

**University of São Paulo  
“Luiz de Queiroz” College of Agriculture**

**Tnt1 retrotransposon expression and ethylene phytohormone  
interplay mediates tobacco (*Nicotiana tabacum*) defense responses**

**Danielle Maluf Quintanilha**

Thesis presented to obtain the degree of Doctor in  
Science. Program: International Plant Cell and  
Molecular Biology

**Piracicaba  
2014**

**Danielle Maluf Quintanilha**  
**Biologist**

**Tnt1 retrotransposon expression and ethylene phytohormone interplay  
mediates tobacco (*Nicotiana tabacum*) defense responses**

Advisor:  
Prof. Dr. **MARIE-ANNE VAN SLUYS**

Thesis presented to obtain the degree of Doctor in  
Science. Program: International Plant Cell and  
Molecular Biology

**Piracicaba**  
**2014**

**Dados Internacionais de Catalogação na Publicação**  
**DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP**

Quintanilha, Danielle Maluf

Tnt1 retrotransposon expression and ethylene phytohormone interplay  
mediates tobacco (*Nicotiana tabacum*) defense responses / Danielle Maluf  
Quintanilha. - - Piracicaba, 2014.

103 p: il.

Tese (Doutorado) - - Escola Superior de Agricultura "Luiz de Queiroz", 2014.

1. Tnt1 2. Retrotransposon 3. *Nicotiana tabacum* 4. Etileno 5. Respostas de defesa  
6. Pequenos RNAs I. Título

CDD 633.71  
Q7t

**"Permitida a cópia total ou parcial deste documento, desde que citada a fonte -O autor"**

## ACKNOWLEDGEMENTS

I would like to acknowledge and dedicate this thesis to my parents, José Francisco and Gisleine. They have worked hard to provide me everything I needed to get to this point of my education. I hope I have made you proud, and I am thankful for everything you have done for me. This thesis is in your honor.

I would also like to register here that I am deeply thankful to my best friend (and soon to be husband) Massimo for his unconditional love, support and kindness. Thank you for making me so happy and for cheering me up when I needed, teaching me how to be a better person and for being so understanding and thoughtful. You are the love of my life.

I must thank my dear friend and research project partner Erika. Thank you for training me so patiently, for all our endless talks, confessions and existential crisis. They were very important for my personal growth and also for the conclusion of this work. I will miss you!

I am thankful to my co-advisor Dr. Pal Maliga, his wife Zora and to my friend Tarinee for their warm welcome at the time I spent in the USA. I am even more thankful for all the knowledge I was able to absorb daily at their lab, a truly great environment to work in. It was the most enriching experience of my life both personal and professionally. I was lucky to be able to exchange cultures and get to know such amazing people. You will always be very dear to me for that.

I cannot forget to thank Professor Luciano Freschi and Professor Lazaro E. P. Peres and their lab staff, always willing to help me with ethylene experiments. I am also thankful to my own lab staff. Thanks for the support, prolific discussions and for the daily company throughout these six years.

Last but not least, I must thank my advisor, Prof. Marie-Anne Van Sluys. Without her this project wouldn't be possible. Thank you for the opportunity, trust and intellectual, financial and moral support. I have learned that we can always become better versions of ourselves, that perseverance pays back and that being a positive person makes life easier. There were times when I have doubted I would be able to conclude this PhD, but you haven't questioned my ability once. And today it is finished. Thank you!



## SUMMARY

RESUMO.....	7
ABSTRACT.....	9
LIST OF ABBREVIATIONS.....	11
1 INTRODUCTION.....	13
1.1 Transposable elements.....	13
1.1.1 What are transposable elements and how are they classified?.....	13
1.1.2 Transposition impacts on genome size, mutations and gene expression.....	19
1.1.3 Gene regulatory networks wired by transposable elements.....	24
1. 2 Plant small RNAs.....	26
1.2.1 Small RNAs.....	26
1.2.2 Short interfering RNAs.....	27
1.2.3 Micro RNAs and Polycomb Group proteins (PcGs).....	29
1.3 Tnt1.....	31
1.3.1 Tnt1 discovery, structure and life cycle.....	31
1.3.2 Tnt1 is transcriptionally active and responsive to stress signals.....	33
1.3.3 Retrotransposon control.....	36
1.4. Ethylene.....	38
1.4.1 Ethylene and plant defense responses.....	38
1.4.2 Ethylene synthesis, perception and responses.....	39
1.4.3 Ethylene triple response.....	41
1.5 Objectives.....	41
1.5.1 Motivation and previous results in our laboratory.....	41
1.5.2 General objective .....	42
1.5.3 Specific objectives .....	42
References .....	42
2 EVIDENCING GENE REGULATORY NETWORKS FROM TRANSPOSABLE ELEMENT EXPRESSION .....	51
Abstract.....	51
2.1 Introduction.....	51
2.2 Development.....	52
2.2.1 Methods.....	52
2.2.1.1 Plant samples.....	52
2.2.1.2 Microscopy.....	53
2.2.1.3 RNA sequencing.....	53
2.2.1.4 Bioinformatics and statistical Analysis.....	53
2.2.1.5 Gene regulatory networks.....	54
2.2.1.6 Real time PCRs.....	54

2.2.2 Results.....	54
2.3 Conclusion/ Final considerations.....	74
References.....	75
3 INTERPLAY BETWEEN RETROELEMENT EXPRESSION AND ETHYLENE MEDIATES PLANT DEFENSE RESPONSES.....	79
Abstract.....	79
3.1 Introduction.....	79
3.2 Development.....	80
3.2.1 Methods.....	80
3.2.1.1 Plant samples.....	80
3.2.1.2 Online bioinformatics tools.....	80
3.2.1.3 Ethylene treatment and ethylene gas chromatography.....	81
3.2.1.4 RNA isolation and Real Time PCR (qPCR).....	81
3.2.1.5 Histochemical GUS assay.....	82
3.2.2 Results .....	83
3.3 Conclusion/ Final considerations .....	91
References.....	93
4 THESIS FINAL DISCUSSION.....	97
Reference.....	103

## RESUMO

### **A dinâmica entre a expressão do retrotransposon Tnt1 e o fitormônio etileno envolvida nas respostas de defesa em tabaco (*Nicotiana tabacum*)**

Tnt1 é um retrotransposon com LTR transcricionalmente ativo, e está presente em mais de 600 cópias no genoma de *Nicotiana tabacum*. Em condições normais de crescimento Tnt1 é expresso em níveis basais. No entanto, sua expressão é induzida pelo estímulo de estresses bióticos e abióticos. Plantas de tabaco transgênicas (chamadas de HP) expressando um grampo da transcriptase reversa de Tnt1 foram geradas. Estas apresentaram fenótipos como: pontos de morte celular e deposição de calose nas folhas e severas anomalias de desenvolvimento severas nas porções aérea e radicular das plantas. Sequenciamento de RNA de folhas com os pontos de morte celular revelou uma reorganização de redes de regulação transcricional relacionadas a resposta a estresses. Essas novas redes surgiram exclusivamente nas plantas HP. Entre os genes modulados positivamente estavam genes de síntese e de resposta ao etileno. O presente trabalho teve como objetivo elucidar a relação observada entre Tnt1 e o fitormônio etileno gerando um modelo de atuação. Os resultados obtidos permitiram demonstrar que plântulas e plantas HP adultas tem um aumento na síntese de etileno quando comparadas à selvagem. A predição do dobramento do RNA mensageiro de Tnt1 permitiu a identificação de sequências responsivas ao etileno localizadas em posição potencial para formar grampos. Desta forma, é possível que a expressão de Tnt1 leve à produção de pequenos RNAs que tem como alvo sequências responsivas a etileno presentes tanto no próprio elemento quanto em regiões promotoras de outros genes. A quantificação da expressão de Tnt1 versus genes relacionados ao etileno revelou um padrão em “oposição de fase” nas HPs, o que nos levou a hipotetizar que talvez ocorra uma relação antagonista entre a expressão de Tnt1 e a expressão de genes responsivos ao etileno envolvidos em respostas de defesa vegetais. Nossos resultados sugerem que Tnt1 pode gerar pequenos RNAs que exercem controle transcricional sobre Tnt1 e outros genes endógenos. Nosso modelo estabelece um novo papel biológico para um retrotransposon: Tnt1 agiria como um modulador da indução de genes mediada por etileno nas respostas de defesa de tabaco, trazendo o sistema de volta à condição homeostática e encerrando as respostas de defesa.

Palavras-chave: Tnt1; Retrotransposon; *Nicotiana tabacum*; Etileno, Respostas de defesa; Pequenos RNAs





## ABSTRACT

### **Tnt1 retrotransposon expression and ethylene phytohormone interplay mediates tobacco (*Nicotiana tabacum*) defense responses**

Tnt1 is a transcriptionally active LTR-retrotransposon, present in over 600 copies in the *Nicotiana tabacum* genome. Under normal growth conditions, Tnt1 expression is limited to basal levels, but its expression is further induced under biotic and abiotic stresses. Transgenic tobacco plants (HP plants) expressing a Tnt1 reverse transcriptase hairpin were generated. These showed pleiotropic phenotypes such as cell death spots on the leaves and callose deposition and other severe abnormal development in aerial and underground portions. RNA sequencing of leaves with cell death spots revealed a rewiring of transcriptional regulatory networks related to stress responses exclusive to HPs. Among the positively modulated genes were ethylene synthesis and response cascade genes. The objective of the present work was to unravel the relation observed between Tnt1 and ethylene, generating a model. The results obtained suggest that HP seedlings and plants have increased ethylene synthesis when compared to the wildtype. Folding prediction of Tnt1 messenger RNA allowed the identification of ethylene-responsive sequences in putative stem loop locations. Thus it is possible that Tnt1 expression can produce small RNAs targeted to sequences present in the Tnt1 retrotransposon itself as well as at the promoter region of other ethylene responsive genes. Quantification of the expression of Tnt1 and ethylene related genes revealed “phase opposition” expression kinetics in the HPs, which led us to hypothesize that there might be an antagonistic relationship between the expression of Tnt1 and the expression of ethylene responsive genes involved in plant defense responses. Our findings suggest that Tnt1 could generate sRNAs that exerts transcriptional control over itself as well as other genes. Our model establishes a completely new biological role for a retrotransposon: Tnt1 would provide feedback control to ethylene-mediated gene regulation in tobacco defense responses, bringing the system back to a homeostatic condition and turning the defense responses down.

**Keywords:** Tnt1; Retrotransposon; *Nicotiana tabacum*; Ethylene; Defense responses; Small RNAs



## LIST OF ABBREVIATIONS

AGO = ARGONAUTE  
 cDNA = Complementary DNA  
 CDS = Coding DNA Sequences  
 DBP = DNA Binding Protein  
 DCL = DICER-LIKE  
 dsRNA = Double-stranded RNA  
 ER = Endoplasmic Reticulum  
 ERE = Ethylene Responsive Element  
 EREBP = Ethylene Responsive Element Binding Protein  
 ERF = Ethylene Responsive Factor  
 HP = Hairpin  
 Hypersensitive Response = HR  
 INT= Integrase  
 LTR = Long Terminal Repeat  
 MCE = Mean Conditional Entropy  
 miRNA = micro RNA  
 mRNA = Messenger RNA  
 NLS = Nuclear Localization Signal  
 NR = Nitrate Reductase  
 ORF = Open Reading Frame  
 PBS = Primer Binding Site  
 PcG = Polycomb Group  
 PCR = Polymerase Chain Reaction  
 Pathogenesis-Related = PR  
 PHD = Plant Homeo Domain  
 Po Lv = RNA Polymerase V  
 Pol II = RNA Polymerase II  
 Pol IV = RNA Polymerase IV  
 PPT = Polypurine Tract  
 PR = Protease  
 PRC = Polycomb Repressive Complex  
 PTGS = Post Transcriptional Gene Silencing

qPCR = Quantitative PCR

R region = Repeated RNA region

RBP = RNA Binding Protein

RdDM = RNA-directed DNA Methylation

RDRP = RNA-dependent RNA Polymerase

RISC = RNA-Induced Silencing Complex

RITS = RNA-Induced Transcriptional gene Silencing

RNAi = RNA Interference

RNA-seq = RNA sequencing

RNP = RiboNucleoprotein Particle

RT = Reverse Transcriptase

ShRNA = Short Hairpin RNA

siRNA = short interfering RNA

sRNA = small RNA

Systemic Acquired Response = SAR

TE = Transposable Element

TGS = Transcriptional Gene Silencing

TIR = Terminal Inverted Repeat

TrxG = Trithorax Group proteins

U3 = Unique 3' RNA

U5 = Unique 5' RNA

VIN3 = VERNALIZATION INSENSITIVE 3

VLP = Virus-Like Particle

WT = Wildtype

## **1 INTRODUCTION**

### **1.1 Transposable Elements**

#### **1.1.1 What are transposable elements and how are they classified?**

There are genes capable of moving from one position to another in the genome of a cell. Those genes or “genetic entities” were named Transposable Elements (TEs). Their discovery was made in the early 1940s when the researcher Barbara McClintock observed a correlation between the movement of a locus, which led to chromosomal breakage and the appearance of spontaneous and unstable spot mutations in corn kernels. Due to the ability to alter the plants characteristics, changing the functioning of genes to which they are associated with, she initially named them “controlling elements of the gene” (MCCLINTOCK, 1984). Although her work was seen with great skepticism by the scientific society at the time, as the years went by several articles corroborated her observations. In 1983 she was awarded with the Nobel prize in Physiology or Medicine for the discovery of the transposable elements through her studies in corn.

The current technologies allow sequencing of whole genomes fast and at low costs. With the whole sequence of a genome in hands, it is possible to identify practically all the TEs present. Mobilome is the entire set of transposable elements present in a genome. However, the first TEs described were discovered because they generated phenotypes or a mutant characteristic. The identification of new TEs was made possible through the ability to recover DNA fragments that contained the mutated gene with a TE insertion. One example of that is the retrotransposon Tnt1. This retrotransposon was discovered when researchers were studying the enzyme nitrate reductase. When they sequenced the gene that codes for this enzyme in two different mutants, they discovered that a retrotransposon had inserted itself into a nitrate reductase (NR) gene. As the gene was disrupted, the enzyme wasn't produced (GRANDBASTIEN; SPIELMANN; CABOCHE, 1989).

The transposable element Ac, described by McClintock, was discovered because of the phenotypes it causes due to chromosomal breakage. Those breaks occur on the short arm of the chromosome 9, within a region that harbors genes which confer characteristics easily observable on the corn kernels. One of the genes present in that region is involved in the biosynthesis of pigment, conferring color to the kernels. When the pigment gene is disrupted after a TE insertion inside its

sequence, no pigment is produced, generating white colored kernels. The variegated phenotype occurs when the TE excises from the pigment gene during the kernel development, restoring the gene. Thus, some cells will be able to produce the pigment (creating sectors which have the red pigment) while others don't (remaining white) (McClintock 1950) (Figure 1).

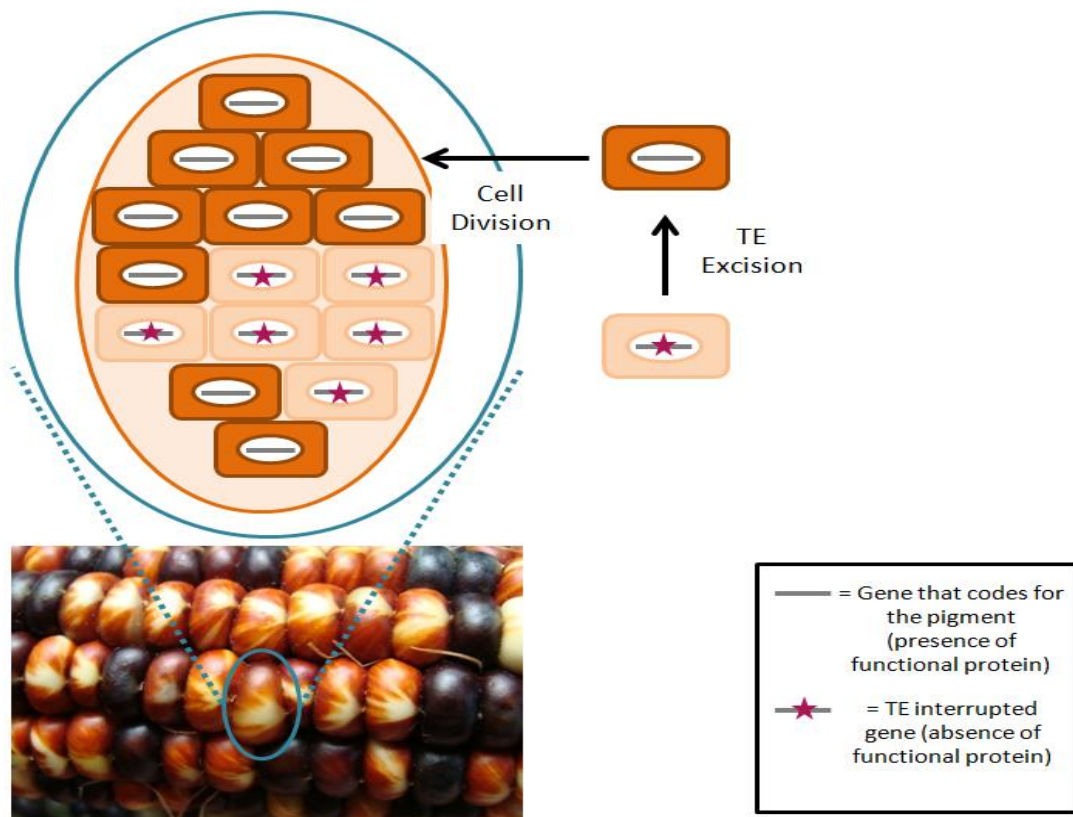


Figure 1 – Transposable elements are responsible for the variegated phenotype in corn. The picture shows mosaic pigmentation in the corn kernels derived from insertion or excision of a transposable element (TE). The illustration shows different cells in a kernel (the rectangles) and the DNA fragment that corresponds to the pigment gene is represented by the gray bar. When the TE is inserted within the gene it becomes interrupted and lack of pigmentation occurs, generating lighter colored sector. When the TE transposes out of the pigment gene sequence, its sequence is restored as well as the synthesis of pigment, generating darker colored spots

(Adapted from FESCHOTTE; JIANG; WESSLER, 2002)

Transposable elements have a great variety of structures. Some of them code for all the enzymes needed for transposition and are known to be autonomous elements. Terminal repeats that flank the coding region can be reverse complementary or direct, contain the promoter region (complete or partial) and are important for the recognition of the TE at the moment of transposition. When the TE is inserted in the genome the target site is duplicated, thus creating a “scar” even if

the TE is later on excised. TEs are present in the genomes as two forms: autonomous elements, which have all the proteins needed for transposition coded in their sequence and non-autonomous elements which are defective, usually coding for non-functional proteins. Non autonomous elements can still transpose by the action *in trans* of enzymes produced by an autonomous element of the same family that recognize its terminal repeats. The great number of TE families described so far reflects its extensive structural diversity. TEs are considered part of the same family when they share terminal repeats, which allows them to be mobilized by the same enzyme. It is also accepted to consider TEs from a same family when they share at least 80% of nucleotide identity (WICKER et al., 2007).

Transposable elements are classified in two major classes based on the mechanism of transposition: class I (also called retroelements or “copy-and-paste”) transpose via an RNA intermediate used for reverse transcription into cDNA further creating a new copy of the retroelement. Class II (also called DNA transposons and some called “cut-and-paste”) transpose directly through a molecule of DNA (WICKER et al., 2007). These TEs are present in practically all living forms, including bacteria, algae, fungi, plants and animals (BENNETZEN, 2000) (Figure 2).



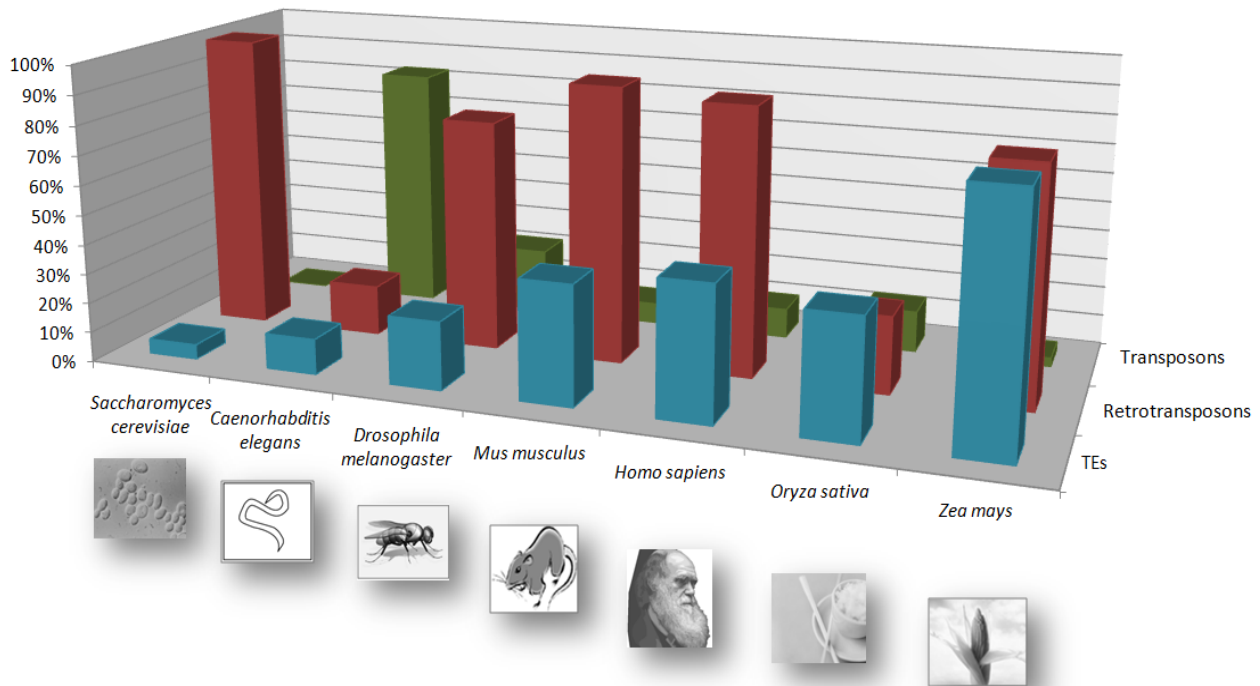


Figure 2 – Transposable element content in different eukaryotic genomes. The bar graph depicts the total TE percentage (in blue) as well as its distribution in transposons (in green) and retrotransposons (in red) in yeast (*Saccharomyces cerevisiae*), a nematode (*Caenorhabditis elegans*), fruit fly (*Drosophila melanogaster*), a rodent (*Mus musculus*), human (*Homo sapiens*), rice (*Oryza sativa*) and corn (*Zea mays*) (Adapted from CHENAIS et al., 2012; FESCHOTTE; PRITHAM, 2007; PATERSON et al., 2009)

Some bacterial genomes are composed up to 10% of insertion sequences (IS elements), and TE content in eukaryotes can reach much higher percentage (MILLER; CAPY, 2004). The human genome, for instance, is made up 45% of TEs (INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM, 2001), although some new studies claim this percentage might be even higher, up to 69% (KONING et al., 2011). Some grasses like maize can have surprising 80% of TEs in their genomes (PATERSON et al., 2009), and usually retrotransposons make up a larger fraction in the genomes than transposons. On the other side, there are organisms with fewer amounts of TEs, such as *Arabidopsis thaliana*, in which TEs are only about 14% of the genome (Arabidopsis genome initiative, 2000). The nematode *Caenorhabditis elegans* is another unusual organism when it comes to TEs: not only does it have little amount of TEs (12%), but also it is much more abundant in transposons (85%) than retrotransposons (15%) (BIÉMONT; VIEIRA, 2006; CHENAIS et al., 2012; FESCHOTTE; PRITHAM, 2007; PATERSON et al., 2009). This broad distribution is evidence that they have appeared early in evolution.

Class I retroelements are composed by retrotransposons with LTR (Long Terminal Repeats) and by retrotransposons without LTR (also known as retroposon). Retrotransposons with LTR are divided in two groups, *Ty3/gypsy* and *Ty1/copia*. Non-LTR retrotransposons are divided in LINEs (long interspersed nuclear elements) and SINEs (short interspersed nuclear elements) (XIONG; EICKBUSH, 1990). All these elements have in common the fact that they are transposed by the transcription of a messenger RNA (mRNA) molecule which is reverse transcribed into a cDNA molecule (performed by the activity of a reverse transcriptase). This cDNA molecule, a copy of the original element, will thus complete the replicative retrotransposition through the insertion of a new copy of the retroelement in the genome (Figure 3). This replicative mode of retrotransposition is the reason they are called “copy-and-paste”. Retrotransposons are the most abundant TEs in eukaryote genomes. LTR retrotransposons are majority in plant genomes (ZHAO; MA, 2013), while non-LTR retrotransposons are majority in mammal genomes (DEININGER; BATZER, 2012). Retroviruses have similar structure and protein domains to the *Ty3/gypsy*, but they have one extra domain that codes for the viral envelope, which allows them to infect other cells (KIM et al., 1994).

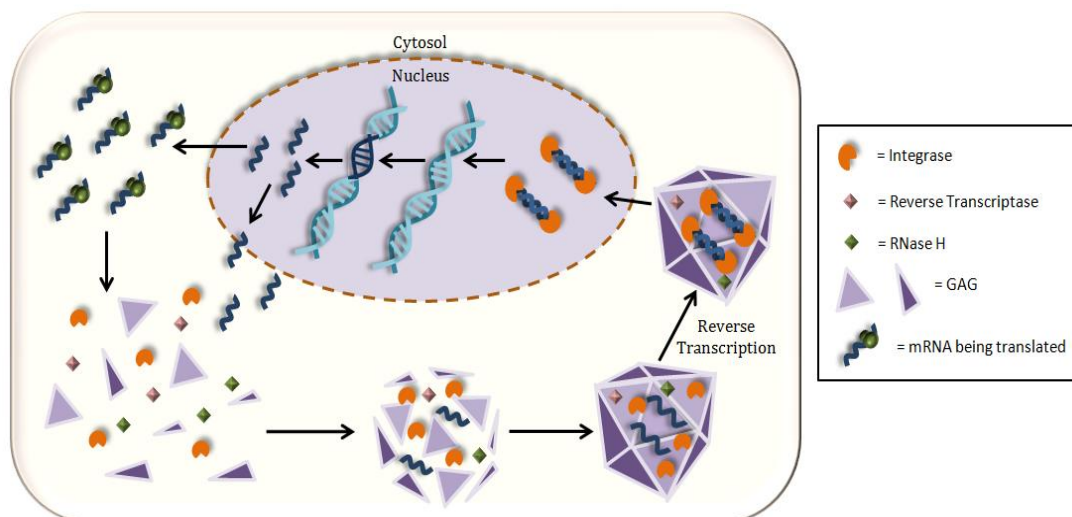


Figure 3 - Life cycle of a retrotransposon with Long Terminal Repeats inside the eukaryotic cell. Initially the retrotransposon is transcribed. The messenger RNAs are directed to the cytoplasm, where some molecules are translated and others will be packaged with Integrase, RNase H and Reverse Transcriptase in a Virus Like Particle (VLP) by Gag proteins. Inside the VLP the messenger RNA is reverse transcribed into a double stranded cDNA, and this cDNA with the integrase are relocated back to the nucleus. There the new retrotransposon copy is then inserted in a new locus of the genome (Modified from SABOT; SCHULMAN, 2006)

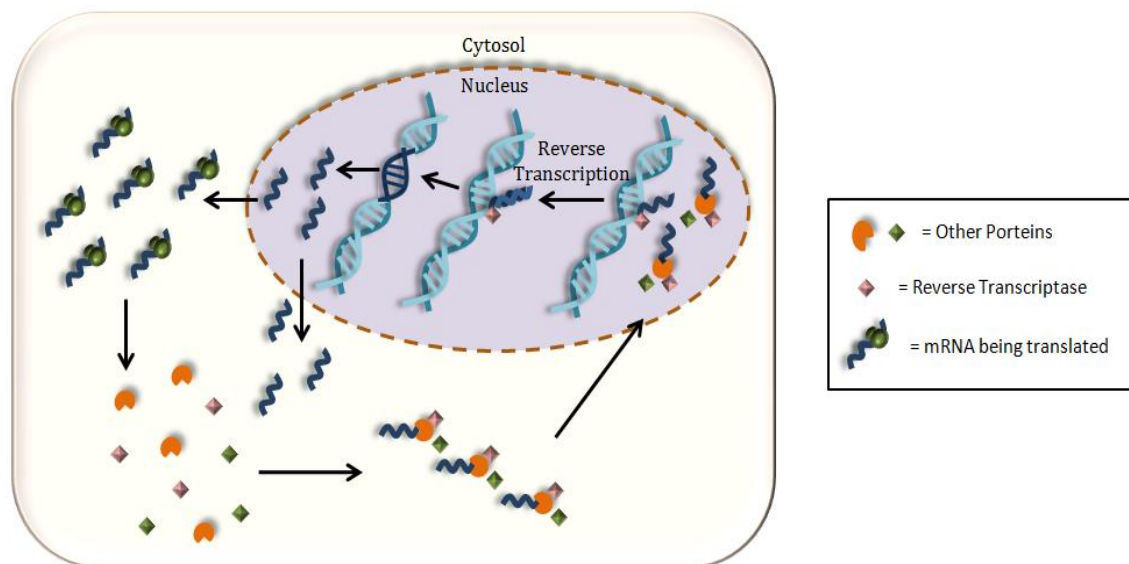


Figure 4 - Life cycle of a non-LTR retrotransposon inside the eukaryotic cell. Like retrotransposons that do have LTRs, initially the retrotransposon is transcribed and the messenger RNAs are directed to the cytoplasm, where some molecules are translated and others will be packaged with proteins into a ribonucleoprotein particle (RNP). The ribonucleoprotein complex is then relocated back to the nucleus, where the messenger RNA is reverse transcribed into a double stranded cDNA directly upon insertion in a new locus of the genome

(Modified from HAN, 2010)

Class II DNA transposons move directly through a DNA molecule. The transposon is transcribed and the protein needed for transposition (transposase) is translated. The transposon is then recognized through its terminal inverted repeats (TIRs) and excised from its original locus by the transposase, thus being reinserted in a new locus in the genome. Because this type of transposition does not imply in higher number of copies inserted like retrotransposons, transposons are present in fewer copy number (WICKER et al., 2007). Transposons can be autonomous or non-autonomous. The first ones code for a functional transposase, and the second ones present a defective ORF and rely on a functional transposase coded by another transposon from the same family to mobilize it *in trans* (MASSON et al., 1989). Figure 5 illustrates the general mechanism by which transposons move.

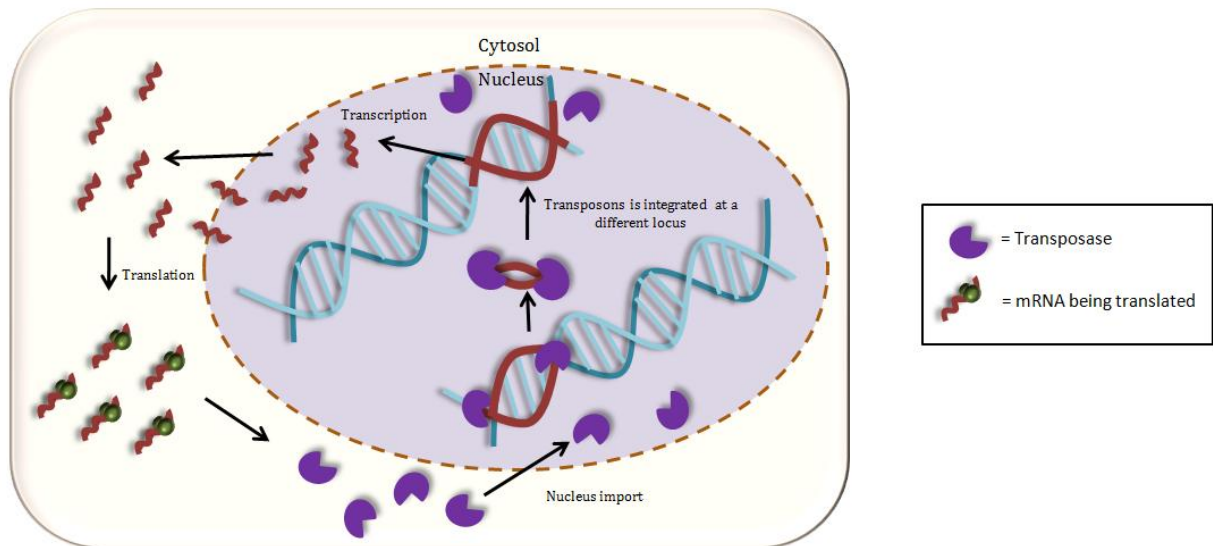


Figure 5 – Transposon life cycle. Transposons are mobilized by a cut-and-paste mechanism, encode a transposase and have terminal inverted repeats (TIRs) at their ends. It is initially transcribed, then the messenger RNA is translated into transposase proteins. Those transposases are relocated to the nucleus where it recognizes and binds to the TIRs and subterminal regions. Once the transposon is excised from its original position, transposase catalyzes its reinsertion in a different locus of the genome (Modified from FEDOROFF, 2012)

### 1.1.2 Transposition impacts on genome size, mutations and gene expression

The various copies of a TE generated during its propagation can have different outcomes. Copies of elements of the same family tend to accumulate mutations derived from errors during its synthesis, transposition or DNA replication. Another possible outcome is the degradation of signaling sequences for the mobilization (such as transposase recognition sites of transposons or integrase of retrotransposons). These mutations can lead to TE inactivation due to erroneous synthesis of the proteins needed for transposition or inability of recognition of the TE sequences by its own proteins. Ultimately, TEs can be eliminated completely by genetic drift or natural selection. Alternatively, TEs can be domesticated (ALZOHAIRY et al., 2013), which involves the recruitment of whole TEs or parts of it in the performance of new functions in the genome. Whichever is the TE destiny, upon transposition they can cause mutations, structural and functional variability to the genome. TE mutagenesis potential is related to their insertion in new genome loci, illegitimate recombination events of non-homologous regions and also epigenetic regulation of other genes (CHENAIS et al., 2012, SLOTKIN; MARTIENSSEN, 2007).

Kobayashi, Yamamoto and Hirochika published (2004) the mechanisms through which transposable element activity can induce notable phenotypes. The accumulation of anthocyanins (a plant pigment) determines the color of grape (*Vitis vinifera*) skin. White cultivars of grape are thought to have arisen from different red cultivars by independent mutations. They describe that the insertion of the LTR retrotransposon *Grapevine retrotransposon 1* or *Gret1* (not present in the Cabernet cultivar) resulted in loss of function of a Myb-related gene (*VvmybA1*), leading to loss of pigmentation in the white cultivar Chardonnay. A subsequent excision of *Gret1* through recombination of its LTRs reverted the effect and resulted in red colored grapes such as the cultivar Ruby Okuyama (as shown in Figure 6).

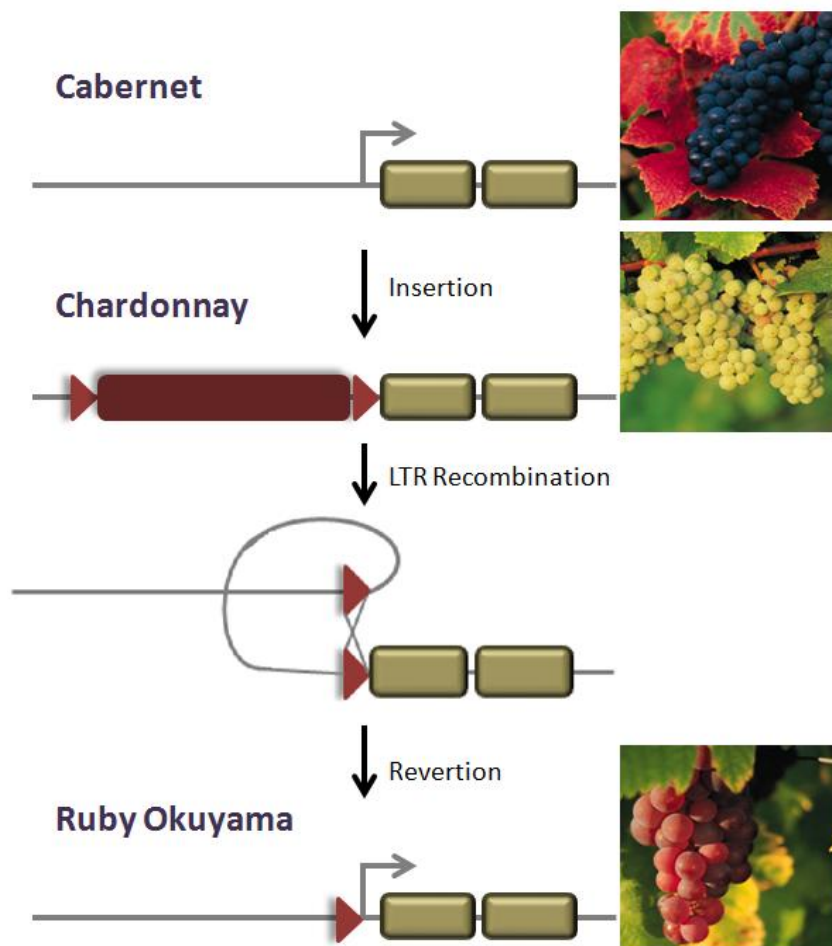


Figure 6 - Transposable element insertions associated with changes in grape color. An initial insertion of a *Gret1* long terminal repeat (LTR) retrotransposon (not present in the Cabernet variety) resulted in a loss-of-function allele of the *Vvmyb1A* gene, leading to a loss of color in the fruit of the Chardonnay variety. A subsequent rearrangement in *Gret1* results in revertant, colored grapes in varieties such as Ruby Okuyama. Exons are depicted as golden boxes. The LTRs flanking *Gret1* just upstream of the *Vvmyb1A* gene are depicted as red triangles (Modified from LISCH, 2013)

Plant genome size is variable, ranging from 0.125 Gbp of *Arabidopsis thaliana* to over 8 Gbp of *Secale cereale*. The size is affected by polyploidization, and also by TE insertion and elimination. TEs comprise variable proportions ranging from 16% of total genome of *Arabidopsis thaliana* up to 88% of *Hordeum vulgare* (CIVAN; SVEC; HAUPTVOGEL, 2011). Even though transposition might be replicative, genome size changes are not unidirectional. Remnants of removed TEs (fragments and solo LTRs) also constitute a significant fraction of repetitive DNA.

TEs were discovered originally by Barbara McClintock (MCCLINTOCK, 1950) because of their potential to cause chromosomal rearrangements which can generate observable phenotypes. Since TEs are repetitive sequences, their spread through the genome provides sites for recombination facilitating chromosomal changes such as deletions, inversions and reshuffling of the intertransposon segment (CHENAIS et al., 2012; CIVAN; SVEC; HAUPTVOGEL, 2011). This phenomenon occurs in different species of invertebrates and vertebrates animals and plants (BOURQUE, 2009; MILLER; CAPY, 2004; PARISOD et al., 2010).

Beyond the “macroscopic” effect of TEs on genome size and chromosome rearrangement, TE insertion may have a more subtle effect by affecting single gene expression. TEs can regulate gene expression in several different ways and numerous examples of such gene expression alteration have been found in all living organisms (CHENAIS et al., 2012).

Transposition into the coding regions or promoter region of genes (gene disruption) is perhaps the most evident effect of TE insertion. However, those transpositional events that passed through the sieve of selection can induce other genetic changes. These changes include: affecting expression levels of adjacent genes (by functioning as enhancers or promoters); altering splicing (TE exonization); deleterious premature termination of the peptide sequence (through modifications of the polyadenylation patterns) and generating antisense reads of a gene when inserted on the opposite DNA strand (triggering RNA interference silencing mechanisms). Also, down-regulating genes by siRNA-guided DNA methylation, which is primarily directed to TEs, but can affect nearby genes too. Lastly, whole-genome differences in TE-siRNA interactions have such dramatic effects on gene expression patterns that they might contribute to speciation (CHENAIS et al., 2012; CIVAN; SVEC; HAUPTVOGEL, 2011; FESCHOTTE, 2008; GIRARD; FREELING, 2009;



SLOTKIN; MARTIENSSEN, 2007). Figure 7 depicts some genetic changes discussed.

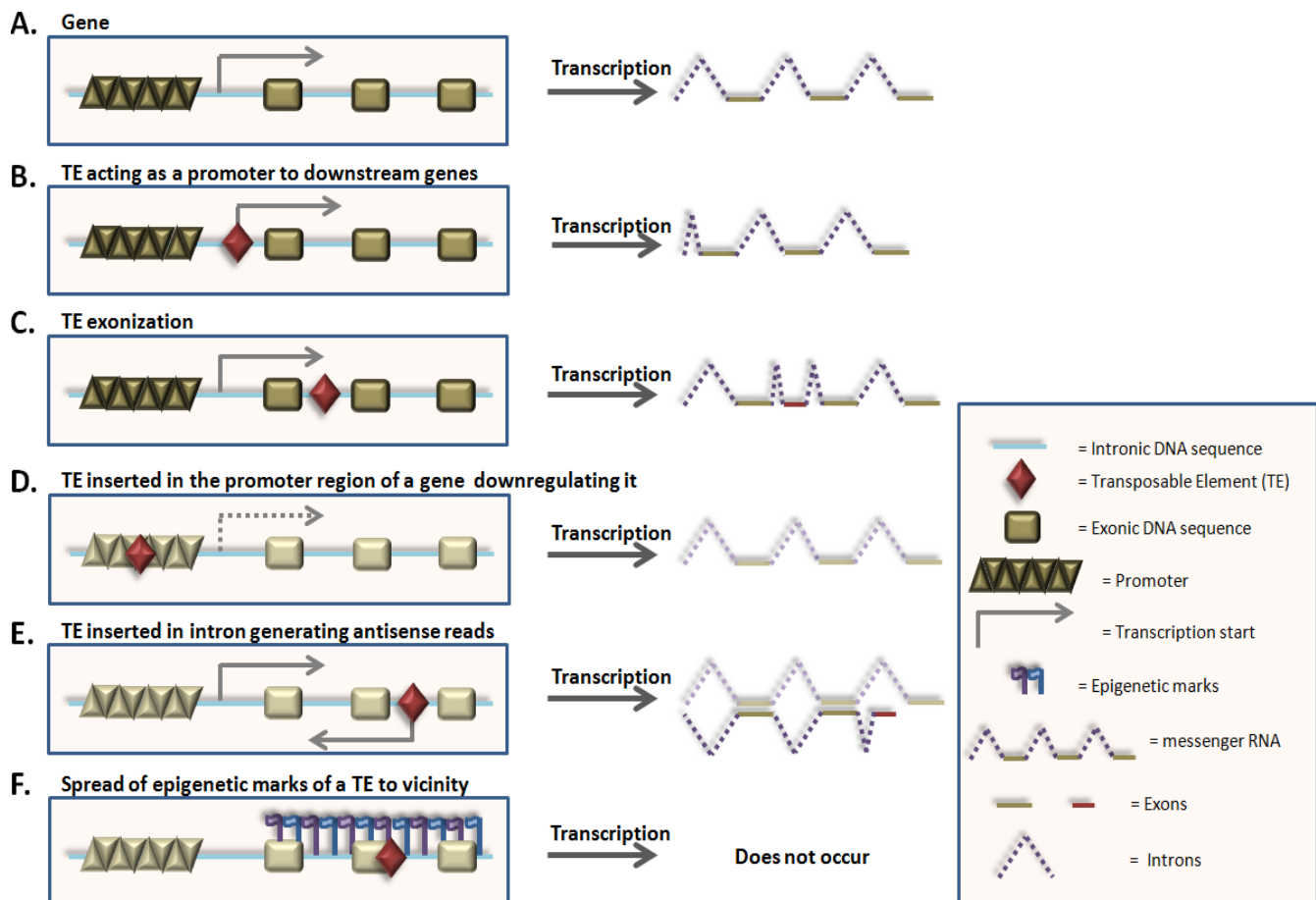


Figure 7 - The influence of transposable elements in gene expression. A. A gene to the left and its resulting transcript to the right (full line represent exons, interrupted lines represent introns). B. A newly inserted TE acts as promoter to a downstream gene, but this insertion did not imply in a modification of the exons present in the gene transcript. C. Newly inserted TE in the coding sequence is exonized as an alternative exon. D. When a TE is inserted in a promoter region of the gene, its transcription might be interrupted if the existing cis-regulatory element is disrupted. E. TE can also be inserted within an intron, drive antisense transcription and potentially interfere with the gene sense transcription. F. TEs may serve as a nucleation center for the formation of heterochromatin potentially silencing the transcription of an adjacent gene (Modified from FESCHOTTE, 2008)

The C-value enigma arose from the perception that genome size does not necessarily correspond to the degree of complexity of the organisms. The complexity of humans when compared to nematodes such as *Caenorhabditis elegans* (which has approximately 19,000 genes) led to unrealistic estimative of our protein coding gene content before the human genome was sequenced. It was expected that the human genome would comprise from 60,000 to 150,000 genes. However, once the latest numbers were published revealing our mere 22,333 genes (which corresponds

to only 2% of protein coding sequences in the genome), the C-value enigma became very obvious (PERTEA; SALZBERG, 2010). See Figure 8 for more species total number of genes.

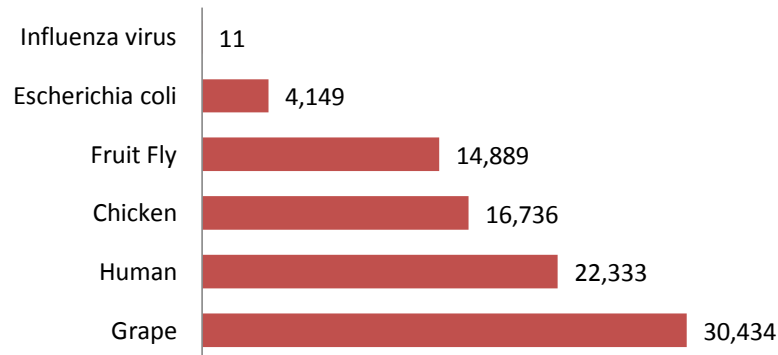


Figure 8 - Gene count in a variety of species. Viruses, the simplest living entities, have only about 11 genes, but are very well adapted to their environments. *Escherichia coli* bacteria have about four thousand genes, whereas multicellular plants and animals have three to eight times more genes. Thus the number of genes in a species bears little relation to its size or to intuitive measures of complexity (Adapted from PERTEA; SALZBERG, 2010)

How to explain that the complexity of living beings is not directly related to total gene number? Several factors account for this discrepancy, but most importantly one: insertion of TEs into exons or introns can in turn result in new patterns of pre-mRNA processing (alternative splicing and premature polyadenylation), enabling the generation of multiple mRNA isoforms from a single gene. Alternative splicing can produce unstable mRNA isoforms (which can be degraded), or produce alternate functional mRNAs that can be translated into proteins which in turn can perform new functions. It was shown that more than 95% of human multi-exonic genes are alternatively spliced (COWLEY; OAKLEY, 2013; PAN et al., 2008) and 61% in *Arabidopsis thaliana* (MARQUEZ et al., 2012), while it is only 25% in *C. elegans*. TEs have been related to the introduction of novel splice sites in human (COWLEY; OAKLEY, 2013) and in plant genes (VARAGONA; PURUGGANAN; WESSLER, 1992).



### 1.1.3 Gene regulatory networks wired by transposable elements

The control and coordination of eukaryotic gene expression rely on transcriptional and post-transcriptional regulatory networks. The co-option (or exaptation) of TEs has been described to serve the purpose of gene regulation and the orchestration of a number of vital processes. TEs can spread regulatory motifs throughout the genome. Jordan et al. (2003) reported that nearly 25% of human promoters contain TE-derived sequences, including *cis*-regulatory sequences. It was also demonstrated that many promoters in the human genome and polyadenylation signals in mouse genes derive from primate and rodent-specific TEs respectively.

The binding of transcription factors to their recognition motifs in eukaryotic gene promoters can activate the expression of genes dispersed in the genome in a co-ordinate fashion. These linked genes can be considered part of a same transcriptional network. TE (a transposon or a retrotransposon) wiring of transcriptional networks can occur through the following model: TE can be inserted in the 5' flanking sequence of a gene, placing its element promoters near the TATA box of the downstream gene. If the TE has a motif recognized by a DNA Binding Protein (DBP), and this DBP is a transcription factor, then genes in the vicinity and the TE can be co-transcribed. Thus, two genes that had no connection before will be activated coordinately, participating of the same transcriptional regulatory network as a result of the TE spread. It could happen, for instance, that if the TE is responsive to an external signal, the nearby gene becomes responsive too (such as shown in Figure 9).

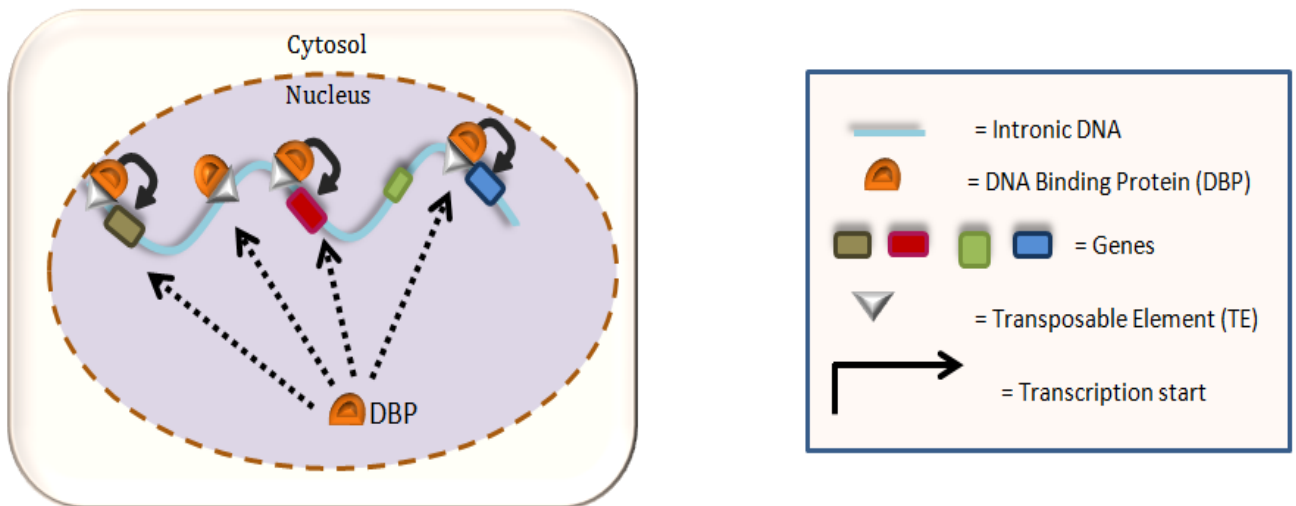


Figure 9 – Transcriptional regulatory network wired by a transposable element-derived *cis*-elements. TEs are shown as white triangles interspersed with genes (color boxes) in the genome. If the TE has a binding site for a DNA binding protein (DBP) (orange semi-circles), this binding site can get dispersed throughout the genome as part of the TE. If the DBP is a transcription factor, then when it binds to a TE adjacent to a gene it may influence the expression of that gene (the arrow points at the gene). Thus multiple genes are brought simultaneously under the control of the transcription factor through their association with different copies of the same TE family  
(Modified from FESCHOTTE, 2008)

A TE (transposon or retrotransposon) can also wire a post-transcriptional network: if a certain TE that is spread through the genome is co-transcribed along with its nearby genes, there will be the generation of chimerical mRNAs containing at least a TE part. If the TE has a binding domain for a RNA-binding protein (RBP), then the mRNAs containing the TE portion will undergo the same post-transcriptional pathway of gene regulation. This way, for example, if the TE is target for posttranscriptional gene silencing, then the nearby gene will also be a target (Figure 10).

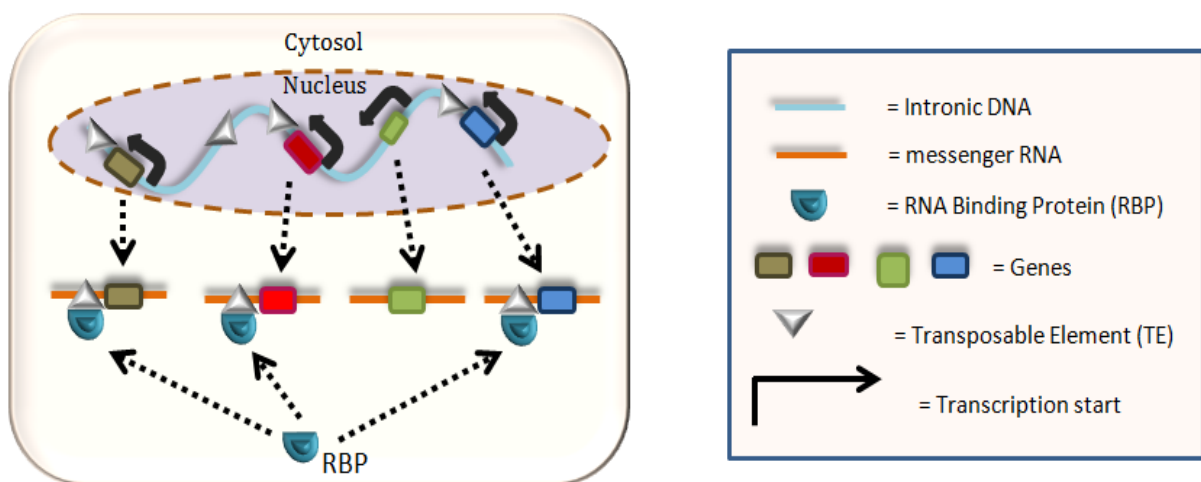


Figure 10 - Transposable elements regulation can be extended to nearby genes. TEs are shown as white triangles interspersed with genes (color boxes) in the genome. Several TE copies positioned downstream of genes can be co-transcribed along with their neighboring gene, resulting in the production of different mRNAs containing similar TEs. If the TE contains a binding site for a RNA-binding protein (RBP), it may engage the different mRNAs in the same post-transcriptional pathway of gene regulation (Modified from FESCHOTTE, 2008)

## 1. 2 Plant small RNAs

### 1.2.1 Small RNAs

In the past two decades the idea that RNA molecules are just passive carriers of DNA information has been discredited. Small non-coding RNA molecules (also known as small RNAs) are now known to be important regulators of gene expression and genome integrity. Small RNA is a general term to designate all types of 18 to 40-nucleotide-long RNAs, they are present in all eukaryotes and regulate genes through RNA-triggered gene silencing (AXTELL, 2013).

The two major classes of plant small RNAs are short interfering RNAs (siRNAs) and microRNAs (miRNAs), and they impact several developmental and physiological processes by regulating gene expression (CHEN, 2012). Among the functions performed by small RNAs in plants are: (1) regulation of transposon activity; (2) pathogenic defense; (3) regulation of intrinsic pathways, such as development and the response to environmental stresses. These processes are generally called RNA interference (RNAi) (HICKS; LIU, 2013).

The activity of small RNAs relies on proteins like DICER (or DICER-LIKE “DCL”) and ARGONAUTE, as well as RNA-dependent RNA Polymerase (RDRP) in the case of some heterochromatic siRNAs. Despite the differences between the

distinct small RNA pathways, they are interconnected: the pathways compete and collaborate as they regulate genes and protect the genome from external and internal threats (GHILDIYAL; ZAMORE, 2009; XIE et al., 2004).

### 1.2.2 Short Interfering RNAs

In 1998, Fire and Mello established that double-stranded RNA (dsRNA) is a signal for gene-specific silencing of expression in *Caenorhabditis elegans*. After dsRNA was injected into the worm, the corresponding gene products were eliminated from both the somatic cells of the organism as well as in the F1 progeny (FIRE et al., 1998). Double-stranded RNA can be originated by aberrant transcription products, produced by the transcription of inverted-repeat TEs or be derived from viral sequences. Once the dsRNA is detected by the plant cell, it is processed into short interfering RNAs that can result in: in the cytoplasm (1) RNA interference of complementary messenger RNAs leading to their cleavage and degradation; (2) RNA interference of complementary messenger RNAs resulting in translation repression; in the nucleus (3) siRNAs can direct repressive epigenetic modifications to homologous regions of the genome (which can be gene promoters) (FERRER; VOINNET, 2009; GHILDIYAL; ZAMORE, 2009; MATZKE; MOSHER 2014; VAUCHERET, 2006). Processes one and two are referred as Post-Transcriptional Gene Silence (PTGS), and process three is known as Transcriptional Gene Silencing (TGS). Interestingly, silencing siRNAs can travel to different plant parts via plasmodesmata and the phloem (MOLNAR et al., 2010).

Post-transcriptional gene silence (PTGS) is started when a double-stranded RNA is recognized and cleaved by a member of the RNase-III-like Dicer family of enzymes into short interfering RNA (siRNA) duplexes of 20-25 base pairs. Dicers act as a molecular ruler that determines the length of the resulting siRNA. The accurate localization in the plant cell of the processing step by DICER hasn't been established, but apparently it can occur both in the nucleus and cytoplasm (PAPP et al., 2003). After the siRNA duplexes are generated they are sent to the cytoplasm where their loading into an appropriate ARGONAUTE (AGO) protein is mediated by heat-shock protein 90 (HSP90). One strand of the siRNA duplex will be selected and remain bound to AGO (the guide strand), while the other is degraded. The combination of the guide strand of RNA, the AGO and other proteins is called the RISC Complex

(RNA-induced silencing complex). The siRNA directs RISC to bind to specific messenger RNAs, and the targeting is precise because it is determined by perfect base pair complementarity. Once bound to the target, AGO catalyzes cleavage of the mRNA target or repress their translation (FERRER; VOINNET, 2009; GHILDIYAL; ZAMORE, 2009), as shown in Figure 11 A.

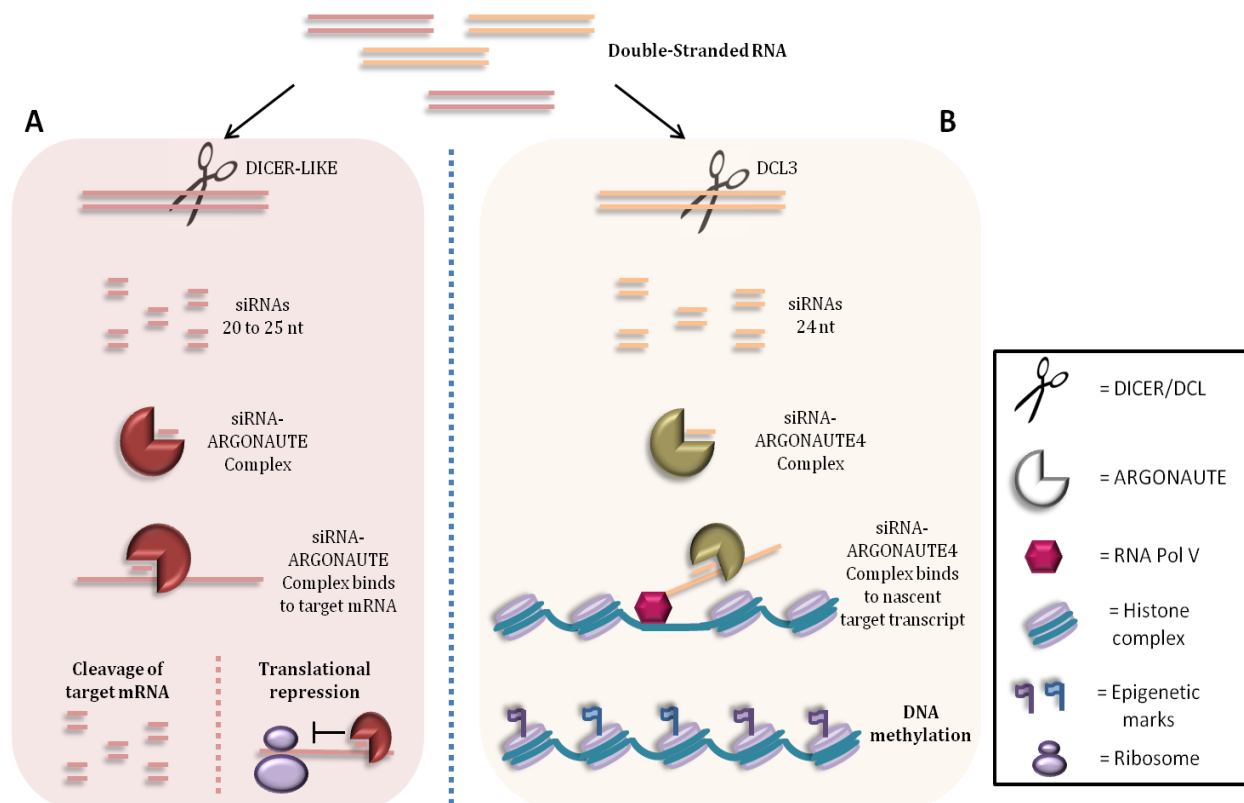


Figure 11 – Presence of double-stranded RNA in the plant cell triggers RNA interference. A. Post-transcriptional gene silencing can lead to the cleavage of the target messenger RNA or its translational repression. B. long siRNAs can transcriptionally silence target genes through DNA methylation

RNA-mediated transcriptional gene silencing (TGS) is required for vital cellular functions such as transposon silencing, genome stability, cell identity maintenance, defense against exogenous DNAs, stress responses and reproduction, as well as in interallelic and intercellular communication (MATZKE; MOSHER 2014). TGS in plants is called RNA-directed DNA methylation (RdDM) and requires two multisubunit plant-specific RNA polymerases (Pol IV and Pol V), 24-nucleotide small-interfering RNAs, DICER-Like 3, ARGONAUTE 4 (as well as a growing number of accessory proteins) and directs *de novo* cytosine methylation of complementary DNA sequences (HAAG; PIKAARD, 2011; ZHANG; HE; ZHU, 2013). Firstly, RNA Polymerase IV catalyzes transcription of single-stranded RNA transcripts from

heterochromatic repeated regions. Next RNA-Dependent RNA Polymerase (RDR2) converts the transcripts onto double-stranded RNAs, which are recognized and cleaved by DCL3 into 24-nucleotide siRNA that will be loaded into the AGO4 and thus form the RITS (RNA-induced transcriptional gene silencing) complex. Lastly, the RITS complex binds to Pol V nascent target transcripts, ultimately guiding the *de novo* DNA methylation as shown in Figure 11B (DOUET; TUTOIS; TOURMENTE, 2009).

### 1.2.3 Micro RNAs and Polycomb Group proteins (PcGs)

Plant microRNAs (miRNAs) are 21 to 24 nucleotide long endogenous small RNAs able to repress gene expression at the transcriptional and posttranscriptional level. Plants miRNAs control the expression of transcription factor genes, stress response proteins, and other proteins that impact development, growth, and physiology of plants (VAZQUEZ; LEGRAND; WINDELS, 2010).

Plant miRNAs are encoded by nuclear microRNA genes (miR genes). When a miR gene is transcribed by RNA Polymerase II, the long primary transcript generated contains the miRNA hairpins embedded in it. Due to the presence of those hairpins (double-stranded RNA), the primary transcript is processed by a Dicer-like enzyme (DCL1 or DCL3) in the nucleus. Next, in the cytoplasm miRNAs are loaded into Argonaute (AGO1 or AGO4) to form the RISC protein complex. After RISC assembly, plant miRNAs can repress transcriptional or post-transcriptionally gene expression. It can direct mRNA cleavage or even block its translation (post-transcriptional gene silencing). A second miRNA-mediated silencing pathway was described, in which miRNAs can promote methylation in *cis* (when their own loci, the miR genes, have their sequence methylated) or in *trans* (methylation of target gene sequences) (LI et al., 2013; ROGERS; CHEN, 2013; WU et al., 2010).

As it was mentioned before, both siRNA and miRNA pathways share the same cellular machinery to recognize double-stranded RNAs (DICERs), process (DICERs) and bind to them (ARGONAUTES) in order to interfere with gene expression. Both mechanisms are vital plant gene regulatory mechanisms.

Two groups of protein have also been described to play important roles in the transcriptional regulation of eukaryotic genes: Polycomb group (PcG) and Trithorax group (TrxG) proteins. These two display antagonistic modes of action. While PcG

promotes repression, TrxG promotes activity of gene expression, and they are required for stable and heritable maintenance of gene expression states. It was also shown that PcG preferentially bind to core promoter regions of genes, sometimes even miR gene promoters, unraveling an additional layer of epigenetic regulatory capacity (ENDERLE et al., 2011).

In Zardo et al. (2012) demonstrated that in humans the silencing of a gene can result from coordinated epigenetic events triggered by the promoter recognition and transcriptional targeting activity of Polycomb groups associated to a miRNA. This means that miRNAs can complex with Polycomb group proteins and act directly at the promoter region of target genes leading to their silencing (using the RNAi machinery) without requesting prior transcription of those genes to trigger different PTGS or TGS pathways discussed earlier. Such interaction between miRNAs and Polycomb proteins has not been described in plants so far.

Polycomb group proteins can form at least two multiprotein complexes called Polycomb Repressive Complex 1 (PRC1) and Polycomb Repressive Complex 2 (PRC2). In animals PRC2 is the site where Polycomb group proteins interact with small RNAs (DAVIDOVICH et al., 2013; HE; HUANG; XU, 2013). PcG proteins have been described as repressor of many plant genes involved in development through histone modification. PRC1 is absent in plants, however, PRC2 is present and harbors the histone methyltransferase that confers modifications to histone tails (specifically histone 3 lysine 27 trimethylation, H3K27me3) (LUCIA et al., 2008).

Polycomb silencing was found to be the mechanistic basis of vernalization (which consists in the acceleration of flowering induced by prolonged exposure to low temperatures). The *VERNALIZATION INSENSITIVE 3* (VIN3) is the most upstream gene in the vernalization pathway, and it is induced by cold temperatures. This gene, a component of the VEL gene family, codes for a PHD finger domain protein that is incorporated into the Polycomb Response Complex 2 (PRC2), which mediates the subsequent promotion of flowering in response to prolonged cold (FINNEGAN et al., 2011). Thus, polycomb proteins have important roles in the epigenetic gene regulation of both plants and animal, and the mechanistic component of this regulation is slowly being understood.

### 1.3 Tnt1

#### 1.3.1 Tnt1 discovery, structure and life cycle

Tnt1, the first retrotransposon discovered in *Solanaceae* isolated from tobacco (*Nicotiana tabacum*), is one of the best characterized plant retrotransposons. Tnt1 isolation was performed after its transposition in the nitrate reductase gene of tobacco (GRANDBASTIEN; SPIELMANN; CABOCHE, 1989).

Tnt1 superfamily is present in high copy number in the genus *Nicotiana*. *Nicotiana tabacum* has approximately 673 insertions of the retrotransposon, which is also present in its ancestor species: *Nicotiana sylvestris* has approximately 579 insertions and *Nicotiana tomentosiformis* has 383 insertions of Tnt1 (GRANDBASTIEN et al., 2005). The superfamily is composed of three subfamilies: Tnt1A, Tnt1B and Tnt1C (VERNHETTES; GRANDBASTIEN; CASACUBERTA et al., 1998), distinct among themselves in their promoter region sequence (U3) (CASACUBERTA; GRANDBASTIEN, 1993; CASACUBERTA, VERNHETTES, GRANDBASTIEN, 1995; POUTEAU et al., 1991; VERNHETTES; GRANDBASTIEN; CASACUBERTA et al., 1998). Seventy-five percent of Tnt1 insertions in *Nicotiana tabacum* correspond to Tnt1A, while Tnt1B and Tnt1C are majority among Tnt1 insertions in *Nicotiana plumbaginifolia* (CASACUBERTA, VERNHETTES, GRANDBASTIEN, 1995; VERNHETTES; GRANDBASTIEN; CASACUBERTA, 1997; VERNHETTES; GRANDBASTIEN; CASACUBERTA, 1998).

Tn1 (accession number X13777.1) is 5334 nucleotides long and its coding region contains two flanking Long Terminal Repeats (LTRs) of 610 nucleotides each (GRANDBASTIEN; SPIELMANN, CABOCHE, 1989). Tnt1 is a Ty1-copia retrotransposon and its structure in between the Target Site Duplications (TSD) is depicted in Figure 12. The LTRs do not code for protein, but contain the promoter region U3 ("Unique 3' RNA"), the R region ("Repeated RNA") and the terminator U5 ("Unique 5' RNA"). Between the LTRs and coding region are PBS (Primer Binding Site) and PPT (Polypurine Tract), which play a key role in the reverse transcription of the retrotransposon. The coding region is composed of major genes called *gag* (which codes for the Gag capsid protein) and *pol* (which codes for a polyprotein that contains a protease PR, integrase INT, reverse transcriptase RT and RNase H RNH).



The proteins they code for are sufficient for the successful replication of Tnt1 (KUMAR; BENNETZEN, 1999).

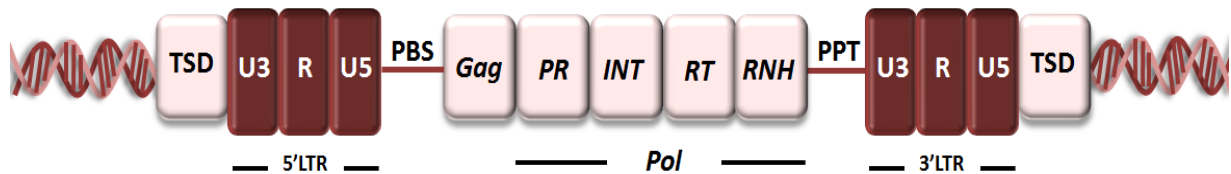


Figure 12 - Tnt1 retrotransposon structure. TSD: Target Site Duplication; U3: Unique 3' RNA; R: Repeated RNA; U5: Unique 5'RNA; LTR: Long Terminal Repeat; PBS: Primer Binding Site; *Gag*: Gag capsid gene; *PR*: protease gene; *INT*: integrase gene; *RT*: reverse transcriptase gene; *RNH*: RNase H gene; *Pol*: polyprotein gene; PPT: Polypurine Tract

Tnt1 messenger RNAs are transcribed as single molecules; they start at the 5' LTR R region and stop at the other R region at the 3' LTR. As a result of that, Tnt1 messenger RNAs contain the 3' U3 promoter region that can lead to read-through transcription of downstream gene sequences. Transcription produces RNAs lacking complete LTRs at either end, because the promoter and terminator are within the 5' and 3' LTRs respectively. While this does not interfere with translation, both LTRs must be restored to produce a Tnt1 full length cDNA that will be integrationally and replicationally competent. Polyadenylated and non-polyadenylated messenger RNAs are produced, and these transcripts must serve both translation and reverse transcription (CHANG et al., 2013; SCHULMAN, 2013).

After Tnt1 is transcribed by the RNA Polymerase II, the transcripts are directed to the cytoplasm. Proteins encoded by *gag*, *pol* and *int* are translated as a polyprotein that is cleaved into functional peptides by a protease (PR) (encoded by *pol*). Translated and processed Gag capsid proteins binds and packages two Tnt1 messenger RNAs as well as INT and RT-RNaseH into a virus-like particle (VLP). A complex reverse transcription process mediated by the reverse transcriptase and RNase H proteins is responsible for the completion of the missing LTR fragments in the messenger RNA and the synthesis of a double-stranded cDNA copy of Tnt1 inside the VLP (SCHULMAN, 2013). In order for the new copy of Tnt1 to be integrated in a new chromosome site, it needs to be directed back to the nucleus. Nuclear entry is moderated by the Nuclear Localization Signal (NLS), but for plant retrotransposons little information is available about which proteins present NLS. Yeast retrotransposons bare the NLS in both Gag and Int proteins (KIM et al., 2005). The Int protein allows the retrotransposon cDNA to be integrated in a new site of the

genome through trimming and joining of the retrotransposon DNA stretches to the target region (SCHULMAN, 2013).

Under certain conditions, retrovirus-derived Gag protein was detected not only in the cytoplasm, but also in the nuclei (STAKE et al., 2013). It was also demonstrated that a fraction of Gag proteins of *Drosophila* retrotransposons are moved into the nuclei (RASHKOVA; KARAN; PARDUE, 2002). Such find hasn't been described in plant retrotransposons yet. Considering the closeness between retrovirus and retrotransposons (that are only different in the fact that retrotransposons don't code for the Envelope protein and therefore are unable to exit the cell), and the example demonstrated in *Drosophila*, it is possible that Gag capsid protein derived from plant retrotransposon could be internalized in the plant cell nuclei too.

### **1.3.2 Tnt1 is transcriptionally active and responsive to stress signals**

Tnt1 is a retrotransposon transcriptionally active in somatic tissues: under normal growth conditions Tnt1 messenger RNA is detected in tobacco mesophyll protoplasts and in basal levels in the roots of wild type tobacco plants (POUTEAU et al., 1991). Likewise other transposable elements have been described transcriptionally active, such as the members of *Copia/Ty1* and *Gypsy/Ty3* in sugarcane (ROSSI; ARAUJO; VAN SLUYS, 2001). However, it is important to emphasize that retrotransposon transcription does not necessarily imply in retrotransposition and consequent in the genome, which is a tightly controlled event.

Plants, like humans, can be stressed by challenging living conditions. Stresses can originate from the surrounding environment (called abiotic stresses) or from living organisms that can cause disease or damage (called biotic stresses). Tnt1 is transcriptionally active in basal levels in somatic tissues. However, under different stress stimuli its expression is even further induced. Table 1 shows different transposable elements likewise induced under stress and reveals that the majority of the responsive elements are *Copia* retrotransposons (the same kind as Tnt1) (MANSOUR, 2007; SCHULMAN, 2013). Retrotransposon activation by stresses seems to be a frequent feature of eukaryotic genomes. This is in agreement with McClintock's original model which postulates that transposable elements are involved in genome restructuring in response to environmental challenges, referred then by

McClintock as “genomic shocks” (GRANDBASTIEN et al., 1994; MCCLINTOCK, 1984).

Table 1 – Description of some stress activated transposable elements. (Modified from MANSOUR, 2007.)

Stress type	Elements	Element type
Tissue Culture	Tto1, Tos17	<i>Copia</i>
Cell-wall hydrolases	Tnt1	<i>Copia</i>
Chilling	Tos17	<i>Copia</i>
Adenine starvation	Ty1	<i>Copia</i>
Wounding	TLC1	<i>Copia</i>
Protoplast preparation	TLC1	<i>Copia</i>
High salt concentrations	TLC1	<i>Copia</i>
Interspecific hybridization	Wis 2-1A	<i>Copia</i>
Adaptation to a moisture gradient	Bare-1	<i>Copia</i>
Microbial factors	Tnt1	<i>Copia</i>
Mechanical damage	Tnt1	<i>Copia</i>
<i>In vitro</i> regeneration	Tnt1	<i>Copia</i>
Viral infection	Tos17	<i>Copia</i>
Cytosine demethylation	Tos17	<i>Copia</i>
UV light	Reme1	<i>Copia</i>
Resistance to bacterial blight and plant development	Tos17	<i>Copia</i>
Fungal infection	Romani PP and Erika LTR	<i>Gypsy</i>
Heat shock	MAGGY	<i>Gypsy-like</i>
Viral infection	HERV-W	LTR-retrotransposon
Trichothecene mycotoxin deoxynivalenol (DON)	Retrotransposon-like homologue	NA
Cold stress	LTR-Retrotransposon	NA
Environmental hydrocarbon	LINE-1 (L1)	Non-LTR
Common chemotherapeutic drugs and gamma-radiation	SINES	Retroposon
Hybridization and generation of double haploids	mPing	Transposon MITEs

The responsiveness of those transposable elements is due to the presence of *cis*-elements in their promoter region. Upstream of the TATA box of Tnt1 (at the U3 LTR region) there are two different box sequences: BI and BII, composed of different activator elements which play a key role in the induction of Tnt1 transcription

(CASACUBERTA; GRANDBASTIEN, 1993). The localization of these boxes in Tnt1A U3 LTR is depicted in the Figure 13.

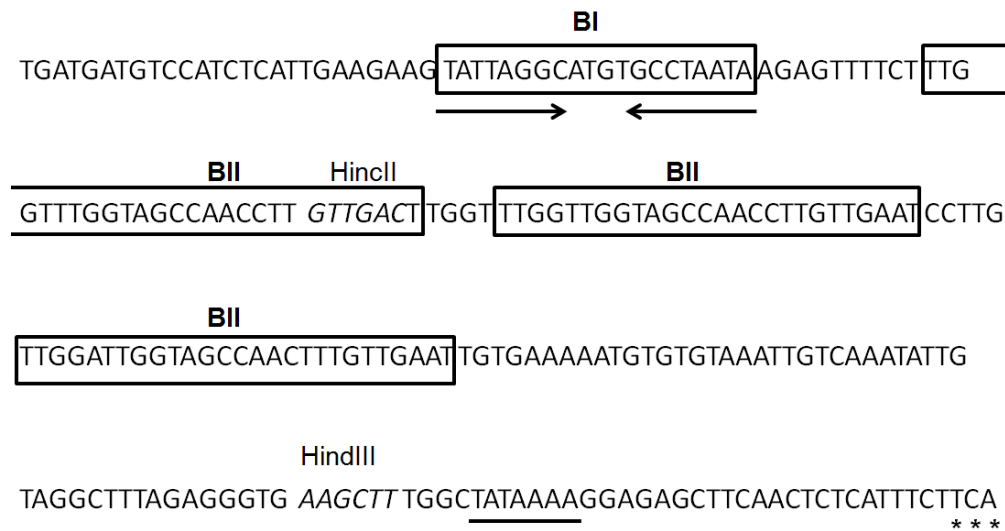


Figure 13 - Nucleotide sequence of the Tnt1A LTR U3 region. The putative TATA box is underlined and the major transcriptional start site is shown by asterisks. Sequences corresponding to the BI and BII elements are boxed and the dyad symmetry of the BI sequence is shown by arrows. Restriction sites (HincII and HindIII) defining the proximal and the distal regions of the upstream LTR sequences are shown in italics. The putative TATA box is underlined and the asterisks at the end of the sequence indicate the major start transcriptional site. (Adapted from CASACUBERTA; GRANDBASTIEN, 1993.)

In order to prepare mesophyll protoplasts, tobacco leaves are treated with hydrolases which loosen the leaf tissue and thus release cell-wall free protoplasts. Pouteau and collaborators (1991) described that upon the presence of cell wall hydrolases, tobacco leaf-derived protoplasts presented higher concentration of Tnt1 mRNA. Casacuberta and collaborators (1993) demonstrated that this Tnt1 induction during protoplast preparation is due to the interaction between the BII sequences and tobacco nuclear proteins. The BII region contains a short (31 bp) tandemly repeated sequence (three or four repeats) that interacts with protoplast-specific nuclear factors, and is a good candidate to explain the induction of Tnt1 expression associated with the presence of cell wall hydrolases (GRANDBASTIEN et al., 1997).

Vernhettes and collaborators (1998) showed that Tnt1 subfamilies (Tnt1A, B and C) possess divergent regulatory box sequences in their U3 promoter region. The rest of Tnt1 sequence, however, is conserved among the subfamilies. Reverse transcriptase coding sequences, for instance, are almost identical between the three subfamilies (100% identity between Tnt1A and Tnt1B and 88% identity between

these and Tnt1C). However, box regions of the three subfamilies are highly divergent (there is less than 35% identity between any pair of subfamilies). This divergence of the transcriptional regulatory region suggests that Tnt1 subfamilies are differentially induced under environmental influences.

To date, several stresses have been reported to induce Tnt1 transcription: presence of cryptogein, capsicein and culture supernatants of the bacterium *Erwinia chrysanthemi* in tobacco cell cultures (the first two are elicitors produced respectively by *Phytophthora cryptogea* and *Phytophthora capsici*, which correlates with the biological necrotic activity of both elicitors) (POUTEAU; GRANDBASTIEN; BOCCARA, 1994); mechanical injury to tobacco leaves (MHIRI et al., 1997); salicylic acid, methyl jasmonate and 2,4-dichlorophenoxyacetic acid (2,4-D, a synthetic auxin) (BEGUIRISTAIN et al., 2001). Interestingly, Tnt1 induction is subfamily-specific: cryptogein and salicylic acid (salicylate) induce Tnt1A in leaves (VERNHETTES; GRANDBASTIEN; CASACUBERTA, 1997); methyl jasmonate induces Tnt1B and salicylic acid and 2,4-D induce Tnt1C (BEGUIRISTAIN et al., 2001).

This leads to the conclusion that Tnt1 expression is activated by abiotic stresses and by compounds involved in early stress signal transduction and that the expression of the Tnt1 promoter is a sensitive indicator of the plant defense response (MHIRI et al., 1997). Tnt1 induction correlates with necrotic symptoms that accompany the Hypersensitive Reaction when this defense mechanism is triggered. The hypersensitive reaction is characterized by the death of individual plant cells which come into contact with pathogenic organisms (creating a “barrier” to prevent the spread of the contaminant), and is generally associated with disease resistance of the whole plant to the pathogen. Cellular death is not necessary for Tnt1 activation, it actually precedes necrotic symptoms in most cases. Tnt1 expression is thus most probably linked to earlier events that result from elicitation or pathogen attacks, which can also lead to cellular death (GRANDBASTIEN et al., 1997).

### 1.3.3 Retrotransposon control

TEs have contributed widely to genome evolution. However, genetic mutations such as insertions and rearrangements caused by TE mobilization are generally regarded to be deleterious to genome integrity. Maintaining the integrity of the genomes depends on mechanisms to regulate TEs, avoiding eventual damage that

can result from their activity. To suppress the eventual harmful activity of TEs, plant cell has mechanisms to control its activity transcriptional and translationally, and recent progress in the study of plants and other organisms has revealed that both TGS and PTGS might be responsible for the silencing of TEs (NAKAYASHI et al., 2011).

To initiate silencing, it is first necessary for the cell machinery to recognize transcripts originated from TEs, which are basically no different from other gene transcripts. The mechanism by which the plant cell can differentiate a retrotransposon-derived transcript from other gene-derived transcripts is the detection of double-stranded RNA molecules that can be produced by retrotransposons. Such double-stranded RNAs can originate from similar copies of the same retrotransposon at different positions which will likely produce both sense and antisense transcripts because of read-through transcription from neighboring genes (HERNANDEZ-PINZON; DE JESUS, 2009, NAKAYASHI et al., 2011).

Once detected, double-stranded RNA triggers the post-transcriptional gene silencing (PTGS) of the retrotransposon that originated it. As described previously, this can lead to the cleavage of target messenger RNAs or translation inhibition (LISCH, 2009). After PTGS is established, it can lead to transcriptional gene silencing (TGS): the population of small RNAs targeting active transposons is amplified through an RNA-Dependent RNA Polymerase mechanism. The double stranded small RNAs will then recruit a specific ARGONAUTE complex and result in epigenetic gene silencing effects that suppress transcription at the targeted locus. This mechanism typically involves histone modifications such as histone deacetylation, methylation of histone H3 lysine 9, and DNA methylation. In plants, cytosine methylation is controlled by specific DNA methyltransferases and daughter cells can inherit the epigenetic marks established (GIRARD; HANNON, 2007).

Even though retrotransposons are targets of silencing in plants, it has been described that Tnt1 can escape this control and be expressed in tobacco (HERNANDEZ-PINZON et al ., 2012). The mechanism through which Tnt1 and other retrotransposon remain active despite the silencing directed towards them is still unclear.

## 1.4 Ethylene

### 1.4.1 Ethylene and plant defense responses

Ethylene (C<sub>2</sub>H<sub>4</sub>) is a gaseous phytohormone that affects a variety of plant processes throughout its development including: seed germination, apical hook formation, growth, organ senescence, fruit ripening, abscission, gravitropism and plant defense responses. Its physiological effects are detectable at ambient levels as low as 0.1 uL/L (VANDENBUSSCHE et al., 2012). Lag times for ethylene responses in tomato range from 10 to 15 min (inhibition of seedling growth) to hours (enhancement of enzyme activities), and even to days (promotion of leaf senescence). Analysis of mRNA accumulation in tomato fruits indicated a 30-min lag from the time of ethylene application to the time of measurable increases in mRNA abundance (BLEECKER; KENDE, 2000).

During plant-pathogen interaction, a rapid increase in ethylene biosynthesis is one of the earliest detectable events. The ethylene synthesis is increased as a result of an increase in the activity of the rate-limiting ethylene biosynthetic enzyme, ACC synthase (S-adenosyl-L-methionine methylthioadenosine-lyase, EC 4.4.1.14). Ethylene is a plant stress hormone that acts as a signal for plants to activate defense mechanisms against invading pathogens such as bacteria, fungi and viruses (ECKER; DAVIS, 1987). Interestingly, ethylene has also been described to induce the expression of TLC1.1 retrotransposon in *Solanum chilense*. The Open Reading Frame (ORF) of this retrotransposon and Tnt1 shares 87% similarity in sequence (TAPIA et al., 2005).

Pathogen attacks and other external stimuli such as wounding and chemicals can trigger plant defense responses. These defense responses are characterized by the derepression of several genes, leading to the production of different compounds that contribute to restricting pathogen development such as phytoalexins (plant antibiotics) and Pathogenesis-Related (PR) proteins (GRANDBASTIEN et al., 1994). In the case of an attack by bacteria or fungal pathogens the initial response is called hypersensitive response (HR), which is the formation of necrotic lesions at the site of infection. A long-lasting secondary response called systemic acquired resistance (SAR) develops throughout the plant in infected and uninfected tissues. Plant tissues infected by pathogens or wounded also have elevated levels of ethylene. At the

molecular level, plants mount chemical and physical barriers through the production of antimicrobial agents and structural proteins such as callose to avoid the spread of the infection (JOHNSON; ECKER, 1998).

#### **1.4.2 Ethylene synthesis, perception and responses**

The diversity of the roles performed by ethylene and the specificity of its action suggest the complexity in the regulation of ethylene synthesis. Methionine is the precursor of ethylene, and nearly 80% of its total is converted to S-adenosylmethionine (SAM) by SAM synthetase at the expense of ATP utilization. SAM is used by the enzyme 1-aminocyclopropane-1-carboxylic acid (ACC) synthase (ACS) to form ACC, ethylene's direct precursor. This step is rate-limiting for ethylene biosynthesis. ACC is then oxidized to ethylene, CO<sub>2</sub> and cyanide (later detoxified) by ACC oxidase (ACO). Once ethylene gas is produced, it freely diffuses to surrounding cells and tissues. Without the possibility of an actively regulated transport mechanism, control of cells sensitivity to the hormone is critical for limiting the response to the targeted cells (ALONSO; ECKER, 2001; HUA; MEYEROWITZ, 1998; WANG; LI; ECKER, 2002; WANG et al., 2013).

Ethylene signaling starts with its perception, performed by a family of receptor proteins located at the Golgi and Endoplasmic Reticulum (ER) membranes. Two tobacco ethylene receptor genes are *NTHK1* and *NTHK2*. The receptors are negative regulators of the signaling pathway and the family members function redundantly. Upon ethylene binding, the receptors become inactive. They physically interact with CONSTITUTIVE TRIPLE RESPONSE 1 (CTR1), a negative regulator of ethylene signaling and with ETHYLENE INSENSITIVE 2 (EIN2), a positive regulator of ethylene signaling and an N-RAMP metal transporter-like protein anchored at the ER membrane. In the absence of ethylene, CTR1 is closely associated with the receptor complex, blocking signal transfer to EIN2. When ethylene is present and binds to the ethylene receptor, CTR1 changes its conformation allowing interaction of EIN2 with the kinase domain of the receptor. Next, the signal must be passed on to the nucleus, but the mechanisms involved in this step are still unclear. EIN3 and EIN3-like (EIL, or in the case of tobacco, TEIL) are transcription factors that induce the expression of ethylene response factor (*ERF*) genes (VANDENBUSSCHE et al., 2012). ERF proteins function as ethylene responsive element binding proteins



(EREBPs) that stimulate the transcription of target genes responsive to ethylene. The *cis*-acting GCC box (the core sequence of the ethylene-responsive element –ERE– in tobacco) is commonly found in the 5' upstream region of ethylene-responsive genes (FUJIMOTO et al., 2000; SHINSHI; USAMI; OHME-TAKAGI, 1995). Ethylene-Responsive Element Binding Proteins (EREBPs) have conserved DNA-binding domains that interact with GCC box acting as transcriptional activators or repressors of gene expression (OHME-TAKAGI; SHINSHI, 1995). The ethylene signaling cascade is represented in Figure 14.

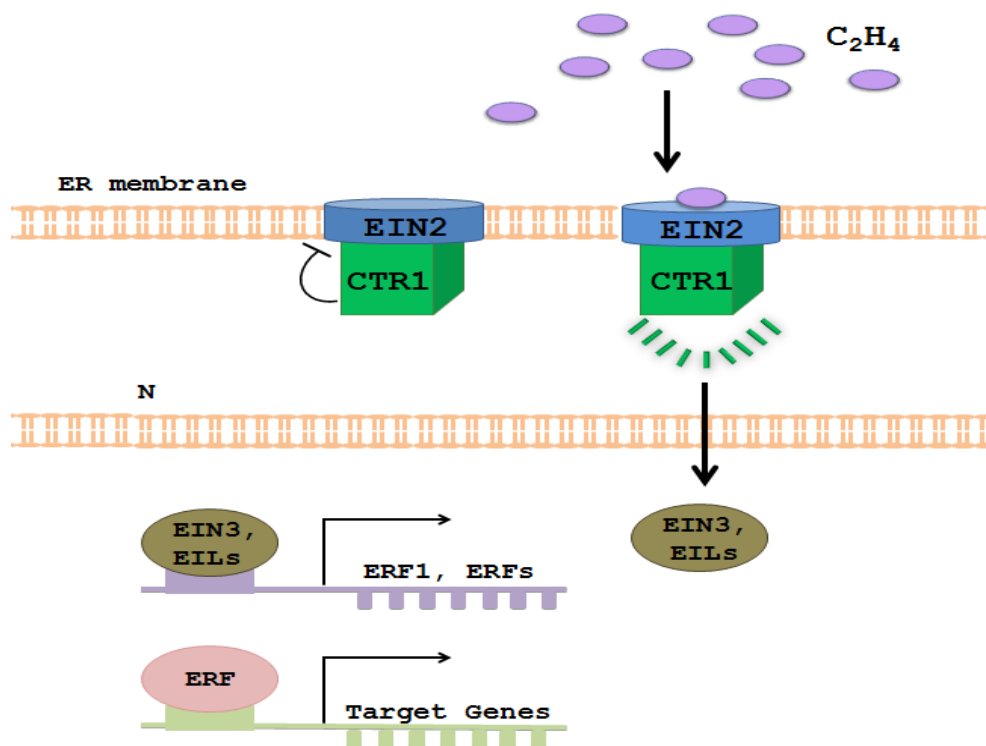


Figure 14 – Model of ethylene signaling pathway. To the left, in the absence of ethylene, CONSTITUTIVE TRIPLE RESPONSE 1 (CTR1) is tightly associated with the receptor complex, preventing signal transfer to ETHYLENE INSENSITIVE 2 (EIN2). To the right, in the presence of ethylene, the receptor/CTR1/EIN2 complex changes conformation, allowing EIN2 to transfer the signal to the downstream components belonging to the EIN3 transcription factor family. The EIN3 family of transcription factors binds promoters of ETHYLENE RESPONSE FACTORS (ERFs), thus stimulating their expression. ERFs further bind target promoters carrying an ethylene responsive element (ERE). ER = endoplasmic reticulum, N = nucleus (Modified from VANDENBUSSCHE et al., 2012)

Ethylene can exert positive or negative feedback regulation of its own biosynthesis, depending on tissue/organ or developmental stage. Typically, in vegetative tissues the rate of ethylene production is low and is inhibited by exogenous ethylene (enhanced hormone signaling often leads to a reduction in its biosynthesis, termed auto-inhibition or negative feedback) (CHANG et al., 2008).

### **1.4.3 Ethylene triple response**

Neljubov and collaborators (1901) described that when germinating in the dark, impeded pea seedlings presented a characteristic “triple response” as a result of ethylene accumulation: the seedlings present hypocotyl and root reduced elongation, apical hook exaggeration as well as hypocotyl swelling. This seedling phenotype was successfully used for the identification of both ethylene-insensitive and constitutive response mutants in mutagenized populations, which yielded a large amount of information about the key genes and proteins in ethylene responses and signaling pathway.

## **1.5 Objectives**

### **1.5.1 Motivation and previous results in our laboratory**

When taken into consideration the importance and the impact of transposable elements in genome structure, chromosome rearrangement and gene expression, study of TEs makes sense because it is a way to better understand genome itself, especially due to some genomes being composed 88% of TEs. The study of Tnt1 retrotransposon is relevant because of its particularities: it is a multicopy active retrotransposon that is up-regulated under very specific conditions. Tnt1 increased expression under environmental challenging situations is at the same time intriguing and unexpected, since TEs are thought to be selfish entities over which the cell has to keep a tight control. Therefore, it is relevant to gain knowledge on the interaction of the tobacco genome and its retrotransposon and try to understand why natural selection maintained some retrotransposons stress-inducible.

In 2006 the Nobel Prize was awarded to Andrew Z. Fire and Craig C. Mello for their discovery in 1998 of gene silencing by double-stranded RNA (RNA interference) (Fire et al., 1998). A genetic revolution had started, making possible the accurate gene silencing and thus fueling the reverse genetic field with a powerful tool. This is when our research group saw the opportunity of doing a functional study of Tnt1. With the possibility of silencing the retrotransposon, maybe its real biological role would be unraveled.

Bruno Karolski obtained in our lab during his masters (2007) transgenic tobacco lines expressing a hairpin construct (to induce RNA interference) of a Tnt1 reverse transcriptase portion. He described the phenotypes the plants displayed, ranging from cell death spots on leaves to severe developmental abnormalities.

### 1.5.2 General objective

The general objective of this PhD project was to study the retrotransposon Tnt1 in *Nicotiana tabacum* and try to unravel its putative biological role in tobacco (other than its self-propagation). The hypothesis was that Tnt1 is used as a genetic tool to structure regulatory networks involved in tobacco defense responses.

### 1.5.3 Specific objectives

(1) The macroscopic and microscopic phenotypic characterization of the transgenic lines generated by Bruno Karolski;

(2) Generation of more independent tobacco transgenic lines expressing the Tnt1 reverse transcriptase hairpin in order to corroborate the phenotypes observed. A collaboration with the State University of New Jersey Rutgers in Dr. Pal Maliga's lab was made to achieve this objective;

(3) Normalization and analysis of the RNA sequencing results of leaves of 45-day-old *in vitro* grown plants of transgenic lines, wildtype and a control with the empty vector (the sequencing was performed by Post-Doc Erika Maria de Jesus, and normalization and analysis of the data was done in collaboration with her);

(4) Study of putative new regulatory gene networks wired by Tnt1 revealed by the RNA-seq and to further exploration of them.

## References

ALONSO, J.M.; ECKER, J.R. The ethylene pathway: a paradigm for plant hormone signaling and interaction. **Science's Signal Transduction Knowledge Environment**, New York, v. 70, p. 1-11, Feb. 2001.

ALZOHAIY, A.M.; GYULAI, G.; JANSEN, R.K.; BAHIELDIN, A.. Transposable elements domesticated and neofunctionalized by eukaryotic genomes. **Plasmid**, Amsterdam, v. 69, p. 1-15, Jan. 2013.

AXTELL, M.J. Classification and comparison of small RNAs from plants. **Annual Review of Plant Biology**, Palo Alto, v. 64, p. 137-159, Nov. 2013.

BEGUIRISTAIN, T.; GRANDBASTIEN, M.A.; PUIGDOMÈNECH, P.; CASACUBERTA, J.M. Three Tnt1 subfamilies show different stress-associated patterns of expression in tobacco: consequences for retrotransposon control and evolution in plants. **Plant Physiology**, Rockville, v. 127, p. 212–221, Jan. 2001.

BENNETZEN, J.L. Transposable element contributions to plant gene and genome evolution. **Plant Molecular Biology**, New York, v. 42, p. 251-269, Feb. 2000.

BIÉMONT, C.; VIEIRA, C. Genetics: junk DNA as an evolutionary force. **Nature**, London, v. 443, n. 5, p. 521-524, Oct. 2006

BLEECKER, A.B.; KENDE, H. Ethylene: a gaseous signal molecule in plants. **Annual Review in Cell Developmental Biology**, Palo Alto, v. 16, p.1–18, Apr. 2000.

BOURQUE, G. Transposable elements in gene regulation and in the evolution of vertebrate genomes. **Current Opinion in Genetic Development**, Amsterdam, v. 19, n. 6, p. 607-612, Dec. 2009.

KAROLSKI, B. **Efeito da inativação do retroelemento Tnt1 na regeneração de plantas de fumo**. 2007. 81 p. Dissertação (Mestrado em Ciências Botânicas) – Faculdade de Ciências Biológicas, Universidade de São Paulo, São Paulo, 2007.

CASACUBERTA, J.M.; GRANDBASTIEN, M.A. Characterization of LTR sequences involved in the protoplast specific expression of the tobacco Tnt1 retrotransposon. **Nucleic Acids Research**, Oxford, v. 21, n. 9, p. 2087-2093, Nov. 1993.

CASACUBERTA J.M.; VERNHETTES, S.; GRANDBASTIEN, M.A. Sequence variability within the tobacco retrotransposon Tnt1 population. **EMBO Journal**, Heidelberg, v. 14, n. 11, p. 2670-2678, June 1995.

CHANG, S.; LU, L.; WANG, N.N.; CHARNG, Y. Negative feedback regulation of system-1 ethylene production by the tomato 1-aminocyclopropane-1-carboxylate synthase 6 gene promoter. **Plant Science**, Amsterdam, v. 175, p. 149–160, Jan. 2008.

CHANG, W.; JÄÄSKELÄINEN, M.; LI, S.; SCHULMAN, A.H. BARE Retrotransposons are translated and replicated via distinct RNA Pools. **PLOS One**, San Francisco, v. 8, n. 8, Apr. 2013. Disponível em: <<http://www.plosone.org>>. Accessed in November 1<sup>st</sup> 2013.

CHEN, X. Small RNAs in development: insights from plants. **Current Opinion in Genetic Development**, Amsterdam, v. 22, p. 361–367, Jan. 2012.

CHENAIS, B.; CARUSO, A.; HIARD, S.; CASSE, N. The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to stressful environments. **Gene**, Amsterdam, v. 509, p. 7-15, Aug. 2012.

CIVÁN, P.; SVEC, M.; HAUPTVOGEL, P. On the coevolution of transposable elements and plant genomes. **Journal of Botany Article**, Cairo, v. 2011, p. 1-9, Apr. 2011.

DAVIDOVICH, C.; ZHENG, L.; GOODRICH, K.J.; CECH, T.R. Promiscuous RNA binding by Polycomb repressive complex 2. **Nature Structural and Molecular Biology**, London, v. 20, n. 11, p. 1250-1261, Feb. 2013.

DEININGER, P.L.; BATZER, M.A. Mammalian retroelements. **Genome Research**, Cold Spring Harbor, v. 12, p. 1455-1465, Aug. 2012.

DOUET, J.; TUTOIS, S.; TOURMENTE, S. A Pol V-mediated silencing, independent of RNA-directed dna methylation, applies to 5S rDNA. **PLOS One**, San Francisco, v. 5, n. 10, Oct. 2009. Disponível em: <<http://www.plosone.org>>. Accessed in November 1st 2009.

ECKER, J.R.; DAVIS, R.W. Plant defense genes are regulated by ethylene. **Proceedings of the National Academy of Sciences of the USA**, Washington, v. 84, p. 5202-5206, Oct. 1987.

ENDERLE, D.; BEISEL, C.; STADLER, M.B.; GERSTUNG, M.; ATHRI, P.; PARO, R. Polycomb preferentially targets stalled promoters of coding and noncoding transcripts. **Genome Research**, Cold Spring Harbor, v. 21, p. 216-226, Feb. 2011.

FEDOROFF, N.V. Transposable elements, epigenetics and genome evolution. **Science**, New York, v. 338, n. 6108, p. 758-767, Nov. 2012.

FERRER, V.; VOINET, O. Roles of plant small RNAs in biotic stress responses. **Annual Review of Plant Biology**, Palo Alto, v. 60, p. 485-513, Apr. 2009.

FESCHOTTE, C. The contribution of transposable elements to the evolution of regulatory networks. **Nature Review in Genetics**, London, v. 9, n. 5, p. 397-405, Dec. 2008.

FESCHOTTE, C.; PRITHAM, E.J. DNA transposons and the evolution of eukaryotic genomes. **Annual Review of Genetics**, Palo Alto, v. 41, p. 331-68, Aug. 2007.

FESCHOTTE, C.; JIANG, N.; WESSLER, S.R. Plant transposable elements: where genetics meets genomics. **Nature Review in Genetics**, London, v. 3, n. 5, p. 329-341, Oct. 2002.

FINNEGAN, J.; BOND, D.M.; BUZAS, D.M.; GOODRICH, J.; HELLIWELL, C.A.; TAMADA, Y.; YUN, J.Y.; AMASINO, R.M.; DENNIS, E.S. Polycomb proteins regulate the quantitative induction of vernalization insensitive 3 in response to low temperatures. **The Plant Journal**, Rockville, v. 65, n. 3, p. 382-391, Feb. 2011.

FIRE, A.; XU, S.; MONTGOMERY, M.K.; KOSTAS, S.A.; DRIVER, S.E.; MELLO, C.C. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. **Nature**, London, v. 392, n. 6669, p. 806-811, Feb. 1999.

FUJIMOTO, S.Y.; OHTA, M.; USUI A.; SHINSHI, H.; OHME-TAKAGI, M. *arabidopsis* ethylene-responsive element binding factors act as transcriptional activators or repressors of gcc box-mediated gene expression. **The Plant Cell**, Rockville, v. 12, p. 393–404, Mar. 2000.

GHILDIYAL, M.; ZAMORE, P.D. Small silencing RNAs: an expanding universe. **Nature Review**, London, v. 10, p. 94-108, June 2009.

GIRARD, L.; FREELING, M. Regulatory changes as a consequence of transposon insertion. **Developmental Genetics**, New York, v. 25, p. 291–296, Apr. 1999.

GIRARD, A.; HANNON, G.J. Conserved themes in small-RNA mediated transposon control. **Trends in Cell Biology**, Cambridge, v. 18, n. 3, p. 136-148, Dec. 2007.

GRANDBASTIEN, M.A.; LUCAS, H.; MOREL, J.B.; MHIRI, C. The expression of the tobacco Tnt1 retrotransposon is linked to plant defense responses. **Genetica**, New York, v. 100, p. 241-252, Mar. 1997.

GRANDBASTIEN, M.A.; SPIELMANN, A.; CABOCHE, M. Tnt1, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. **Nature**, London, v. 337, p. 376-380, Jan. 1989.

GRANDBASTIEN, M.A.; AUDEON, C.; CASACUBERTA, J.M.; GRAPPIN, P.; LUCAS, H.; MOREAU, C.; POUTEAU, S. Functional analysis of the tobacco Tnt1 retrotransposon. **Genetica**, New York, v. 93, p. 181-189, Jan. 1994.

GRANDBASTIEN, M.A.; AUDEON, C.; BONNIVARD, E.; CASACUBERTA, J.M.; CHALHOUB, B.; COSTA, A.P.; LE, Q.H.; MELAYAH, D.; PETIT, M.; PONCET, C.; TAM, S.M.; VAN SLUYS, M. Stress activation and genomic impact of Tnt1 retrotransposon in *Solanaceae*. **Cytogenetics Genome Research**, Allschwilerstrasse, v. 110, p. 229-241, Feb. 2005.

HAAG, J.R.; PIKAARD, C. Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. **Nature Review**, London, v. 12, p. 483-492, Apr. 2011.

HAN, J.S. Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions. **Mobile DNA**, London, v. 1, n. 15, p. 1-12, Nov. 2010.

HE, C.; HUANG, H.; XU, L. Mechanisms guiding Polycomb activities during gene silencing in *Arabidopsis thaliana*. **Frontiers in Plant Science**, Bern, v. 4, p. 1-6, Aug. 2013.

HERNANDEZ-PINZON, I.; DE JESUS, E.; SANTIAGO, N.; CASACUBERTA, J.M. The frequent transcriptional read through of the tobacco Tnt1 retrotransposon and its possible implications for the control of resistance genes. **Journal of Molecular Evolution**, New York, v. 68, p. 269–278, Apr. 2009.

HERNANDEZ-PINZON, I.; CIFUENTES, M.; HENAFF, E.; SANTIAGO, N.; ESPINAS, M.L.; CASACUBERTA, J.M. The Tnt1 retrotransposon escapes silencing in tobacco, its natural host. **PLOS One**, San Francisco, v. 7, n. 3, Nov. 2012. Disponível em: <<http://www.plosone.org>>. Accessed in December 1st 2012.

HICKS, J.; LIU, H. Involvement of eukaryotic small RNA pathways in host defense and viral pathogenesis. **Viruses**, Basel, v. 30, n. 11, p. 2659-2678, Oct. 2013.

HUA, J.; MEYEROWITZ, E.M. Ethylene responses are negatively regulated by a receptor gene family in *Arabidopsis thaliana*. **Cell**, Cambridge, v. 94, p. 261–271, Aug. 1998.

JOHNSON, P.R.; ECKER, J.R. The ethylene gas signal transduction pathway: a molecular perspective. **Annual Review of Genetics**, Palo Alto, v. 32, p. 227-254, Jan. 1998.

JORDAN, I.K., ROGOZIN, I.B.; GLAZKO, G.V.; KOONIN, E.V. Origin of a substantial fraction of human regulatory sequences from transposable elements. **Trends in Genetics**, Cambridge, v. 19, n. 2, p. 68-72, Feb. 2003.

KIM, A.; TERZIAN, C.; SANTAMARIA, P.; PELISSON, A.; PRUD'HOMME, N.; BUCHETON, A. Retroviruses in invertebrates: The gypsy retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*. **Proceedings of the National Academy of Sciences of the USA**, Washington, v. 91, p. 1285-1289, Feb. 1994.

KIM, M.; CLAIBORN, K.C.; LEVIN, H.L. The long terminal repeat-containing retrotransposon Tf1 possesses amino acids in gag that regulate nuclear localization and particle formation. **Journal of Virology**, Amsterdam, v. 79, n. 15, p. 9540–955, Aug. 2005.

KOBAYASHI, S.; GOTO-YAMAMOTO, N.; HIROCHIKA, H. Retrotransposon-induced mutations in grape skin color. **Science**, New York, v. 304, p. 982, May 2004.

KONING, A.P.J.; GU, W.; CASTOE, T.A.; BATZER, M.A.; POLLOCK, D.D. Repetitive elements may comprise over two-thirds of the human genome. **PLOS one**, San Francisco, v. 7, n. 12, Dec. 2011. Disponível em: <<http://www.plosone.org>>. Accessed in January 1<sup>st</sup> 2012.

KUMAR, A.; BENNETZEN, J.L. Plant retrotransposons. **Annual Review of Genetics**, Palo Alto, v. 33, p. 479–532, Jan. 1999.

INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM. Initial sequencing and analysis of the human genome. **Nature**, London, v. 409, p. 860-921, Apr. 2001.

LI, S.; LIU, L.; ZHUANG, X.; YU, Y.; LIU, X.; CUI, X.; JI, L.; PAN, Z.; CAO, X.; MO, B.; ZHANG, F.; RAIKHEL, N.; JIANG, L.; CHEN, X. MicroRNAs inhibit the translation of target mRNAs on the endoplasmic reticulum in *Arabidopsis*. **Cell**, Cambridge, v. 153, p. 526-574, Nov. 2013.

- LISCH, D. Epigenetic regulation of transposable elements in plants. **Annual Review of Plant Biology**, Palo Alto, v. 60, p. 43-66, Apr. 2009.
- LISCH D. How important are transposons for plant evolution? **Nature Review Genetics**, London, v. 14, n. 1, p. 49-61, Dec. 2013.
- LUCIA, F.; CREVILLEN, P.; JONES, A.M.; GREB, T.; DEAN, C. A PHD-polycomb repressive complex 2 triggers the epigenetic silencing of FLC during vernalization. **Proceedings of the National Academy of Sciences of the USA**, Washington, v. 105, n. 44, p. 16831–16836, Sept. 2008.
- MANSOUR, A. Epigenetic activation of genomic retrotransposons. **Journal of cell and Molecular Biology**, Oxford, v. 6, n. 2, p. 99-107, Mar. 2007.
- MARQUEZ, Y.; BROWN, J.W.; SIMPSON, C.; BARTA, A.; KALYNA, M. Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis. **Genome Res.**, Cold Spring Harbor, v. 22, n. 6, p. 1184-1195, Dec. 2012.
- MASSON, P.; RUTHERFORD, G.; BANKS, J.A.; FEDOROFF, N. Essential large transcripts of the maize Spm transposable element are generated by alternative splicing. **Cell**, Cambridge, v. 58, n. 4, p. 755-765, Aug. 1989.
- MATZKE, M.A.; MOSHER, R.A. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. **Nature Review Genetics**, London, v. 15, n. 6, p. 394-408, June 2014.
- MCCLINTOCK, B. The origin and behavior of mutable loci in maize. **Proceedings of the National Academy of Sciences of the USA**, Washington, v. 36, n. 6, p. 344–355, Aug. 1950.
- MCCLINTOCK, B. The significance of responses of the genome to challenge. **Science**, New York, v. 226, p. 792-801, Feb. 1984.
- MHIRI, C.; MOREL, J.B.; VERNHETTES, S.; CASACUBERTA, J.M.; LUCAS, H.; GRANDBASTIEN, M.A. The promoter of the tobacco Tnt1 retrotransposon is induced by wounding and by abiotic stress. **Plant Molecular Biology**, New York, v. 33, n. 2, p. 257-266, Aug. 1997.
- MILLER, W.J.; CAPY, P. Mobile genetic elements as natural tools for genome evolution. **Methods in Molecular Biology**, New York, v. 260, p. 1-20, Dec. 2004.
- MOLNAR, A.; MELNYK, C.W.; BASSETT, A.; HARDCASTLE, T.J.; DUNN, R.; BAULCOMBE, D.C. Small silencing RNAs in plants are mobile and direct epigenetic modification in recipient cells. **Science**, New York, v. 328, p. 872-875, May 2010.
- NAKAYASHIKI, H. The Trickster in the genome: contribution and control of transposable elements. **Genes to Cells**, New York, v. 16, p. 827–841, Apr. 2011.



NELJUBOV, D. Ueber die horizontale Nutation der Stengel von *Pisum sativum* und einiger anderen. **Journal of Plant Growth Regulation**, Amsterdam, v. 10, p. 128-139, Jan. 1901.

OHME-TAKAGI, M.; SHINSHI, H. Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element. **The Plant Cell**, Rockville, v. 7, p. 173-182, Nov. 1995.

PAN, Q.; SHAI, O.; LEE, L.J.; FREY, B.J.; BLENCOWE, B.J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. **Nature Genetics**, London, v. 40, n. 12, p. 1413-1415, Dec. 2008.

PAPP, I.; METTE, M.F.; AUFSATZ, W.; DAXINGER, L.; SCHAUER, S.E.; RAY, A.; VAN DER WINDEN, J.; MATZKE, M.; MATZKE A.J.M. Evidence for nuclear processing of plant micro RNA and short interfering RNA precursors. **Plant Physiology**, Rockville, v. 132, p. 1382-1390, Jan. 2003.

PARISOD, C.; ALIX, K.; JUST, J.; PETIT, M.; SARILAR, V.; MHIRI, C.; AINOUCHE, M.; CHALHOUB, B.; GRANDBASTIEN, M.A. Impact of transposable elements on the organization and function of allopolyploid genomes. **New Phytologist**, Lancaster, v. 186, n. 1, p. 37-45, Apr. 2010.

PATERSON, A.H.; BOWERS, J.E.; BRUGGMANN, R. The *Sorghum bicolor* genome and the diversification of grasses. **Nature**, London, v. 457, p. 551-556, Sept. 2009.

PERTEA, M.; SALZBERG, S. Between a chicken and a grape: estimating the number of human genes. **Genome Biology**, Oxford, v. 11, n. 206, p. 1-7, Aug. 2010.

POUTEAU, S.; GRANDBASTIEN, M.A.; BOCCARA, M. Microbial elicitors of plant defense responses activate transcription of a retrotransposon. **The Plant Journal**, Rockville, v. 5, n. 4, p. 535-542, Aug. 1994.

POUTEAU, S.; SPIELMANN, A.; MEYER, C.; GRANDBASTIEN, M.A.; CABOCHE, M. Effects of Tnt1 tobacco retrotransposon insertion on target gene transcription. **Molecular Gene Genetics**, v. 228, n. 1/2, p. 233-239, Aug. 1991.

RASHKOVA, S.; KARAM, S.E.; PARDUE, M.L. Element-specific localization of Drosophila retrotransposon Gag proteins occurs in both nucleus and cytoplasm. **Proceedings of the National Academy of Sciences of the USA**, Washington, v. 99, n. 6, p. 3621-3626, Mar. 2002.

ROGERS, K.; CHEN, X. Biogenesis, turnover, and mode of action of plant microRNAs. **The Plant Cell**, Rockville, v. 25, p. 2383-2399, Jan. 2013.

ROSSI, M.; ARAUJO, P.G.; VAN SLUYS, M.A. Survey of transposable elements in sugarcane expressed sequence tags (ESTs). **Genetics and Molecular Biology**, Brasília, v. 24, n. 1/4, p. 147-154, Mar. 2001.

SABOT, F.; SCHULMAN, A.H. Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome. **Heredity**, London, v. 97, n. 6, p. 381-388, Dec. 2006.

SCHULMAN, A.H. Retrotransposon replication in plants. **Virology**, Amsterdam, v. 3, p. 604-614, Apr. 2013.

SHINSHI, H.; USAMI, S.; OHME-TAKAGI, M. Identification of an ethylene-responsive region in the promoter of a tobacco class I chitinase gene. **Plant Molecular Biology**, New York, v. 27, p. 923-932, Feb. 1995.

SLOTKIN, R.K.; MARTIENSSEN, R. Transposable elements and the epigenetic regulation of the genome. **Nature Review Genetics**, London, v. 8, p. 272-285, Aug. 2007.

STAKE, M.S.; BANN, D.V.; KADDIS, R.J.; PARENT, L.J. Nuclear trafficking of retroviral RNAs and gag proteins during late steps of replication. **Viruses**, Basel, v. 5, p. 2767-2795, May 2013.

TAPIA, G.; VERDUGO, I.; YANEZ, M.; AHUMADA, I.; THEODULOZ, C.; CORDERO, C.; POBLETE, F.; GONZALEZ, E.; RUIZ-LARA, S. Involvement of ethylene in stress-induced expression of the TLC1.1 retrotransposon from *Lycopersicon chilense* Dun. **Plant Physiology**, Rockville, v. 138, p. 2075-2086, Aug. 2005.

VANDENBUSSCHE, F.; VASEVA, I.; VISSENBERG, K.; VAN DER STRAETEN, D. Ethylene in vegetative development: a tale with a riddle. **New Phytologist**, Lancaster, v. 194, p. 895-909, Jan. 2012.

VARAGONA, M.J.; PURUGGANAN, M.; WESSLER, S.R. Alternative splicing induced by insertion of retrotransposons into the maize waxy gene. **Plant Cell**, New York, v. 4, n. 7, p. 811-820, Jul. 1992.

VAUCHERET, H. Post-transcriptional small RNA pathways in plants: mechanisms and regulations. **Genes Development**, Cold Spring Harbor, v. 20, n. 7, p. 759-771, Apr. 2006.

VAZQUEZ; LEGRAND; WINDELS. The biosynthetic pathways and biological scopes of plant small RNAs. **Trends in Plant Science**, Cambridge, v. 15, n. 6, p. 337-345, Nov. 2010.

VERNHETTES, S.; GRANDBASTIEN, M.A.; CASACUBERTA, J.M. In vivo characterization of transcriptional regulatory sequences involved in the defense - associated expression of the tobacco retrotransposon Tnt1. **Plant Molecular Biology**, New York, v. 35, n. 5, p. 673-679, Nov. 1997.

VERNHETTES, S.; GRANDBASTIEN, M.A.; CASACUBERTA, J.M. The evolutionary analysis of the Tnt1 retrotransposon in *Nicotiana* species reveals the high variability of its regulatory sequences. **Molecular Biology Evolution**, Oxford, v. 15, n. 7, p. 827-836, Dec. 1998.

WANG, F.; CUI, X.; SUN, Y.; DONG, C. Ethylene signalling and regulation in plant growth and stress responses. **Plant Cell Reports**, New York, v. 32, p. 1099–1109, Jan. 2013.

WANG, K.L.C.; LI, H. ECKER, J.R. Ethylene biosynthesis and signalling networks. **The Plant Cell**, New York, v. suppl., p. S131–S151, Apr. 2002.

WICKER, T.; SABOT, F.; HUA-VAN, A.; BENNETZEN, J.L.; CAPY, P.; CHALHOUB, B.; FLAVELL, A.; LEROY, P.; MORGANTE, M.; PANAUD, O.; PAUX, E.; SANMIGUEL, P.; SCHULMAN, A.H. A unified classification system for eukaryotic transposable elements. **Nature Review Genetics**, London, v. 8, n. 12, p. 873–982, Dec. 2007.

WU, L.; ZHOU, H.; ZHANG, J.; NI, F.; LIU, C.; QI, Y. DNA methylation mediated by a microRNA pathway. **Molecular Cell**, Rockville, v. 38, n. 3, p. 465–475, May 2010.

XIE, Z.; JOHANSEN, L.K.; GUSTAFSON, A.M.; KASSCHAU, K.D; LILLIS, D.; ZILBERMAN, D.; JACOBSEN, S.E.; CARRINGTON, J.C. Genetic and functional diversification of small RNA pathways in plants. **PLOS Biology**, San Francisco, v. 2, n. 5, May 2004. Disponível em: <<http://www.plosone.org>>. Accessed in November 1<sup>st</sup> 2012.

XIONG, Y.; EICKBUSH, T.H. Origin and evolution of retroelements based upon their reverse transcriptase sequences. **EMBO Journal**, Heidelberg, v. 9, n. 10, p. 3353–3362, Oct. 1990.

ZARDO, G.; CIOLFI, A.; VIAN, L.; STARNES, L.M.; BILI, M.; RACANICCHI, S.; MARESCA, C.; FAZI, F.; TRAVAGLINI, L.; NOGUERA, N.; MANCINI, M.; NANNI, M.; CIMINO, G.; LO-COCO, F.; GRIGNANI, F.; NERVI, C. Polycombs and microRNA-223 regulate human granulopoiesis by transcriptional control of target gene expression. **Blood**, Washington, v. 119, n. 17, p. 4034–4046, Jan. 2012.

ZHANG, H.; HE, X.; ZHU, J. RNA-directed DNA methylation in plants where to start? **RNA Biology**, Austin, v. 10, n. 10, p. 1593–1596, Oct. 2013.

ZHAO, M.; MA, J. Co-evolution of plant LTR-retrotransposons and their host genomes. **Protein and Cell**, New York, v. 4, n. 7, p. 493–501, July 2013.

## 2 EVIDENCING GENE REGULATORY NETWORKS FROM TRANSPOSABLE ELEMENT EXPRESSION

### Abstract

In the early 1940s McClintock defined transposable elements (TEs) as controlling elements of the gene. Features as stress responsiveness and high copy number in plant genomes made them suitable to be incorporated in gene regulation networks during evolution. In order to investigate a role for TEs in the regulation of gene circuits, we generated transgenic tobacco lines carrying a hairpin construct targeting the reverse transcriptase of Tnt1 retrotransposon (HP) and carried phenotypic characterization, transcriptome sequencing and gene expression networks inference. A diverse array of phenotypes was observed in HP plants and ninety-seven genes were found modulated in HP plants. Among these, genes coding for proteins involved in photosynthesis, defense and stress response and cell death, as well as transcription factors were identified. Thirty-five gene networks were established de novo in HP plants, suggesting the emergence of novel connections through gene regulation re-wiring. The evidences observed indicate the direct association of TE expression with the emergence of new gene circuits. Different from others, the present work is the first functional study that reveals a general genome expression landscape modification as a result of targeted TE RNA interference (RNAi), and provides direct evidence of the biological function of retrotransposons in eukaryotic genomes.

Keywords: Tnt1, retrotransposon, tobacco, RNA interference, plant defense response, transcriptome

### 2.1 Introduction

What if transposable elements had a genuine cellular function like any other gene? And how could we experimentally address this question, as retrotransposons are abundant constituents of eukaryotic genomes? Indeed, they can constitute up to 80% of the genome in grasses (PATERSON; BOWERS; BRUGGMANN, 2009). In the human genome, approximately 25% of the genes contain transposable elements derived sequences in their promoter (VAN DE LAGEMAAT et al., 2003). They are recognized as an important source of genetic variability, essential for gene evolution and genomic structure of eukaryotes. It is known that these elements affect gene regulation and genome structure through transposition, recombination of non-homologous regions, and their involvement in epigenetic regulation mechanisms (FEDOROFF, 2012; SLOTKIN; MARTIENSSSEN, 2007; WESSLER, 2006). In humans, somatic transposition may occur and new insertions generate tissue mosaics that may be related to some neurological syndromes, as well as some types

of tumors (BAILLIE; BARNETT; UPTON, 2011). Moreover, it was demonstrated that while only a small part of the human genome correspond to the functional protein-coding genes, a larger part participate in gene regulation (THE ENCODE CONSORTIUM, 2012). Therefore, many regulators present in the genome non-coding fraction remain to be identified. Likewise, only recent evidence has put forward the fact that transposable elements are fully expressed in several tissues both in plants and in animals (ARAUJO; ROSSI; DE JESUS, 2005; JAASKELAINEN et al., 2013; MUOTRI et al., 2009).

To challenge the speculative question raised above, Tnt1 retrotransposon is used as a case study to answer whether interfering in retrotransposon expression would result in phenotypic changes and/or changes in expression patterns of known cellular genes.

## 2.2 Development

### 2.2.1 Methods

#### 2.2.1.1 Plant samples

To achieve the hairpin construction, a 271bp fragment of the Tnt1 (accession number X13777.1) reverse transcriptase was amplified from *N. tabacum* cv Xanthi XHFD8 genomic DNA and cloned in the expression vector pHANNIBAL (accession number AJ311872.1), in both sense and antisense directions, flanking an intron, in order to express a Tnt1 hairpin. The sequence of primers used to amplify the reverse transcriptase fragment are:

	Forward	Primer	1:
CGGGATCCATCTCAGCAGAAGTACAT;	Reverse	Primer	1:
CCATCGATACTTCCCAATGTTCC;	Forward	Primer	2:
CCGCTCGAGATCTCAGGAGAAGTACA;	Reverse	Primer	2:

GGGGTACCACTGCTTCCCAATGTTCC. Expression cassettes were excised from pHANNIBAL and transferred to two binary vectors: pCAMBIA1201 (accession number AF234293), generating pCAMBIA-Tnt1-RT, and pPZP221 (accession number U10491.1), generating pPZP-Tnt1-RT. Control plants were transformed only with the pCAMBIA1201 backbone containing the hygromycin resistance gene. *Nicotiana tabacum* and *Nicotiana benthamiana* foliar discs were transformed with pCAMBIA-Tnt1-RT through *Agrobacterium tumefaciens* (LBA4404) co-culture, according to a previously established method (GALLOIS; MARINHO, 1995).

*Nicotiana tabacum* was transformed also with pPZP-Tnt1-RT through the same method. Regenerated *in vitro* transgenic plants were cultivated in MS20 media with the proper antibiotic, under a 14-hour photoperiod at 24°C. The samples to the RNA-seq were frozen from tobacco 45-day-old leaves.

### **2.2.1.2 Microscopy**

Leaves of different lineages were photo-documented using a Multiphoton Confocal Microscope Zeiss LSM-780 NLO. Trypan blue staining was made according to (KOCH; SLUSARENKO, 1990) and aniline blue according to (DIETRICH et al., 1994).

### **2.2.1.3 RNA sequencing**

RNA was extracted using a modified LiCl method (SAMBROOK; FIETSCH; MANIATIS; 1989), treated with DNaseI (Ambion) and ribosomal RNA was depleted using Ribominus Plant Kit (Invitrogen) following the manufacturer's instructions. The cDNA libraries were made using SOLID Total RNA-seq Kit (Ambion), according to Whole Transcriptome Library Preparation for SOLID Sequencing Protocol (Life Technologies). The handling of the beads for sequencing was done strictly according to the SOLID 3 System Templated Bead Preparation Guide (Life Technologies). The run of the samples followed the SOLID 3 System Instrument Operation Guide (Life Technologies). All sequences are made available at GenBank GEO accession GSE44027.

### **2.2.1.4 Bioinformatics and Statistical Analysis**

Raw sequences were analyzed through CLC Genomics Workbench program. All sequences containing bases with ambiguous identification were removed (sequences with dots instead of color values) and this represented 2.4% of each library. The normalization of transcriptome data was done using the CLC Genomics Workbench program 4.0.3, and consisted of the RPKM expression measure (MORTAZAVI et al., 2008), square root, and Bonferroni's correction. Statistical analysis was performed using multiple t-test. Classification of the modulated genes

was made using BLASTX and the browser AmiGO in the Gene Ontology website (<<http://www.geneontology.org/>> Accessed in May 10 2013).

#### **2.2.1.5 Gene regulatory networks**

The connections between genes presented in this work were identified by the method of inference (LOPES; CESAR, 2008), which is based on a feature selection approach and takes as input data the gene expression profiles. As a result infers connections between genes, generating a gene network.

#### **2.2.1.6 Real time PCRs**

cDNA was synthesized using SuperScript First-Strand Synthesis System for RT-PCR (Life Technologies) following the manufacturer's instructions. The real time PCR reactions were performed using the SYBR Green Real-Time PCR Master Mix according to manufacturer's instructions.

#### **2.2.2 Results**

Tnt1 is present in the tobacco genome in hundreds of copies, it is transcriptionally active and induced by biotic and abiotic stresses (GRANDBASTIEN; AUDEON; BONNIVARD, 2005). Those characteristics make it a likely retrotransposon to have a potential role in gene regulation networks. Thus, we first developed a system aiming to interfere with Tnt1 transcription. We generated transgenic plants of two *Nicotiana* species (*N. tabacum* and *N. benthamiana*) expressing a hairpin (HP) of a Tnt1 reverse transcriptase fragment 273 bp long. Control plants (C) transformed with the backbone cassette, expressing the hygromycin resistance gene, were also generated through the same transformation system. The integrity of the transgene was checked through PCR (Figure 1). Based on classical genetics rationale, the null hypothesis was that if no phenotype was observed, TEs are selfish genes and only occasional insertion outcomes are selected. The H1 hypothesis is that by producing a genetic mark affecting TE expression, phenotypes and changes in expression profile would occur. Thus, the

experimental strategy was to compare HP, C and wild-type (WT) plants transcriptional profile and phenotypes.

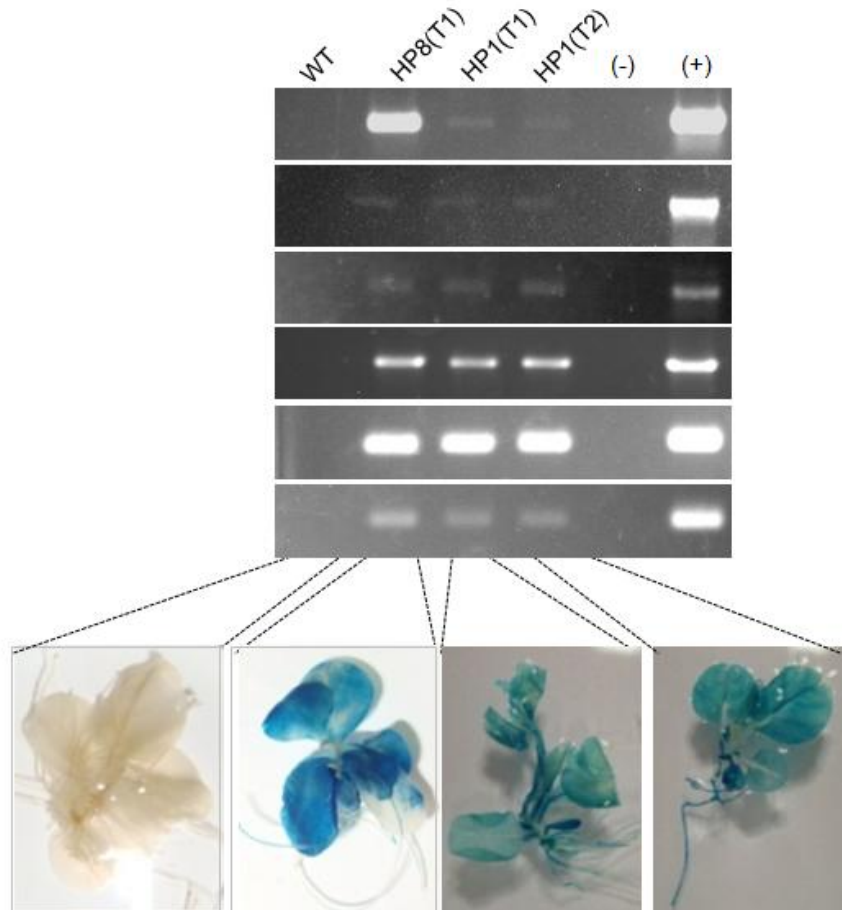
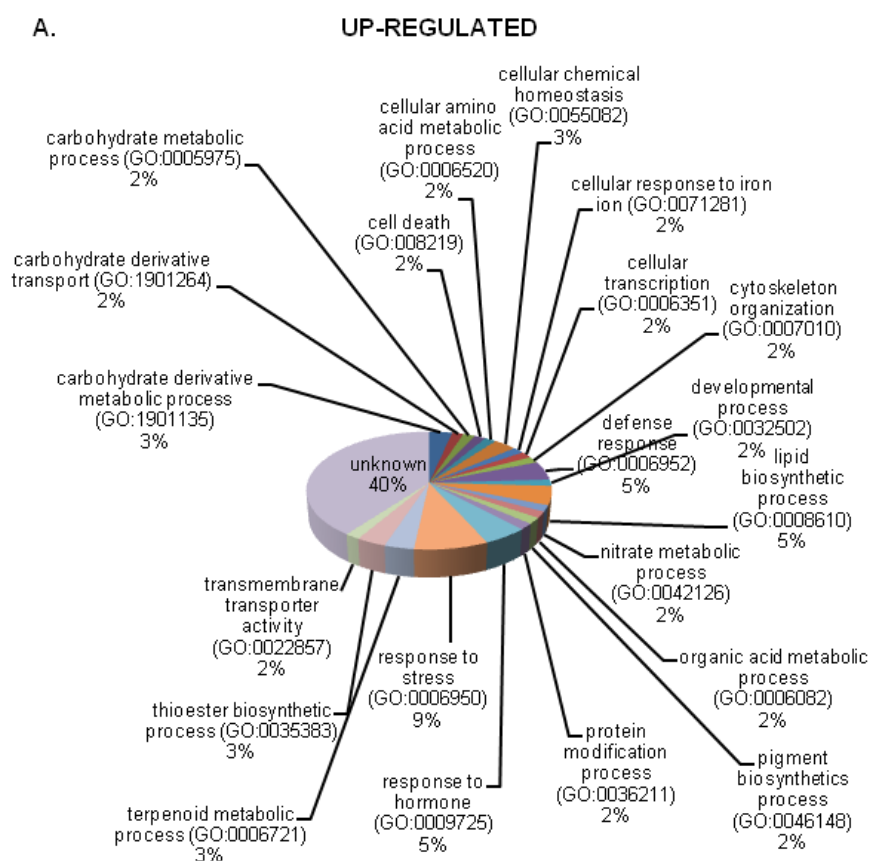


Figure 1 - Checking of the transgene integrity in HP plants. The main elements of the construction were amplified through PCR performed on genomic DNA of wild-type (WT) and HP lineages HP8(T1), HP1(T1) and HP1(T2). The structures amplified are, from top to bottom: CaMV 35S promoter (amplicon 597bp), Tnt1 sense reverse transcriptase fragment (1059bp), intron (615bp), Tnt1 antisense reverse transcriptase fragment (726bp), OCS terminator (355bp) and GUS (240bp). (-) and (+) are negative and positive PCR controls respectively. Below, X-gluc histochemical assay to check the expression of the reporter gene GUS contained in the construction

In order to compare the transcriptional profiles, RNA-seq was performed on leaves of WT, C and three independent HP lines (HP8(T1), HP1(T1) and HP1(T2)) presenting cell death spots - one of the observed phenotypes which will be presented below. The RNA-seq produced a total of 602,744,341 reads (GEO accession number GSE44027), mapped against 24,069 unigenes from the tobacco database in Genbank. Using statistical t-test with stringent p-value < 0.001 we generated two subsets of modulated genes, the first comparing the three HP in contrast to WT lines, and only the modulated in the three HP were considered. The second subset



comprises the genes modulated comparing HPs with C, in order to filter out genes possibly modulated due to the transgenesis process. Overlapping these two subsets we found 97 genes modulated in HP, reflecting significant changes in gene expression. These genes were categorized in biological processes according to the Gene Ontology (<<http://www.geneontology.org/>> Accessed in May 10 2013). Figures 2A and 2B depict genes up-regulated or turned on and genes down-regulated or turned off, respectively (Table 1). In the categorization of the up-regulated genes, the most representative biological processes were: response to stress (9%), response to hormone (5%), lipid biosynthetic process (5%) and defense response (5%). Forty percent of the up-regulated genes were classified as unknown (Figure 2A). Among the down-regulated genes, the most representative biological processes were: photosynthesis (23%), cellular amino acid metabolic process (8%), reactive oxygen species metabolic process (5%), microtubule-based process (5%) and cellular chemical homeostasis (5%). Twenty-six per cent were classified as unknown (Figure 2B).



## B. DOWN-REGULATED

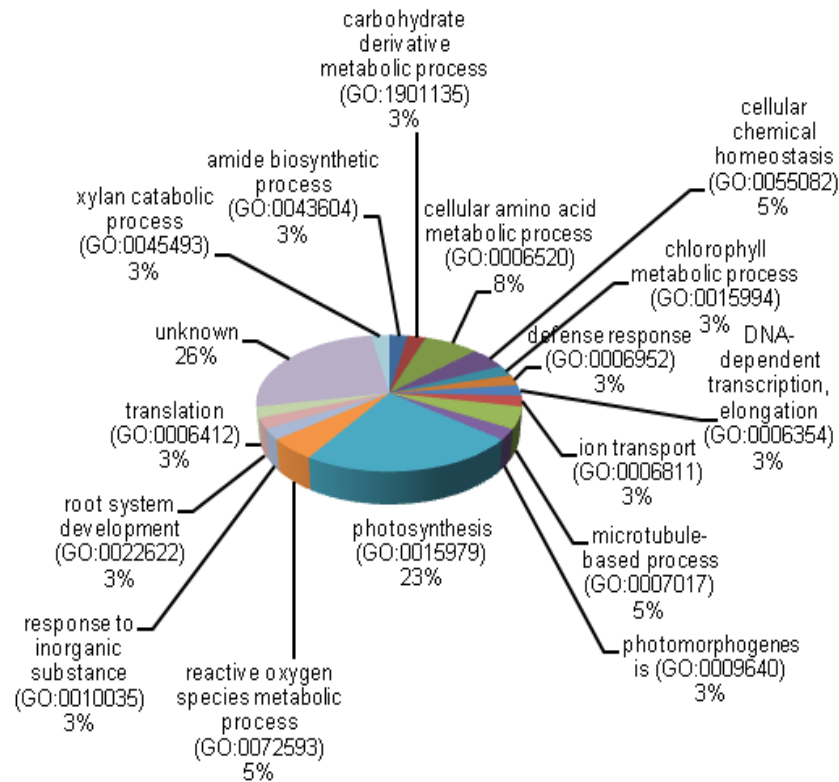


Figure 2 - Global view of the 97 genes modulated in HP leaves under a p-value of 0.001. (A) Categorization of the 58 genes up-regulated or activated in leaves of tree independent HP plants HP8(T1), HP1(T1) and HP1(T2) in comparison to wild-type and control by biological processes (the numbers in parentheses are Gene Ontology terms) (B) Categorization of the 39 genes down-regulated or turned off in HP leaves by biological processes (Gene Ontology terms in parentheses)

Table 1 – Genes modulated in HP lines compared to wild-type and control. Ninety seven genes were modulated in HP plants comparing to WT with a stringent e-value > 0.001. These genes were up or down-regulated, and in some cases switched from no detectable expression to detectable. The first column presents the Genbank accession numbers. Fold change column represents the fold change expression of HPs compared to WT. The genes assigned as “induced” in the fold change column presented no detectable expression in the WT sample and were induced in the three HPs. The ones assigned as “repressed” presented expression in WT and became not detectable in HP. The gene name column shows the BLASTX based annotation

(continues)

GB accession #	Fold change	gene name
EV849906	-20,30156758	[light harvesting chlorophyll a/b-binding protein [Nicotiana sylvestris] (gi:3036953/e-value: e-153)]
EV849907	-15,62836198	[RecName: Full=Chlorophyll a-b binding protein 3C, chloroplastic; AltName: Full=LHCII type I CAB-3C; Short=LHCP; Flags: Precursor protein 3C, chlorophyll binding (gi:115825/e-value: e-106)]
DV998840	-13,50186081	[RecName: Full=Chlorophyll a-b binding protein 13, chloroplastic; AltName: Full=LHCII type III CAB-13; Flags: Precursor LHCII type III [Solanum lycopersicum] (gi:115794/e-value: e-137)]
X52742	-13,4130905	[RecName: Full=Chlorophyll a-b binding protein 50, chloroplastic; AltName: Full=LHCII type I CAB-50; Short=LHCP; Flags: Precursor unnamed protein product [Nicotiana tabacum] (gi:115810/e-value: e-147)]
EV849909	-13,23246289	[light harvesting chlorophyll a/b-binding protein [Nicotiana sylvestris] (gi:3036955/e-value: e-147)]
DV162523	-13,08716429	[RecName: Full=Chlorophyll a-b binding protein 13, chloroplastic; AltName: Full=LHCII type III CAB-13; Flags: Precursor LHCII type III [Solanum lycopersicum] (gi:115794/e-value: e-141)]
X52744	-11,67194942	[RecName: Full=Chlorophyll a-b binding protein 40, chloroplastic; AltName: Full=LHCII type I CAB-40; Short=LHCP; Flags: Precursor unnamed protein product [Nicotiana tabacum] (gi:115805/e-value: e-153)]
AY219853	-10,25775914	[chlorophyll a/b binding protein [Nicotiana tabacum] (gi:29123378/e-value: e-153)]
EV849912	-9,967117026	[RecName: Full=Chlorophyll a-b binding protein 16, chloroplastic; AltName: Full=LHCII type I CAB-16; Short=LHCP; Flags: Precursor unnamed protein product [Nicotiana tabacum] (gi:115781/e-value: e-138)]
DW000203	-9,314287875	[chlorophyll a/b binding protein [Solanum tuberosum] (gi:100801744/e-value: e-135)]
DW000189	-4,066550076	[glycine-rich protein precursor [Nicotiana tabacum] (gi:270313172/e-value: 2e-06)]
EB424812	-2,704509186	[CPRD49-like protein [Capsicum annuum] (gi:60459405/e-value: 6e-88)]
EB451549	-2,690746193	[putative delta TIP [Nicotiana glauca] (gi:12006841/e-value: 2e-98)]
EB424981	-2,19595428	[meloidogyne-induced giant cell protein-like protein [Solanum tuberosum] (gi:76573349/e-value: e-118)]
CB322075	-2,095645826	[No hits in NR]
EB451689	-2,087813378	[PREDICTED: hypothetical protein [Vitis vinifera] (gi:225445774/e-value: 1e-43)]
EB680120	-2,020918162	[PREDICTED: hypothetical protein [Vitis vinifera] unnamed protein product [Vitis vinifera] (gi:225427218/e-value: e-100)]
DW000430	-1,984812055	[conserved hypothetical protein [Ricinus communis] conserved hypothetical protein [Ricinus communis] (gi:255537843/e-value: 1e-31)]
EB441754	-1,924803509	[unknown [Solanum lycopersicum] (gi:643469/e-value: 1e-45)]
DV161022	-1,893331491	[PREDICTED: hypothetical protein [Vitis vinifera] unnamed protein product [Vitis vinifera] (gi:225461306/e-value: 6e-33)]
DV160478	-1,856723145	[DnaJ-like protein isoform [Solanum phureja] (gi:113374278/e-value: 6e-73)]
FG636365	-1,852434675	[1-amniocyclopropane-1-carboxylate oxidase [Nicotiana tabacum] (gi:602586/e-value: 2e-56)]
DW000473	-1,851831657	[sigma factor sigb regulation protein rsbq, putative [Ricinus communis] sigma factor sigb regulation protein rsbq, putative [Ricinus communis] (gi:255567977/e-value: e-112)]
EB678821	-1,836317118	[beta-galactosidase [Solanum lycopersicum] ss-galactosidase [Solanum lycopersicum] (gi:7939619/e-value: 5e-81)]
EB450703	-1,834044756	[chloroplast-specific ribosomal protein [Solanum lycopersicum] (gi:51457950/e-value: e-108)]
AY061820	-1,824871949	[asparagine synthetase [Nicotiana tabacum] (gi:17025882/e-value: 9e-80)]
DV999555	-1,791543605	[beta-ketoacyl-ACP synthase II [Perilla frutescens] (gi:2921211/e-value: 1e-68)]
EB449422	-1,788824817	[PREDICTED: hypothetical protein [Vitis vinifera] (gi:225427690/e-value: e-139)]
DV160279	-1,777701877	[LEXYL2 [Solanum lycopersicum] (gi:37359708/e-value: 4e-98)]
FG640918	-1,75525839	[unnamed protein product [Vitis vinifera] (gi:297744764/e-value: 2e-68)]
FG641183	-1,707414588	[No hits in NR]
EB683915	-1,680229422	[PREDICTED: hypothetical protein [Vitis vinifera] (gi:225468496/e-value: 2e-48)]
DV999145	-1.677224419	[unnamed protein product [Nicotiana tabacum] (gi:8096273/e-value: 4e-80)]

Table 1 – Genes modulated in HP lines compared to wild-type and control. Ninety seven genes were modulated in HP plants comparing to WT with a stringent e-value > 0.001. These genes were up or down-regulated, and in some cases switched from no detectable expression to detectable. The first column presents the Genbank accession numbers. Fold change column represents the fold change expression of HPs compared to WT. The genes assigned as “induced” in the fold change column presented no detectable expression in the WT sample and were induced in the three HPs. The ones assigned as “repressed” presented expression in WT and became not detectable in HP. The gene name column shows the BLASTX based annotation

(conclusion)

GB accession #	Fold change	gene name
U52078	-1,672260725	[kinesin-like protein [Nicotiana tabacum] (gi:1491931/e-value: 0.0)]
AB009885	-1,619975248	[unnamed protein product [Nicotiana tabacum] (gi:8096273/e-value: e-170)]
DV160610	1,400694196	[acyl carrier protein [Capsicum chinense] (gi:4457219/e-value: 2e-45)]
EH622610	1,455686792	[prb-1b [Nicotiana tabacum] (gi:19970/e-value: e-106)]
EH618558	1,578196416	[5-epi-aristolochene synthase [Nicotiana tabacum] (gi:1620988/e-value: 6e-58)]
AY748245	1,612708483	[cytosolic acetoacetyl-coenzyme A thiolase [Nicotiana tabacum] (gi:53854350/e-value: 0.0)]
DV162389	1,652920749	[RecName: Full=Pleiotropic drug resistance protein 1; AltName: Full=NpPDR1 ABC1 protein [Nicotiana plumbaginifolia] (gi:75331994/e-value: e-154)]
AB012857	1,656879315	[ACC oxidase [Nicotiana tabacum] (gi:5751171/e-value: e-171)]
EH622095	1,677147437	[PREDICTED: hypothetical protein [Vitis vinifera] (gi:225424329/e-value: 6e-62)]
AB176524	1,807743174	[glycosyltransferase [Nicotiana tabacum] (gi:62241067/e-value: 0.0)]
AF039921	1,841944194	[ethylene receptor ERS homolog [Nicotiana tabacum] (gi:7652766/e-value: 0.0)]
EH621848	1,872625984	[mRNA inducible by salicylic acid or by TMV during Systemic Acquired Resistance response.' [Nicotiana tabacum] Sar8.2b protein [Nicotiana tabacum] (gi:170335/e-value: 1e-33)]
AF272244	1,875259203	[5-epi-aristolochene synthase [Nicotiana tabacum] (gi:10444007/e-value: 0.0)]
EB426351	1,881210529	[farnesyl diphosphate synthase [Panax ginseng] (gi:68185941/e-value: e-148)]
X06134	1,892694217	[unnamed protein product [Nicotiana tabacum] (gi:929750/e-value: 0.0)]
AB176525	1,925666099	[malonyltransferase [Nicotiana tabacum] (gi:62241069/e-value: 0.0)]
EB444508	1,961058275	[unnamed protein product [Vitis vinifera] (gi:297745162/e-value: 0.002)]
AY138861	2,165321456	[phospholipase D beta 1 isoform [Nicotiana tabacum] (gi:22770428/e-value: 0.0)]
Z29529	2,287637792	[ethylene forming enzyme (EFE) [Nicotiana tabacum] (gi:450357/e-value: e-171)]
AF004232	2,316222988	[hydroxy-methyl-glutaryl-coenzyme A reductase [Nicotiana tabacum] (gi:18000042/e-value: 0.0)]
AF004233	2,350873288	[hydroxy-methyl-glutaryl-coenzyme A reductase [Nicotiana tabacum] (gi:18000044/e-value: 0.0)]
AY286010	2,408735243	[callus-expressing factor [Nicotiana tabacum] (gi:30961941/e-value: 0.0)]
U60452	2,661731421	[hydroxy-methylglutaryl-coenzyme A reductase [Nicotiana tabacum] (gi:2654423/e-value: 0.0)]
FG643486	2,874230565	[No hits in NR]
EH618301	2,893734099	[unnamed protein product [Vitis vinifera] (gi:296081945/e-value: 1e-79)]
EF051128	3,061189787	[pyruvate decarboxylase [Citrus sinensis] (gi:70609688/e-value: 0.0)]
EH623458	3,106000795	[epi-aristolochene synthase 110 [Nicotiana tabacum] (gi:32328283/e-value: 4e-78)]
AF203476	4,147433458	[putative ethylene receptor [Nicotiana tabacum] (gi:22209014/e-value: 0.0)]
DQ131887	5,841481934	[cytochrome P450 monooxygenase CYP82E2 [Nicotiana tabacum] (gi:126669246/e-value: 0.0)]
X51426	8,224590037	[acidic chitinase PR-P [Nicotiana tabacum] (gi:19771/e-value: e-139)]
M29274	11,86677248	[RecName: Full=Probable glutathione S-transferase parA; AltName: Full=Auxin-regulated protein parA; AltName: Full=STR246C protein par peptide [Nicotiana tabacum] STR246C [Nicotiana tabacum] (gi:114208/e-value: e-126)]
EH623372	14,60007808	[tetrahydroxychalcone 2'-glucosyltransferase [Cyclamen persicum] (gi:156138807/e-value: 5e-70)]

Fold changes of specific genes in the following categories response to stress, defense response, response to hormone, photosynthesis and lipid biosynthetic processes are presented in Figure 3 as changes in expression only occurred in the HPs. Several genes were up-regulated, as follows: the stress response transcription factor *Jerf1*, known to be responsive to ethylene and methyljasmonate and considered an integrating factor across distinct signal transduction pathways (ZHANG; HUANG, 2004); acidic endochitinase PR-R which is involved in defense response (CORNELISSEN et al., 1987; LINTHORST; VAN LOON; VAN ROSSUM, 1990); phospholipase D which participates in signal transduction cascades in stress responses (BARGMANN; MUNNIK, 2006); a wound inducible mRNA (accession number AB009885) and the pathogenesis-related protein 1A (PR-1A). Senescence is marked by a decline of the photosynthetic apparatus and mobilization of nutrients from senescing leaves to growing tissues, culminating in cell death (QUIRINO et al., 2000). We identified the repression of several chlorophyll a/b binding proteins in HP leaves. Likewise, hydroxy-methylglutaryl-coenzyme A reductase (HMGR)-like genes were up-regulated in HP plants. This enzyme participates in the steroid derivatives synthesis that follows pathogen infections (CHOI et al., 1994) (Figure 3).

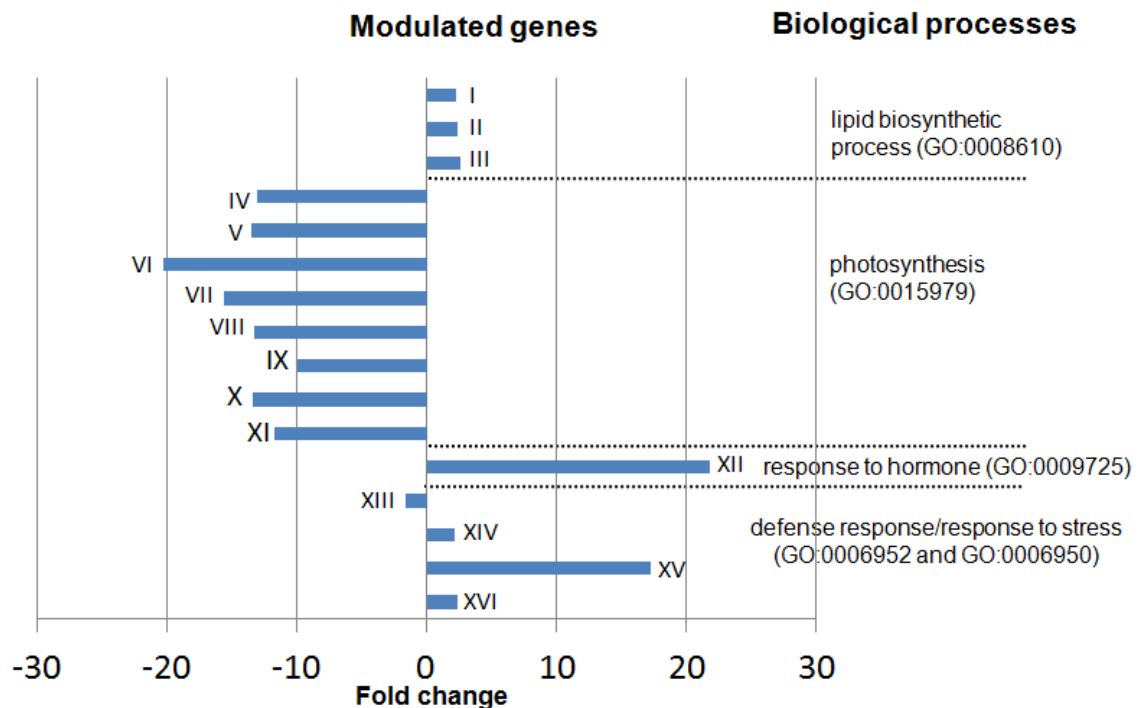


Figure 3 – Histogram showing the fold change of 16 genes belonging to the best represented biological processes (according to AmiGO BLAST) presented in the pie charts previously shown. The X axis indicates the fold change expression relative to WT. Positive values represent induced genes and negative, repressed genes. The genes are identified with roman numerals next to the respective bars. Lipid biosynthetic process (GO:0008610) I: hydroxy-methylglutaryl-coenzyme A reductase (AF004232); II: hydroxy-methylglutaryl-coenzyme A reductase (AF004233); III: hydroxy-methylglutaryl-coenzyme A reductase (U60452); Photosynthesis (GO:0015979) IV: chlorophyll a-b binding protein 13 (DV162523); V: chlorophyll a-b binding protein 13 (DV1998840); VI: chlorophyll a-b binding protein 7 (EV849906); VII: chlorophyll a-b binding protein 3C (EV849907); VIII: chlorophyll a-b binding protein 50 (EV849909); IX: chlorophyll a-b binding protein 16 (EV849912); X: chlorophyll a-b binding protein 50 (X52742); XI: chlorophyll a-b binding protein 40 (X52744); Response to hormone (GO:0009725) XII: pathogenesis-related protein 1A (PR-1A) (EH621793); Defense response (GO:0006952) XIII: wound inductive mRNA (AB009885); XIV: phospholipase D beta 1 isoform (AY138861); XV: pathogenesis-related protein R minor form (PR-R) (X03913); Response to stress (GO:0006950) XVI: transcription factor JERF1 (AY286010)

Other up-regulated genes were those encoding ACC oxidase (AB012857), ethylene forming enzyme (EFE) (Z29529) and ethylene receptors (AF039921 and AF203476). The induction of ethylene synthetic genes and signaling pathway was already identified in senescent leaves of *Arabidopsis* (VAN DER GRAAF; SCHWACKE; SCHNEIDER, 2006). Glutathione S-transferase, contributor to defense signaling in plants (GHANTA; BHATTACHARYYA; CHATTOPADHYAY, 2011) and *Sar8.2*, induced during the pathogen-induced systemic acquired resistance (SONG;

GOODMAN, 2002) were also up-regulated. This possibly corresponds to the spread of the initial stress signal throughout the plant.

Moreover some genes switched from no detectable transcription in WT to transcriptionally active in HPs such as genes encoding 1,3- $\beta$ -glucanase (BAB17320.1) and NAC (NAM, ATAF1/2 and CUC2) transcription factor (BAA78417.1). While the first is described as a stress-inducible gene in tobacco (TAKEMOTO et al., 2003), NAC is a transcription factor family that is related to the control of leaf senescence (GUO; GAN, 2006) among other functions. PR-P and NAC induction were validated by qPCR (Figure 4). Although at this point is not possible to distinguish between primary and secondary effects, we propose that not only the retrotransposon is responsive to some biotic and abiotic stresses, but also that it is fine-tuning their occurrence and progression.

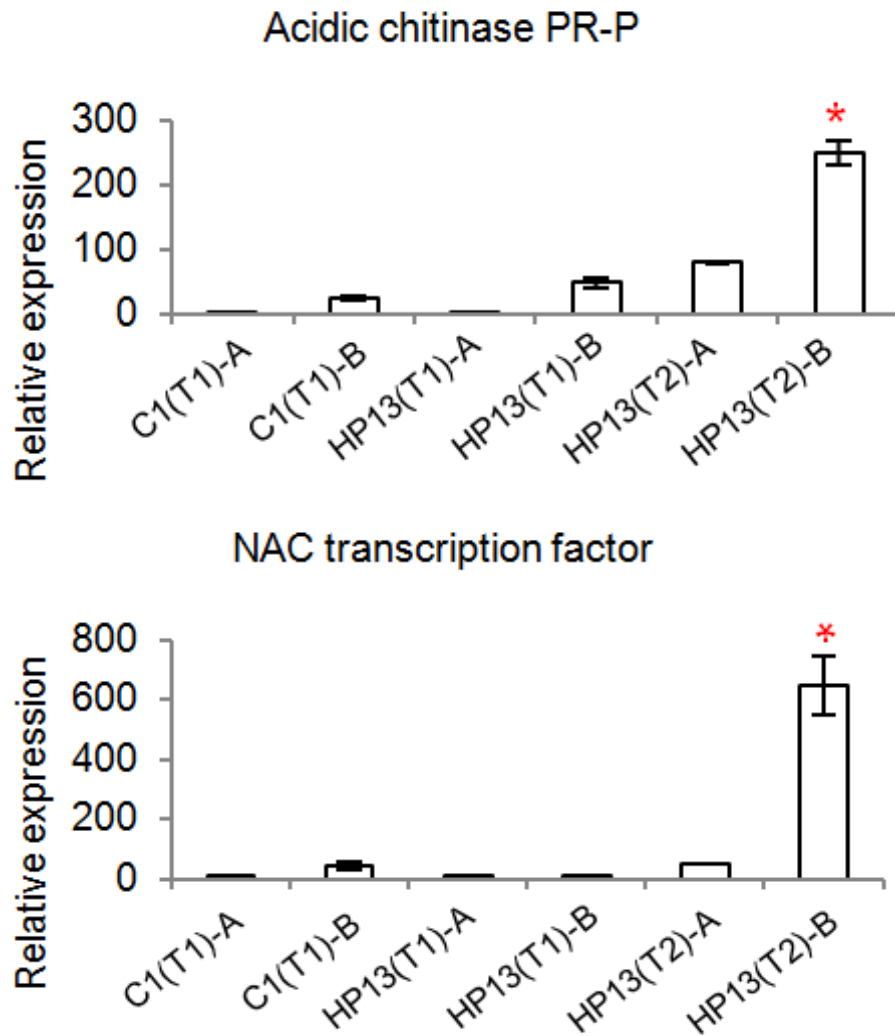


Figure 4 - Validation of acidic chitinase PR-P and NAC transcription factor induction through qPCR analysis, respectively. Leaf samples were taken from in vitro 45-day-old plants. C1(T1) is the plant transformed with the empty vector. HP1(T1) e HP1(T2) refer to first and second generation of a hairpin line. A and B are samples from the same plant and indicate younger and older leaves respectively. Error bars represent the standard deviation of three technical replicates. Leaves presenting the cell death spots phenotype are indicated with a red asterisk

The 97 genes modulated in HPs were used as seeds to infer gene regulatory networks using entropy based approach from gene expression patterns. The connections of the inferred networks were obtained by observing the expression data of the 24,069 unigenes and by applying the penalized mean conditional entropy (MCE), as described in [16]. In the present work, only the inferred connections that had no uncertainty (MCE=0) were selected in the prediction process. The adopted model for prediction is deterministic, so there is only one network that satisfied the



adopted criterion function, i.e. MCE, with best values. In this way, all the predictors that have tied with best criterion function values were considered for building the networks.

Thirty-five networks were identified revealing gene circuits in HP plants that were not identified in WT. From these, the most representative network and which had the highest number of nodes connecting 44 genes is presented in figure 5. Remarkably, 13 genes out of 44 that composed the network were the seeds themselves. Even though the seeds were compared to a universe of 24,069 genes, the emergence of connections occurred among the same core of genes modulated in the RNA-seq. These results using different approaches strengthen the emergence of a new pattern of expression and gene regulation in the HP plants, and uncovered similar biological processes (defense and stress response, indicated in Figure 3) as the main changes in HP lines. In conclusion, the most frequent term associated to changes in gene expression was related to the biological term “response” (defense, stress and hormone) thus supporting that TEs are sensors of the environment (internal or external). Also, placing them early in the hierarchy of perception and transduction pathway.

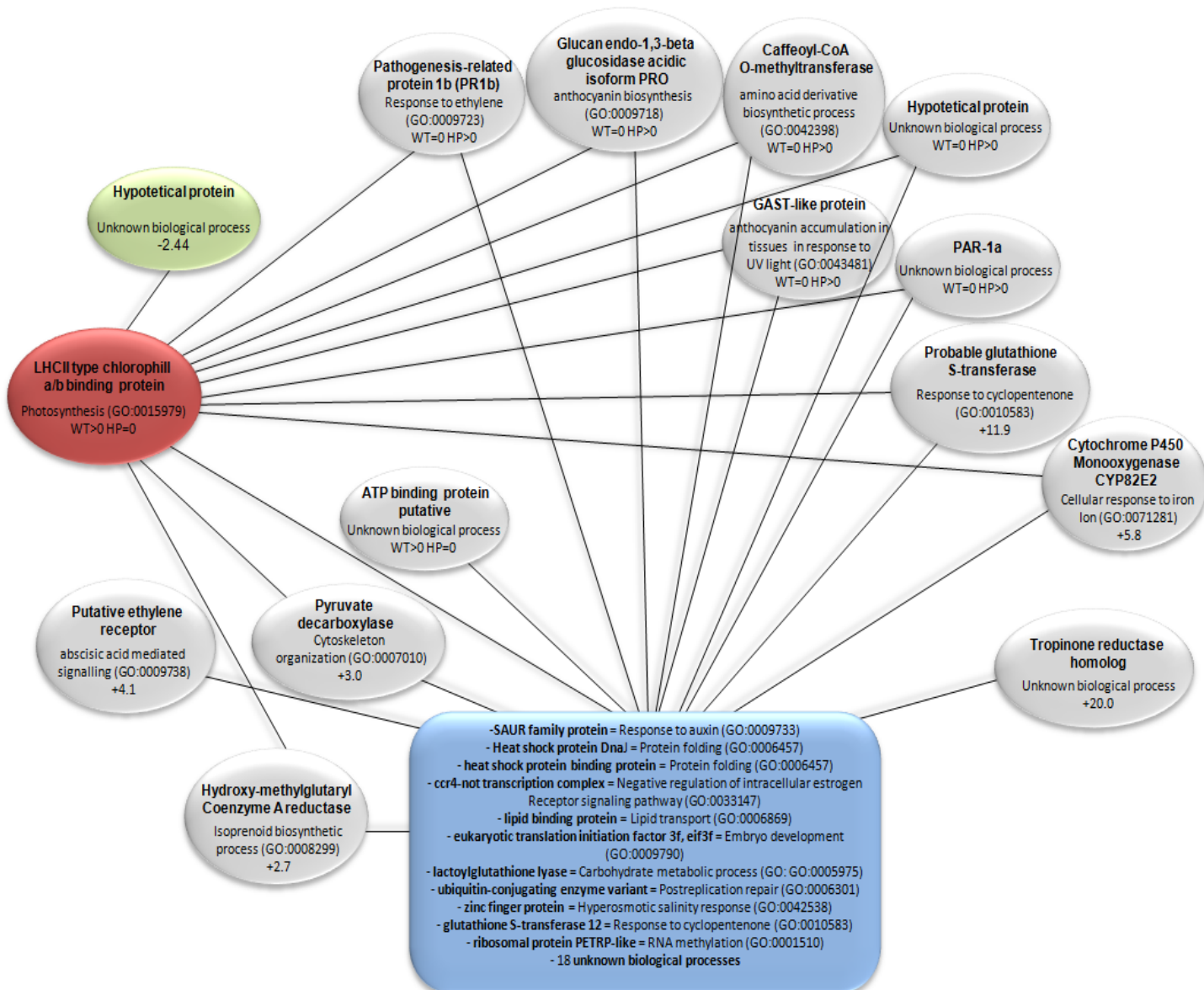


Figure 5 - Diagram of a gene regulatory network observed exclusively in HP plants. The network is composed of forty-four genes of which we show fifteen. Each node (green, red and grey ellipses) denotes a gene. The blue box gathers a cluster of genes that connects to the red and grey nodes. From these eleven have a known biological function and eighteen which function still unknown. The black lines that connect the genes represent a statistically supported link between changes in expression. Each node presents the gene annotation (bold), the major biological process to which it is related and the fold change. WT=0 HP>0 means induction of the gene in HP, with no detectable expression in WT. WT>0 HP=0 means repression of the gene in HP, with detectable expression in WT

To check the effect of the construction on Tnt1 expression, we quantified the RNA-Seq reads mapped to distinct Tnt1 protein domains and normalized them according to the length of each domain. The transcript amount did not decrease in HP plants. However, when accumulation of transcripts occurs, as in the case of the sample HP1(T2), the RT domain (target of the expressed shRNA) remained at a

lower level compared to the other domains (Figure 6A). Because Tnt1 is present in the tobacco genome in hundreds of copies, including full-length and truncated versions, it is not surprising that the silencing strategy adopted was not able to block all Tnt1 copies, but only the ones having the RT domain. Endogenous promoter methylation was not detected (data not shown), indicating that TGS (transcriptional gene silencing) is not predominating. Alternatively, PTGS (post-transcriptional gene silencing), in which the translation is blocked, could be the mechanism to explain our data, since there would be an accumulation of transcripts in the cytoplasm (VOINNET, 2008). However, conclusions on the effects of the expression of the shRNA on Tnt1 are not yet completely understood. We also mapped the reads against Tto1 and Tnt2 (two tobacco retroelements, Tto1 being closely related to Tnt1), and both presented no expression changes (Figure 6B). The alignment of the RNA-seq reads against Tnt1 RT domain contained in the hairpin revealed a different distribution pattern among the samples (Figure 7). HP8(T1) and HP1(T1) presented a number of mapped reads in sense direction similar to the WT, while HP1(T2) presented a higher number of these reads. However, when looking in detail to the region of the hairpin, the three HP samples present a higher number of antisense reads when compared to WT. In HP8(T1) these reads are mapped only in the region of the hairpin, and in HP1(T1) and HP1(T2) they spread to the RT neighbor region. Although we did not detect a common pattern of mapped reads among the HPs, WT presented only few antisense reads. This difference is possibly related to the effect of the hairpin.

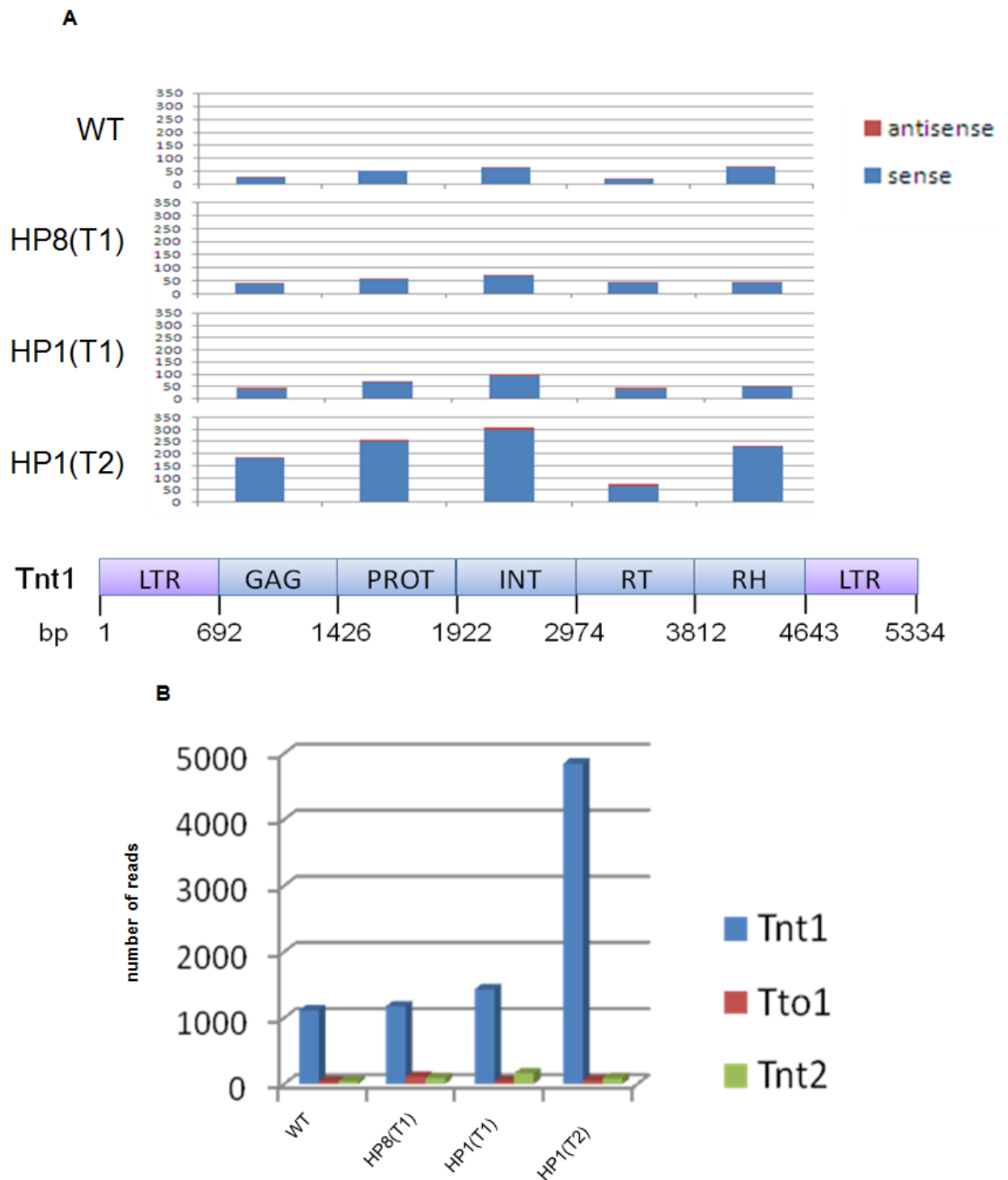


Figure 6 - Expression of Tnt1 by protein domains and number of reads mapped against Tnt1 and related retrotransposons Tto1 and Tnt2. a, Tnt1 expression compared among the lineages WT and HP by protein domain. Bars indicate the number of RNA-Seq reads mapped against every Tnt1 domain, normalized by length. Sense reads are represented in the bars in blue and antisense in red. The diagram below represents Tnt1 regions, including 5' LTR, protein domains GAG, PROT (protease), INT (integrase), RT (reverse transcriptase), RH (RNase H) and LTR 3'. b, Comparative expression of the retrotransposons Tnt1, Tto1 and Tnt2 among the analyzed tobacco lineages. Measurements were based on the number of reads mapped against the nucleotide sequence of any element, normalized by length. While Tnt1 presented detectable expression even in WT, the other elements did not presented significant expression in any lineage

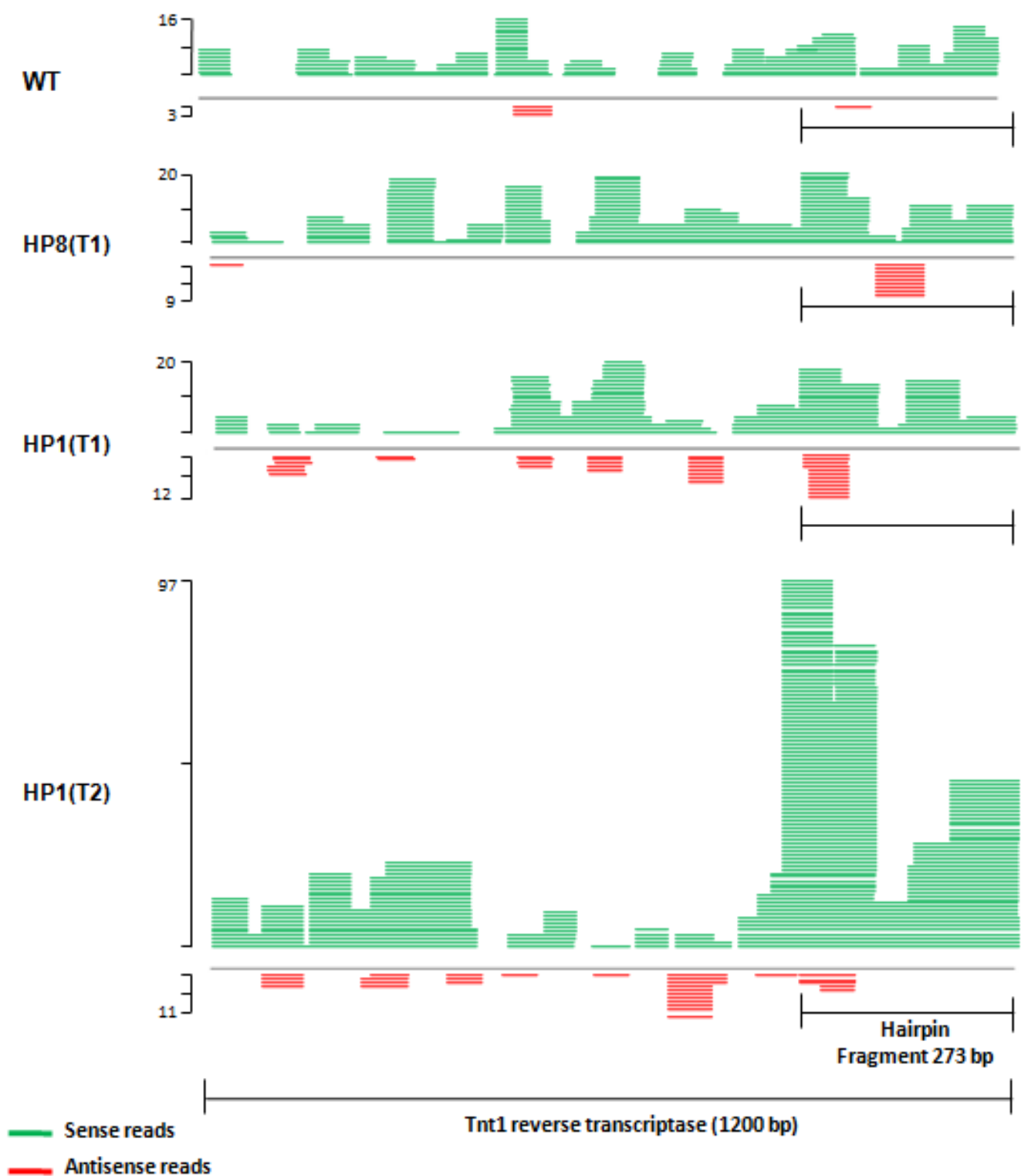


Figure 7 - Alignment of the RNA-seq reads against Tnt1 RT domain. The diagram presents the lineages analyzed, WT, HP8(T1), HP1(T1) and HP1(T2). Each X axis represents the extension of the Tnt1 reverse transcriptase domain (RT). Green bars refer to the reads in sense direction mapped on the RT domain. Red bars refer to reads mapped in antisense direction. The horizontal thick black bars at right indicate the RT fragment inserted in the hairpin

Having reported those results, it is highly important to interpret them in the light of the standard expression of Tnt1 in tobacco plants. The data herein described are in accordance with previous publications that report transcriptional induction of Tnt1 under stress stimuli. Basal levels of Tnt1 transcripts were also identified in WT root tissues (GRANDBASTIEN; AUDEON; BONNIVARD, 2005). Here we first describe the presence of these transcripts in tobacco leaves, also in basal levels of expression. This denotes the endogenous control over Tnt1 expression, keeping the level of expression low (except under specific conditions such as biotic and abiotic stresses). If this is so, difficulties in the interpretation of data regarding the transcription quantitation of Tnt1 under the expression of the hairpin were expected. To date, it is not possible to explain in detail the effect of the hairpin over all Tnt1 copies in the genome. However, not being able to fully comprehend the mechanisms does not invalidate the observations that link Tnt1 to the wide gene reprogramming events observed, mainly if we consider that the set of 97 genes modulated in the three independent HP lines.

In addition, transgenic HP *Nicotiana tabacum* plants displayed a varied range of phenotypes, ranging from cell death spots on leaves (transmissible to the next generation) (Figure 8A) to severe developmental abnormalities (Figure 8B). Several plant lines exhibited aberrant phenotypes at the photosynthetic parts or roots (Figure 8C). The same phenotypes were observed in *N. benthamiana* (Figure 8D-F).

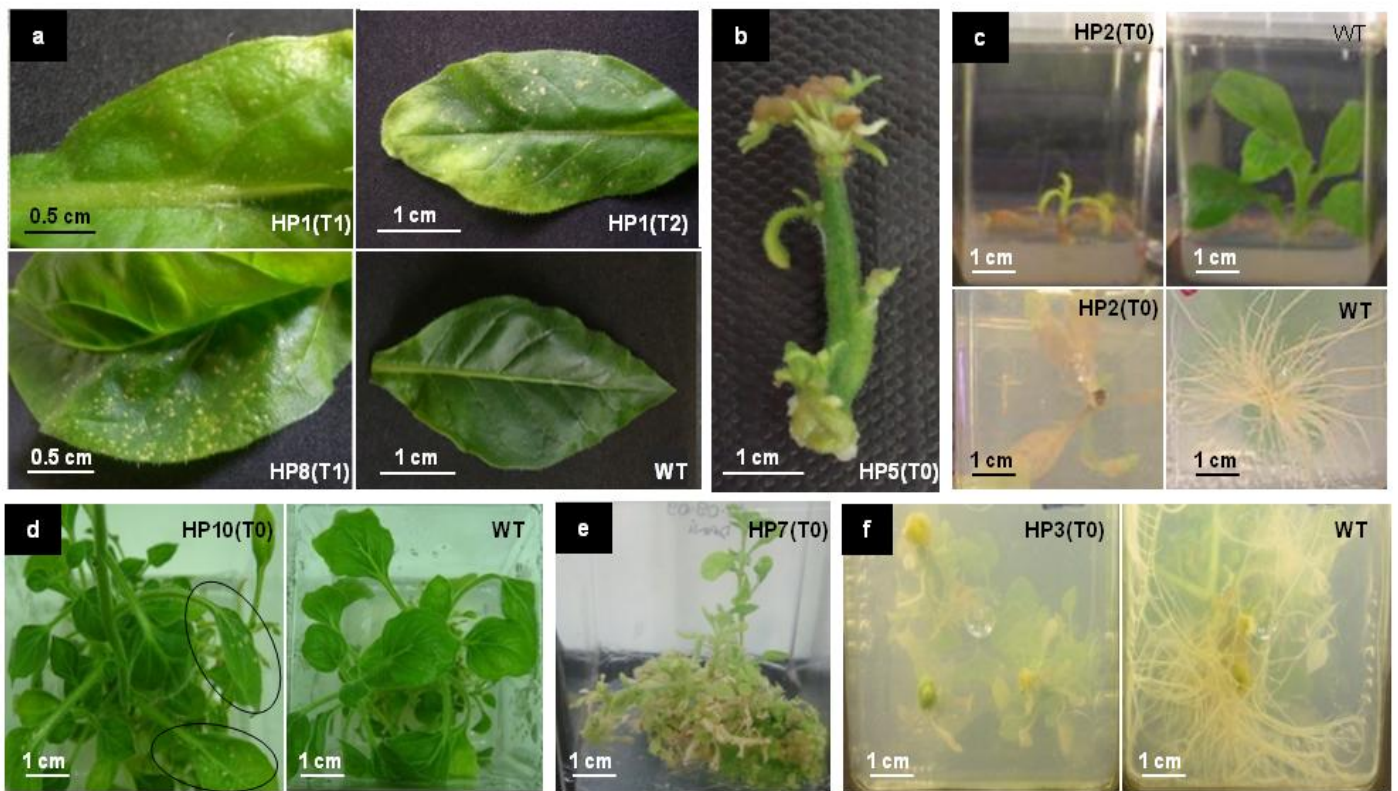


Figure 8 - Phenotype characterization of distinct tobacco lineages. a-c, Phenotypes displayed by in vitro plants of *Nicotiana tabacum*. a, 45-day-old plant leaves. Wild type presents normal leaves while HP lineages present cell death spots. b, Severe developmental abnormality in HP5(T0). c, HP plants presenting limitations such as no development of aerial parts and roots, compared to WT. d-f, The same phenotypes reproduced in *N. benthamiana*

In order to get a more detailed characterization of the cell death spots on leaves we photo-documented the material using confocal microscopy. HP leaves showed a lower cell density in the mesophyll in the region of the spots (Figure 9A). It is possible to see a green fluorescence inside the spots, and the chloroplasts of the cells (in red) around it (Figure 9B-D). A 43µm depth scan revealed that the spots are not only at the surface of the leaf, but spans transversely through the mesophyll (Figure 9D). Callose deposition was initially detected in the regions emitting green fluorescence. Although callose participates of multiple processes in the plant cell, its involvement in responses to biotic and abiotic stresses has been reported (LUNA et al., 2011).



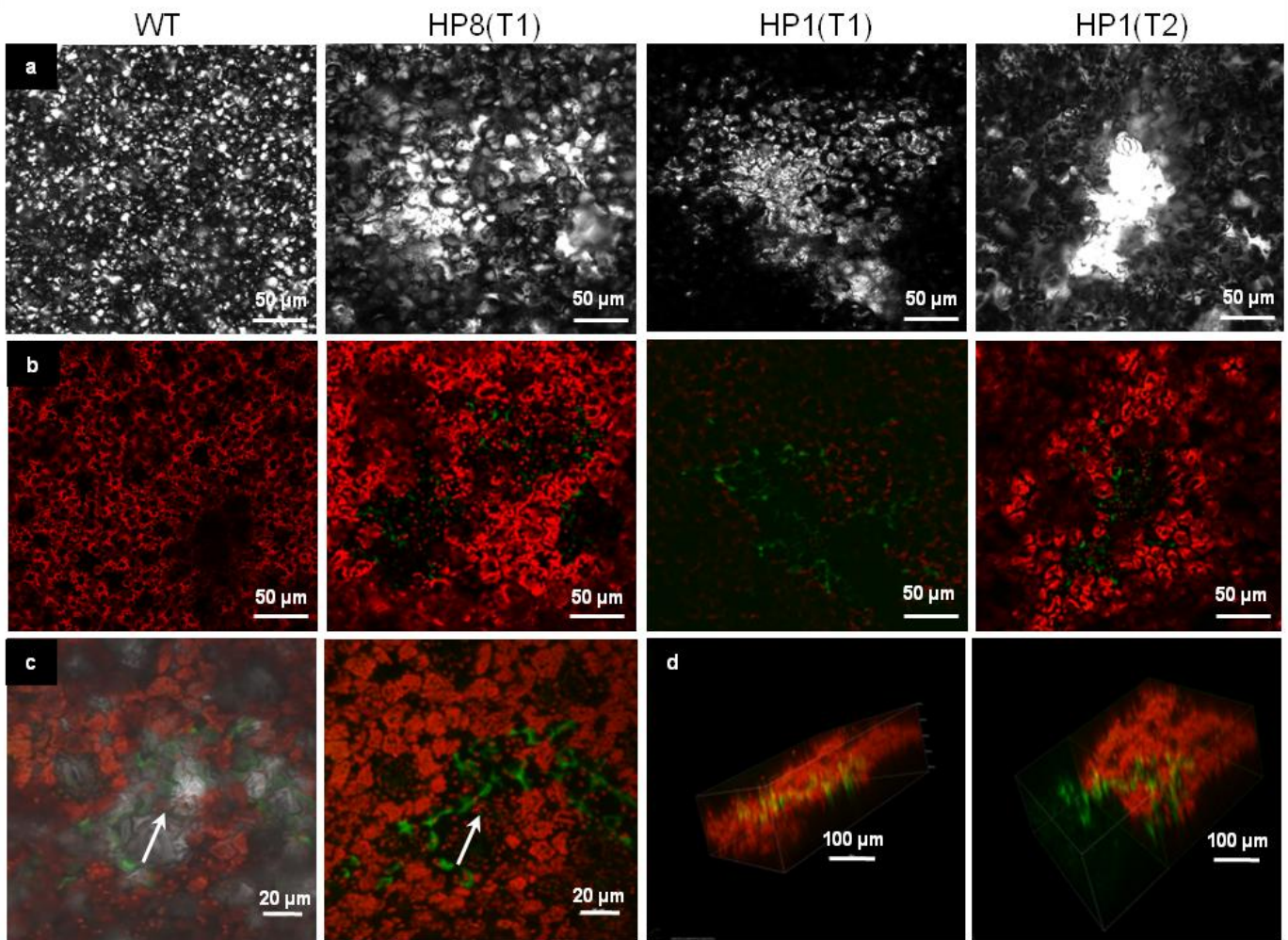


Figure 9 - Microscopic characterization of distinct tobacco lineages. a-d, confocal microscopy. a-b, columns from left to right: WT, HP8(T1), HP1(T1) and HP1(T2). a, Microscopic characterization of the cell death spot leaf phenotype. Leaves from 45-day-old plants presenting the cell death spots phenotype were captured through confocal microscopy in order to compare wild-type and the lineages HP8(T1), HP1(T1) and HP1(T2) using transmitted white light. b, red and green emitted spectra. Chloroplasts are in red and the region of the spots in green. c, HP5(T1) lineage in transmitted white light merged with red and green emitted spectra. White arrows indicate the region of one spot. d, 3D scan of HP5(T1) leaf in the region of a spot, with 43µm depth. To the left the whole area scanned and to the right the same area in cross-section. Chloroplasts are in red and the region of the spots in green

Aiming to exclude pleiotropic effects of the original HP constructs, novel HP plant lines were produced after the RNAi cassette was transferred to a second binary vector backbone that confers different antibiotic resistance to the plants. The *N. tabacum* transgenic lineages (HP14(T0), HP15(T0) and HP16(T0)) presented similar range of phenotypes and particularly the spotted leaves. Aniline blue staining confirmed callose deposition (Figure 10). It is known that hypersensitive responses (PONTIER et al., 1999) as well as leaf senescence (LIM; KIM; NAM, 2007) are



implicated in cell death. We conclude that the reproducibility of the phenotypes in these plants and in *N. benthamiana* makes it unlikely to be an unintended effect of the RNAi cassette.

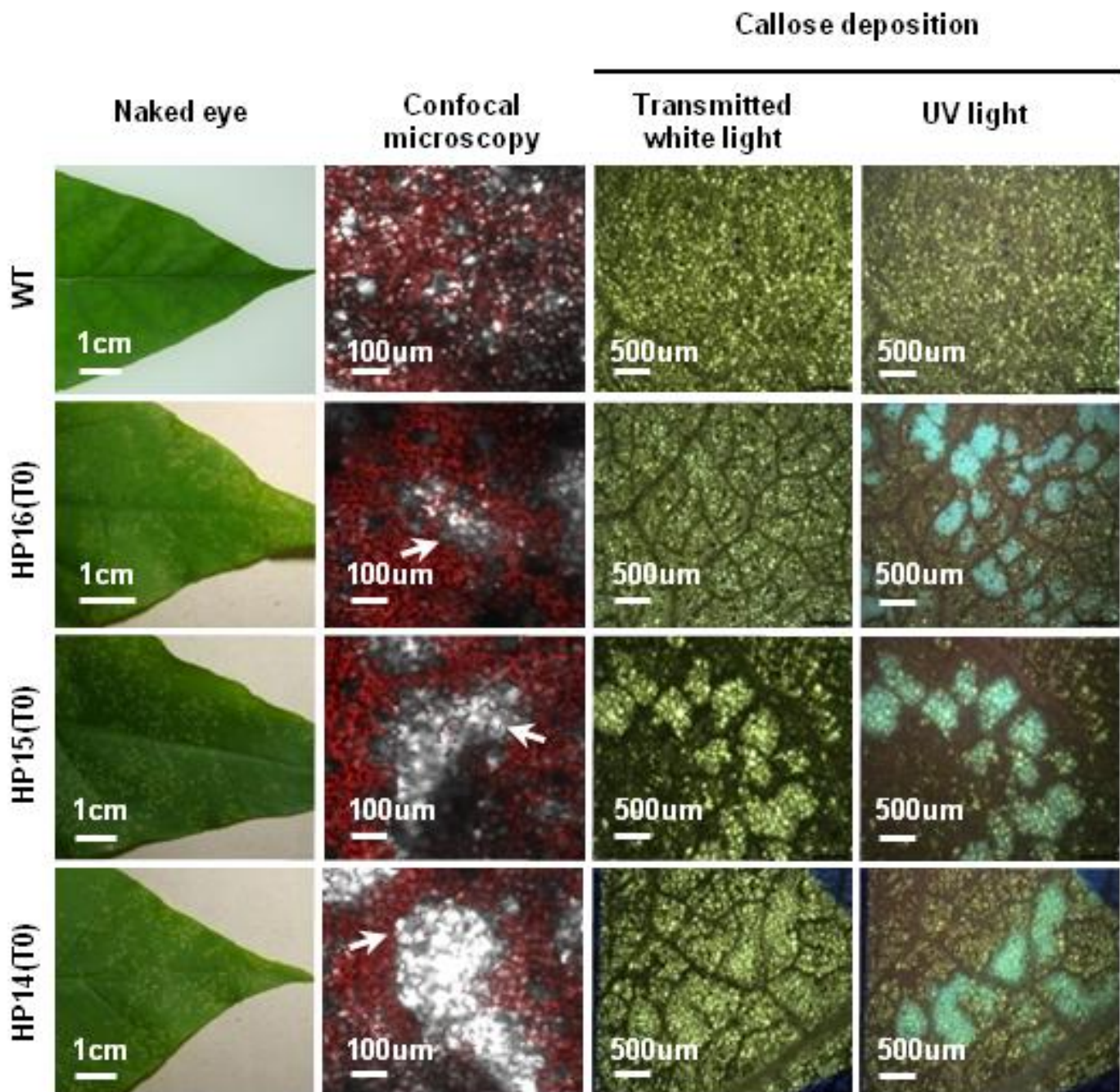


Figure 10 - Phenotypes and microscopic characterization of soil grown *Nicotiana tabacum* plants. Leaves from transformant plants (T0) presenting the cell death spots phenotype visible to the naked eye were captured in three different HP lines and WT. Confocal microscopy compares the healthy WT leaf tissue to the lesioned regions of HP leaves. Red spectra indicated chloroplasts. Callose deposition is demonstrated through aniline blue assay. To the left are the leaves under transmitted white light and to the right under UV light that shows callose deposited in the spots

The cell death phenotype is thus supported by changes in gene expression and gene networks. Based on these results, we propose a model that improves the current knowledge concerning the induction of Tnt1, ethylene signaling/responsive genes and other defense response genes by external biotic or abiotic stresses. Our data suggest a hierarchy of these events rather than parallel activation, placing Tnt1 in an upstream position in the regulation cascade (Figure 11).

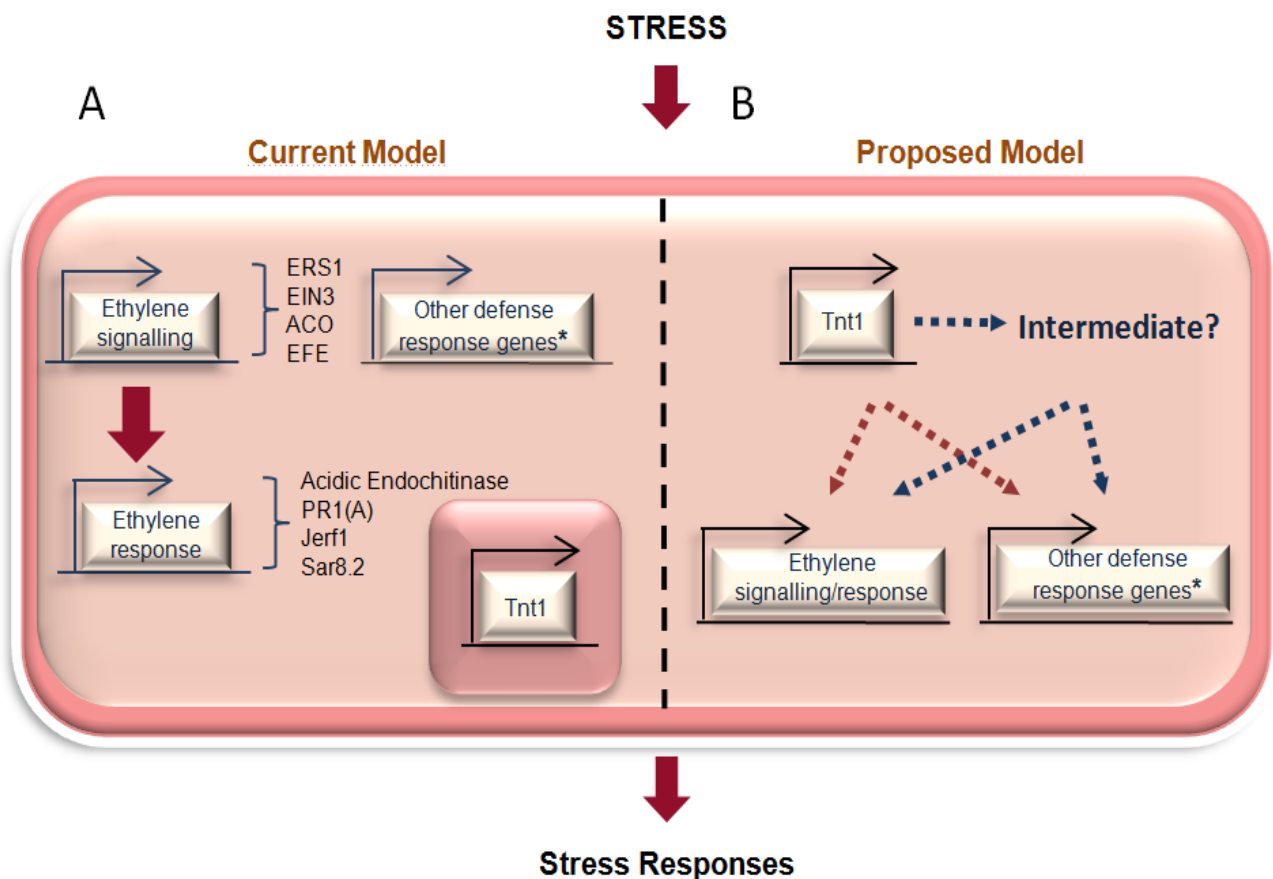


Figure 11 - Proposed model for a role of Tnt1 in stress and pathogen responses. (A) State of the art of the knowledge regarding the transcriptional response of Tnt1, ethylene signaling/response genes and other defense genes upon biotic or abiotic stresses. To date no hierarchic relationship has been established for these events. (B) Proposed model, in which Tnt1 has an upstream position in the process of stress and defense response. Transcriptional changes upon Tnt1 RT hairpin expression. The same genes responsive to biotic and abiotic stresses are modulated. Red and dark blue arrows represent two possibilities: red arrows indicate a cascade in which Tnt1 directly regulates the downstream genes while dark blue arrows assume the existence of an intermediate between Tnt1 and the downstream affected genes

## 2.3 Conclusions/Final considerations

Retrotransposons are abundant components of plant genomes, able to generate structural and functional variability. These elements can affect gene regulation and genome structure through transposition, recombination of non-homologous regions, and by being targets of epigenetic regulation mechanisms. On a global scale, retrotransposons can spread regulatory motifs through the genome, leading to the formation of new gene regulation networks (COWLEY; OAKEY, 2013; FESCHOTTE, 2008). Due to these features retrotransposons can be considered important players in genomes evolution. While deleterious effects are negatively selected, advantageous changes could be incorporated as mechanisms for gene regulation.

Our experimental design revealed that by introducing a shRNA construct from Tnt1 reverse transcriptase a wide gene reprogramming was observed. The interference with Tnt1 led to the emergence of pleiotropic phenotypes subjected to Mendelian segregation with the presence of the construct. One possible scenario is that different Tnt1 copies could be linked to the regulation of distinct *loci*, through insertion inside or near the promoter regions, or forming read-through transcripts leading to PTGS pathways (HERNANDEZ-PINZON et al., 2009), for example. Our results point to a Tnt1-mediated regulation of several concerted sets of genes. This reflects a remarkable gene reprogramming in an apparently complex network and put forward that TEs could control gene regulatory circuits. This was first conceived by Barbara McClintock in the early 40s, when transposable elements were defined as controlling elements of the gene suggesting that they could be active players in responses to environmental factors (MCCLINTOCK; 1984).

Furthermore, those genes relate to processes like defense response and response to stress, revealing a complex and mutualistic relationship between the retrotransposon and plant defense responses. Therefore we propose that Tnt1 has an active molecular role in gene circuitry in tobacco. Such evidences reinforce the idea that these retrotransposons are not always just genetic fossils or selfish genes, and apparently have the potential to evolve functions that are essential for plant growth and development. This work introduces a new coherent perspective on the study of transposable elements, aggregating evidences of the mobilome influence on

the global transcriptome regulation, and deepening the comprehension of the process through which retrotransposons impact their genomes.

## References

ARAUJO, P.G. de; ROSSI, M.; DE JESUS, E.M.; SACCARO, N.L.; KAJIHARA, D.; MASSA, R.; DE FELIX, J.M.; DRUMMOND, R.D.; FALCO, M.C.; CHABREGAS, S.M.; ULIAN, E.C.; MENOSSE, M.; VAN SLUYS, MA. Transcriptionally active transposable elements in recent hybrid sugarcane. **Plant Journal**, Oxford, v. 44, p. 707-717, Aug. 2005.

BAILLIE, J.K.; BARNETT, M.W.; UPTON, K.R.; GERHARDT, D.J.; RICHMOND, T.A.; DE SAPIO, F.; BRENNAN, P.M.; RIZZU, P.; SMITH, S.; FELL, M.; TALBOT, R.T.; GUSTINCICH, S.; FREEMAN, T.C.; MATTICK, J.S.; HUME, D.A.; HEUTINK, P.; CARNINCI, P.; JEDDELOH, J.A.; FAULKNER, G.J. Somatic retrotransposition alters the genetic landscape of the human brain. **Nature**, London, v. 479, p. 534-537, Oct. 2011.

BARGMANN, B.O.R.; MUNNIK, T. The role of phospholipase D in plant stress responses. **Current Opinion in Plant Biology**, Amsterdam, v. 9, p. 515-522, Oct. 2006.

CHOI, D.; BOSTOCK, R. M.; AVDIUSHKO, S.; HILDEBRAND, D. F. Lipid-derived signals that discriminate wound- and pathogen-responsive isoprenoid pathways in plants: methyl jasmonate and the fungal elicitor arachidonic acid induce different 3-hydroxy-3-methylglutaryl-coenzyme A reductase genes and antimicrobial isoprenoids in *Solanum tuberosum*. **Proceedings of National Academy of Sciences of the USA**, Washington, v. 91, n. 6, p. 2329-2333, Mar. 1994.

CORNELISSEN, B.J.C.; HOROWITZ, J.; KAN, J.A.L.; GOLDBERG, R.B.; BOL, J.F. Structure of tobacco genes encoding pathogenesis-related proteins from the PR-1 group. **Nucleic Acids Research**, Oxford, v. 15, n. 17, p. 6799-6811, Sept. 1987.

COWLEY, M.; OAKEY, R.J. Transposable elements re-wire and fine-tune the transcriptome. **PLOS Genetics**, San Francisco, v. 9, n. 1, Jan. 2013. Disponível em: <<http://www.plosone.org>>. Accessed in February 1st 2012.

DIETRICH, R.A.; DELANEY, T.P.; UKNES, S.J.; WARD, E.R.; RYALS, J.; DANGL, J.L. Arabidopsis mutants simulating disease resistance response. **Cell**, Cambridge, v. 77, p. 565-577, May 1994.

THE ENCODE CONSORTIUM. An integrated encyclopedia of DNA elements in the human genome. **Nature**, London, v. 489, p. 57-74, Dec. 2012.

FEDOROFF, N.V. Transposable elements, epigenetics and genome evolution. **Science**, New York, v. 338, n. 6108, p. 758-767, Nov. 2012.

FESCHOTTE, C. The contribution of transposable elements to the evolution of regulatory networks. **Nature Review Genetics**, London, v. 9, n. 5, p. 397–405, Dec. 2008.

GALLOIS, P.; MARINHO, P. Leaf disk transformation using *Agrobacterium tumefaciens*-expression of heterologous genes in tobacco. **Methods in Molecular Biology**, Totowa, v. 49, p. 3948, Nov. 1995.

GHANTA, S.; BHATTACHARYYA, D.; CHATTOPADHYAY, S. Glutathione signaling acts through NPR1-dependent SA-mediated pathway to mitigate biotic stress. **Plant Signal Behavior**, Austin, v. 6, n. 4, p. 607-609, Jan. 2011.

GRANDBASTIEN, M.A.; AUDEON, C.; BONNIVARD, E.; CASACUBERTA, J.M.; CHALHOUB, B.; COSTA, A.P.; LE, Q.H.; MELAYAH, D.; PETIT, M.; PONCET, C.; TAM, S.M.; VAN SLUYS, M. Stress activation and genomic impact of Tnt1 retrotransposon in *Solanaceae*. **Cytogenetic Genome Research**, Allschwilerstrasse, v. 110, p. 229-241, Feb. 2005.

GUO, Y.; GAN, S. AtNAP, a NAC family transcription factor, has an important role in leaf senescence. **Plant Journal**, Oxford, v. 46, p. 601-612, May 2006.

HERNANDEZ-PINZON, I.; DE JESUS, E.; SANTIAGO, N.; CASACUBERTA, J. M. The frequent transcriptional read-through of the tobacco Tnt1 retrotransposon and its possible implications for the control of resistance genes. **Journal of Molecular Evolution**, Berlin, v. 68, p. 269-278, Dec. 2009.

JÄÄSKELÄINEN, M.; CHANG, W.; MOISY, C.; SCHULMAN, A. H. Retrotransposon BARE displays strong tissue-specific differences in expression. **New Phytologist**, Lancaster, v. 200, p. 1000-1008, Oct. 2013.

KOCH, E.; SLUSARENKO, A. *Arabidopsis* is susceptible to infection by a downy mildew fungus. **Plant Cell**, Amsterdam, v. 2, p. 437-445, Feb. 1990.

LIM, P.O.; KIM, H.J.; NAM, H.G. Leaf senescence. **Annual Review of Plant Biology**, Palo Alto, v. 58, p. 115–136, Jan. 2007.

LINTHORST, H.J.M.; VAN LOON, L.C.; VAN ROSSUM, C.M.A.; MAYER, A.; BOL, J.F.; VAN ROEKEL, J.S.; MEULENHOF, E.J.; CORNELISSEN, B.J. Analysis of acidic and basic chitinases from tobacco and Petunia and their constitutive expression in transgenic tobacco. **Molecular Plant Microbe Interaction**, St. Paul, v. 3, n. 4, p. 252-258, July 1990.

LOPES, F.M.; MARTINS JR, D.C.; CESAR JR, R.M. Feature selection environment for genomic applications. **BMC Bioinformatics**, Berlin, v. 9, p. 451-460, Feb. 2008.

LUNA, E.; PASTOR, V.; ROBERT, J.; FLORS, V.; MAUCH-MANI, B.; TOM, J. Callose deposition: a multifaceted plant defense response. **Molecular Plant Microbe Interaction**, St. Paul, v. 24, n. 2, p. 183-193, Aug. 2011.

MCCLINTOCK, B. The significance of responses of the genome to challenge. **Science**, New York, v. 226, p. 792-801, Feb. 1984.

MORTAZAVI, A.; WILLIAMS, B.A.; MCCUE, K.; SCHAEFFER, L.; WOLD, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. **Nature Methods**, London, v. 5, n. 7, p. 621-628, Nov. 2008.

MUOTRI, A.R.; ZHAO, C.; MARCHETTO, M.C.N.; GAGE, F.; H. Environmental influence on L1 retrotransposons in the adult hippocampus, **Hippocampus**, Hoboken, v. 19, p. 1002-1007, May 2009.

PATERSON, A.H.; BOWERS, J.E.; BRUGGMANN, R. The *Sorghum bicolor* genome and the diversification of grasses. **Nature**, London, v. 457, p. 551-556, Sept. 2009.

PONTIER, D.; GAN, S.; AMASINO, R.M.; ROBY, D.; LAM, E. Markers of hypersensitive response and senescence show distinct patterns of expression. **Plant Molecular Biology**, New York, v. 39, p. 1243-1255, Apr. 1999.

QUIRINO, B.F.; NOH, Y.; HIMELBLAU, E.; AMASINO, R.M. Molecular aspects of leaf senescence. **Trends in Plant Science**, Cambridge, v. 5, n. 7, p. 278-282, Jan. 2000.

SAMBROOK, J.; FRIETSCH, E.F.; MANIATIS, T. Molecular cloning: a laboratory manual. **Cold Spring Harbor Laboratory**, Cold Spring Harbor, v. 3, p. 323-330, Feb. 1989.

SLOTKIN, R.K.; MARTIENSSEN, R. Transposable elements and the epigenetic regulation of the genome. **Nature Review Genetics**, London, v. 8, p. 272-285, Aug. 2007.

SONG, F.; GOODMAN, R.M. Cloning and identification of the promoter of the tobacco *Sar8.2b* gene, a gene involved in systemic acquired resistance. **Gene**, Amsterdam, v. 290, p. 115-124, Sept. 2002

TAKEMOTO, D.; YOSHIOKA, H.; DOKE, N.; KAWAKITA, K. Disease stress-inducible genes of tobacco: expression profile of elicitor-responsive genes isolated by subtractive hybridization. **Physiologia Plantarum**, Amsterdam, v. 118, p. 545-553, Nov. 2003.

VAN DE LAGEMAAT, L.N.; LANDRY, J.R.; MAGER, D.L.; MEDSTRAND, P. Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. **Trends in Genetics**, Amsterdam, v. 19, n. 10, p. 530-536, Feb. 2003.

VAN DER GRAAFF, E.; SCHWACKE, R.; SCHNEIDER, A. Transcription analysis of Arabidopsis membrane transporters and hormone pathways during developmental and induced senescence. **Plant Physiology**, Rockville, v. 141, p. 776-792, Apr. 2006.

VOINNET, O. Post transcriptional RNA silencing in plant-microbe interactions: a touch of robustness and versatility. **Current Opinion in Plant Biology**, Amsterdam, v. 11, n. 4, p. 464-70, Dec. 2008.

WESSLER, S.R. Transposable elements and the evolution of eukaryotic genomes. **Proceedings of National Academy of Sciences of the USA**, Washington, v. 103, n. 47, p. 17600-17601, Jan. 2006.

ZHANG, H.; HUANG, Z.; XIE, B. The ethylene-, jasmonate-, abscisic acid- and NaCl-responsive tomato transcription factor Jerf1 modulates expression of GCC box-containing genes and salt tolerance in tobacco. **Planta**, New York, v. 220, p. 262-270, Nov. 2004.

### 3 INTERPLAY BETWEEN RETROTRANSPOSON EXPRESSION AND ETHYLENE MEDIATES PLANT DEFENSE RESPONSES

#### Abstract

Retrotransposons are major components of eukaryotic genomes involved in their structural and functional evolution (NAKAYASHI, 2011) and yet no biological role has been attributed to them other than its own propagation or retrotransposon-derived domesticated genes (KOKOSAR; KORDIS, 2013). Tnt1 is a transcriptionally active tobacco (*Nicotiana tabacum*) retrotransposon responsive to a range of stress-associated signaling molecules (BEGUIRISTAIN et al., 2001). It is closely related TLC1.1 retrotransposon from *Solanum chilense*, which expression is induced by ethylene due to responsive elements present in its promoter region (TAPIA et al., 2005). Plant defense response genes are regulated by ethylene (ECKER; DAVIS, 1987). Our data is the first to make a correlation between ethylene and Tnt1, and we here propose a new biological role for the retrotransposon Tnt1 in cooperation with ethylene in tobacco defense responses as a homeostatic controller. It acts through a Tnt1-derived small RNA to down-regulate ethylene-responsive genes involved in the defense response back to their normal expression pattern after the environmental challenge has been overcome. Our pioneer work breaks a paradigm from 1980, when transposable elements were stated as selfish DNA which hosts couldn't benefit from. However, some genomes are composed 88% of those same "parasitic elements" (CIVAN; SVEC; HAUPTVOGEL, 2011), contradicting evolutionary fitness.

Keywords: Tnt1, retrotransposon, tobacco, RNA interference, plant defense responses, ethylene, gene regulatory network

#### 3.1 Introduction

Tnt1, a *Nicotiana tabacum* retrotransposon is composed of coding domains flanked by two long terminal repeats (LTRs), one at the 5' and the other at the 3' of Tnt1 sequence. In each of these LTRs are located Tnt1 promoter regions, called U3. Tnt1 is classified in three subfamilies (Tnt1A, TTnt1B and Tnt1C) which differ in their U3 sequences (VERNHETTES; GRANDBASTIEN; CASACUBERTA, 1998). When Tnt1 is transcribed by Pol II, the 3' promoter U3 is included in the transcript (KUMAR; BENNETZEN, 1999).

As described in De Jesus-Quintanilha et al. 2014 (submitted), we have generated transgenic RNA interference plants for the reverse transcriptase (RT) of Tnt1 (accession number X13777), and the transgenic plants (HP and AS plants) had a variety of phenotypes. We have also profiled the transcriptome of 45 days old HPs and wild-type plants (WT), which revealed that defense response related genes,



ethylene biosynthesis and downstream signaling cascade genes are upregulated in HP plants (GEO accession number GSE44027). In addition, we found 35 new gene networks established in the HP plants (which were not observed in the WT).

Gaseous phytohormone ethylene plays several roles in plant growth, development and defense (ECKER; DAVIS, 1987; WANG et al., 2013). The cis-acting GCC box is commonly found in the 5' upstream region of ethylene-responsive genes (SHINSHI; USAMI; OHME-TAKAGI, 1995). Ethylene-Responsive Element Binding Proteins (EREBPs) have conserved DNA-binding domains that interact with GCC box acting as transcriptional activators or repressors of gene expression (OHME-TAKAGI; SHINSHI 1995). In 2005 Grandbastien (GRANDBASTIEN et al., 2005) and collaborators reported the presence of repeats (AGCCAACCT) 71% similar to the GCC core (AGCCGCC) in the promoter at the U3 region of Tnt1 A subfamily.

The present work further explores the transcriptional dynamic of Tnt1, ethylene synthesis and ethylene response genes. We also aim to unravel how this transcription interplay that takes place is related to plant defense responses, and a biological model is proposed.

## **3.2 Development**

### **3.2.1 Methods**

#### **3.2.1.1 Plant samples**

*Nicotiana tabacum* HP transgenics were used (description in DE JESUS-QUINTANILHA et al., 2014, submitted). Plants were cultivated in MS20 media with the proper antibiotic, under a 14-hour photoperiod at 24°C.

#### **3.2.1.2 Online bioinformatics tools**

For the analysis used in the experiments we used the following tools available online:

tBLASTn:

<[https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=tblastn&PAGE\\_TYPE=BlastSearch&LINK\\_LOC=blasthome](https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=tblastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome)>. Accessed in December 26 2013

miRBase:<<http://www.mirbase.org/>>. Accessed in January 07 2014.

The mfold web server:<<http://mfold.rna.albany.edu/?q=mfold>>. Accessed in January 10 2014.

### **3.2.1.3 Ethylene treatment and ethylene gas chromatography**

For Tnt1 induction by ethylene plants were placed in sealed containers and then ethylene was taken from a concentrated stock (Alltech, Deerfield, IL) and injected into the containers using a syringe to give a final concentration of 10 mL/L. This concentration was monitored by gas chromatography every 6 h and remained stable throughout the treatment. Control plants were incubated in sealed containers without ethylene injection. The containers were opened after 24 h, leaf samples (300 mg) were collected and processed for total RNA isolation.

### **3.2.1.4 RNA isolation and Real Time PCR (qPCR)**

RNA was extracted using a modified LiCl method (SAMBROOK et al., 1989) and treated with DNaseI (Ambion). cDNA was synthesized using SuperScript First-Strand Synthesis System for RT-PCR (Life Technologies) following the manufacturer's instructions. The real time PCR reactions were performed using the SYBR Green Real-Time PCR Master Mix according to manufacturer's instructions. All the primers used for Real Time PCRs are described in Table 1.

Table 1 – GB accession number and primer sequences used in the quantitative PCRs to determined transcript levels of the genes in Figure 4

GB accession number	Gene name	Gene name in fig. 4	Forward primer	Reverse primer
AF026267.3	Ethylene receptor homolog (NTHK1)	NTHK1	GTG TGC AGA AAG CTG GTT CA	CCG CAG TTG AAA CCC AAG AA
AF203476.2	Putative ethylene receptor (NTHK2)	NTHK2	TGC AGC AGG CTA AGG AGA AT	CTG AGT GCA TTG GCT GTC TC
AB012857.1	ACC oxidase	ACC Oxidase 1	GCC AAA GAG CCA AGG TTT GA	AAA CCC CAT CCT TCC TCT CG
X98493.1	ACC-oxidase (clone TEFE)	ACC Oxidase 2	GCAACTATCCACCATGCCCAA	TGG AAG AGA AGG ATT ATG CCA CCG
DV162128.1	JERF1	JERF1	GACCTTGACGACGAATTTGAAGCTGACTTTTCAG	CCGTAAGCGAAAGCCTTGACTTCATCAACATC
FG642155.1	multi-protein bridging factor ER24	ER24	CAT TGC GAA CCG GAG TAC AG	AAC ATT AGC AGC AGC GGA AG
X51426.1	<i>Nicotiana tabacum</i> acidic chitinase PR-P	Acidic Chitinase PR-P	GGAAGTAGTGGTGATGATACTGCCCGTAGG	GCCTAACAAAGCAATATCCTCCTGTAAAAGGTTCTG
EH621848.1	<i>Nicotiana tabacum</i> Sar8.2b	Sar8.2b	GGTAATATCCTCACAAGCTGATGCAAGGGA	GGTATCTTACGGATGATTCTGCACCCG
AB015855.1	Transcription factor TEIL	TEIL	AAATGGGGTTCTGTGGCGAT	TCCTGCTCAATCTGCGAAGC
X16938.1	Endochitinase CHN48	CHN48	GCCTTTATCAATGCTGCTCGG	GAGCAAAGAAAGCCGCGATT
X13777.1	Tnt1 Reverse Transcriptase	Tnt1 Reverse Transcriptase	GCG ATT CCT TCA AGA GCT TGG ATT GCA TCA	CAT AGA GTT CTT GCT AAG GTC TAT TGC ACT TTG ACT
X13777.1	Tnt1 Integrase	Tnt1 Integrase	GTG AAG CAG TTC AGA CAG CCT GTT ACC TGA TC	CCT TCA GAT GCG AGT AGG ACA CCT CCT TG
X13777.1	Tnt1 Gag	Tnt1 Gag	TTGCCAACCTCGGAGTGAAA	GCAGGATGGTTGTTGCCAAAT
L18908.1	TOBRPL25A <i>Nicotiana tabacum</i> ribosomal protein L25	L25 (Constitutive)	CAGCTAAAGCTGATCCGTCCA	TGATCCTGACTTGACAGCCTTG

### 3.2.1.5 Histochemical GUS assay

Histochemical GUS assays were performed on transgenic plants. In vitro Grown in the dark seedlings collected 18 days. Histochemistry was performed by placing whole seedlings in the appropriate staining buffer containing X-Gluc (0.5 mg mL<sup>-1</sup> 5-bromo-4-chloro-3-indolyl-b-D-glucuronide in 50 mM sodium phosphate, pH 7, 0.05% Triton X-100, 0.1 mM K<sub>4</sub>Fe(CN)<sub>6</sub>·3H<sub>2</sub>O, 0.1 mM K<sub>3</sub>Fe(CN)<sub>6</sub>) and incubated in the dark at 37°C for 48 hours. After staining seedlings were preserved in glycerol (100%), and photographed.

### 3.2.2 Results

There are 2 GCC-like motifs in the 5' U3 and 3 GCC-like motifs in the 3' U3 of Tnt1A (those motifs were not found in Tnt1B or Tnt1C subfamilies). The location of the three GCC-like motifs as well as the putative TATA box of the 3' promoter region are depicted in figure 1b. When Tnt1 is transcribed by Pol II, the 3' promoter U3 is included in the transcript (KUMAR; BENNETZEN, 1999). Folding prediction of part of the Tnt1A messenger RNA (mRNA) which includes three 3' U3 GCC-like motifs was obtained using the platform The mFold Web Server (<<http://mfold.rna.albany.edu/?q=mfold/rna-folding-form>>. Accessed in January 10 2014). The folding of the mRNA has potential to form stem-loop secondary structures where GCC-like motifs would be located in the arms of the putative hairpin loops (Figure 1c). This led us to hypothesize that Tnt1 could be a putative source of GCC small RNAs (sRNAs), supported by the fact that the folded RNA free energy ( $dG = -26.10$ ) is within the range associated to known plant miRNAs (THAKUR et al., 2011). We then searched the plant mature miRNA database (<<http://www.mirbase.org/>>. Accessed in January 10 2014) for sequences similar to this putative sRNA that includes the 3' U3 GCC-like motifs. The mature miRNA bdi-miR7783 expressed in *Brachypodium distachyon* under drought stress (BERTOLINI et al., 2013) showed 75% similarity to the second GCC-like box found in Tnt1 3'U3 (Figure 1c), further supporting our hypothesis.

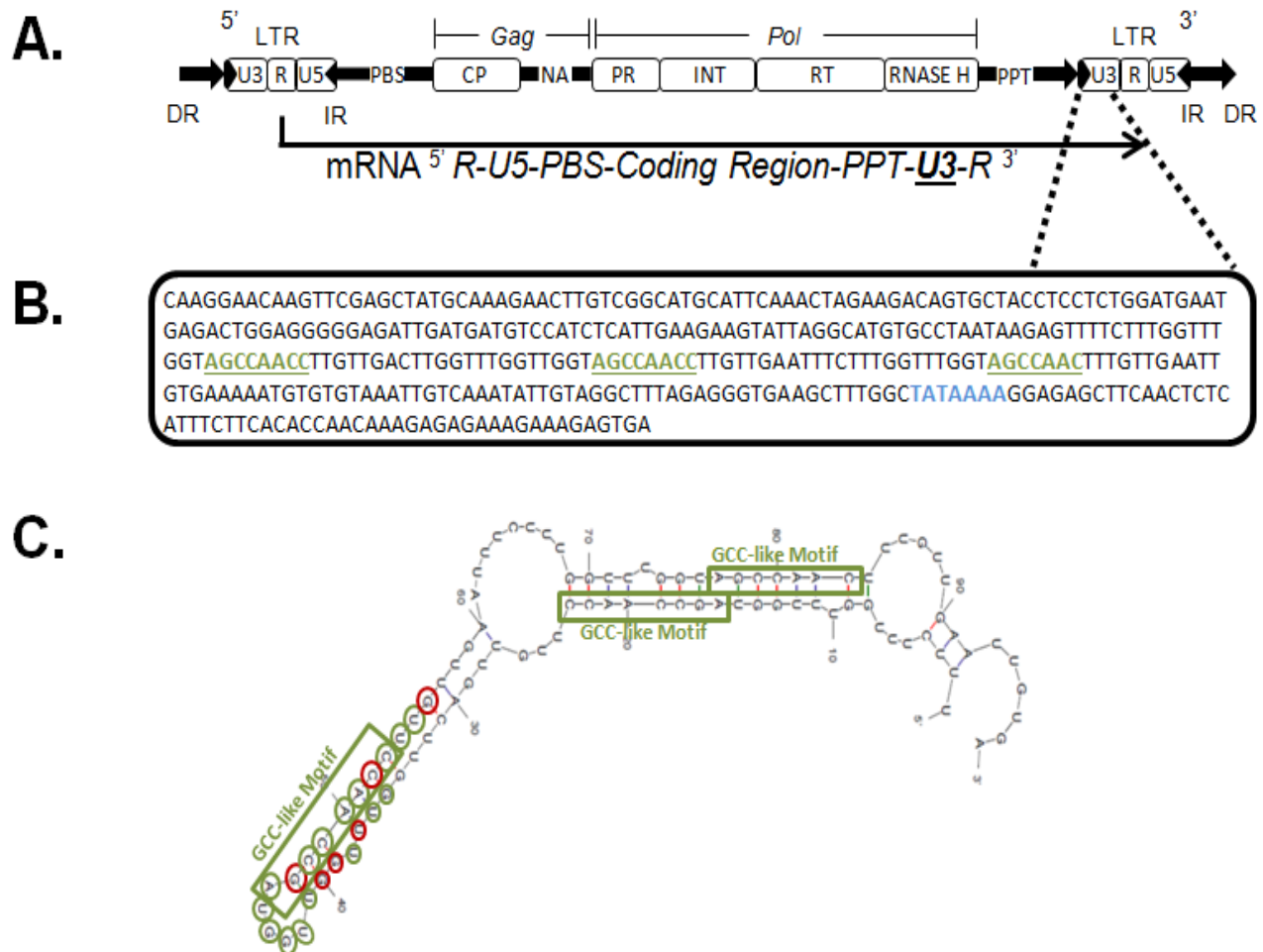


Figure 1 - The discovery of a putative Tnt1-derived GCC-box small RNAs in the promoter region of Tnt1A. (a) Schematization of Tnt1 retrotransposon structure, with 5' LTR divided in U3 (unique 3') R (repeated RNA) and U5 (unique 5'), PBS (primer binding site), the protein domains contained in the element, capsid-like protein (CP), protease (PR), integrase (INT), reverse transcriptase (RT) and RNase H (RNASEH), PPT (polypurine tract) and the 3' LTR. IR are inverted terminal repeats inside the LTRs, and DR are direct repeats generated by the insertion of the element in the genome. The transcript initiates at 5' end of R in the 5' LTR and terminates at 3' end of R in the 3' LTR, as shown by the black arrow. (b) 3' U3 region of Tnt1A in detail. Green underlined words indicate 3 GCC-like motifs and TATA box is in blue. (c) Folding of a 98 bp part of the Tnt1A mRNA sequence containing the three GCC-like boxes. The GCC-like boxes are highlighted inside the green rectangles. Green spheres around a nucleotide represent the similarity to the mature miRNA bdi-miR7783, and red spheres represent the mismatches

Since there are ethylene responsive elements (GCC-like motifs) in the promoter region of the retrotransposon, we then proceeded to test if Tnt1 is transcriptionally activated upon ethylene stimulus. We treated 2 WT plants and 2 HP1(T2) with 10  $\mu\text{L/L}$  of ethylene, and other 2 individuals of each genotype did not receive the treatment (control group). After 24 hours of treatment we quantified the expression (through quantitative PCR) of 3 different Tnt1 coding domains in all

individuals. Among plants that didn't receive ethylene treatment, HP1(T2) plants had higher expression level of all three Tnt1 domains than the WT (Figure 2a). Upon ethylene treatment both WT and HP1(T2) plants overexpressed the three Tnt1 domains when compared to the control group (Figure 2b). However, for ethylene-treated plants, WT had higher expression levels of Tnt1 than HP1(T2). This could be explained by the expression of the T-DNA hairpin of Tnt1 impairing expression upon ethylene induction.

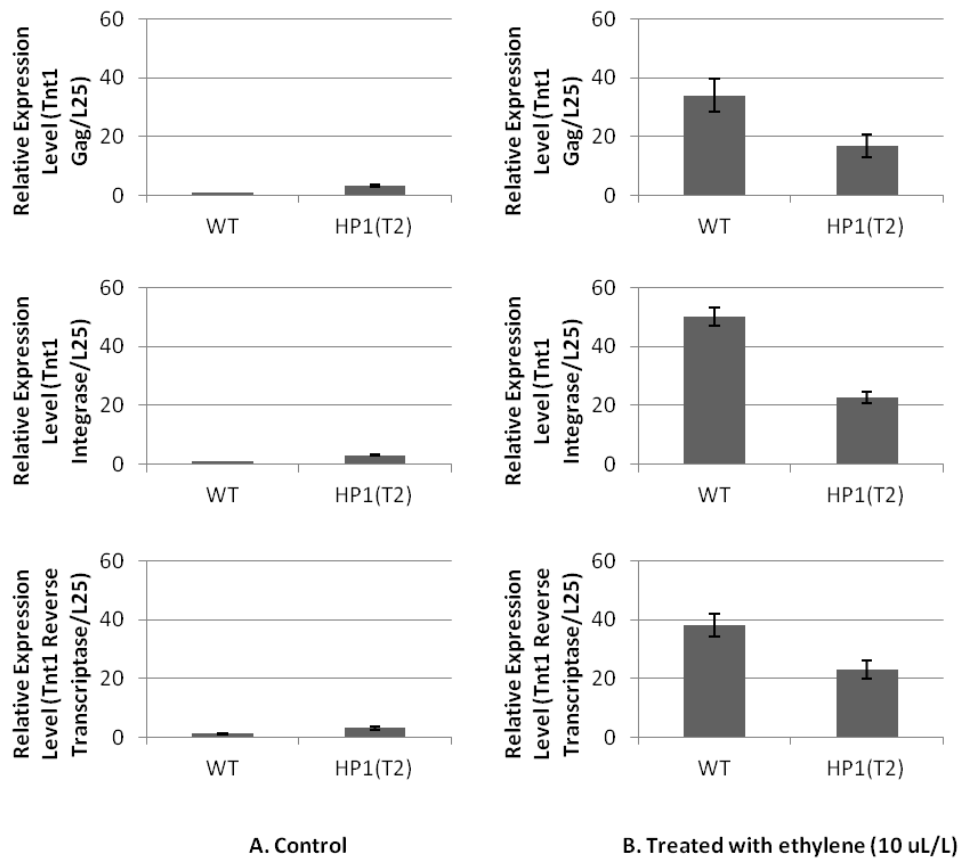


Figure 2 - Expression of three Tnt1 protein domains is induced upon exogenous ethylene application. The expression of Gag, Integrase and Reverse Transcriptase was measured through qRT-PCR, carried using two biological replicates, with 4 technical replicates each. Values are means, with bars showing standard errors. (a) Plants treated with ethylene. The plants were placed in sealed containers and then ethylene was injected to the final 10 uL/mL concentration, monitored by gas chromatography. (b) Control plants were incubated in sealed containers without any chemical

Once genes related to the ethylene biosynthesis and ethylene response were positively modulated in HP plants (DE JESUS-QUINTANILHA et al., 2014, submitted), we tested HP plants for ethylene production. When germinating in the

dark, WT seedlings produce more ethylene, which confers a characteristic “triple response”: hypocotyl and root reduced elongation, apical hook exaggeration and hypocotyl swelling (NELJUBOV, 1901). Thus we germinated seeds of WT and four independent third generation HP lines (HP1, HP5, HP8 and HP13) in non-selective plant media in the dark. Histochemical assay for *uidA* (GUS) reporter gene was performed to select the transgenic HPs after plants were grown in the dark for 18 days (Figure 3a). The ratio of root length to total seedling length (Figure 3b) was significantly smaller in all HP lines when compared to the WT, and apical hook exaggeration was observed in some HP seedlings (Figure 4a). Since we have identified two of triple response phenotypes in HP seedlings, it is likely that these HP lines overproduce ethylene from a very early stage of development. Next we tested adult HP plants to check if they keep the overproduction of ethylene at an older age. Ninety days old WT, HP1(T1), HP1(T2), HP8(T1) and AS11(T1) plants were used to detect ethylene production using gas chromatography. All HP lines showed significant ethylene overproduction when compared to WT (Figure 4b). With these data we conclude that not only are the genes related to ethylene biosynthesis upregulated in adult 45 days old HP plants (DE JESUS-QUINTANILHA et al., 2014, submitted), but also that HP plants are overproducing the phytohormone.

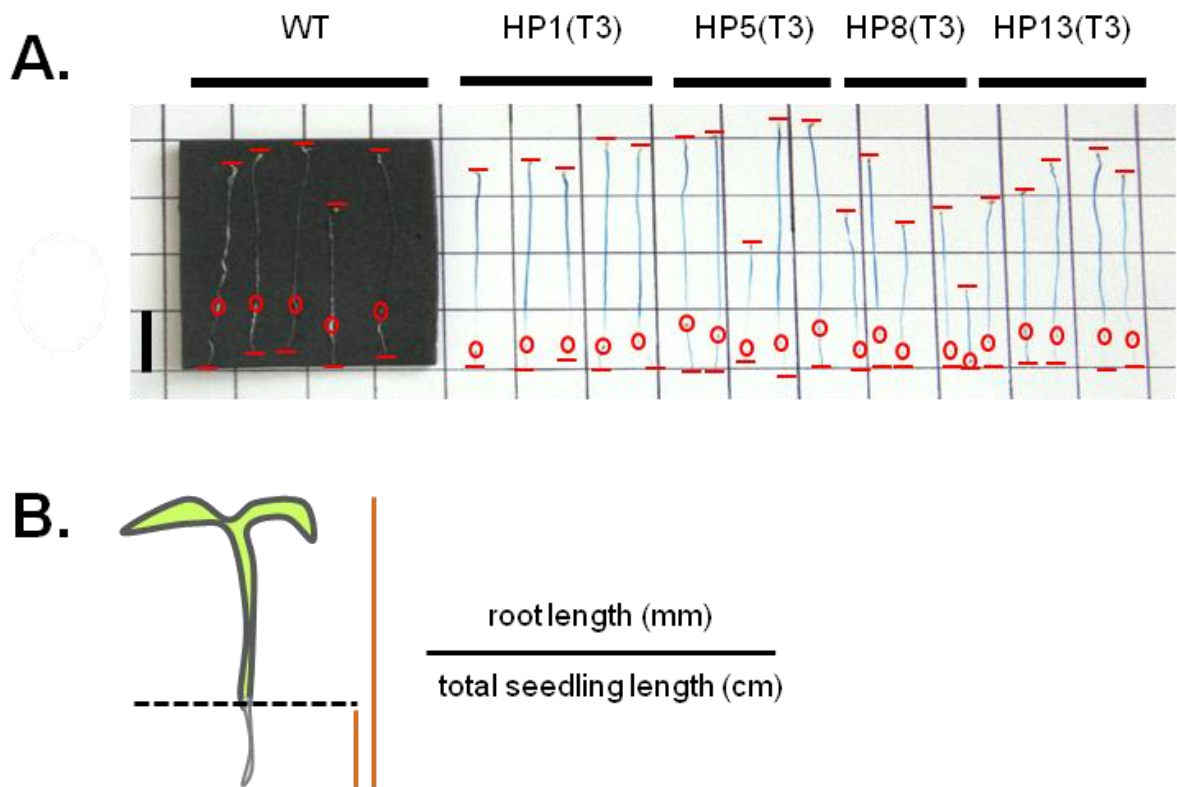


Figure 3 - Ethylene triple response in HP seedlings grown in the dark. (a) GUS assay to verify transgenic HP seedlings. Measurement revealed a reduced elongation of transgenic dark-grown HP seedlings compared to WT, corroborating the triple response and indicating ethylene production. Black bar represents 1 cm. Red circles indicate the transition of the hypocotyl to root, and the red horizontal lines delimit the total seedling length (b) scheme of how the ratio calculated for the box plot in figure 3 was calculated. The ratio was the root length (in mm) per total seedling length (in cm)



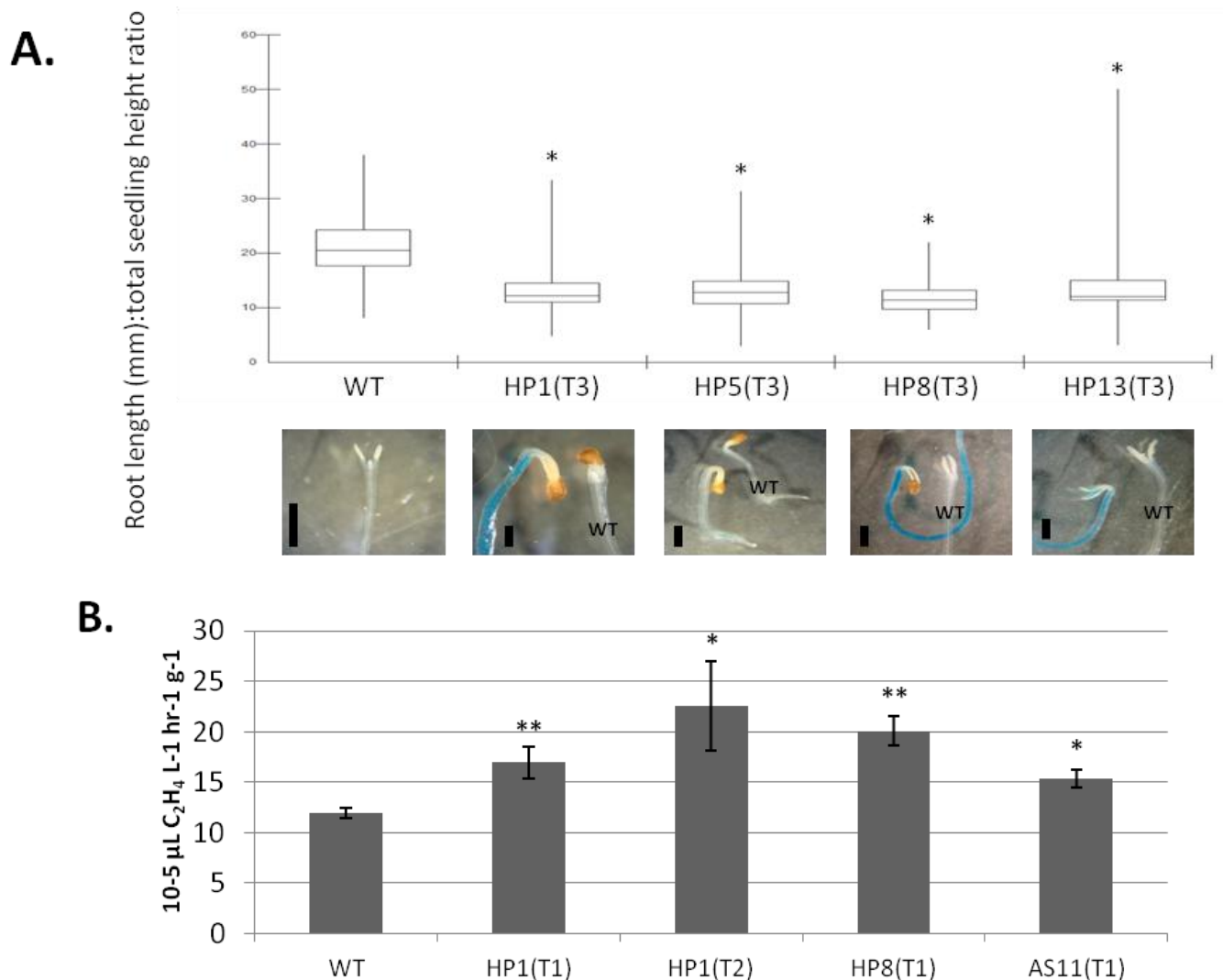


Figure 4 - Ethylene triple response in 10 days old HP seedlings and ethylene production in 90 days old HP plants. (a) Box plot representing the the ratio of root length (in mm) to the total seedling length (in cm) of 50 seedlings from each line. Significance of the difference between HP and WT was assessed by a one-tailed Student's t-test: (\*) significant at the 1.0E-06 level. Bellow the box-plot the photos show the apical hook phenotype in GUS-positive seedlings (black bar represents 1 millimeter). (b) Ethylene production in 90 days old HP plants. The plants were sealed in containers, the air was reseted and then ethylene concentration after 24 hours was measured through gas chromatography. Values are means, with bars showing standard errors. Significance of the difference between HP and WT was assessed by a one-tailed Student's t-test: (\*\*) significant at the 0.004 level; (\*) significant at the 0.017 level. Ethylene production was higher in HP plants, according to the expected after the triple response was confirmed

To further analyze the correlation between Tnt1 expression and ethylene biosynthetic and responsive genes, 90 days old WT and HP plants (the same used for measurement of ethylene in gas chromatography) were also used for quantitative PCR. Two tobacco ethylene receptor homolog genes NTHK1 and NTHK2, which are

involved in ethylene perception and negatively regulate ethylene responses (HUA; MEYEROWITZ, 1998), were upregulated in HP plants (Figure 5). All the other genes assayed related to ethylene synthesis and ethylene response were downregulated in HP plants when compared to the WT (Figure 5), and they are: two ACC oxidase genes, which code for the enzyme involved in the last step of ethylene synthesis (HAMILTON; BOUZAYEN; GRIERSON, 1991); JERF1, an ethylene responsive factor that can bind to GCC motifs and enhance expression of plant defense response genes (ZHANG et al., 2004); ER24, an ethylene-regulated gene that codes for a TATA Binding Protein-associated factors complex (TAF) which might be a bridging protein between EREBPs and TATA box binding proteins required for transcription initiation (ZEGZOUTI, 1999); Sar8.2 gene that is expressed in an ethylene-dependent way upon tobacco mosaic virus infection and following induction of the Systemic Acquired Resistance (SAR) (GUO et al., 2000); TEIL, the EIN3 tobacco homolog, which belongs to family of nuclear proteins involved in the ethylene responses (SOLANO et al., 1998) and lastly CHN48 which is an ethylene induced tobacco gene that codes for a class I basic chitinase8. The presence of the transgene in heterozygosity could have interfered in a copy number-dependent manner in gene expression among different first generation (T1) HP lines. We also assayed three different Tnt1 coding domains (Gag, Integrase and Reverse Transcriptase), and they were downregulated in all HP lines (Figure 5).

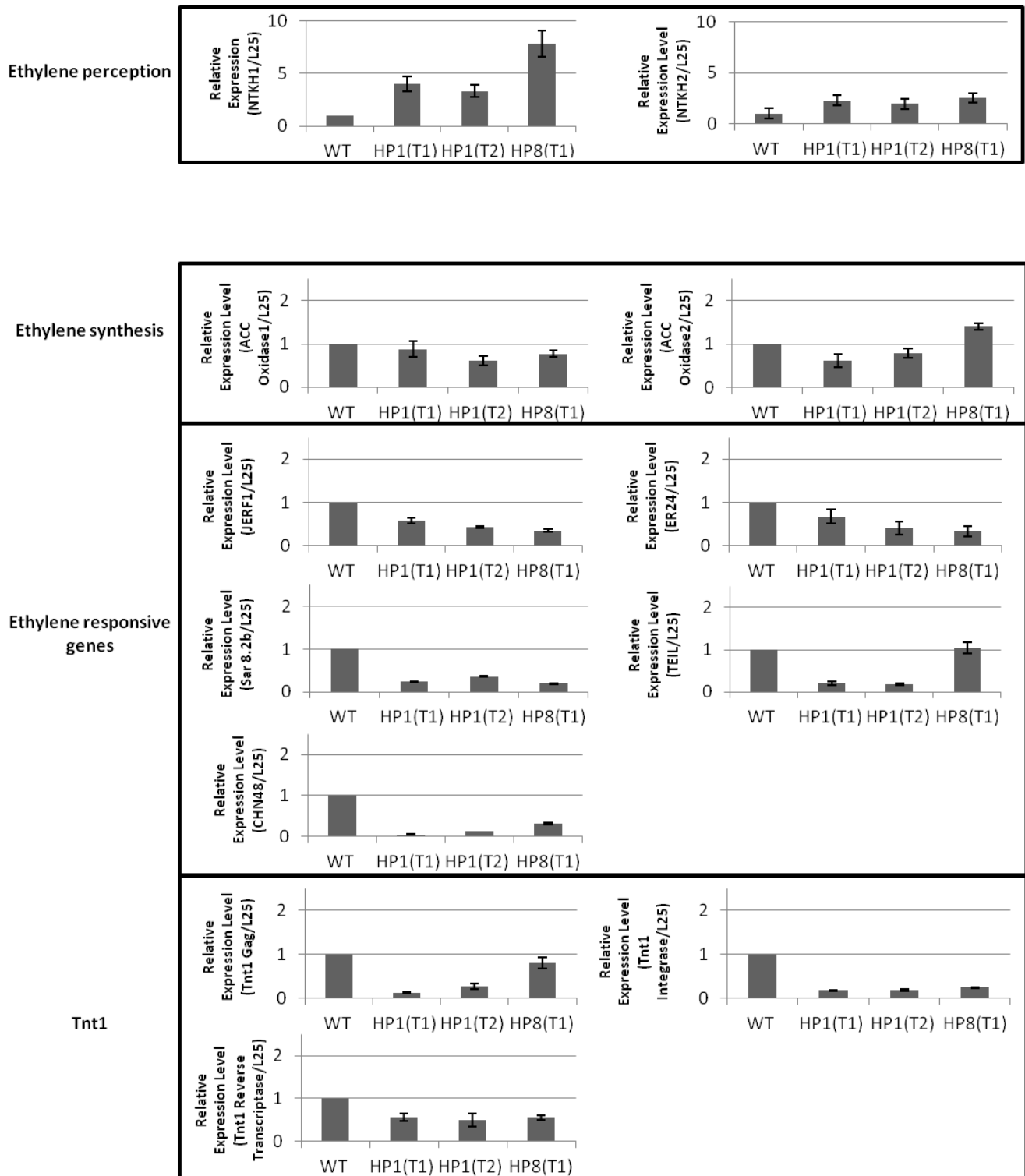


Figure 5 - Quantitative expression of ethylene-related genes and Tnt1 in 90 days old HP plants. Expression levels were measured through qRT-PCR. Genes related to ethylene perception (NTKH1 and NTKH2), ethylene biosynthesis (ACC Oxidase1 and ACC Oxidase2) as well as ethylene responsive genes (JERF1, ERF4, Sar8.2b, TEIL and CNH48) were analyzed. Tnt1 coding regions Gag, Reverse Transcriptase and Integrase were also evaluated in WT, HP1(T1), HP1(T2) and HP8(T1). Two biological replicates were used for each line (each with four technical replicates), and the expression level of the WT was set as 1. Values are means, with bars showing standard errors. The leaves used to extract the RNA for the quantitative PCRs were from the same plants used for the ethylene gas chromatography

### 3.3 Conclusions/Final considerations

It is interesting that the expression kinetics of these genes in 90 days old plants which overproduce ethylene differ from the expression kinetics of 45 days old plants used in RNA sequencing (DE JESUS-QUINTANILHA et al. 2014, submitted). RNA sequencing of 45 days old plants revealed little difference in Tnt1 expression pattern among HPs and the WT (except for an uneven increase in transcript number in all Tnt1 domains except for the Reverse Transcriptase in HP1(T2)). It also revealed up-regulation of genes involved in the ethylene biosynthesis and response. These results in a “phase opposition” fashion led us to hypothesize that there might be an antagonistic relationship between the expression of Tnt1 and the expression of ethylene responsive genes involved in plant defense responses.

We have incorporated in our model (Figure 6) two different conditions: a period of time when a WT plant exists under ideal conditions for normal growth and another under a stress pressure, seeking explanation of the interaction between the expression of Tnt1, ethylene biosynthesis genes and ethylene responsive genes that contain GCC motifs in their promoter. Under normal growth conditions, WT tobacco plants express Tnt1 in a basal level (we refer to that basal level of expression as “range of homeostasis”) (Figure 6a). Genes involved in ethylene biosynthesis and other ethylene responsive genes that have GCC motifs in their promoters are also expressed within a homeostasis range (since ethylene is required in various moments of plant development and not only during plant defense responses) (Figure 6b and c). Upon a stress stimulus, ethylene biosynthesis is increased, and this is one of the events that define the commencement of the defense responses (Figure 6d). The overproduction of ethylene triggers the up-regulation of genes that contain GCC motifs in their promoter (such as Tnt1 and other responsive genes) (Figure 6e and f). According to our analysis, it is possible that Tnt1A mRNA can be a source of small RNAs that target GCC motifs. Thus, as Tnt1 is overexpressed in response to ethylene, there is a turning point (Figure 6g) in which the Tnt1A sRNAs 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 (Figure 6h). We find examples in the literature of miRNAs that generate small RNAs able to promote DNA methylation of target genes in *trans* (WU et al., 2014). Interestingly, in humans a miRNA was described to form a complex with

Polycomb proteins, return to the nucleus and successfully target the promoter of a gene and repress its transcription (ZARDO et al., 2012). So we reckon it is conceivable that Tnt1A sRNAs can direct transcriptional gene silencing of target genes through homology recognition and attachment to the DNA sites of the GCC motifs in the promoter region of ethylene responsive genes, thus downregulating these 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 (Figure 6i). 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.

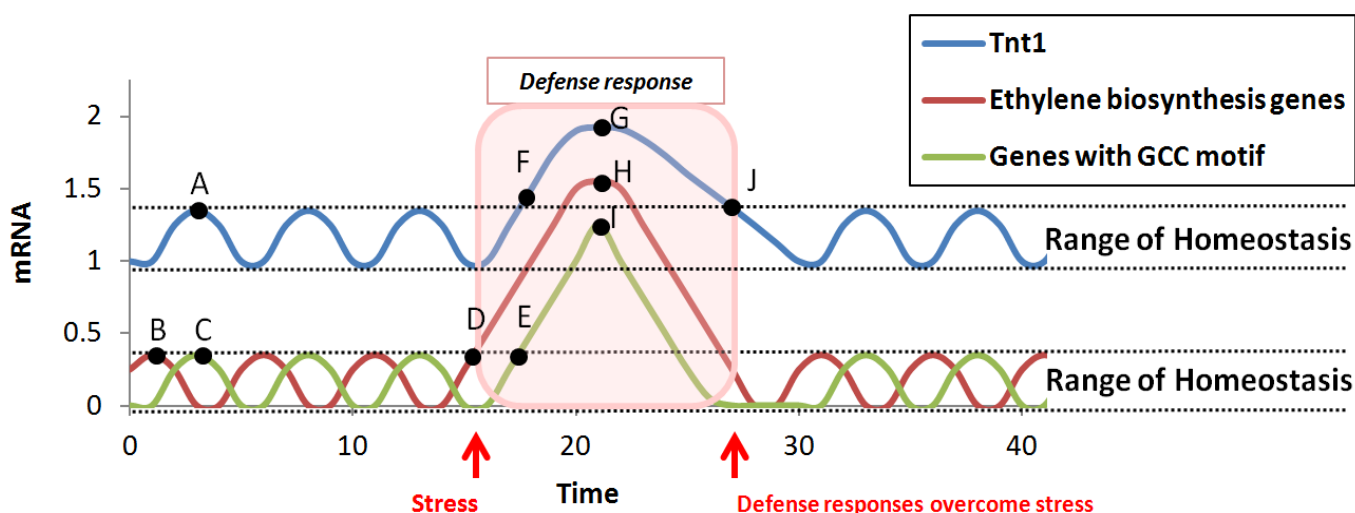


Figure 6 - Model proposing a dynamic equilibrium between expression of defense response genes and Tnt1 in WT. X axis represent time while Y axis the level of transcripts of the analyzed genes. (a) Tnt1 presents a basal level of expression within a range of homeostasis. (b) Genes of the ethylene biosynthetic and response pathway also present a basal level of expression in leaves. (c) In response to the oscillation of the genes belonging to the ethylene biosynthetic pathway, GCC containing genes expression also oscillate within a range of homeostasis induced by the presence of the ethylene. (d) Stress stimulus induces ethylene biosynthesis. This represents the start of the defense response. (e, f) Genes that contain GCC motifs in their promoter are up-regulated in the presence of ethylene. (g) Tnt1-U3A GCC-like motifs miRNAs overcomes ethylene induction and targets Tnt1 messenger RNA (which contain the U3 region). (h) Tnt1-U3A GCC-like motifs miRNAs overcomes ethylene induction and promotes transcriptional inhibition of GCC motifs in other ethylene-responsive gene promoters. (i) Defense responses are at its highest and probably the stress has been overcome. Thus, ethylene biosynthesis starts to return to levels within the range of homeostasis. (j) Tnt1 returns to normal levels of expression after reduction in ethylene levels initiated in (h) and also due to the repression mediated by its own miRNA

Our findings suggest that an active retrotransposon can generate sRNAs that exerts transcriptional control over itself as well as other genes. In our model we thus establish a completely new biological role for Tnt1: 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, resuming the defense responses which would no longer be needed.

## References

- BEGUIRISTAIN, T.; GRANDBASTIEN, M.A.; PUIGDOMÈNECH, P.; CASACUBERTA, J.M. Three Tnt1 subfamilies show different stress-associated patterns of expression in tobacco: consequences for retrotransposon control and evolution in plants. **Plant Physiology**, Rockville, v. 127, p. 212–221, Jan. 2001.
- BERTOLINI, E.; VERELST, W.; HORNER, D.S.; GIANFRANCESCHI, L.; PICCOLO, V.; INZÉ, D.; PÈ, M.E.; MICA, E. Addressing the role of microRNAs in reprogramming leaf growth during drought stress in *Brachypodium distachyon*. **Molecular Plant**, Oxford, v. 6, p. 423–443, Dec. 2013.
- CIVÁN, P.; SVEC, M.; HAUPTVOGEL, P. On the coevolution of transposable elements and plant genomes. **Journal of Botany Article**, Cairo, v. 2011, p. 1-9, Apr. 2011.
- ECKER, J.R.; DAVIS, R.W. Plant defense genes are regulated by ethylene. **Proceedings of the National Academy of Sciences of the USA**, Washington, v. 84, p. 5202-5206, Oct. 1987.
- GRANDBASTIEN, M.A.; AUDEON, C.; BONNIVARD, E.; CASACUBERTA, J.M.; CHALHOUB, B.; COSTA, A.P.; LE, Q.H.; MELAYAH, D.; PETIT, M.; PONCET, C.; TAM, S.M.; VAN SLUYS, M. Stress activation and genomic impact of Tnt1 retrotransposon in *Solanaceae*. **Cytogenetics Genome Research**, Allschwilerstrasse, v. 110, p. 229-241, Feb. 2005.
- GUO, A.; SALIH, G.; KLESSIG D.F. Activation of a diverse set of genes during the tobacco resistance response to TMV is independent of salicylic acid; induction of a subset is also ethylene independent. **The Plant Journal**, Rockville, v. 21, n. 5, p. 409-418, Apr. 2000.
- HAMILTON, A.J.; BOUZAYEN, M.; GRIERSON, D. Identification of a tomato gene for the ethylene-forming enzyme by expression in yeast. **Proceedings of the National Academy of Sciences of the USA**, Washington, v. 88, p. 7434-7437, Sept. 1991.
- HUA, J.; MEYEROWITZ, E.M. Ethylene responses are negatively regulated by a receptor gene family in *Arabidopsis thaliana*. **Cell**, Cambridge, v. 94, p. 261–271, Aug. 1998.

KOKOSAR, J.; KORDIS, D. Genesis and regulatory wiring of retrotransposon-derived domesticated genes: a phylogenomic perspective. **Molecular Biology Evolution**, Oxford, v. 30, n. 5, p. 1015–1031, Feb. 2013.

KUMAR, A.; BENNETZEN, J.L. Plant retrotransposons. **Annual Review of Genetics**, Palo Alto, v. 33, p. 479–532, Jan. 1999.

NAKAYASHIKI, H. The Trickster in the genome: contribution and control of transposable elements. **Genes to Cells**, New York, v. 16, p. 827–841, Apr. 2011.

NELJUBOV, D. Ueber die horizontale Nutation der Stengel von *Pisum sativum* und einiger anderen. **Journal of Plant Growth Regulation**, Amsterdam, v. 10, p. 128–139, Jan. 1901.

OHME-TAKAGI, M.; SHINSHI, H. Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element. **The Plant Cell**, Rockville, v. 7, p. 173–182, Nov. 1995.

SAMBROOK, J.; FRIETSCH, E.F.; MANIATIS, T. Molecular cloning: a laboratory manual. **Cold Spring Harbor Laboratory**, Cold Spring Harbor, v. 3, p. 323–330, Feb. 1989.

SHINSHI, H.; USAMI, S.; OHME-TAKAGI, M. Identification of an ethylene-responsive region in the promoter of a tobacco class I chitinase gene. **Plant Molecular Biology**, New York, v. 27, p. 923–932, Feb. 1995.

SOLANO, R.; STEPANOVA, A.; CHAO, Q.; ECKER, J.R. Nuclear events in ethylene signalling: a transcriptional cascade mediated by ethylene-insensitive3 and ethylene-response-factor1. **Genes & Development**, Woodbury, v. 12, p. 3703–371, Jan. 1998.

TAPIA, G.; VERDUGO, I.; YANEZ, M.; AHUMADA, I.; THEODULOZ, C.; CORDERO, C.; POBLETE, F.; GONZALEZ, E.; RUIZ-LARA, S. Involvement of ethylene in stress-induced expression of the TLC1.1 retrotransposon from *Lycopersicon chilense* Dun. **Plant Physiology**, Rockville, v. 138, p. 2075–2086, Aug. 2005.

THAKUR, V.; WANCHANA, S.; XU, M.; BRUCKIEWICH, R.; QUICK, W.P.; MOSIG, A.; ZHU, X.G. Characterization of statistical features for plant microRNA prediction. **BMC Genomics**, Berlin, v. 12, p. 108–111, Aug. 2011.

VERNHETTES, S.; GRANDBASTIEN, M.A.; CASACUBERTA, J.M. The evolutionary analysis of the Tnt1 retrotransposon in Nicotiana species reveals the high variability of its regulatory sequences. **Molecular Biology Evolution**, Oxford, v. 15, n. 7, p. 827–836, Dec. 1998.

WANG, F.; CUI, X.; SUN, Y.; DONG, C. Ethylene signaling and regulation in plant growth and stress responses. **Plant Cell Reports**, New York, v. 32, p. 1099–1109, Jan. 2013.

WU, L.; ZHOU, H.; ZHANG, J.; NI, F.; LIU, C.; QI, Y. DNA methylation mediated by a microRNA pathway. **Molecular Cell**, Cambridge, v. 38, p. 465-475, Nov. 2014.

ZARDO, G.; CIOLFI, A.; VIAN, L.; STARNES, L.M.; BILI, M.; RACANICCHI, S.; MARESCA, C.; FAZI, F.; TRAVAGLINI, L.; NOGUERA, N.; MANCINI, M.; NANNI, M.; CIMINO, G.; LO-COCO, F.; GRIGNANI, F.; NERVI, C. Polycombs and microRNA-223 regulate human granulopoiesis by transcriptional control of target gene expression. **Blood**, Washington, v. 119, n. 17, p. 4034-4046, Jan. 2012.

ZEGZOUTI, H. Ethylene-regulated gene expression in tomato fruit: characterization of novel ethylene-responsive and ripening-related genes isolated by differential display. **The Plant Journal**, Rockville, v. 18, n. 6, p. 589-600, Dec. 1999.

ZHANG, H.; HUANG, Z.; XIE, B. The ethylene-, jasmonate-, abscisic acid- and NaCl-responsive tomato transcription factor Jerf1 modulates expression of GCC box-containing genes and salt tolerance in tobacco. **Planta**, New York, v. 220, p. 262-270, Nov. 2004.





#### 4 THESIS FINAL DISCUSSION

The following main conclusions can be drawn from this thesis: (1) the expression of a Tnt1 reverse transcriptase hairpin messenger RNA in *Nicotiana tabacum* plants generates diverse, consistent and inheritable phenotypes that range from cell death spots and callose deposition on leaves to severe abnormal development in aerial and root portions. Those phenotypes in tobacco were corroborated in different transformation events using different vector backbones as the pCAMBIA1201 binary vector and the pPZP221 binary vector. Moreover, transformation events in *Nicotiana benthamiana* validated similar phenotypes. Control plants transformed with the empty vector and wildtype plants grown under the same conditions didn't show the phenotypes, further supporting the correlation of the phenotypes and the expression of the Tnt1 reverse transcriptase hairpin. (2) The effect of the hairpin expression on Tnt1 transcriptional activity remains unclear. RNA sequencing of the HP, Control and WT leaves of 45-day-old *in vitro* grown plants results didn't corroborate Tnt1 silencing in HP lines. However, an increased transcript number in all Tnt1 coding domains except for the Reverse Transcriptase was shown in HP1(T2). Pinzon and collaborators (2012) described that although Tnt1 is a target of silencing in tobacco, it can escape this control and remain transcriptionally active. Perhaps the RNA interference of Tnt1 is only effective in terms of lowering its amount of messenger RNA once Tnt1 surpasses a threshold of expression beyond which the cell "interprets" it as deleterious and RNA interference takes place. It is important to say that we did not expect to completely silence Tnt1 with the expression of the RT hairpin because of the high number of Tnt1 copies in the tobacco genome (over 600). We intended (and achieved) to alter Tnt1 expression pattern so we could analyze the magnitude of its consequences. (3) RNA-seq revealed molecular phenotypes exclusive to HP plants such as an extended modulation of defense response-related and ethylene-related genes and the rewiring of gene regulatory networks. The molecular phenotype corroborated the macroscopic phenotypes typical of plants under environmental adversities. (4) Expression quantification of Tnt1 (three fragments of the coding domain) in 90 days old *in vitro* grown plants revealed different results than the RNA-seq: HPs had less Tnt1 transcripts than the WT plants. Ethylene synthesis and ethylene responsive genes also showed the opposite of the RNA-seq results, they were downregulated in 90 days old HP plants. It was clear to

us then that different development stages showed different expression patterns of Tnt1 and other defense related genes, reflecting a dynamic transcriptional scenario. Our interpretation is that the expression pattern is under an oscillatory rhythm. (5) Adult 90 days old *in vitro* grown HP plants as well as young HP seedlings overproduce ethylene. This data is in agreement with the dynamic transcriptional scenario, since different development stages revealed up and down regulation of ethylene synthesis genes and Tnt1. (6) Tnt1 has sequences similar to ethylene responsive elements (GCC) in the promoter region and is responsive to the ethylene stimulus. Furthermore, these responsive elements are contained in Tnt1A messenger RNA, which folds into strong RNA binding stem-loop secondary structures. The folding of Tnt1 mRNAs into hairpins (with double-stranded RNA portions) could result in their processing into small RNAs, and these small RNAs would then target the GCC-like motifs present in other ethylene-responsive genes.

The results presented here allowed us to come up with a biological model to explain the transcriptional dynamic we consider happens during tobacco defense responses. Wildtype plants growing under normal conditions would keep a basal level of expression of Tnt1 as described by Pouteau and collaborators (1991) (also referred here as “range of homeostasis”). Upon a stress stimulus, ethylene biosynthesis is increased, up-regulating Tnt1 and other ethylene-responsive genes involved in plant defense response. This initiates the necessary metabolic changes for the defense response to take place. Increased Tnt1A messenger RNA would generate small RNAs that target and transcriptionally silence genes that have GCC motifs in their promoter region such as ethylene-responsive genes and Tnt1 itself in a feedback regulatory loop. Assuming the defense response was successful and the stress has been “overcome”, the stress stimulus is removed and ethylene synthetic genes are down-regulated. With less ethylene being produced, Tnt1 is no longer up-regulated, and with less transcription of Tnt1, less Tnt1A-derived small RNAs are produced. This is what we would expect to happen to wildtype plants. In the case of HP plants, we believe that expressing a Tnt1 RT hairpin has affected the inner balance between Tnt1 and defense response gene expression. The disturbance of this inner balance results in a transcriptional oscillation of those genes, which would explain our results in phase opposition. Therefore, we believe that Tnt1 biological role is to 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, resuming the defense responses which would no longer be needed.

The conclusions of this thesis are consistent with McClintock's theory (1984) of "genomic shock" which postulates that specific activation of plant transposable elements upon environmental challenges would act as important mechanisms to induce genome restructuration. This genomic shock is particularly important for plants because of their susceptibility to environmental fluctuations as sessile organisms.

In 1997, Grandbastien and collaborators discussed in their publication: "As the link between Tnt1 expression and plant defense responses becomes clearer, it raises more fundamental and exciting questions. (...) How could sequences so strikingly similar to regulatory sequences of cellular genes have emerged in a retrotransposon such as Tnt1? How have these sequences and features been maintained during evolution, and what is the biological impact, if any, of Tnt1-specific regulation?" (GRANDBASTIEN, 1997, p. 249). At that time, the link between Tnt1 expression and plant defense responses was limited to the observation of Tnt1 responsiveness to stresses, and there were no other evidences of "biological impacts" caused by Tnt1-specific regulation. The questioning concerning how Tnt1 evolved to share regulatory sequences with other cellular genes was and remains very pertinent, but still unanswered. The possible explanation discussed in the 1997 paper is still the most likely: regulatory motifs in Tnt1 were probably originated from the ancestral capture of cellular sequences by Tnt1.

The most important conclusion brought by this work is that plant defense response genes are regulated by the interplay between two intimately involved entities: ethylene induced transcription factors and another unexpected transcriptional regulator: Tnt1A small RNAs. But the mechanism through which Tnt1A small RNAs would perform such task is still unclear. Taking into consideration that Polycomb proteins can bind to small RNAs, be relocated to the nucleus and interfere with the expression of other genes by altering methylation patterns (ZARDO et al., 2012); and that our data suggests that Tnt1A-derived small RNAs could interfere with the expression of other genes (as well as its own); we searched for a putative enzyme that could interact with Tnt1A-derived small RNAs and perform the role of "carrier" of this complex to the nucleus where it could act as a negative transcriptional regulator.

Among Tnt1 proteins, Gag was the best candidate for the “carrier” role. In the introduction of the thesis (section 1.3.1), it was mentioned that under certain conditions retrovirus-derived Gag proteins were detected not only in the cytoplasm, but also in the nuclei (STAKE et al., 2013). Rashkova and collaborators (2002) demonstrated that a fraction of Gag proteins of *Drosophila* retrotransposons are moved into the nuclei as well. It is possible that Tnt1 Gag protein could perform the role of “carrier” of the Tnt1A-derived small RNAs, since it already does so during the retrotransposon replicative life cycle. To check the feasibility of this hypothesis, an alignment was made of the Tnt1 Gag protein sequence with several genes that code for Polycomb proteins and genes that code for Polycomb-Associated proteins (which are part of Polycomb Complexes). The alignment was performed using tBLASTn tool, and the result is shown in Figure 1.

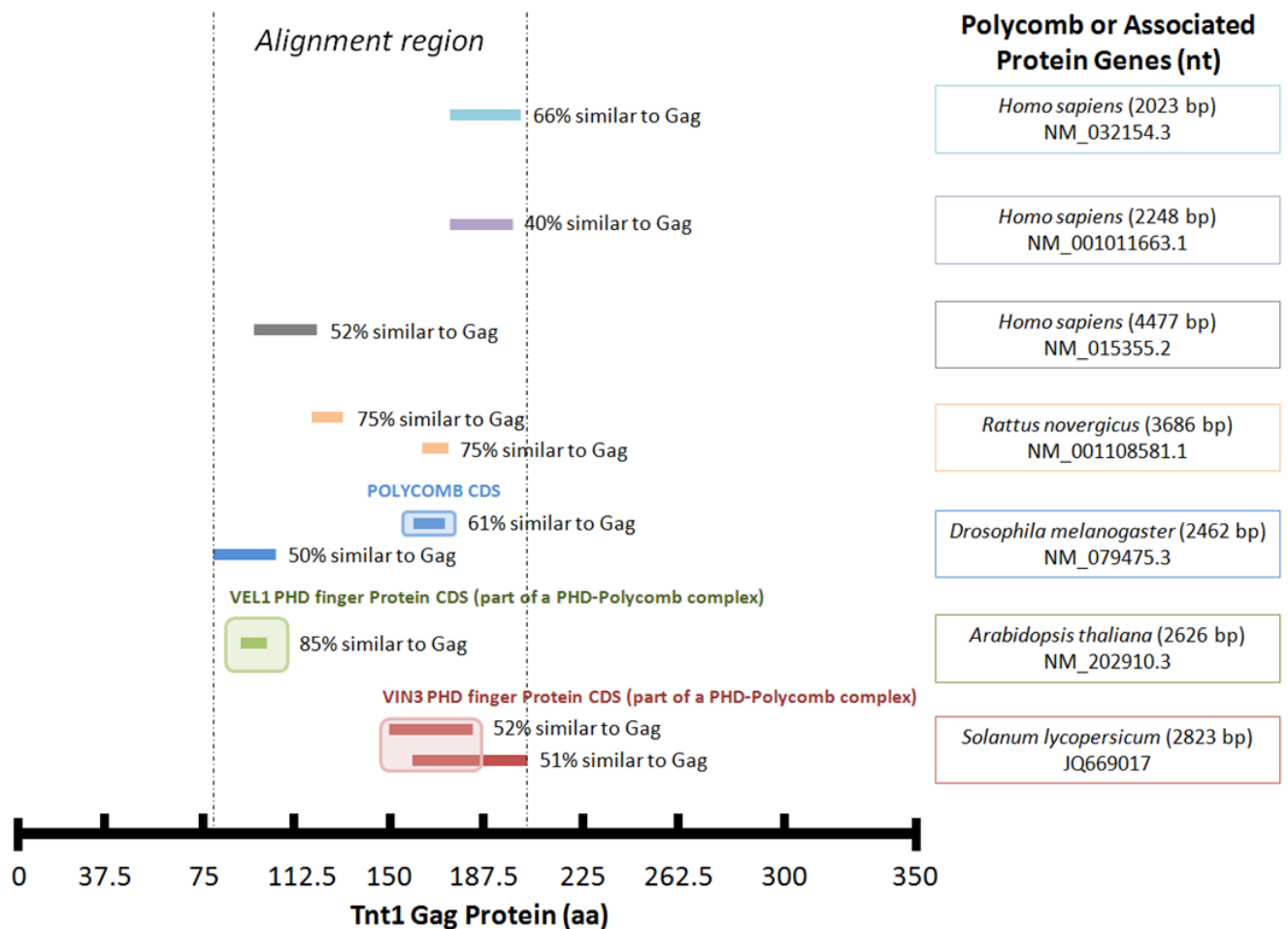


Figure 1 - Alignment of the Gag protein sequence from Tnt1 and Polycomb genes or genes that code for Polycomb Associated Proteins. On the bottom of the figure is the Gag protein sequence of Tnt1 (354 aminoacids long). The colored boxes to the right represent one Polycomb gene sequence from each species, and the gene accession numbers are listed below the species name. Numbers appearing in parentheses are the length of each gene. Colored rectangles on top of the Gag protein sequence are the corresponding portions of the Polycomb genes that showed similarity to Gag (these portions are represented in scale with the length of the Gag protein). To the right of each colored rectangle is the percentage of the similarity between Gag aminoacids and the Polycomb gene nucleotides. Around the rectangles corresponding to *Solanum lycopersicum*, *Arabidopsis thaliana* and *Drosophila melanogaster* Polycomb genes is a matching colored box. The region this colored box delimits corresponds to the Coding DNA Sequences (CDS) of the genes

The aligned region for all the Polycomb genes runs from Gag's 75th to 205th aminoacid. *Solanum lycopersicum* and *Arabidopsis thaliana* Polycomb-Associated protein genes and *Drosophila melanogaster* Polycomb gene have partially aligned to Gag inside their own Coding DNA Sequences (CDS). VIN3 and VEL1 proteins contain PHD (Plant Homeo Domain) finger domains that have been described as part of PHD-Polycomb Repressive Complex 2 in plants (LUCIA et al., 2008). PHD domains bind molecules of methylated histone H3 (MANSFIELD et al., 2011). For all

the other Polycomb gene sequences, the aligned region to Gag is out of their CDS (Figure 1). Nonetheless, it is notable that Gag has similarity with Polycomb and associated proteins, and it becomes possible that Gag and Tnt1A small RNAs could complex with Polycombs and alter the methylation scenario of target genes. This could be the mechanism through which Tnt1 small RNAs could target the GCC motifs on the promoter region of other ethylene responsive genes, further supporting our model for Tnt1 biological role (Figure 2 A).

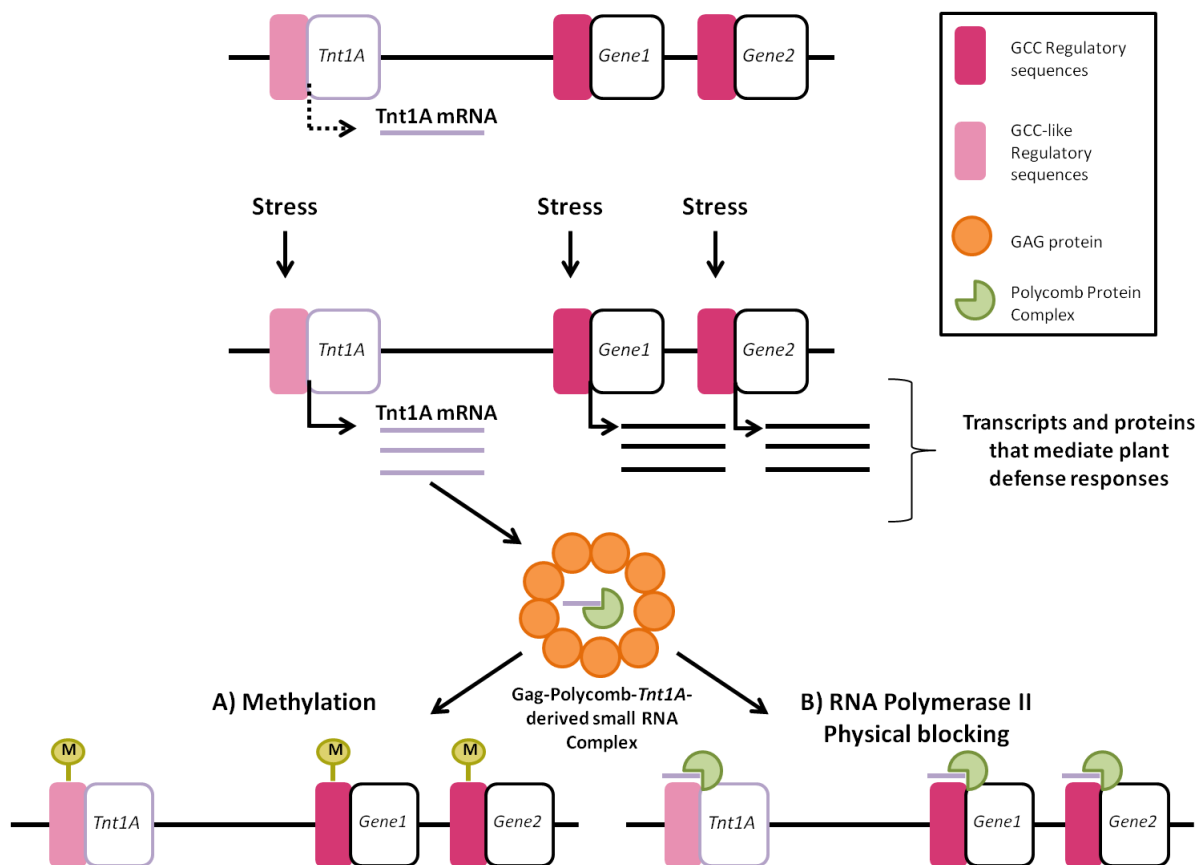


Figure 2 - The possible mechanisms through which Tnt1A-derived small RNAs could silence its own expression as well as the expression of genes that contain GCC motifs in the promoter region upon a stress signal. On the top of the figure, Tnt1 is expressed in basal levels and contain GCC-like motif in its promoter region. Gene 1 and gene 2 are two ethylene-responsive genes that have GCC motifs in their promoter as well. Then below that the figure illustrates a stress condition that induces the expression of Tnt1 as well as the ethylene responsive genes that are involved in the defense responses. As a result of Tnt1 induction, GAG proteins are translated in the cytoplasm and compartmentalize Polycomb proteins and a Tnt1A-derived small RNA molecule. This GAG-Polycomb-Tnt1A-derived small RNA complex is redirected to the nucleus. Once in the nucleus, we present two possibilities: A. The Polycomb Complex mediates epigenetic modifications that transcriptionally repress complementary DNA sequences to the GCC-like motifs in the small RNAs associated in the complex (including Tnt1 and other responsive genes). B. Tnt1A small RNA-Polycomb Complex attaches to its complementary DNA sequences thus creating a physical barrier to transcription by RNA Polymerase II. It is unclear to this date if GAG proteins are internalized in the plant nucleus or not

The existence of such transcriptional repressing mechanism is only one of the possible modes of action of Tnt1A-derived small-RNAs. The existence of other mechanisms through which complementary small RNAs could be relocated to a gene promoter and thus repress transcription are expected, but not yet described. For example: it was described that plant miRNAs act with ARGONAUTE1 and other proteins to inhibit the translation of target messenger RNAs at the Endoplasmic Reticulum without involving mRNA cleavage (Li et al., 2013). We consider possible that small RNAs could bind through similarity to promoter region of genes and thus act as a physical barrier to transcription by Pol II in a similar way that miRNAs can inhibit translation (Figure 2 B). It is also important to stress that our work unravels one of the various possible ways through which TEs can perform new biological roles, and that at the same time that we do not expect all transposable elements to perform biological roles such as Tnt1, we propose that many other TEs do and are yet to be discovered.

## Reference

MANSFIELD, R.E.; MUSSELMAN, C.A.; KWAN, A.H.; OLIVER, S.S.; GARSKE, A.L.; DAVRAZOU, F.; DENU, J.M.; KUTATELADZE, T.G.; MACKAY, J.P. Plant homeodomain (PHD) fingers of CHD4 are histone H3-binding modules with preference for unmodified H3K4 and methylated H3K9. **Journal of Biology Chemistry**, Rockville, v. 386, n. 13, p. 11779-11791, Apr. 2011.