signaling or an elevated ethylene concentration.

Likewise, expression of Tnt1A, as indicated by quantification of both the RT domain and the specific U3A region, oscillate in a non-circadian fashion and greatly differs between biological replicates growing under the same controlled conditions. These results suggest that, under normal conditions, Tnt1A expression is not dictated by ethylene but rather by unknown, possibly stochastic, factors. In this scenario, ethylene induction of Tnt1A expression seems to be subject of additional regulation and may require exposure to a minimum threshold of the hormone.

Perturbations in Tnt1 expression lead to phenotypic and transcriptomic changes.

Since Tnt1 has basal expression in different tissues [20], we sought to understand whether its suppression would culminate in detectable phenotypes. Thus, we generated transgenic RNAi lines (herein referred as HP lines) expressing a hairpin targeting the RT domain of Tnt1. With this approach, Tnt1 transcripts were consistently downregulated upon ethylene treatment.

Different HP lines displayed recurring phenotypes, namely lesion mimicry in leaves, underdevelopment of the root system, overproduction of root hairs and early loss of seed viability. Most of these developmental aspects are regulated by ethylene to some extent. For instance, root hair production is induced by ethylene in *Arabidopsis* [34–36]. Interestingly, mutation in the *Arabidopsis RHD6* gene, which mediates ethylene-response during root hair

328 formation, leads not only to a decrease in root hair production but also to a shift in their
329 initiation site towards the root base [35], which is the opposite effect observed in HP lines.

330 In addition, the germination process is mediated by counteracting effects of ethylene
331 and ABA in numerous species, including tobacco (see 37 for a review). It is possible that HP
332 seeds lose viability via early partial release of dormancy by an overproduction of ethylene.
333 Accordingly, we demonstrated that HP plants overproduce ethylene when compared to the WT.

334 Functional categorization of the 932 DEGs in HP lines reveals enrichment of stress-
335 related processes among upregulated genes, while metabolic and homeostatic processes in the
336 downregulated set of genes. This pattern is supported in a second and more strict analysis that
337 pointed to 97 modulated genes. Among these DEG, several genes related to ethylene
338 biosynthesis and plant defense response were upregulated in all HP lines but not in the WT as
339 follows: two ACC oxidase/ethylene forming enzymes (EFEs), involved in the last step of
340 ethylene synthesis (accession numbers AB012857 and Z29529)[38]; five pathogenesis-related
341 protein (PR) family members (EH620111, EH621793, X03913, M29868 and X51426)[39–41];
342 phospholipase D which participates in signal transduction cascades in stress responses [42]; and
343 an ethylene responsive gene induced during the pathogen-induced systemic acquired resistance,
344 Sar8.2b (EH621848) [43]. In addition, we identified the repression of several chlorophyll a/b
345 binding proteins in HP leaves. Senescence is marked by a decline of the photosynthetic
346 apparatus and mobilization of nutrients from senescing leaves to growing tissues, culminating in
347 cell death [44]. Likewise, hydroxy-methylglutaryl-coenzyme A reductase (HMGR)-like genes
348 were up-regulated in HP plants. This enzyme participates in the steroid derivatives synthesis
349 that follows pathogen infections [45]. This possibly corresponds to the spread of the initial
350 stress signal throughout the plant. Although at this point it is not possible to distinguish between
351 primary and secondary effects, we propose that the retrotransposon not only responds to some
352 biotic and abiotic stresses, but also fine-tunes their occurrence and progression. Likewise, our
353 network analysis strengthened the emergence of a new pattern of expression and gene regulation

in the HP plants, and uncovered similar biological processes (e.g. defense and stress response) as the main changes in HP lines.

Tnt1 acting as modulator of plant development and response to stress

Curiously, the expression kinetics differ between samples used for the RNA-seq experiment and those assayed by RT-qPCR, i.e. while ethylene-related genes were upregulated in the former, most of these genes were downregulated in HP lines used for qRT-PCR. Because samples used for qPCR were sealed in a container prior to harvesting, they were exposed to an increased concentration of the hormone. Accordingly, these samples showed downregulation of Tnt1 expression (GAG, INT and RT domains) in comparison to WT plants, much like the results obtained after application of exogenous ethylene (compare Figures 1A and 3). Thus, differently from WT, HP plants growing under normal conditions overproduce ethylene and behave as if constantly responding to stress, while showing a decreased response to stress when exposed to higher concentrations of ethylene. These results are consistent with Tnt1 modulation of two distinct processes: (1) maintaining homeostasis during normal development and (2) fine-tuning stress responses possibly mediated by ethylene.

Finally, we combine our results to propose a model of the interaction between expression of Tnt1, ethylene biosynthesis genes and ethylene responsive genes that contain GCC motifs in their promoter (Figure 10). In this model, we consider two different conditions: a period of time when a WT plant exists under ideal conditions for normal growth and another under a stress pressure. Under normal growth conditions, WT tobacco plants express Tnt1 in a basal level, which fluctuates within a range of up to 8-fold, as observed in our circadian experiment. Genes involved in ethylene biosynthesis and other ethylene responsive genes that have GCC motifs in their promoters are also expressed in a circadian fashion (since ethylene is required in various moments of plant development and not only during plant defense responses). Upon a stress stimulus, ethylene biosynthesis is increased and this is one of the events that define the commencement of the defense responses. The overproduction of ethylene triggers the

upregulation of genes that contain GCC motifs in their promoter (such as Tnt1 and other responsive genes). According to our analysis, it is possible that Tnt1A mRNA can be a source of sRNAs that target GCC motifs in the promoter region of genes, or alternatively, in the 3´UTR of mRNAs. Thus, as Tnt1 is upregulated by ethylene, there is a turning point in which the Tnt1A sRNA production and consequent transcriptional inhibition of Tnt1 and target genes overcomes the ethylene induction, thus lowering the amount of mRNA of Tnt1 and of ethylene responsive genes. After ethylene responsive genes reach their maximum of expression, right before Tnt1A mRNA-derived sRNAs start to inhibit their transcription, it is likely that the defense responses have taken place and were sufficient to overcome the initial stress, thus removing this stimulus and lowering the expression of ethylene biosynthetic genes through ethylene auto-inhibition. The decrease in ethylene production removes the induction signal for Tnt1 transcription, lowering Tnt1 transcription and thus lowering Tnt1A sRNA production also. This way the system is pushed back to its "normality" after the stress has been overcome.

CONCLUSIONS

Although transcriptionally active TEs are commonly taken as a potential threat to their host organisms, there are recent reports of other TEs playing important roles for plant development. Such is the case of the MIKKI retrotransposon modulating root development in rice [19]. Another interesting case is found in *Solanum* species, in which the MESSI retrotransposon family is expressed specifically in the shoot apical meristem, suggesting that these TEs can respond to developmental signals [46]. Our findings suggest that active retrotransposons, not only domesticated transposable elements, can play a significant functional role in their host organisms. We hypothesize that Tnt1 can exert transcriptional control over itself as well as other endogenous genes. In our model we propose a potential new biological role for Tnt1. Upon stress induction, Tnt1 would provide feedback control to ethylene-mediated gene regulation in tobacco defense responses, bringing the system back to a homeostatic condition after the initial stress stimulus has been overcome. Further studies on the progression

of Tnt1 expression during stress response and recovery, as well as small RNA-seq experiments, can validate and bring new insights to the model presented herein.

METHODS

Plant samples and genetic transformation

Plants of *Nicotiana tabacum* cv Xanthi XHFD8 were used for genetic transformation to produce transgenic RNAi lines, herein referred as "HP lines". We design a hairpin construct aiming to target a conserved region of the Tnt1 RT domain, shared by the Tnt1A, Tnt1B and Tnt1C subfamilies. To achieve the hairpin construction, a 273bp fragment of the Tnt1 (accession number X13777.1) reverse transcriptase was amplified from *N. tabacum* genomic DNA (forward primer 5' CGGGATCCATCTCAGCAGAAGTACAT 3', reverse primer 5' CCATCGATACTTCCCAATGTTCC 3'). This fragment was cloned in the expression vector pHANNIBAL (accession number AJ311872.1) in both sense and antisense directions, separated by an intron, in order to express a Tnt1 hairpin. Expression cassettes were excised from pHANNIBAL and transferred to the binary vector pCAMBIA1201, generating pCAMBIA-Tnt1-RT. Control plants were transformed only with the pCAMBIA1201 backbone containing the hygromycin resistance gene. *Nicotiana tabacum* foliar discs were transformed with pCAMBIA-Tnt1-RT through *Agrobacterium tumefaciens* (LBA4404) co-culture, according to a previously established method [47]. Seventeen transgenic HP plants were generated with this cassette. Regenerated *in vitro* transgenic plants were cultivated in MS20 media with the proper antibiotic, under a 14-hour photoperiod at 24^{o}C.

Phenotyping of transgenic lines

For all phenotyping experiments, seeds of WT and HP lines were first primed (24 hours in sterile distilled water at 10ºC) and then germinated and grown *in vitro* in MS20 medium [48] under a 14-hour photoperiod at 24°C. Comparison of leaf area and leaf length was done using pictures taken from 15-day-old seedlings and then used for measurements with Fiji distribution of ImageJ [49]. For root analyzes, plants were grown in 45° inclined Petri dishes. Pictures were

436 taken from plants 15 days after sowing. These pictures were used for measurements with the
437 ImageJ SmartRoot plugin [50]. Germination assay was performed through daily observation of
438 root and cotyledon emergence. Germination Speed Index (GSI) and Mean Germination Time
439 (MGT) were calculated following [51]. All experiments were done using three replicates of 25
440 plants per line.
441
442 **RNA-seq and reads processing**
443 Total RNA from leaves of 45-day-old plants were frozen in liquid nitrogen. Four
444 biological replicates from three independent HP lines, one Control and one WT were harvested.
445 TRI Reagent® (Sigma-Aldrich) was used for RNA isolation according to the manufacturer's
446 instructions. Samples were treated with DNaseI (Ambion) and ribosomal RNA was depleted
447 using Ribominus Plant Kit (Invitrogen) following the manufacturer´s instructions. The cDNA
448 libraries were made using SOLID Total RNA-seq Kit (Ambion), according to Whole
449 Transcriptome Library Preparation for SOLID Sequencing Protocol (Life Technologies). The
450 handling of the beads for sequencing was done strictly according to the SOLID 3 System
451 Templated Bead Preparation Guide (Life Technologies). The run of the samples followed the
452 SOLID 3 System Instrument Operation Guide (Life Technologies). The RNA-seq produced a
453 total of 602,744,341 single-end reads (GenBank GEO accession GSE44027), which were
454 trimmed and quality filtered using Trimmomatic [52] with default settings.
455
456 **Bioinformatics and gene expression analysis of RNA-seq data**
457 The RNA-seq reads were first mapped against the tobacco genomic assemble [53] using
458 HISAT2 [54] set to default parameters. Differentially expressed genes (DEGs) were defined
459 using edgeR [55]. Lowly expressed genes were filtered out based on a minimum of 10 counts
460 per million in at least eight sequenced samples (including replicates). Sample normalization was
461 performed using the trimmed mean of M-values (TMM) method. Threshold for DEGs was set
462 using a false discovery rate (FDR) of < 0.001, yielding 932 genes. In a more stringent approach,
463 reads were also mapped against 24,069 unigenes from the tobacco database in Genbank. The

464 normalization of transcriptome data was done based on RPKM expression measure [56], square
465 root, and Bonferroni's correction. Using statistical t-test with p-value < 0.001 we generated two
466 subsets of modulated genes, the first comparing the three HP in contrast to WT lines, and only
467 the modulated in the three HP were considered. The second subset comprises the genes
468 modulated comparing HPs with Control, in order to filter out genes possibly modulated due to
469 the transgenesis process. This method identified 97 DEGs. Gene ontology (GO) analysis was
470 performed using Blast2GO [57]. Enrichment tests among upregulated and downregulated genes
471 were made comparing the set of expressed genes in all samples with those expressed in HP lines
472 with p-value ≤ 0.01.
473
474 **Gene regulatory networks**
475 Connections between genes were inferred by adopting the mean conditional entropy
476 (MCE) from the observed gene expression data [58]. The MCE is an information measure that
477 for each target gene indicates the contribution of its predictors to correctly detect the target
478 behavior in a multivariate way.
479 **Annotation of Tnt1 genomic insertions**
480 Sequences for each domain of Tnt1A (X13777), as well as known variant sequences of
481 U3A (AJ227998, AJ228000, AJ228002 – AJ228006, AJ228008, AJ228010 – AJ228012,
482 AJ228014 – AJ228017), U3B (AJ227999, AJ228007, AJ228009, AJ228013) and U3C
483 (AJ228001) were imported into Geneious Prime 2020.0.5 software (https://www.geneious.com).
484 Next the "Live Predict & Annotate" tool (which performs a blast-like search) were used with a
485 threshold of 60% to find and annotate each Tnt1 domain in the tobacco genome [53]. The first
486 gene present within a 5kb distance (both upstream and downstream) of each insertion was
487 retrieved, as well as information about its orientation in relation to the insertion (sense or
488 antisense).
489 **Small RNAs target prediction**
490 Prediction of sRNAs putative targets were performed using the PsRNATarget tool [59]
491 with default parameters. For target sites within coding sequences, the build-in *Nicotiana*

492 *tabacum* SGN unigene cDNA library was used. Promoter region of the differentially expressed
493 genes were extracted from the tobacco genome latest release [53], considering 3kb upstream of
494 each gene.
495
496 **Ethylene treatment and ethylene gas chromatography**
497 Tnt1 induction by ethylene was performed by placing plants in sealed containers and
498 then ethylene was taken from a concentrated stock (Alltech, Deerfield, IL) and injected into the
499 containers using a syringe to a final concentration of 10 μL/mL. This concentration was
500 monitored by gas chromatography every 6 h and remained stable throughout the treatment.
501 Control plants were incubated in sealed containers without ethylene injection. The containers
502 were opened after 24 hours, leaf samples (300 mg) were collected and processed for total RNA
503 isolation. We also measured ethylene emission in 90-day-old HP and WT plants by gas
504 chromatography.
505
506 **Gene expression analyzes through quantitative Real Time PCR (qRT-PCR)**
507 For the circadian experiment, plants were grown in a 12 hours light / 12 hours dark
508 regime in MS20 medium at 24°C. Samples were harvested each six hours for 48 hours as
509 follow: 12 PM – midpoint of the light period; 6 PM – start of the dark period; 12 AM –
510 midpoint of the dark period; 6 AM – start of the light period. Each sample comprised a pool of
511 four whole seedlings. All the other qRT-PCR experiments were done using plants grown under
512 a 14-hour photoperiod at 24ºC.
513 RNA was extracted from samples using a modified LiCl method [60] and treated with
514 DNaseI (Ambion). cDNA was synthesized using SuperScript First-Strand Synthesis System for
515 RT-PCR (Life Technologies) following the manufacturer's instructions. The real time RT-PCR
516 reactions were performed using the SYBR Green Real-Time PCR Master Mix according to
517 manufacturer's instructions. All experiments were based on three biological replicates with
518 three technical replicates each.

519

520 **LIST OF ABBREVIATIONS**

521 cDNA: complementary DNA; CPM: Counts per million; DEGs: Differential expression genes;
522 EFEs: Ethylene forming enzymes; FDR: False discovery rates; Gag: Group antigens; GO: Gene
523 ontology; GSI: germination seed index; HP: Hair-pin lines; INT: Integrase; LTR: Long terminal
524 repeats; MCE: Mean conditional entropy; MGT: Mean germination time; miRNA: micro RNA;
525 mRNA: messenger RNA; ORFs: Open reading frames; PR: pathogenesis-related; qPCR:
526 Quantitative polymerase chain reaction; RNAi: RNA interference; RPKM: Reads per kilo base
527 per million mapped reads; RT: Reverse transcriptase; SGN: Sol genomics network; sRNA:
528 small RNA; TE: Transposable Elements; TMM: Trimmed means of M values; WT: Wild type.

529

530 **DECLARATIONS**

531 **Ethics approval and consent of participate**

532 Not applicable

533 **Consent for publication**

534 Not applicable

535 **Availability of data and materials**

536 Raw sequences generated in this study are deposited in GenBank
537 (https://www.ncbi.nlm.nih.gov/genbank/) as GEO accession number GSE44027.

538 **Competing interests**

539 The authors declare that they have no competing interests.

540 **Funding**

541 This study was supported by the following grants: FAPESP 08/55646-4
542 and CNPq 4813322/2009-4, 308197/2010-0 (M.A.V.S.), PNPD/CAPES 1633/04-0 (J.H.L.),
543 FAPESP 2009/50630-5 and PNPD/CAPES 0280/09-0 (E.M.J.), DS CAPES
544 (D.M.Q.), FAPESP 2004/04088-0 (B.K.).

545 **Author's contributions**

Conceptualization MAVS, DQ and JHL; Data curation DQ, EMJ and JHL; Formal analysis DQ, EMJ, JHL, FL, JPK, MAVS; Funding acquisition MAVS; Methodology AAC, RBP, FL, DQ, EMJ, JHL, HMD, BK; Supervision MAVS; Roles/Writing - original draft DQ, JHL, EMJ; Writing - review & editing AAC, TBJ, JHL, MAVS.

Acknowledgements

We thank Dr. Tatiana Corrêa for assistance with experiments, Dr**.** Myna Nakabashi for technical support with plant transformation, Dr. Anna Christina de Mattos Salim for technical assistance with transcriptome sequencing and Dr. Luciano Freschi, Dr. Hana Masuda, Dr. Nathália de Setta, Dr. Walter Colli, Dr. Françoise Simon-Plaz and Dr. Marie-Angele Grandbastien for discussions and comments on the manuscript.

REFERENCES

1. Fedoroff N V. Transposable genetic elements in maize. Scientific American. 1984;250:84–98.

2. Orgel LE, Crick FHC. Selfish DNA: The ultimate parasite. Nature. 1980;284:604–7.

3. Pritham EJ. Transposable Elements and Factors Influencing their Success in Eukaryotes. Journal of Heredity [Internet]. 2009;100:648–55. Available from: https://academic.oup.com/jhered/article-lookup/doi/10.1093/jhered/esp065

4. Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, et al. Initial sequencing and analysis of the human genome. Nature [Internet]. 2001;409:860–921. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11237011?ordinalpos=4&itool=EntrezSystem2.PEntrez.Pubmed.Pubmed_ResultsPanel.Pubmed_RVDocSum

5. de Koning APJ, Gu W, Castoe TA, Batzer MA, Pollock DD. Repetitive elements may comprise over Two-Thirds of the human genome. PLoS Genetics. 2011;7.

6. Negi P, Rai AN, Suprasanna P. Moving through the Stressed Genome: Emerging Regulatory Roles for Transposons in Plant Stress Response. Frontiers in Plant Science [Internet]. 2016;7:1448. Available from: http://journal.frontiersin.org/article/10.3389/fpls.2016.01448/full

574 7. Hirsch CD, Springer NM. Transposable element influences on gene expression in plants.
575 Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms [Internet]. Elsevier B.V.;
576 2017;1860:157–65. Available from: http://dx.doi.org/10.1016/j.bbagrm.2016.05.010
577 8. Feschotte C. The contribution of transposable elements ot the evolution of regulatory
578 networks. Nature Reviews Genetics. 2008;9:397–405.
579 9. Souza FSJ de, Franchini LF, Rubinstein M. Exaptation of transposable elements into novel
580 Cis-regulatory elements: Is the evidence always strong? Molecular Biology and Evolution.
581 2013;30:1239–51.
582 10. Stuart T, Eichten SR, Cahn J, Karpievitch Y V., Borevitz JO, Lister R. Population scale
583 mapping of transposable element diversity reveals links to gene regulation and epigenomic
584 variation. eLife. 2016;5:1–27.
585 11. Slotkin RK, Martienssen R. Transposable elements and the epigenetic regulation of the
586 genome. Nature reviews Genetics [Internet]. 2007;8:272–85. Available from:
587 http://www.ncbi.nlm.nih.gov/pubmed/17363976
588 12. Anwar F, Davenport MP, Ebrahimi D. Footprint of APOBEC3 on the genome of human
589 retroelements. Journal of virology [Internet]. 2013;87:8195–204. Available from:
590 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3700199&tool=pmcentrez&rendert
591 ype=abstract
592 13. Knisbacher BA, Levanon EY. DNA editing of LTR retrotransposons reveals the impact of
593 APOBECs on vertebrate genomes. Molecular Biology and Evolution. 2016;33:554–67.
594 14. Chuong EB, Elde NC, Feschotte C. Regulatory activities of transposable elements: From
595 conflicts to benefits. Nature Reviews Genetics [Internet]. Nature Publishing Group;
596 2017;18:71–86. Available from: http://dx.doi.org/10.1038/nrg.2016.139
597 15. Fedoroff N V. Transposable Elements, Epigenetics, and Genome Evolution. Science
598 [Internet]. 2012;338:758–67. Available from:
599 http://www.sciencemag.org/cgi/doi/10.1126/science.338.6108.758
600 16. Häsler J, Strub K. Alu RNP and Alu RNA regulate translation initiation in vitro. Nucleic
601 Acids Research. 2006;34:2374–85.

602 17. Li Y, Li C, Xia J, Jin Y. Domestication of transposable elements into MicroRNA genes in
603 plants. Schönbach C, editor. PloS one [Internet]. 2011;6:e19212. Available from:
604 https://dx.plos.org/10.1371/journal.pone.0019212
605 18. Wang D, Qu Z, Yang L, Zhang Q, Liu ZH, Do T, et al. Transposable elements (TEs)
606 contribute to stress-related long intergenic noncoding RNAs in plants. Plant Journal.
607 2017;90:133–46.
608 19. Cho J, Paszkowski J. Regulation of rice root development by a retrotransposon acting as a
609 microRNA sponge. eLife. 2017;6:1–21.
610 20. Pouteau S, Huttner E, Grandbastien M a., M.Caboche M. Specific expression of the tobacco
611 Tnt1 in protoplasts. EMBO Journal. 1991;10:1911–8.
612 21. Casacuberta JM, Vernhettes S, Grandbastien M a. Sequence variability within the tobacco
613 retrotransposon Tnt1 population. The EMBO journal [Internet]. 1995;14:2670–8. Available
614 from:
615 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=398381&tool=pmcentrez&renderty
616 pe=abstract
617 22. Grandbastien M-A, Spielmann A, Caboche M. Tnt1, a mobile retroviral-like transposable
618 element of tobacco isolated by plant cell genetics. Nature [Internet]. 1989;337:376–80.
619 Available from: http://www.nature.com/articles/337376a0
620 23. Melayah D, Lim KY, Bonnivard E, Chalhoub B, Dorlhac De Borne F, Mhiri C, et al.
621 Distribution of the Tnt1 retrotransposon family in the amphidiploid tobacco (Nicotiana
622 tabacum) and its wild Nicotiana relatives. Biological Journal of the Linnean Society.
623 2004;82:639–49.
624 24. Araujo PG, Casacuberta JM, Costa APP, Hashimoto RY, Grandbastien MA, Van Sluys MA.
625 Retrolyc1 subfamilies defined by different U3 LTR regulatory regions in the Lycopersicon
626 genus. Molecular Genetics and Genomics. 2001;266:35–41.
627 25. Manetti ME, Rossi M, Costa APP, Clausen AM, Van Sluys M-A. Radiation of the Tnt1
628 retrotransposon superfamily in three Solanaceae genera. BMC evolutionary biology. 2007;7:34.
629 26. Tapia G, Verdugo I, Yañez M, Ahumada I, Theoduloz C, Cordero C, et al. Involvement of

630 Ethylene in Stress-Induced Expression of the TLC1.1 Retrotransposon from Lycopersicon
631 chilense Dun. Plant Physiology [Internet]. 2005;138:2075–86. Available from:
632 http://www.plantphysiol.org/lookup/doi/10.1104/pp.105.059766
633 27. Casacuberta JM, Grandbastien M angèle. Characterisation of LTR sequences involved in the
634 protoplast specific expression of the tobacco Tnt1 retrotransposon. Nucleic Acids Research.
635 1993;21:2087–93.
636 28. Vernhettes S, Grandbastien M a, Casacuberta JM. The evolutionary analysis of the Tnt1
637 retrotransposon in Nicotiana species reveals the high variability of its regulatory sequences.
638 Molecular biology and evolution. 1998;15:827–36.
639 29. Beguiristain T, Grandbastien M-A, Puigdomènech P, Casacuberta JM. Three Tnt1
640 Subfamilies Show Different Stress-Associated Patterns of Expression in Tobacco.
641 Consequences for Retrotransposon Control and Evolution in Plants. Plant Physiology [Internet].
642 2001;127:212–21. Available from: http://www.plantphysiol.org/cgi/doi/10.1104/pp.127.1.212
643 30. Shinshi H, Usami S, Ohme-Takagi M. Identification of an ethylene-responsive region in the
644 promoter of a tobacco class I chitinase gene. Plant Molecular Biology. 1995;27:923–32.
645 31. Grandbastien M-A, Audeon C, Bonnivard E, Casacuberta JM, Chalhoub B, Costa A-PP, et
646 al. Stress activation and genomic impact of Tnt1 retrotransposons in Solanaceae. Cytogenetic
647 and Genome Research [Internet]. 2005;110:229–41. Available from:
648 https://www.karger.com/Article/FullText/84957
649 32. Melayah D, Bonnivard E, Chalhoub B, Audeon C, Grandbastien MA. The mobility of the
650 tobacco Tnt1 retrotransposon correlates with its transcriptional activation by fungal factors.
651 Plant Journal. 2001;28:159–68.
652 33. Grandbastien M a, Lucas H, Morel JB, Mhiri C, Vernhettes S, Casacuberta JM. The
653 expression of the tobacco Tnt1 retrotransposon is linked to plant defense responses. Genetica
654 [Internet]. 1997;100:241–52. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9440277
655 34. Song L, Yu H, Dong J, Che X, Jiao Y, Liu D. The Molecular Mechanism of Ethylene-
656 Mediated Root Hair Development Induced by Phosphate Starvation. PLoS Genetics [Internet].
657 2016;12:1–29. Available from: http://dx.doi.org/10.1371/journal.pgen.1006194

658 35. Masucci JD, Schiefelbein JW. The rhd6 Mutation of Arabidopsis thaliana Alters Root-Hair
659 Initiation through an Auxin- and Ethylene-Associated Process. Plant Physiology [Internet].
660 1994;106:1335–46. Available from:
661 http://www.plantphysiol.org/lookup/doi/10.1104/pp.106.4.1335
662 36. Feng Y, Xu P, Li B, Li P, Wen X, An F, et al. Ethylene promotes root hair growth through
663 coordinated EIN3/EIL1 and RHD6/RSL1 activity in Arabidopsis . Proceedings of the National
664 Academy of Sciences. 2017;114:13834–9.
665 37. Corbineau F, Xia Q, Bailly C, El-Maarouf-Bouteau H. Ethylene, a key factor in the
666 regulation of seed dormancy. Frontiers in Plant Science. 2014;5:1–13.
667 38. Hamilton AJ, Bouzayen M, Grierson D. Identification of a tomato gene for the ethylene-
668 forming enzyme by expression in yeast. Proceedings of the National Academy of Sciences
669 [Internet]. 1991;88:7434–7. Available from:
670 http://www.pnas.org/cgi/doi/10.1073/pnas.88.16.7434
671 39. Cornelissen BJC, Horowitz J, van Kan JA l., Goldberg RB, Bol JF. Structure of tobacco
672 genes encoding pathogenesis-related proteins from the PR-1 group. Nucleic Acids Research
673 [Internet]. 1987;15:6799–811. Available from: https://academic.oup.com/nar/article-
674 lookup/doi/10.1093/nar/15.17.6799
675 40. Linthorst HJM. Analysis of Acidic and Basic Chitinases from Tobacco and Petunia and
676 Their Constitutive Expression in Transgenic Tobacco. Molecular Plant-Microbe Interactions
677 [Internet]. 1990;3:252. Available from:
678 http://www.apsnet.org/publications/mpmi/backissues/Documents/1990Abstracts/Microbe03-
679 252.htm
680 41. van Loon LC, Rep M, Pieterse CMJ. Significance of Inducible Defense-related Proteins in
681 Infected Plants. Annual Review of Phytopathology [Internet]. 2006;44:135–62. Available from:
682 http://www.annualreviews.org/doi/10.1146/annurev.phyto.44.070505.143425
683 42. Bargmann BOR, Munnik T. The role of phospholipase D in plant stress responses. Current
684 Opinion in Plant Biology [Internet]. 2006;9:515–22. Available from:
685 https://linkinghub.elsevier.com/retrieve/pii/S1369526606001178

686 43. Guo A, Salih G, Klessig DF. Activation of a diverse set of genes during the tobacco
687 resistance response to TMV is independent of salicylic acid; induction of a subset is also
688 ethylene independent. The Plant Journal [Internet]. 2000;21:409–18. Available from:
689 http://doi.wiley.com/10.1046/j.1365-313x.2000.00692.x
690 44. Quirino BF, Noh YS, Himelblau E, Amasino RM. Molecular aspects of leaf senescence.
691 Trends in Plant Science. 2000;5:278–82.
692 45. Choi D, Bostock RM, Avdiushko S, Hildebrand DF. Lipid-derived signals that discriminate
693 wound- and pathogen-responsive isoprenoid pathways in plants: methyl jasmonate and the
694 fungal elicitor arachidonic acid induce different 3-hydroxy-3-methylglutaryl-coenzyme A
695 reductase genes and antimicrobial isopre. Proceedings of the National Academy of Sciences
696 [Internet]. 1994;91:2329–33. Available from:
697 http://www.pnas.org/cgi/doi/10.1073/pnas.91.6.2329
698 46. Sanchez DH, Gaubert H, Yang W. Evidence of developmental escape from transcriptional
699 gene silencing in MESSI retrotransposons. New Phytologist. 2019;223:950–64.
700 47. Gallois P, Marinho P. Leaf Disk Transformation Using Agrobacterium tumefaciens-
701 Expression of Heterologous Genes in Tobacco. Plant Gene Transfer and Expression Protocols
702 [Internet]. New Jersey: Humana Press; 2003. p. 39–48. Available from:
703 http://link.springer.com/10.1385/0-89603-321-X:39
704 48. Murashige T, Skoog F. A Revised Medium for Rapid Growth and Bio Assays with Tobacco
705 Tissue Cultures. Physiologia Plantarum [Internet]. 1962;15:473–97. Available from:
706 http://doi.wiley.com/10.1111/j.1399-3054.1962.tb08052.x
707 49. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: An
708 open-source platform for biological-image analysis. Nature Methods. 2012;9:676–82.
709 50. Lobet G, Pagès L, Draye X. A Novel Image-Analysis Toolbox Enabling Quantitative
710 Analysis of Root System Architecture. Plant Physiology. 2011;157:29–39.
711 51. Maguire JD. Speed of Germination—Aid In Selection And Evaluation for Seedling
712 Emergence And Vigor1. Crop Science [Internet]. 1962;2:176. Available from:
713 https://www.crops.org/publications/cs/abstracts/2/2/CS0020020176

714 52. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence
715 data. Bioinformatics. 2014;30:2114–20.
716 53. Edwards KD, Fernandez-Pozo N, Drake-Stowe K, Humphry M, Evans AD, Bombarely A,
717 et al. A reference genome for Nicotiana tabacum enables map-based cloning of homeologous
718 loci implicated in nitrogen utilization efficiency. BMC Genomics. BMC Genomics; 2017;18:1–
719 14.
720 54. Kim D, Langmead B, Salzberg SL. HISAT: A fast spliced aligner with low memory
721 requirements. Nature Methods. 2015;12:357–60.
722 55. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential
723 expression analysis of digital gene expression data. Bioinformatics. 2009;26:139–40.
724 56. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying
725 mammalian transcriptomes by RNA-Seq. Nature Methods. 2008;5:621–8.
726 57. Conesa A, Götz S. Blast2GO: A comprehensive suite for functional analysis in plant
727 genomics. International Journal of Plant Genomics. 2008;2008.
728 58. Lopes FM, Martins DC, Cesar RM. Feature selection environment for genomic applications.
729 BMC Bioinformatics. 2008;9:1–8.
730 59. Dai X, Zhuang Z, Zhao PX. PsRNATarget: A plant small RNA target analysis server (2017
731 release). Nucleic Acids Research. Oxford University Press; 2018;46:W49–54.
732 60. Green MR, Sambrook J. Molecular Cloning: A Laboratory Manual. 4th ed. Cold Spring
733 Harbor Laboratory Press. New York: Cold Spring Harbor; 2012.
734
735
736
737
738
739
740 **TABLES**

741 **Table 1**. Leaf size comparison between 15 days old seedlings of WT and HP lines (T4).

Line	Total leaf area (mm^2)	Total leaf length (mm)
WT	135.26 ± 31.86 [a]	17.20 ± 3.14 [a]
HP8	147.05 ± 44.41 [a]	17.27 ± 3.51 [a]
HP13	175.86 ± 34.71 [b]	19.75 ± 2.68 [b]

742 Letters indicate statistically significant differences (p-value $\leq$ 0.05). Total leaf length was
743 measured as the maximum distance across the longer longitudinal axis of the two leaves.
744

745 **Table 2.** Comparison between the root system of 15 days old WT and HP seedlings.

Sample	Line	Length (cm)	Surface (cm^2)	Volume (cm^3)	Number of lateral roots
Primary root	WT	5.14 ± 0.87 [b]	0.68 ± 0.24 [b]	0.0079 ± 0.0044 [b]	9.30 ± 2.67 [b]
	HP8	3.83 ± 0.58 [a]	0.40 ± 0.07 [a]	0.0036 ± 0.0009 [a]	6.00 ± 2.54 [a]
	HP13	4.22 ± 0.88 [a]	0.46 ± 0.12 [a]	0.0042 ± 0.0015 [a]	7.77 ± 2.86 [ab]
Total root system	WT	11.36 ± 3.65 [b]	1.34 ± 0.52 [b]	0.0140 ± 0.0078 [b]	*NA*
	HP8	7.49 ± 2.31 [a]	0.73 ± 0.24 [a]	0.0060 ± 0.0025 [a]	*NA*
	HP13	8.72 ± 2.89 [ab]	0.92 ± 0.40 [a]	0.0082 ± 0.0047 [a]	*NA*

746 *NA* – Not applicable. Letters indicate statistically significant differences (p-value $\leq$ 0.05).

747

748 **Table 3.** Germination comparison between fresh and six years old seeds from WT and HP lines.

Seed age	Line	Mean germination time (days)	Germination speed index	Germination rate (%)
Six years old	WT	5.68 ± 0.45[c]	4.26 ± 0.64[c]	94.67 ± 9.24[b]
	HP1	6.79 ± 0.19[bc]	3.46 ± 0.30[bc]	89.33 ± 4.62[b]
	HP5	9.26 ± 1.33[a]	2.52 ± 0.21[a]	89.28 ± 6.02[b]
	HP8	7.35 ± 0.44[abc]	3.07 ± 0.55[abc]	85.49 ± 12.27[b]
	HP13	8.53 ± 0.04[ab]	2.00 ± 0.28[ab]	58.93 ± 7.18[a]
Fresh	WT	6.17 ± 0.07[b]	3.81 ± 0.31[b]	98.67 ± 2.31[a]
	HP8	7.53 ± 0.25[a]	3.35 ± 0.15[a]	96.00 ± 4.00[a]
	HP13	7.33 ± 0.38[a]	3.18 ± 0.35[a]	89.33 ± 6.11[a]

749 Letters indicate statistically significant differences (p-value $\leq$ 0.05).

750

751 **ADDITIONAL FILES**

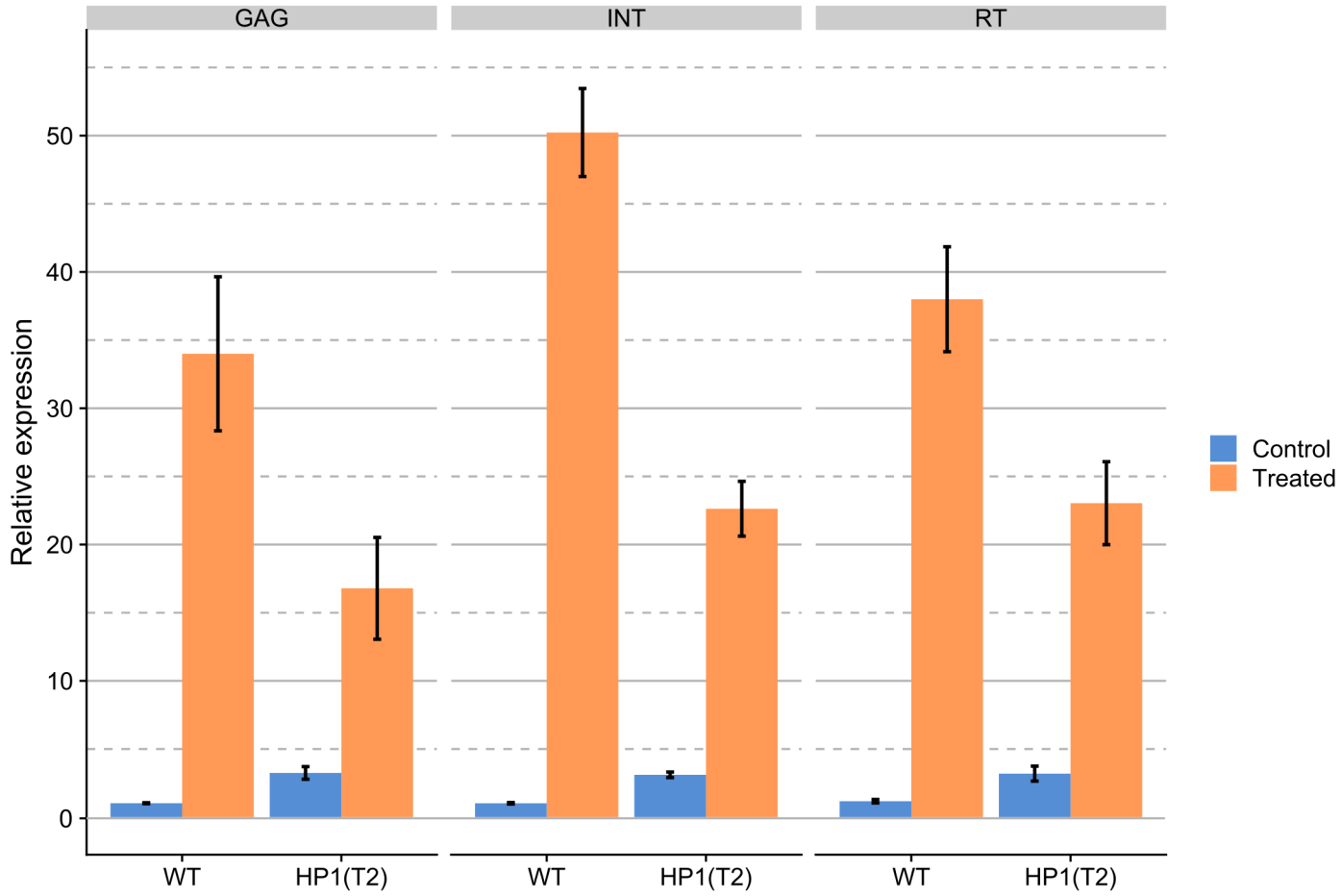
752 **Additional file 1: Figure S1**. Tnt1 retrotransposon structure and RNAi construct used to
753 interfere with Tnt1 levels of transcripts. **Figure S2.** Expression of circadian clock genes and GO
754 categorization of modulated genes in HP lines. **Figure S3.** Diagram of a gene regulatory
755 network observed exclusively in HP plants. **Figure S4.** Phenotypes found in roots of HP lines.
756 **Figure S5.** Germination performance of fresh and 6-years-old seeds.
757
758
759 **FIGURE LEGENDS**
760
761 **Figure 1.** Dynamics of Tnt1 and ethylene emission in WT and HP lines. **(A)** Relative
762 expression of three Tnt1 domains in 60-days-old plants measured by qRT-PCR. Tnt1 expression
763 is induced upon exogenous ethylene application (10 µL/mL for 24hs). Control plants were
764 incubated in sealed containers without input of ethylene. **(B)** Ethylene emission in 90-day-old
765 HP plants measured by gas chromatography. Bars indicate standard error. Significance of the
766 difference between HP and WT was assessed by a one-tailed Student's t-test: (**) significant at
767 the 0.004 level; (*) significant at the 0.017 level.
768
769 **Figure 2.** Relative expression of ethylene-related genes and Tnt1 in 15-days-old WT plants
770 throughout 48 hours period. Plants were grown under a 12 hours light / 12 hours dark regime.
771 Gray areas indicate dark periods. The lowest expression value for each gene was set to one. Bars
772 represent standard deviation between three biological replicates, except for Tnt1 U3A, for which
773 bars represent the standard deviation between three technical replicates.
774
775 **Figure 3.** Relative expression of ethylene-related genes and Tnt1 in 90-day-old WT and HP
776 plants. The expression level of WT was set as one for each gene. Values are linear average, with
777 bars showing standard errors. Leaves used to extract the RNA for the quantitative PCRs were
778 from the same plants used for the ethylene gas chromatography (Figure 1A).
779

780 **Figure 4.** Hierarchical clustering analysis of 932 differentially expressed genes across different
781 tobacco lines. Each column represents a biological replicate. Control line was transformed with
782 a vector containing only the resistance gene.
783
784 **Figure 5.** Gene ontology (GO) enrichment analysis (p-value $\leq$ 0.01) of the 932 differentially
785 expressed genes between wild-type and hairpin lines. Only the 20 most representative biological
786 processes are shown. Red boxes highlight processes related to stress response and homeostasis
787 maintenance. The reference includes all expressed genes in all sequenced samples.
788
789 **Figure 6.** Phenotypes of aerial parts in transgenic RNAi lines. **(A)** Leaves of wild-type (WT)
790 and transgenic (HP) *Nicotiana tabacum* plants growing *in vitro*. Wild type presents normal
791 leaves while HP lineages present cell death spots. Left panel (WT, T1 and T2) displays leaves of
792 45-days-old plants used for the transcriptome sequencing. **(B)** Total leaf area and total leaf
793 length (measured from the first leaf pair) of 15-days-old seedlings. Letters represent statistically
794 significant differences between lineages.
795
796 **Figure 7.** Phenotypes of the root system in transgenic RNAi lines. **(A)** Primary root of wild-
797 type (WT) and transgenic (HP) lines. Root hair production is increased in HP lines. **(B)**
798 Comparison of morphological parameters in the primary roots of WT and HP 15-days-old
799 seedlings. Letters represent statistically significant differences between lineages.
800
801 **Figure 8.** Germination progress in wild-type (WT) and transgenic RNAi (HP) lines. Seeds were
802 observed from 6 to 15 days after sowing. Orange area represents the percentage of seeds
803 displaying root emergence, while green denotes seedlings in which both root and cotyledon
804 emergence had occurred.
805
806 **Figure 9.** The discovery of putative Tnt1-derived GCC-box small RNAs in the promoter region
807 of Tnt1A. Tnt1 retrotransposon genomic and transcript structures. The text box bellow the Tnt1

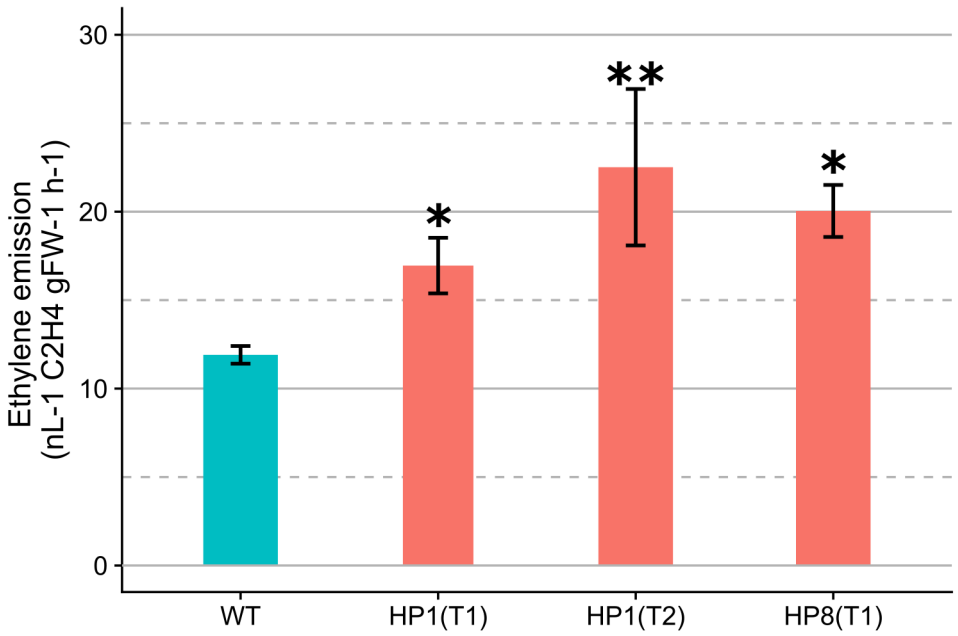
scheme represents the 3' U3 region sequence of Tnt1A. Green underlined words indicate the 3 GCC-like motifs and TATA box is in blue. Bottom figure depicts the folding prediction of a 98 bp part of the Tnt1A mRNA sequence containing the three GCC-like boxes (green rectangles).

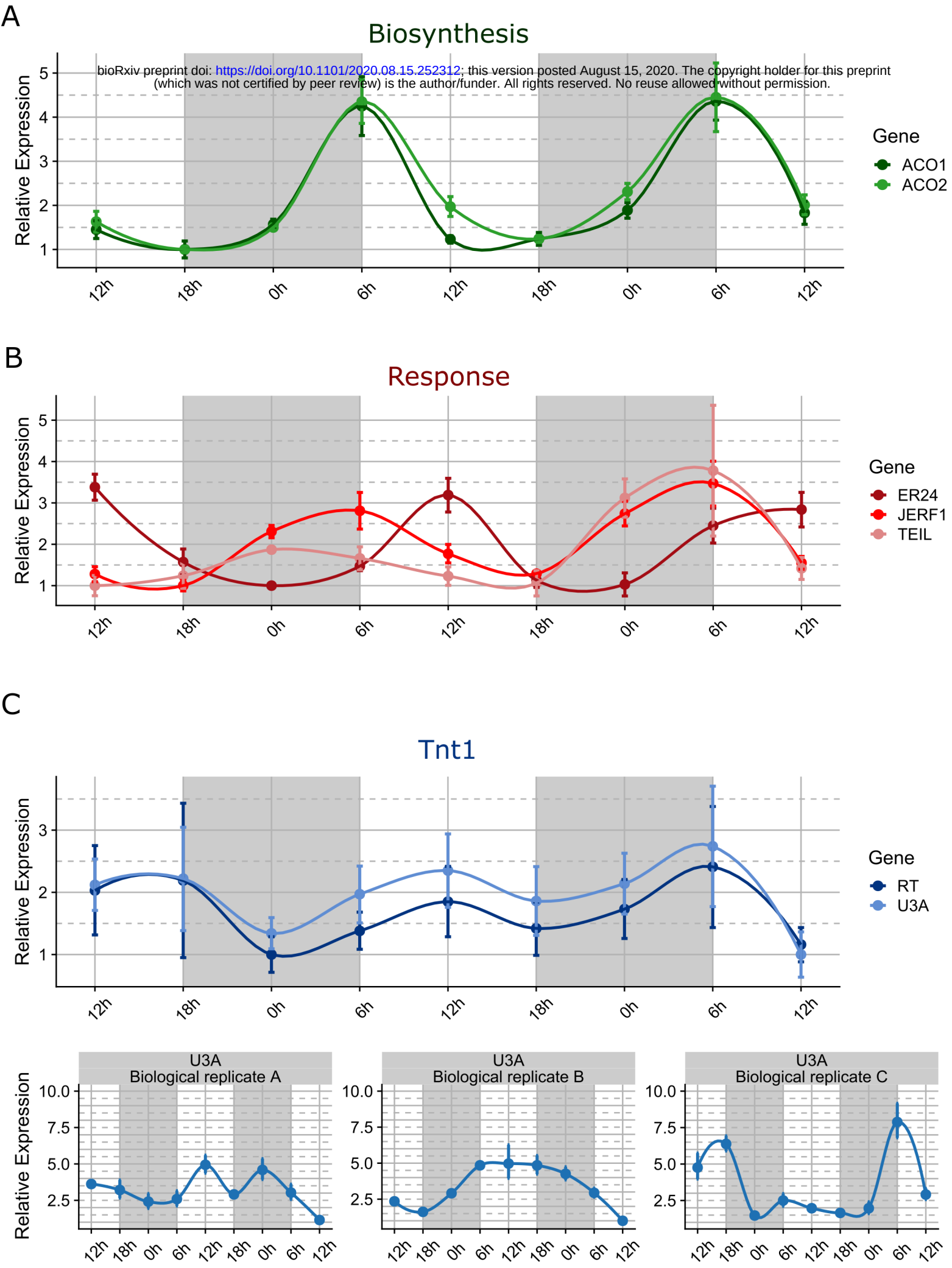
Figure 10. Model proposing a dynamic equilibrium between expression of defense response genes and Tnt1 in WT tobacco. Expression values are based on our observations, except for ethylene biosynthesis and responsive genes during defense/stress response, for which values are hypothetical. During normal development (T0 – T7), ethylene biosynthetic/responsive genes are expressed in a circadian fashion, while Tnt1 fluctuates within basal range with no clearly delimited periodicity (depicted as a blue cloud). Upon stress stimulus, ethylene biosynthesis is upregulated, which also induces expression of genes that contain GCC motifs in their promoter, including Tnt1A (T8). Tnt1-U3A GCC-like motifs sRNAs overcomes ethylene induction and promotes transcriptional inhibition of GCC motifs in other ethylene-responsive gene promoters (T8 – T9). Expression of Tnt1 and ethylene-responsive genes returns to normal levels due to the repression mediated by its own sRNA (T9 – T12).

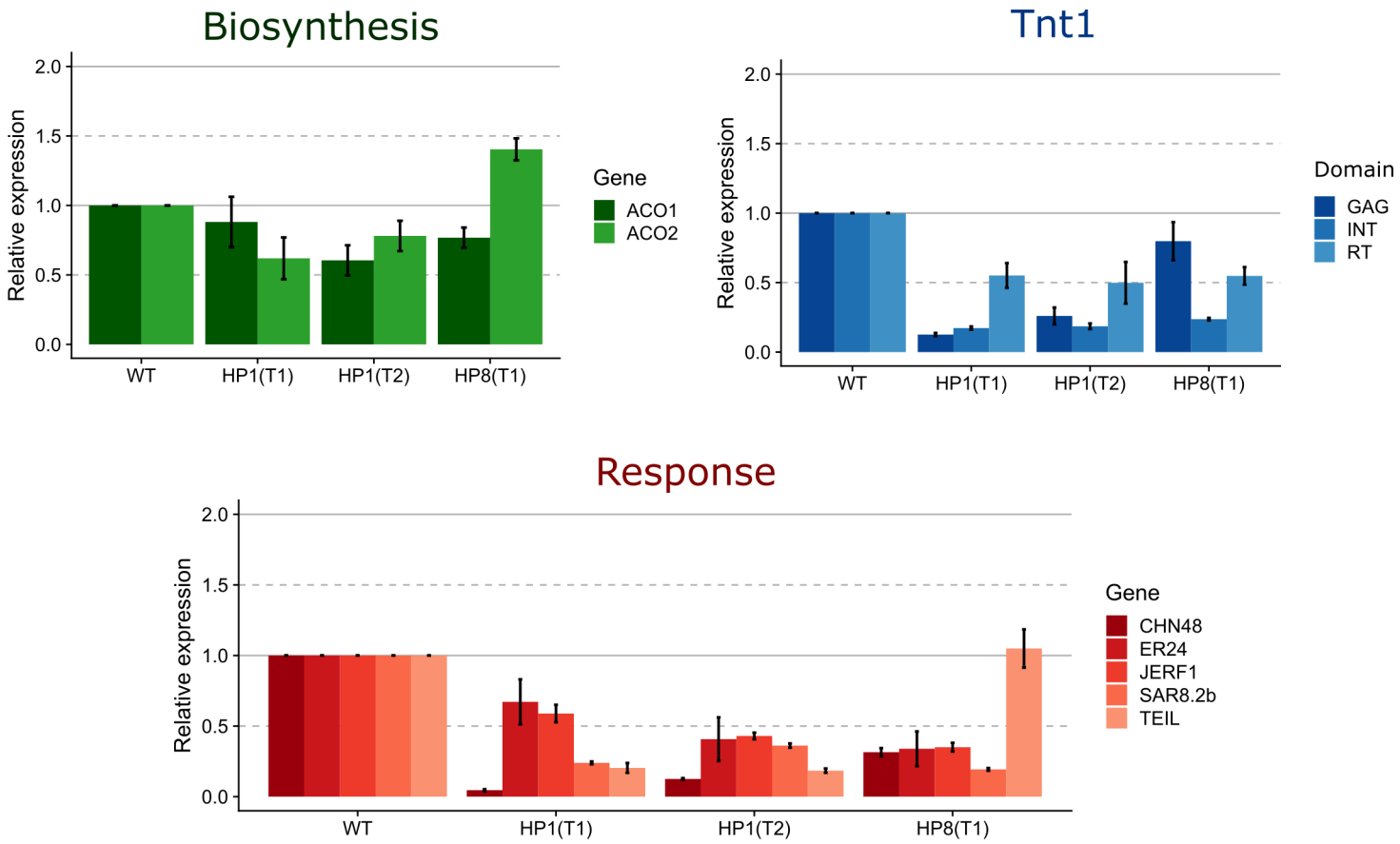
A

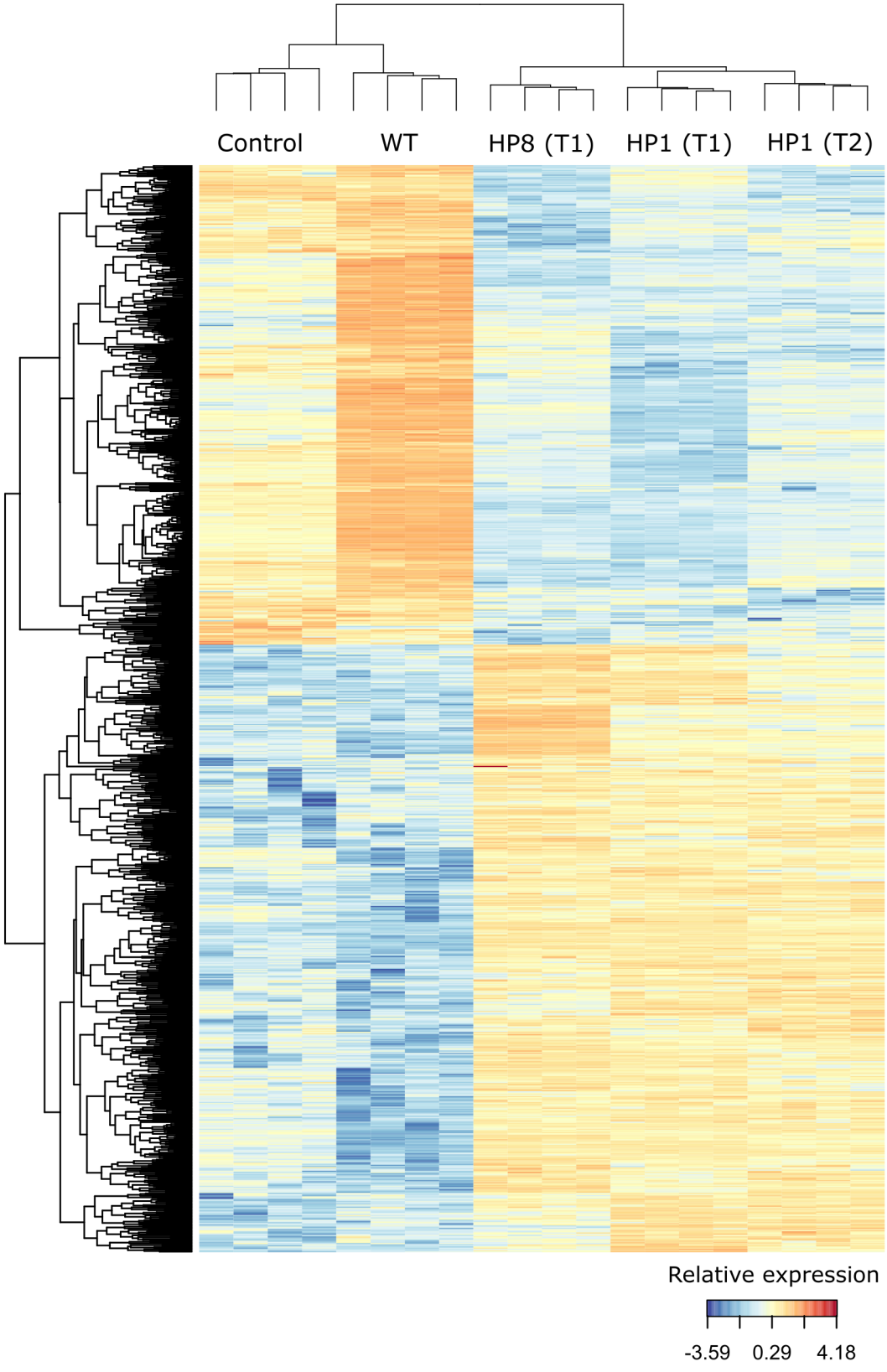


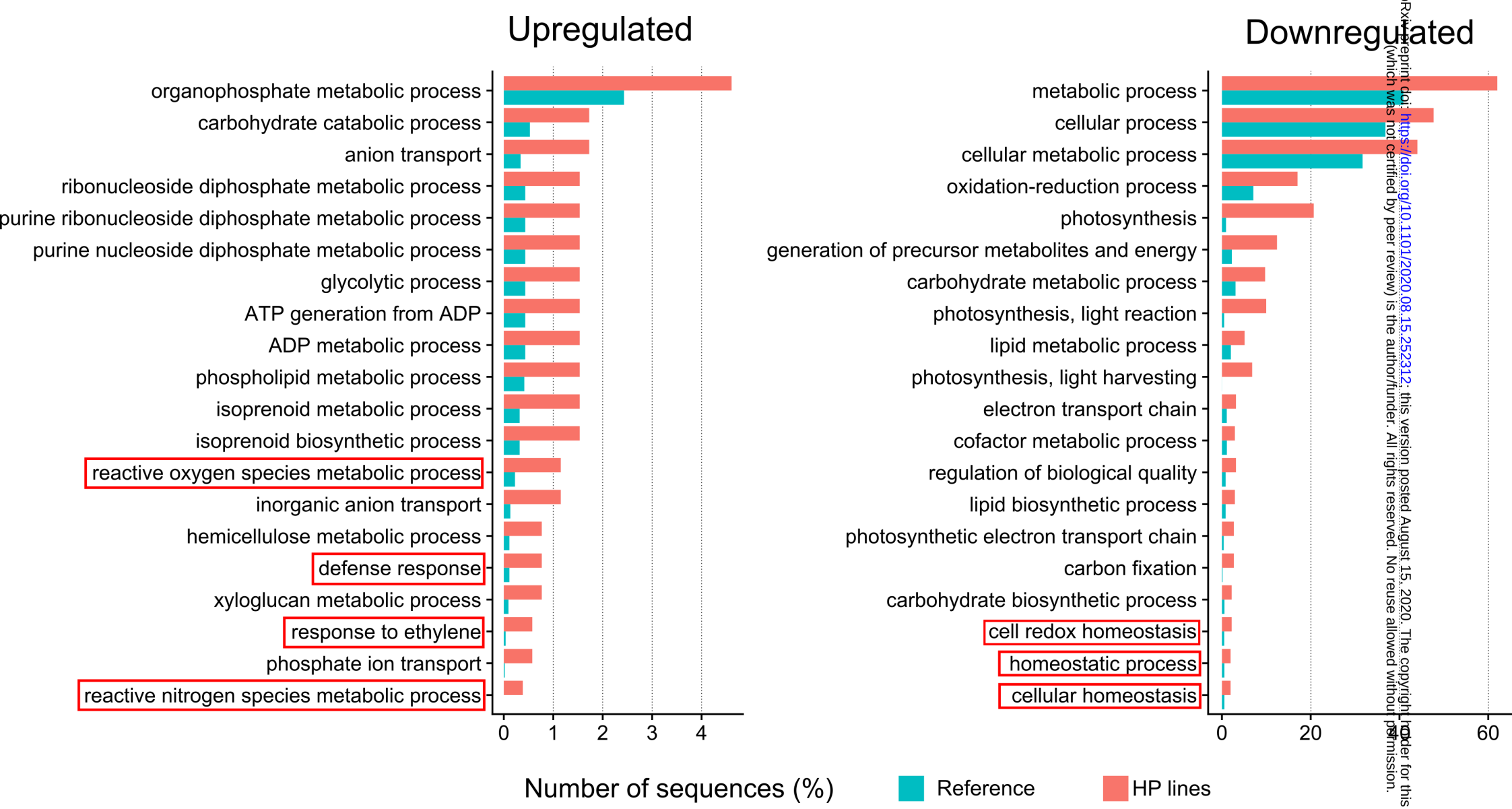
B



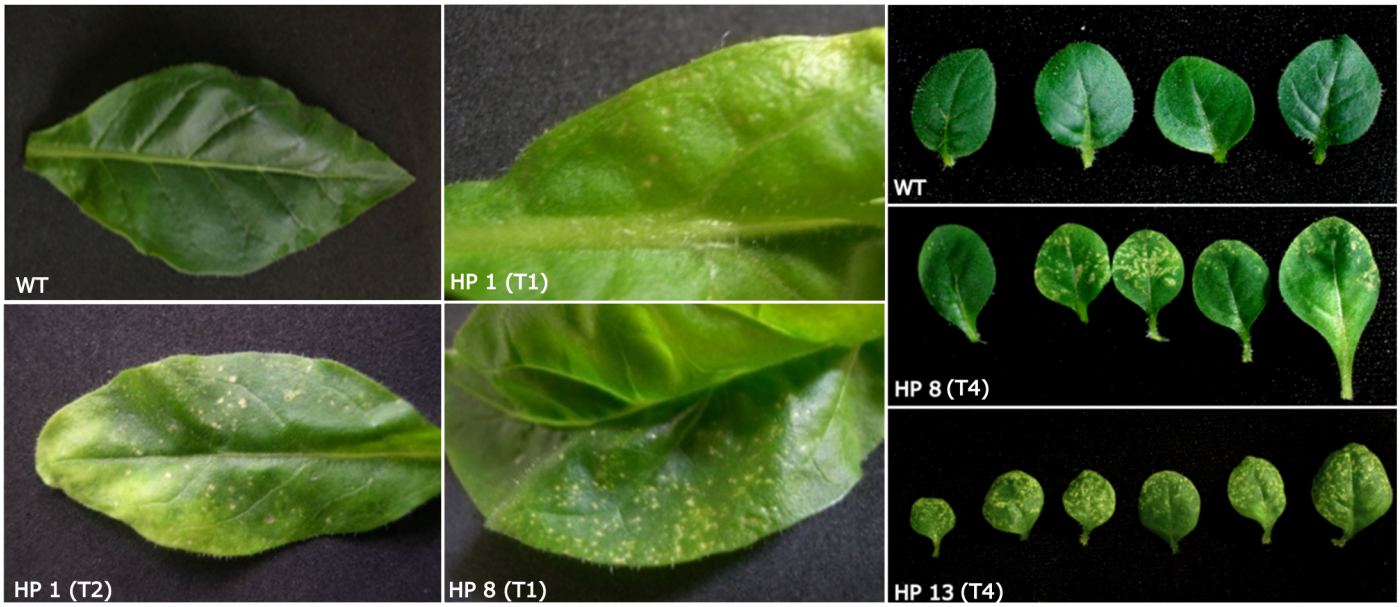




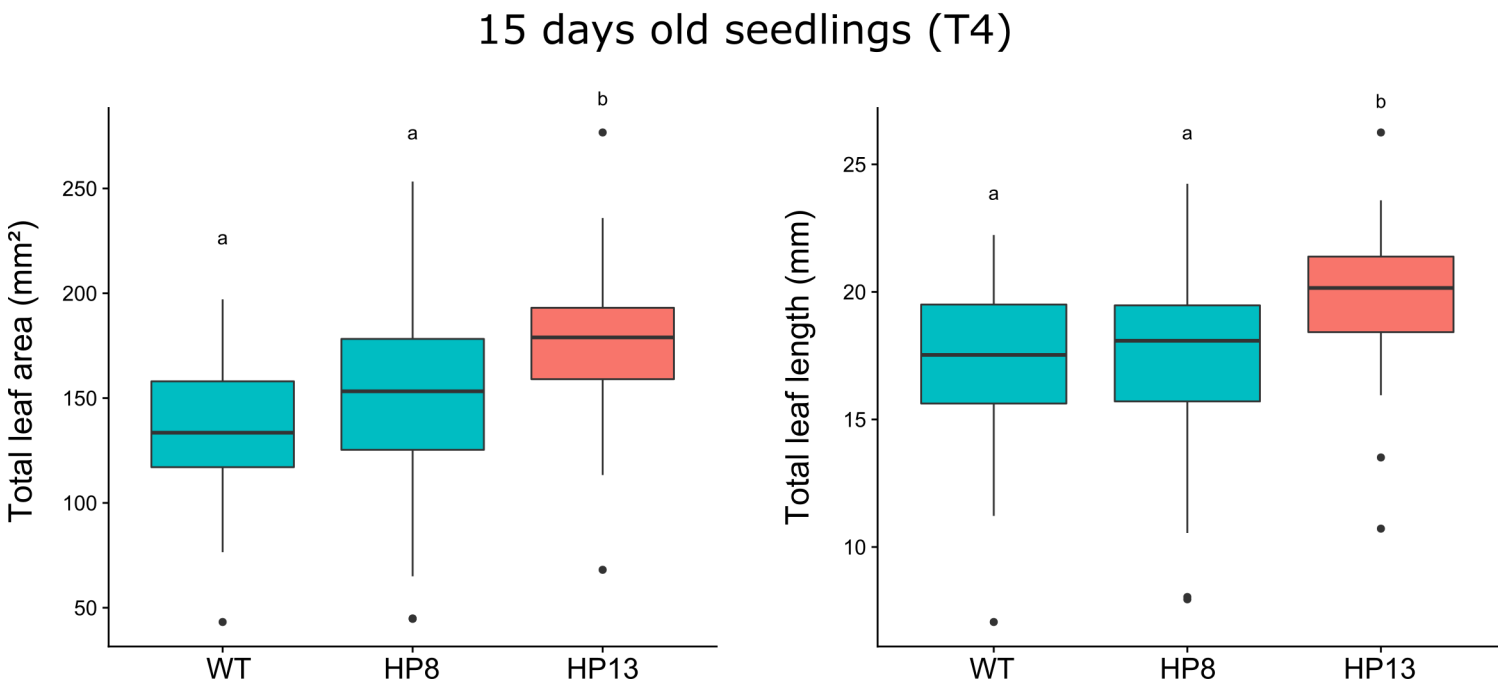




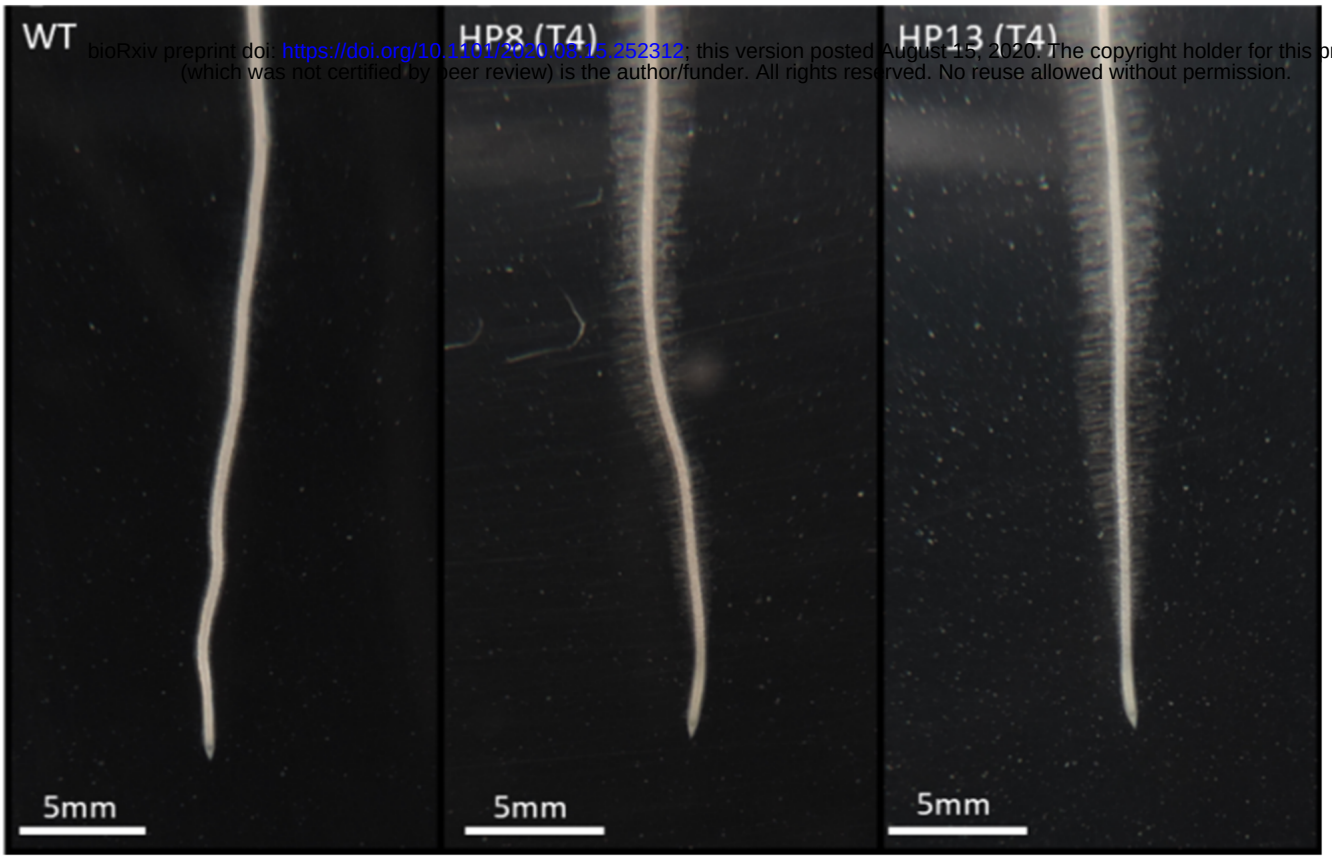
A



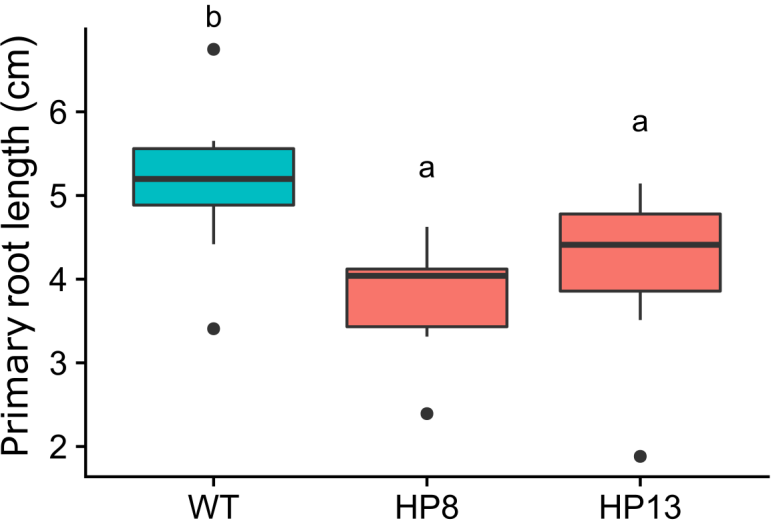
B

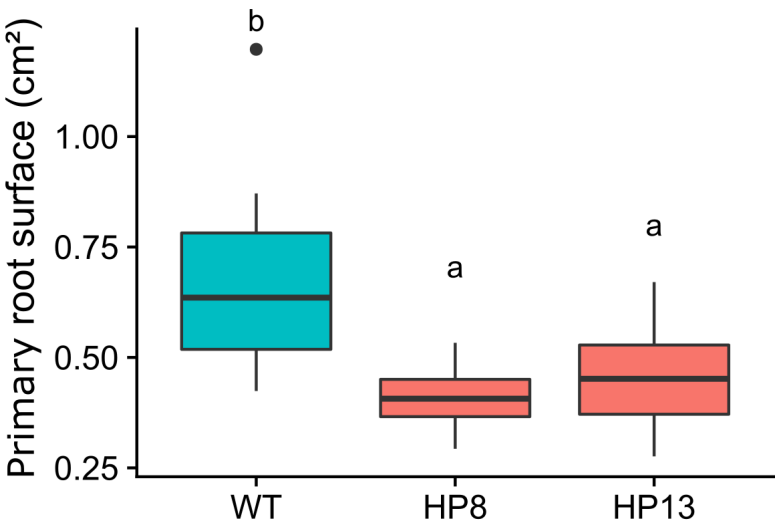


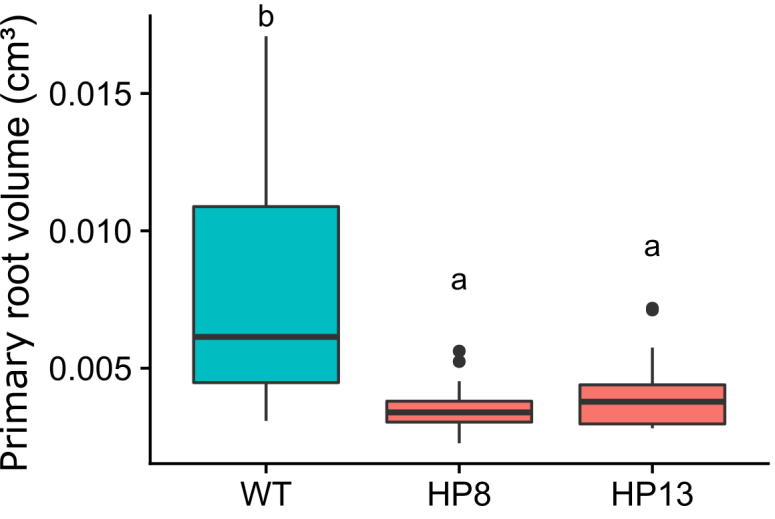
A

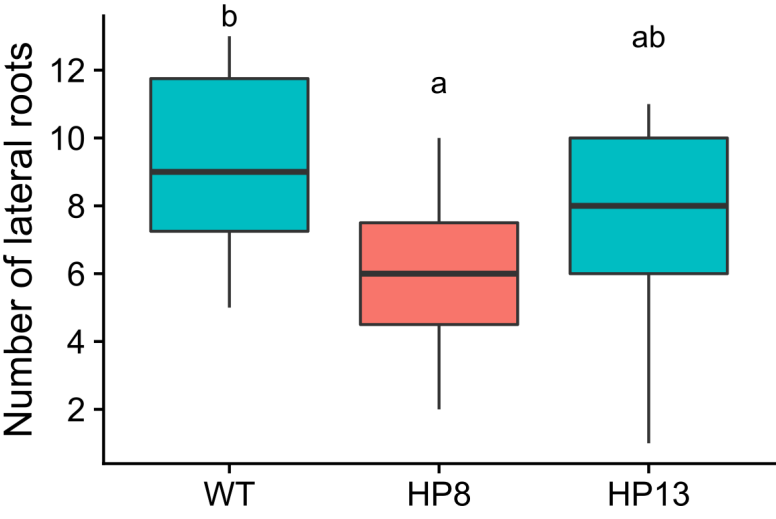


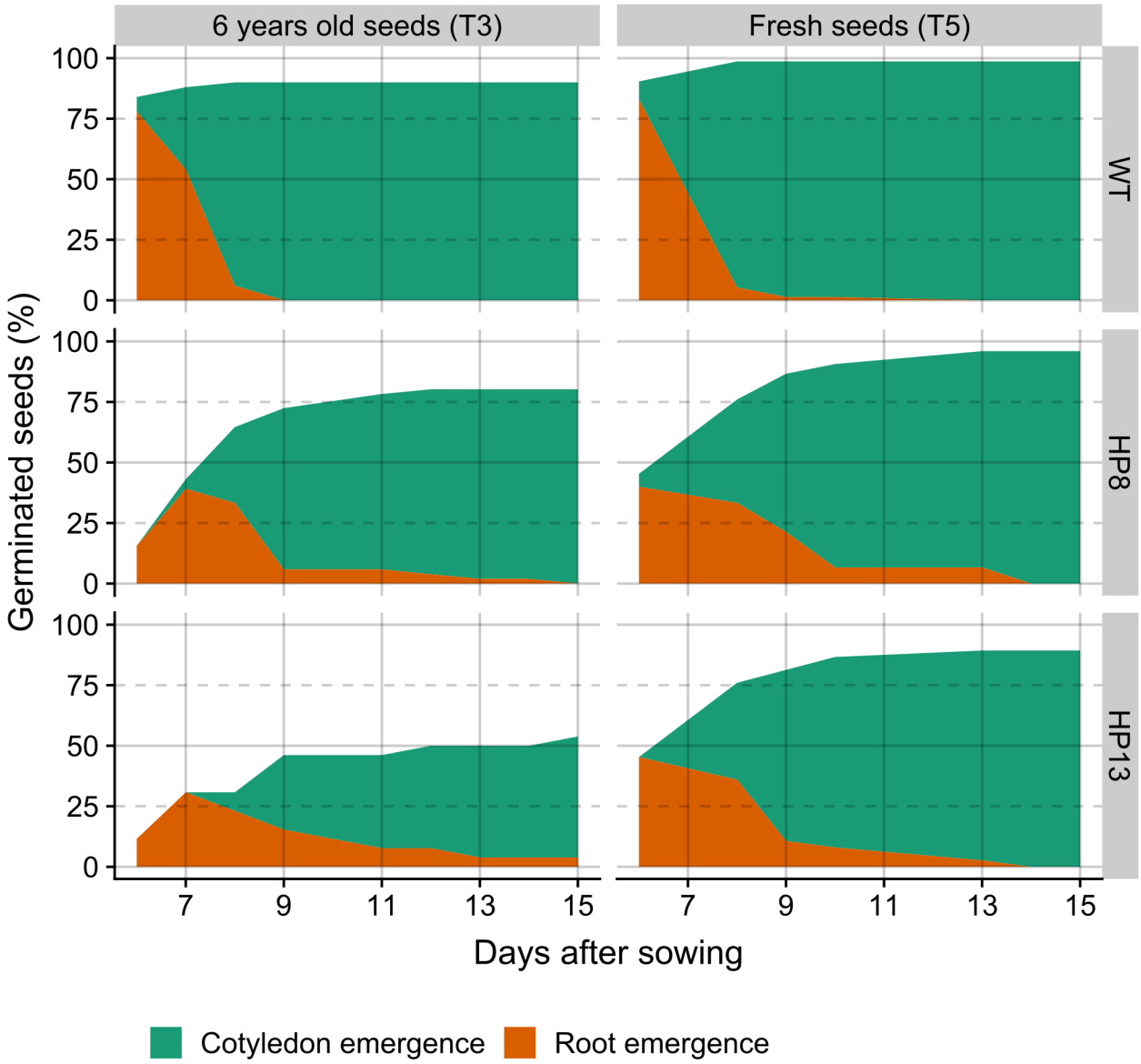
B

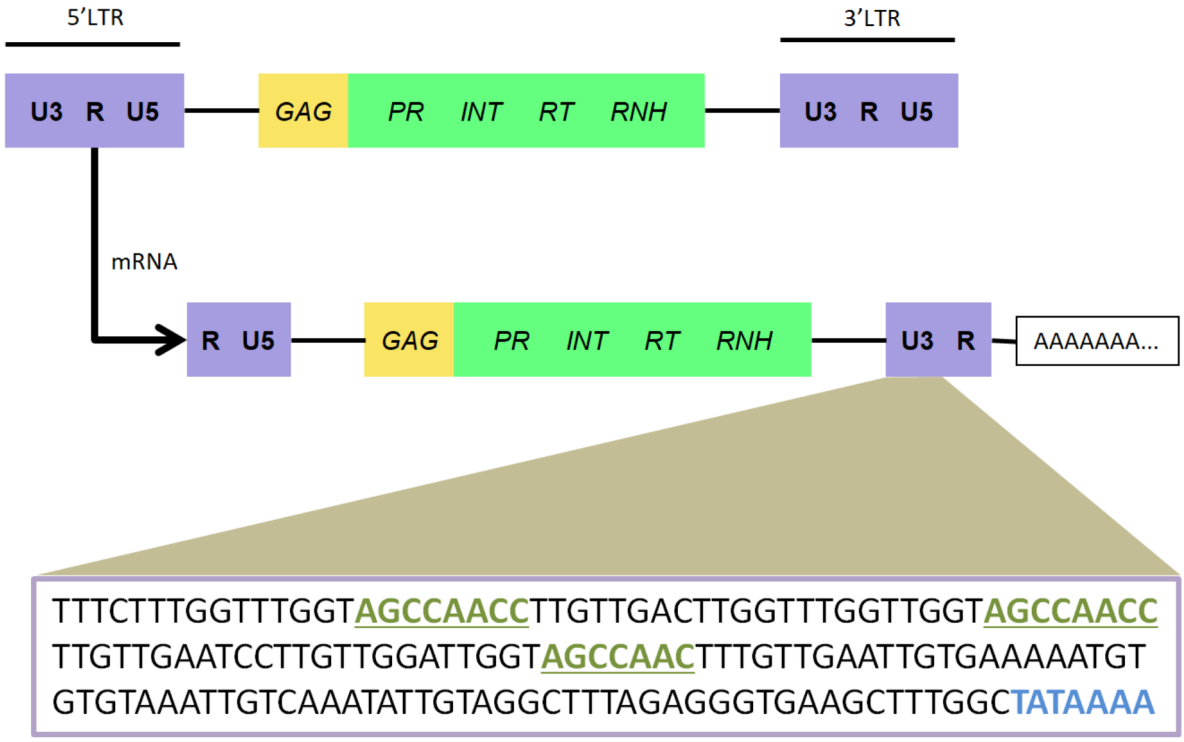




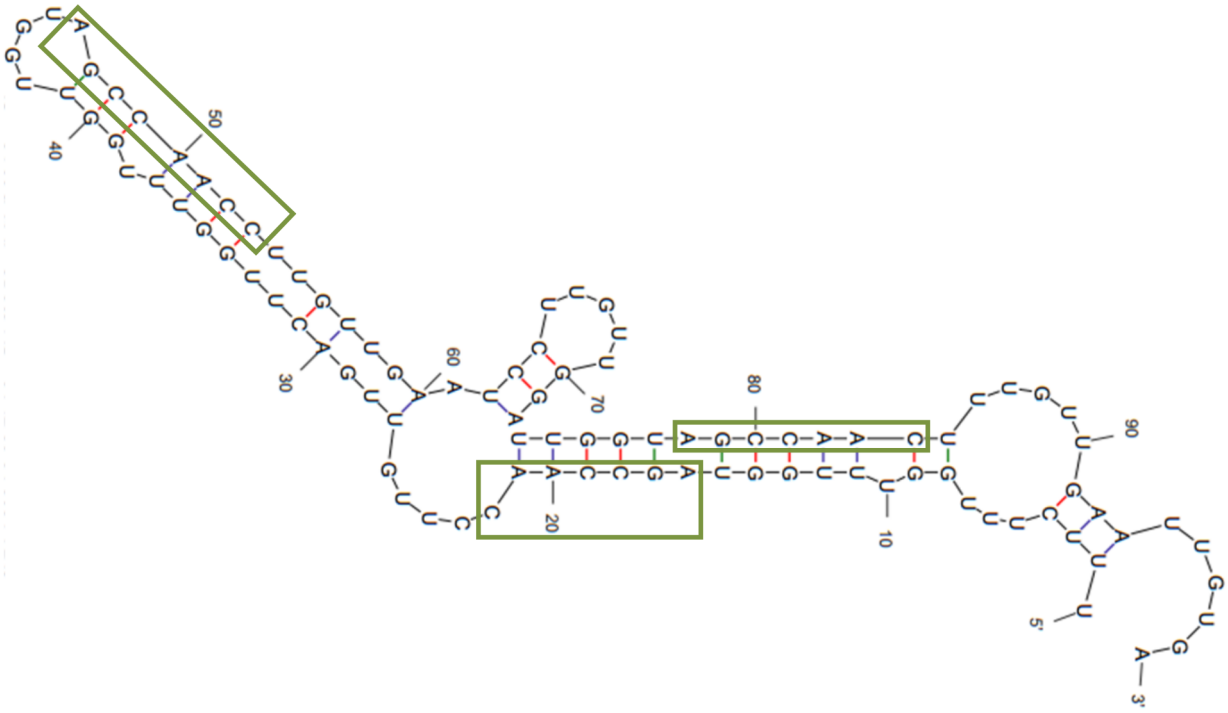


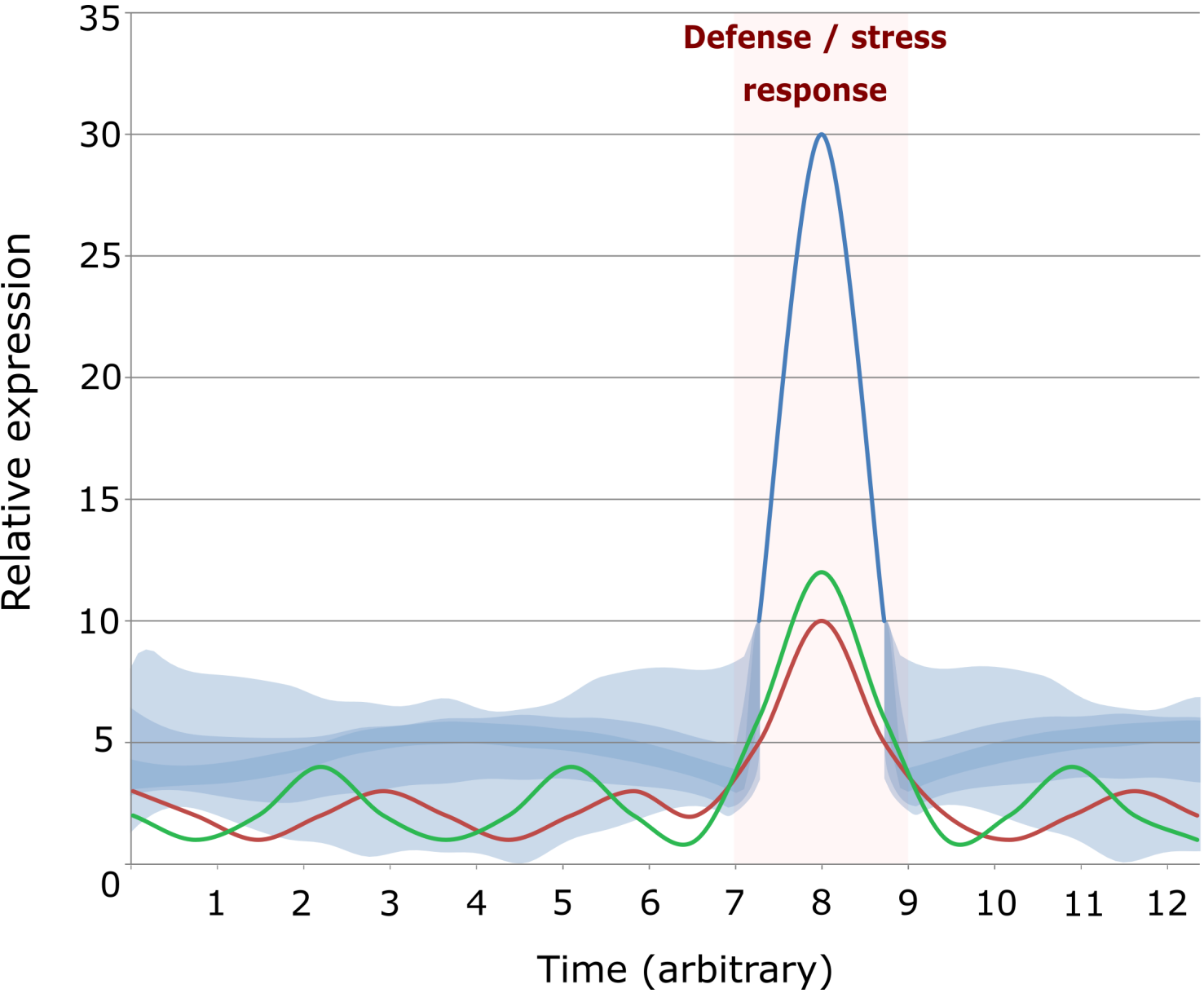






TTTCTTTGGTTTGGTAGCCAACCTTGTTGACTTGGTTTGGTTGGTAGCCAACC
TTGTTGAATCCTTGTTGGATTGGTAGCCAACTTTGTTGAATTGTGAAAAATGT
GTGTAAATTGTCAAATATTGTAGGCTTTAGAGGGTGAAGCTTTGGCTATAAAA





Ethylene biosynthesis genes

Ethylene responsive genes

Tnt1