

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE BIOCÊNCIAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

**Obtenção e análise de sequências similares a Tnt1 em espécies  
nativas de *Petunia* Juss. (Solanaceae)**

RAQUEL ATHAYDE KRIEDT

Dissertação submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS como requisito parcial para a obtenção do título de Mestre em Genética e Biologia Molecular.

Orientação: Loreta Brandão de Freitas

Porto Alegre, março de 2012.

Este trabalho foi realizado no Laboratório de Evolução Molecular do Departamento de Genética da Universidade Federal no Rio Grande do Sul. As reações de sequenciamento foram realizadas no Laboratório de Biologia Genômica e Molecular da Faculdade de Biociências da Pontifícia Universidade Católica do Rio Grande do Sul.

Fontes Financiadoras:

- Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)
- Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior (CAPES)
- Programa Especial em Taxonomia (PROTAX - MCT/CNPq/CAPES);
- Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS)

## **AGRADECIMENTOS**

Agradeço à minha orientadora, Dra. Loreta Brandão de Freitas, pelos ensinamentos e orientação;

Ao Guilherme Cruz pela direção;

Aos meus colegas, pela companhia diária, em especial os amigos Ana Luíza, Aline, Jéferson, e Priscilla, pela paciência e pelos muitos momentos de descontração, desabafo, e incentivo. Só o bar salva, então não me abandonem!

Aos meus amigos queridos que tanto me aguentam, em especial Tielli, Marcela, Bruno, e o mais novo, porém não menos importante, Ismael, pela companhia, paciência, diversão, discussões... Enfim, pela AMIZADE!

Ao Leandro, por pacientemente ouvir minhas lamúrias e dizer “vai dar tudo certo!”; Por me dar o ombro para chorar e estar ali nos meus momentos de dúvida; Por tudo aquilo que é só nosso... Te amo!

Aos meus pais. Meus pais que me ensinaram VALORES que já não encontramos mais por aí... Que me dão a certeza de que quando não houver mais nada, eles estarão lá. Pela imensa paciência, principalmente com minhas dúvidas e “surto” eventuais; Por me incentivarem a fazer o que eu acredito ser certo e não o que é esperado de mim... Desapontar outros é muito menor do que desapontar a si mesmo; Por me assegurarem estrutura e oportunidades nessa longa jornada de autoconhecimento, conhecimento, experiências, e busca pela estabilidade. Tenho muito orgulho dos pais que tenho, e espero superar todas as suas expectativas, me tornando tudo aquilo que posso ser, e mais. Amo vocês!

Ao Elmo Cardoso, pela ajuda, tranquilidade e paciência ao longo desses anos;

À Universidade Federal do Rio Grande do Sul e ao PPGBM pelo ensino gratuito e de qualidade;

Às instituições e fontes financiadoras;

Agradeço a todos que de maneira prática, teórica, ou incentivadora, colaboraram para a realização deste trabalho.

## SUMÁRIO

RESUMO	06
ABSTRACT	07
CAPÍTULO 1: Introdução	08
1.1. Elementos de Transposição	09
1.2. Classificação dos Elementos de Transposição	10
1.3. Origem e Evolução dos Elementos de Transposição	12
1.4. Efeitos dos Elementos de Transposição no Genoma	14
1.5. Retrotransposons com LTR	16
1.5.1. Ty1/ <i>Copia</i>	18
1.6. O Retrotransposon Tnt1	20
1.6.1. Elementos Similares a Tnt1 em Solanaceae	22
1.7. O Gênero <i>Petunia</i>	23
1.7.1. Retrotransposons com LTR em <i>Petunia</i>	25
CAPÍTULO 2: Objetivos	26
CAPÍTULO 3: Novel transposable elements in Solanaceae: evolutionary relationship among Tnt1-related sequences in wild <i>Petunia</i> species	28
CAPÍTULO 4: Considerações finais e Perspectivas	59
CAPÍTULO 5: Referências Bibliográficas	63

## LISTA DE ABREVIATURAS

cDNA: DNA complementar

DIRS: elementos de transposição encontrados em *Dictyostelium*

DNA: ácido desoxirribonucléico

ENV: envelope viral

HT: transferência horizontal

ICTV: *International Committee of Taxonomy on Viruses*

INT: integrase

IRAP: *inter-retrotransposon amplified polymorfism*

LINES: elementos nucleares longos dispersos

LTR: longas repetições terminais

mRNA: RNA mensageiro

ORF: quadro aberto de leitura

PBS: sítio de ligação de *primer*

PLE: elementos do tipo *Penelope*

PPT: trato de polipurinas

PR: protease

R: RNA repetitivo

RBIP: *retrotransposon-based insertional polymorphism*

REMAP: *retrotransposon-microsatellite amplified polymorfism*

RNA: ácido ribonucléico

RNase H: ribonuclease H

RT: transcriptase reversa

RUP: *retrotransposon UTL polymorphism fingerprinting*

SINES: elementos nucleares curtos dispersos

SSAP: *sequence-specific amplified polymorphism*

TAM: *tagged microarray marker*

TEs: elementos de transposição

TIRÇ terminação repetida invertida

TSD: duplicação do sítio alvo

U3: RNA 3' único

U5: RNA 5' único

## RESUMO

Elementos de transposição são caracterizados como segmentos de DNA capazes de mover-se no genoma em que estão inseridos. A diversidade e abundância desses elementos variam de espécie para espécie, sendo sua distribuição nem uniforme, nem aleatória. Com a grande quantidade de informações provenientes de projetos genoma, a busca de um sistema universal para classificação desses elementos é uma constante. Atualmente, elementos de transposição são divididos em duas classes, sendo os elementos da classe I, ou retrotransposons, divididos em cinco ordens. A ordem mais abundante em plantas é a dos LTR retrotransposons, no qual se insere a superfamília Ty1/*Copia*. Essa superfamília apresenta diversas famílias de elementos, dentre elas a família Tnt1. Tnt1 foi o primeiro elemento de transposição caracterizado na família Solanaceae, sendo um dos retrotransposons mais bem estudados em plantas. Originalmente isolado do genoma de *Nicotiana tabacum*, elementos similares a Tnt1 já foram encontrados em pimenta, berinjela, tomate, batata, e em petúnias de jardim. Dessa forma, o objetivo do presente trabalho foi caracterizar a diversidade dos retrotransposons com LTR do clado Tork da superfamília *Copia*, relacionados ao elemento Tnt1, em espécies nativas do gênero *Petunia*. Para tanto, sequências foram obtidas a partir de reação de PCR com primers já descritos na literatura, com posterior clonagem dos fragmentos obtidos, e sequenciamento dos clones. Um total de 148 sequências únicas foram obtidas de 21 *taxa* diferentes, incluindo três espécies de *Calibrachoa* e uma espécie de *Fabiana*, ambos os gêneros relacionados à *Petunia*. Análises de relacionamento entre as sequências sugerem a classificação dessas em dez diferentes famílias de retrotransposons com LTR do tipo *Copia* (RLC), muitas com baixa diferenciação entre as sequências de uma mesma família. Entre essas, uma família de sequências quase que exclusiva de *taxa* encontrados em substrato com baixa umidade. Para uma visão mais completa, representantes dos diferentes grupos de Tnt1, Retrolyc1, e Retrosol foram incluídos na análise. Foi possível observar que as sequências obtidas no presente trabalho não são compartilhadas entre os elementos dos demais gêneros incluídos. As diferentes sequências aqui obtidas possibilitaram estudos de relacionamento e diversidade, e abrem portas para a utilização desses elementos em estudos subsequentes nos gêneros em questão, assim como em outros gêneros relacionados, uma vez que a conservação de partes do elemento é bastante grande.

## ABSTRACT

Transposable elements are characterized as segments of DNA capable of moving across different parts of the host genome. Diversity and abundance of these elements vary from species to species, and distribution within a genome is usually neither random nor uniform. The amount of information resulting from genome projects demands a universal system for transposable elements' classification. Nowadays, these elements are classified into two classes, with class I elements, or retrotransposons, being divided into five orders. The most abundant order in plants is the LTR retrotransposons order, in which the Ty1/*Copia* superfamily is classified. Ty1/*Copia* presents several families, being the Tnt1 family one of them. Tnt1 was the first transposable element characterized in Solanaceae, being the most well known retrotransposon in plants. Originally isolated from *Nicotiana tabacum*, Tnt1-related elements have been reported in peppers, eggplant, tomato, potato, and petunias. Therefore, the aim of this work was to characterize the diversity of LTR retrotransposons from the Tork clade of the *Copia* superfamily, related to Tnt1 elements, in wild *Petunia* species. Thereunto, sequences were retrieved from PCR reactions with primers previously described, cloned into competent cells. A total of 148 unique sequences were obtained from 21 different *taxa*, including three *Calibrachoa* and one *Fabiana* species, both genera closely related to *Petunia*. Relationship analysis among sequences suggests their classification into ten families of retrotransposons with LTR from *Copia* superfamily (RLC), most of low sequence diversity within a family. Among all families, one clade of sequences presented almost exclusively sequences of *taxa* found under substrates of low humidity conditions. For a more complete view, Tnt1, Retrolyc1, and Retrosol representatives were included in the analysis. Sequences retrieved in the present work are not shared among elements of remaining genera included. The diversity of sequences obtained here made relationship and diversity studies possible, and may be used in further investigations of the genera here studied, as well as in other related genera, since there are conserved fragments along these elements.

# INTRODUÇÃO



## 1.1. Elementos de Transposição

Descobertos inicialmente em 1944 por Barbara McClintock, elementos de transposição (TEs - *Transposable elements*) foram inicialmente chamados de “elementos controladores” devido à capacidade de modificar a expressão gênica por sua presença em certo *locus*. Ela notou que esses elementos possuíam a capacidade de moverem-se enquanto as células se dividiam e proliferavam, gerando mutações em apenas algumas delas. Devido à ideia de estabilidade genômica que imperava no contexto científico da época, foi apenas com o acúmulo de evidências a partir de diferentes genomas que Barbara McClintock obteve reconhecimento, sendo finalmente premiada com um Nobel pela descoberta dos elementos transponíveis em 1983. Dessa forma, TEs são caracterizados como segmentos de DNA que possuem a capacidade de se moverem de uma posição de um cromossomo para novas regiões do genoma, e são encontrados em quase todos os organismos já estudados, de bactérias a humanos (Hua-Van *et al.*, 2011).

De maneira geral, TEs possuem genes que codificam os elementos que permitem sua transposição. Porém existem elementos chamados não autônomos, que não possuem essa capacidade e acabam utilizando a maquinaria de transposição de elementos autônomos, que retêm os sinais *cis*-regulatórios necessários para a transposição (Kidwell & Lisch, 1997). A inserção de quase todos os TEs resulta na duplicação de uma sequência curta do genoma no sítio de inserção, chamada duplicação do sítio alvo (TSD - *Target Site Duplication*). Juntos, esses elementos podem compor até 80% de genomas eucarióticos, e 10% de genomas procarióticos, e parece haver uma relação positiva entre a quantidade de elementos e o tamanho do genoma (Kidwell, 2002; Haas *et al.*, 2006).

A diversidade e abundância de TEs são altamente variáveis entre as espécies, refletindo uma relação específica entre o genoma e os TEs que nele se encontram. A distribuição de determinados TEs em um genoma não é uniforme ou aleatória. Alguns elementos são sítio-específicos, outros se encontram em regiões cromossômicas nas quais sua presença não causa efeitos deletérios, como telômeros e regiões heterocromáticas. Outros, ainda, são encontrados em meio a genes. Em algumas espécies de plantas, transposons são encontrados preferencialmente em meio a genes. Essa distribuição não aleatória pode ser resultado de uma possível preferência de inserção, ou apenas ação da seleção (Hua-Van *et al.*, 2011).

## 1.2. Classificação dos Elementos de Transposição

Em 1989, Finnegan propôs o primeiro sistema de classificação de TEs, distinguindo duas classes de acordo com seu intermediário de transposição: RNA ou DNA. Com o aumento da diversidade de elementos descobertos ao longo do tempo, Capy *et al.* (1998) propuseram uma classificação sistemática para os elementos móveis do genoma. Porém, a quantidade de dados e conhecimentos gerados atualmente requer uma classificação complexa e contínua. Assim, Wicker *et al.* (2007) propuseram novas *taxa*, através de uma classificação simples, baseada em um código de três letras que remete à classe, ordem e superfamília, seguido pelo nome da família, e identidade da sequência. Segundo esta classificação, considera-se:

Classe: divisão dos elementos entre aqueles que se utilizam ou não de um intermediário de RNA para a transposição.

Subclasse: separação dos elementos que copiam a si mesmos para uma nova inserção daqueles que deixam o local onde se encontram para serem reintegrados em um novo local.

Ordem: delimitação de acordo com diferenças no mecanismo de inserção, na organização geral, e enzimologia.

Superfamília: discriminação por elementos que compartilham uma mesma estratégia de replicação, mas se diferenciam quanto à estrutura de suas proteínas ou os domínios não-codificadores. Ainda, superfamílias diferem quanto à presença e tamanho da TSD.

Família: identificação pela conservação da sequência de DNA; uma mesma família deve apresentar alta similaridade de sequência (ao menos 80%) em no mínimo 80% da sequência alinhada de região codificadora ou repetições terminais, quando analisando segmentos maiores de 80 pares de bases (pb). Famílias diferentes possuem similaridade apenas em regiões altamente conservadas de regiões codificadoras.

Subfamília: definição em nível filogenético, podem ser populações de elementos não-autônomos dentro de uma mesma família de elementos autônomos, ou subpopulações que podem ser claramente distinguidas dentro de um família.

Inserção: descreve uma cópia individual de um elemento.

A classe I é composta de elementos chamados retrotransposons, que se transpõem a partir de um intermediário de RNA previamente transcrito do DNA, que é então

reversamente transcrito em DNA dupla-fita, e integrado em uma nova posição do genoma. Esse modo de transposição é chamado “copiar-colar”, sendo essa uma classe replicativa (Wicker *et al.*, 2007; Hua-Van *et al.*, 2011). A enzima capaz de gerar o DNA fita-dupla a partir do RNA transcrito é a transcriptase reversa (RT) que, além de ser característica desses elementos, está presente na telomerase eucariótica. Segundo Wicker *et al.*, (2007), essa classe é dividida em cinco ordens:

LTR retrotransposons (do inglês *long terminal repeat*, repetição terminal longa): encontrados em diversos genomas eucarióticos, desde diatomáceas até humanos (Llorens *et al.*, 2009). Apresentam longas repetições terminais (LTRs), de orientação direta em ambas as pontas do elemento e dois genes: *gag* e *pol*. O primeiro produz uma proteína similar a um capsídeo viral, enquanto o segundo codifica uma poliproteína, a qual resulta em: protease (PR), integrase (INT), transcriptase reversa (RT) e ribonuclease H (RNase H) (descrição aprofundada no item 1.5. Retrotransposons com LTR).

Elementos do tipo *DIRS* (do inglês *Dictyostelium transposable elements*): detectados em diversas espécies, desde algas, até fungos e animais, e diferem dos LTR retrotransposons por possuírem um gene da tirosina recombinase no lugar de INT.

Elementos do tipo *Penelope* (PLEs): primeiramente detectados em *Drosophila virilis*, apresentam uma distribuição descontínua, embora já detectados em mais de 50 espécies incluindo animais, fungos e plantas. Esses elementos codificam um RT mais relacionado à telomerase do que com a RT de LTR retrotransposons.

LINES (*Long Interspersed Nuclear Elements* – elementos nucleares longos dispersos): presentes em eucariotos, não possuem LTRs, e codificam pelo menos uma RT e nuclease para transposição. Em sua porção 3' podem apresentar uma região de poli A, outras repetições em *tandem*, ou apenas uma região rica em adeninas.

SINES (*Short Interspersed Nuclear Elements* – elementos nucleares curtos dispersos): são elementos não autônomos originados a partir de retrotransposições acidentais de diversos transcritos de Polimerase III. Esses pseudogenes retroprocessados apresentam promotor para Polimerase III, o que permite sua expressão. Esses elementos dependem da maquinaria de transposição dos LINES para sua dispersão.

A classe II é composta por elementos que se transpõem diretamente, sem necessidade de um intermediário de RNA. Encontrada em procariotos e eucariotos, essa classe é subdividida em duas subclasses que diferem entre si pelo número de fitas de DNA

que são clivadas no processo de transposição. Segundo Wicker *et al.* (2007), representantes da subclasse I estão divididos em duas ordens, e são excisados do local em que se encontram e integrados em uma nova região, através do mecanismo conhecido como “cortar-colar”, que gera uma terminação repetida invertida (TIR) de tamanho variável na extremidade do elemento inserido. A ordem *TIR* apresenta nove superfamílias distinguidas pelo tamanho das sequências TIRs e TSD resultantes de sua inserção. Representam essa subclasse as superfamílias *Tc1-mariner*, *hAT*, *Mutator*, *Merlin*, *Transib*, *P*, *piggyBac*, *PIF-Harbinger*, and *CACTA*. A ordem *Crypton* é encontrada em fungos e apresenta uma tirosina recombinase como os elementos do tipo *DIRS*, mas não possui domínio da RT, o que sugere sua transposição sem intermediário de RNA. Essa ordem não gera TIRs, mas gera TSDs como resultado de transposição.

De acordo com Wicker *et al.* (2007), a subclasse II apresenta TEs que se transpõem de maneira replicativa, sem necessidade de clivagem em ambas as fitas de DNA. Apresentando duas ordens, sua classificação dentro classe II é devido a falta de intermediário de RNA, e não por ancestralidade comum. Elementos da ordem *Helitron*, são encontrados em plantas, fungos e animais. Esses elementos não geram TSD e codificam uma tirosina recombinase que difere das citadas anteriormente. Os elementos da ordem *Maverick* são encontrados em diversos eucariotos, mas não em plantas. Apresentam TIRs e codificam até onze proteínas. *Mavericks* codificam INT relacionada à de elementos de classe I, mas não apresentam RT, o que sugere sua transposição por intermediário de DNA.

### **1.3. Origem e Evolução dos Elementos de Transposição**

O surgimento de um novo elemento em um genoma pode ocorrer de duas maneiras: por transferência horizontal (HT - *Horizontal Transfer*), sendo essa forma mais reconhecida, ou espontaneamente a partir de uma sequência genômica pré-existente (Kidwell & Lisch, 2001). Transferência horizontal é o processo no qual um elemento é adquirido através de cruzamento com uma espécie relacionada ou através de um vetor, como bactérias e vírus (Loreto *et al.*, 2008). Para permanecer no genoma da espécie, esse novo elemento deve chegar até as células germinativas. A partir de então, o elemento passa para as próximas gerações por transferência vertical. O surgimento a partir de uma sequência se refere a elementos que retomam a atividade como resultado de recombinação ectópica entre cópias inativas (Hua-Van *et al.*, 2011), ou ainda aqueles como os elementos *Alu* (representantes de SINES), por exemplo. Esses elementos estão presentes em primatas

e são derivados da duplicação do gene 7SL RNA. 7SL é um componente de RNA do complexo ribonucleoprotéico de partícula de reconhecimento de sinal, e foi constatada grande similaridade entre sequências do gene 7SL RNA de *Drosophila melanogaster*, *Xenopus laevis*, e humanos, demonstrando que sequências *Alu* estão presente nos genomas animais muito antes da radiação dos mamíferos (Ullu & Tschudi, 1984).

Para permanecer no genoma do hospedeiro, o novo elemento precisa se replicar de maneira eficaz. Após uma invasão inicial bem sucedida nesse genoma e na população em que se encontra, torna-se difícil a perda desse elemento. A manutenção do elemento nesse novo ambiente se dá pelo equilíbrio entre transposição e seleção (Nuzhdin, 1999). Esse balanço ocorre devido à seleção purificadora em resposta ao efeito deletério dos TEs, e à atividade do elemento, o qual é capaz de se replicar mais rápido que o genoma em que se encontra. Porém, a persistência de TEs não se dá necessariamente dessa forma, uma vez que esse equilíbrio é frequentemente quebrado pelas explosões de transposição e por vários efeitos deletérios que escapam da seleção.

TEs também estão sujeitos a forças não-adaptativas como o tamanho efetivo da população, visto que populações pequenas tem mais chance de sofrer os efeitos da deriva genética, eliminando ou fixando um elemento, e a história demográfica do hospedeiro, como flutuação no tamanho populacional e *bottlenecks*. Esses e outros mecanismos contribuem para a evolução de TEs como milhares de famílias diferentes, cada uma com sua própria história evolutiva. A diversidade atual pode ser apenas uma parte da história total, e a persistência de alguns elementos e a extinção de outros não são por si só adaptativas, estando sujeitas ao efeito de eventos estocásticos. Assim, a história evolutiva de TEs pode ser explicada sem a necessidade de um equilíbrio do número de cópias à longo prazo (Hua-Van *et al.*, 2011).

A visão do genoma como um ecossistema no qual diferentes cópias de TEs podem ser consideradas membros de uma espécie (Venner *et al.*, 2009) levanta questões quanto à competição a qual TEs estariam sujeitos. Nesse sentido, cópias autônomas e não autônomas de um dado elemento de classe II, por exemplo, competem quanto ao uso da transposase, de maneira que a cópia não autônoma pode ser vista como parasita da cópia autônoma (Hua-Van *et al.*, 2011). Ainda, elementos que se inserem em regiões de recombinação frequente apresentam maiores chances de recombinações desiguais entre diferentes elementos, como ocorre em LTR retrotransposons, gerando novos elementos

com novas propriedades e oportunidades. Também, a inserção de um elemento dentro de outro pode eliminar o elemento interrompido como um potencial competidor para ampliações futuras (Baucom *et al.*, 2009).

Podem existir várias famílias em um mesmo genoma, e a expansão de uma delas pode causar impacto na dinâmica de outras (Hua-Van *et al.*, 2011). De acordo com Baucom *et al.* (2009), quando um ambiente é favorável à proliferação de LTR retrotransposons, alguns elementos bem adaptados se agregam com seus competidores gerando um sistema de grande abundância, mas de baixa diversidade. Já LTR retrotransposons de baixo número de cópias, acabam por se disseminar em menor frequência, mas por diferentes regiões as quais se tornam ricas em diversidade.

#### **1.4. Efeitos dos Elementos de Transposição no Genoma**

A presença de elementos transponíveis no DNA pode acarretar em duas grandes mudanças: inserção do elemento em um locus, e recombinação ectópica que pode gerar diferentes rearranjos. A inserção de um elemento dentro ou próximo de um gene pode alterar atividade do gene gerando mudança na expressão gênica, *splicing* alternativo, mudanças na atividade proteica, ou inativação total do gene. A consequência da inserção de TEs pode ser direta, adicionando nucleotídeos à sequência do gene, ou indireta, como na marcação epigenética do elemento influenciando no gene. Já a recombinação entre elementos de diferentes *loci* pode ocasionar desde inversões e duplicações até deleções e translocações (Delprat *et al.*, 2009; Zhang *et al.*, 2009).

Os TEs são fonte de variação para o genoma, tanto por mutação quanto por recombinação. A transposição é um processo mutagênico, uma vez que muitos elementos de classe II geram deleções ao mover-se de um local para outro, por exemplo. No entanto, a consequência da presença de TEs nem sempre é deletéria, podendo tanto ser neutra como vantajosa, fazendo desses elementos agentes geradores de diversidade (Kidwell & Lisch, 1997; Lukens & Zhan, 2007). Em plantas, por exemplo, a coloração de flores pode ser afetada pela inserção de um TE em genes de coloração. Gerats *et al.* (1990) isolou o elemento de classe II dTph1 (*defective transposable element P. hybrida* 1) na espécie comercial de *Petunia* após sua inserção no gene DFR-C (*dihydroflavonol reductase C*). Sua presença ou ausência coincide com o genótipo do hospedeiro para os alelos do fator de transcrição *an1*, envolvido na cascata de coloração da flor. A presença deste elemento no

genoma de *P. hybrida* foi observada devido à mutação na coloração das flores, resultando em flores variegadas no lugar de flores de coloração uniforme, provocada quando o elemento está ativo. Outras consequências fenotípicas foram observadas por Xiao *et al.*, (2008), que evidenciaram uma duplicação promovida por um retrotransposon LTR que resultou em um aumento na expressão do principal gene do controle do comprimento do fruto em tomate, gerando tomates mais longos em relação ao tipo selvagem, com o gene ancestral. Assim, esses autores concluíram que os retrotransposons podem agir na evolução do genoma através de duplicação gênica, resultando em mudanças fenotípicas.

Estudos posteriores comprovaram a ação de TEs no aumento do tamanho de genomas eucarióticos. Zedek *et al.* (2010) correlacionaram o aumento da densidade LTR retrotransposons e o valor C na filogenia do gênero *Eleocharis*. Os autores encontraram uma nova família de retrotransposon como sendo a responsável pelo aumento da densidade desse tipo de elemento no genoma. Já em procariotos, estudos que comparam a presença ou densidade de TEs e o tamanho dos genomas mostram que esses elementos podem ser responsáveis pela simplificação de genomas procarióticos (Parkhill *et al.*, 2003; Chain *et al.*, 2004).

A presença de TEs também pode acarretar em mudanças estruturais no genoma. Inukai & Sano (2002) puderam sugerir a participação de TEs na compartimentalização do genoma de arroz após a descoberta de TEs em regiões de organização da cromatina (S/MAR - scaffold/matrix attachment regions). TEs também são capazes de aquisição e perda de genes através de recombinação ilegítima. Morgante *et al.* (2005) analisaram diversas sequências do genoma de milho e encontraram fragmentos de genes nas sequências de elementos do tipo *Helitron*. Assim, elementos que contêm genes nucleares, inteiros ou fragmentados, em sua sequência podem espalhar esses genes por todo o genoma, resultando em oportunidades para duplicação gênica e embaralhamento de *exons*. Wang *et al.* (2006), estudando elementos de arroz, sugeriram que a origem de genes por retrotransposição é contínua e frequente, sendo um mecanismo importante na evolução dos genes nesta e em outras gramíneas.

Embora o estado silenciado dos TEs seja o mais comum, explosões de atividade e o alto número de cópias podem acarretar em uma rápida diversificação genômica entre duas espécies relacionadas, através de uma amplificação ou recombinação em uma

linhagem específica, podendo, assim, atuar em processos de especiação (Raskina *et al.*, 2008).

## 1.5. Retrotransposons com LTR

Os retrotransposons com LTR (ou LTR retrotransposons) são a ordem predominante em plantas, chegando a representar 70% do genoma nuclear de algumas angiospermas (Kalendar *et al.*, 2011). Apresentam longas repetições terminais de orientação direta em ambas as pontas do elemento, e quando integrados podem gerar um sítio de duplicação de quatro a seis pares de base para cada lado do elemento (Wicker *et al.*, 2007).

As LTRs podem estender-se por apenas duzentos pares de base, ou por até cinco mil pares de base. Cada LTR inicia com TG em sua porção 5' e termina com CA em sua porção 3' (Wicker *et al.*, 2007). São subdivididas em três regiões: RNA 3' único (U3), RNA repetitivo (R), e RNA 5' único (U5) (Figura 1). Essas regiões contêm promotores e terminadores associados à transcrição do elemento (Kumar & Bennetzen, 1999), e são as regiões do elemento que evoluem de forma mais rápida (Vicient *et al.*, 2005).

Retrotransposons com LTR autônomos apresentam, interno às LTRs, o sítios de iniciação para a transcrição reversa (PBS - *Primer Binding Site*) na extremidade 5', e o trato de polipurinas (PPT - *Polypurine Tract*) na extremidade 3'. Ambas as regiões correspondem a sítios importantes no processo de transposição destes retroelementos com LTRs (Grandbastien *et al.*, 1989; Kumar & Bennetzen, 1999). Internamente a esses sítios estão os genes que codificam as proteínas necessárias para sua proliferação. Essa região é normalmente dividida em duas ORFs: *gag*, que produz uma proteína similar a um capsídeo viral, que empacota o transcrito em uma partícula semelhante a vírus, e *pol*, o qual codifica uma poliproteína. Esta, quando processada, resulta em várias proteínas maduras: a protease (PR), a qual cliva a poliproteína, a integrase (INT), necessária para a integração da cópia de cDNA no genoma hospedeiro, a proteína transcriptase reversa (RT) e ribonuclease H (RNase H), as quais são responsáveis pela síntese de cDNA a partir do mRNA transcrito. Ainda, uma ORF adicional para a proteína do envelope viral (ENV) é encontrada em alguns grupos de uma das superfamílias de LTR retrotransposons (Kumar & Bennetzen, 1999; Kalendar *et al.*, 2011). A transcrição desses elementos inicia na 5'-LTR de um elemento integrado no genoma e produz um transcrito de RNA que compreende quase o



elemento por inteiro. O próprio elemento apresenta em sua 3'-LTR o sítio de poliadenilação que sinaliza o final de sua transcrição. Esse RNA é reversamente transcrito em DNA complementar extra-cromossômico, reconstituindo o elemento por inteiro, o qual é integrado no genoma (Kalendar *et al.*, 2011).

Existem cinco superfamílias na ordem LTR retrotransposons. As duas superfamílias principais, *Copia* (Ty1/*Copia*) e *Gypsy* (Ty3/*Gypsy*) (Figura 1), diferem quanto à ordem dos domínios RT e INT na ORF *pol* (Kumar & Bennetzen, 1999; Wicker *et al.*, 2007). Elementos da superfamília *Bel/Pao* são estruturalmente similares a *Gypsy* ou *Copia*, inclusive quanto à LTR e a geração de TSD. Porém, esses elementos formam um clado distinto quanto à filogenia de sua região codificante (Wicker *et al.*, 2007; Llorens *et al.*, 2009). Outra superfamília é a dos retrovírus. Retrovírus são proximamente relacionados aos LTR-retrotransposons. Esses elementos podem ter evoluídos de elementos *Gypsy* que adquiriram a capacidade de infecção a partir da proteína de envelope viral (ENV), além de outras proteínas e sequências regulatórias. Embora há muito classificados como vírus, os retrovírus entram no sistema de classificação de retrotransposons com LTR devido à estrutura que apresentam. Como retrovírus podem se tornar LTR-retrotransposons pela perda da mobilidade extracelular, os chamados endógenos retrovírus (ERVs) também são considerados membros dos elementos com LTRs. Essa superfamília se assemelha à *Gypsy*, porém apresenta ORF para algumas proteínas virais, características de retrovírus (Wicker *et al.*, 2007).

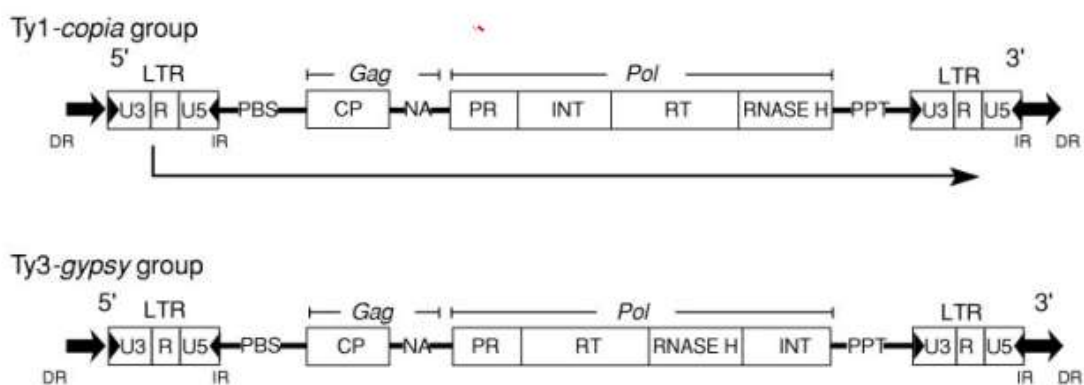


Figura 1: Estrutura geral de duas superfamílias de retrotransposons com LTR. Cada LTR é composta pela U3, R, e U5, regiões que contêm os sinais para início e término da transcrição. O transcrito é indicado pela seta fina abaixo do elemento *Copia*, e se estende da região R da 5'-LTR, até a região R da 3'-LTR. Os genes do elemento codificam a proteína similar a um capsídeo viral (CP), e uma poliproteína que resulta em protease (PR), integrase (INT), transcriptase reversa (RT)

e RNase H (RH). Ainda, as regiões sítios de iniciação para a transcrição reversa (PBS - *Primer Binding Site*) na extremidade 5', e o trato de polipurinas (PPT - *Polypurine Tract*). Ao final de cada LTR encontra-se uma repetição invertida (IR - *Inverted Terminal Repeat*) e uma região de repetição direta (DR - *Target Direct Repeat*), ambas no DNA genômico (Kumar & Bennetzen, 1999).

### 1.5.1. Ty1/Copia

Ty1/Copia representa os elementos com LTRs mais importantes e abundantes em eucariotos. A grande representatividade desses retroelementos em genomas de plantas, fungos, algas e protistas sugere uma coexistência entre seus ancestrais e Ty3/Gypsy, anterior à diversificação entre plantas e animais.

A região interna desses elementos codifica os genes típicos *gag* e *pol*, e a classificação desses elementos é contínua. De acordo com a ICTV (International Committee of Taxonomy on Viruses), Ty1/Copia é uma família que se chama *Pseudoviridae*, sendo originalmente dividida em três gêneros: *Pseudovirus*, *Hemivirus*, e *Sirevirus*. *Pseudovirus* e *Hemivirus* são distinguidos pelo iniciador usado para a transcrição reversa, enquanto *Sirevirus* forma um grupo distinto com base na sequência de aminoácidos de seu domínio da RT, sendo encontrado apenas em plantas.

Estudos recentes demonstraram que a diversidade desses elementos é maior que sua classificação original. Através da análise do gene *pol*, Llorens *et al.* (2009) puderam distinguir dois ramos principais na filogenia de Ty1/Copia. O primeiro ramo (Branch 1) compreende o gênero *Pseudovirus*, juntamente com um clado chamado GalEA, encontrado em alguns organismos marinhos, e LTR retrotransposons do tipo *CoDi*, encontrados em diatomáceas. O segundo ramo (Branch 2) compreende as demais linhagens de retrotransposons com LTR e possíveis retrovírus, incluindo os elementos similares a *Copia*, *Hemivirus* e *Sirevirus*. Ao todo, o segundo ramo apresenta oito cladogramas, sendo Tork o clado de interesse deste trabalho (Figura 2).

Representantes do clado Tork são descritos no genoma de diferentes espécies de plantas: *Zea mays* (SanMiguel *et al.*, 1996), *Solanum lycopersicum* e *Vitis vinifera* (Llorens *et al.*, 2009), *Nicotiana tabacum* (Grandbastien *et al.*, 1989; Hirochika *et al.*, 1996) e *Vigna radiata* (Xiao *et al.*, 2007). Os membros desse clado apresentam variação no tamanho das LTRs de ambos lados.

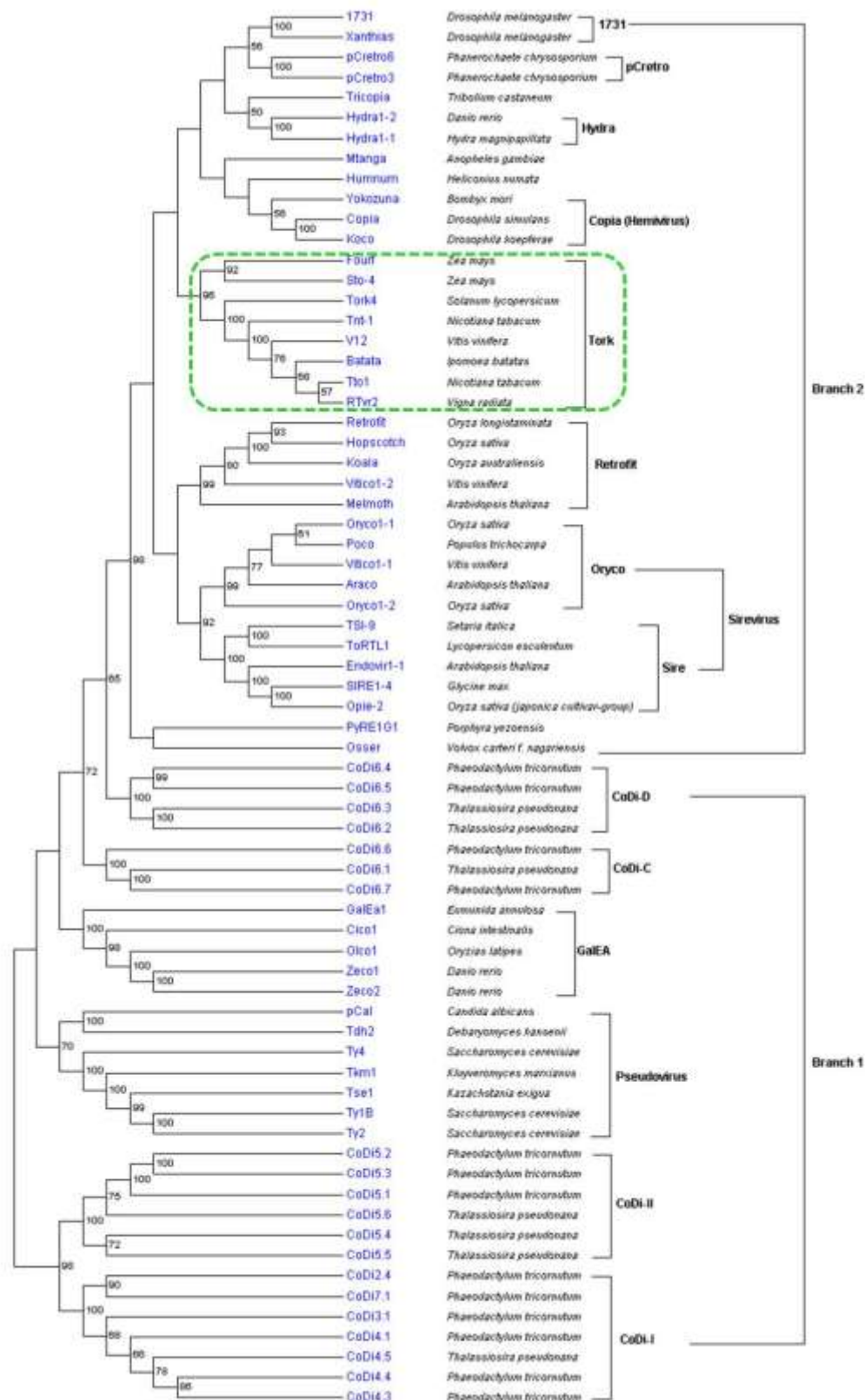


Figura 2: Reconstrução filogenética de Ty1/Copia inferida com base no gene *pol*. Posição do clado Tork está delimitada pelo retângulo tracejado em verde. Árvore consenso resultante de análise realizada com PHYLIP 3.6, com 100 replicações de *bootstrap*, e modelo de substituição de aminoácidos JTT (adaptado de Llorens *et al.*, 2009).

## 1.6. O Retrotransposon Tnt1

Tnt1 foi o primeiro elemento de transposição caracterizado na família Solanaceae, sendo um dos retrotransposons mais bem caracterizados em plantas. Esse retroelemento, mais especificamente Tnt1A, foi isolado do genoma de *Nicotiana tabacum* após sua inserção no gene estrutural nitrato redutase (Grandbastien *et al.*, 1989).

O elemento Tnt1 é um retrotransposon com LTR membro do clado Tork de Ty1/Copia (Llorens *et al.*, 2009), e é encontrado em alto número de cópias em *Nicotiana tabacum*. Possui 5,3 mil pares de bases (kpb), sendo flanqueado por pequenas duplicações de cinco pares de bases em seu sítio de inserção, e é um dos poucos retrotransposons de plantas com transposição demonstrada. Suas LTRs são terminadas pelas sequências palindrômicas 5'-TG...CA-3', comuns a grande maioria dos elementos descritos, e a região interna do elemento apresenta uma única ORF (*gag* e *pol*) de 1.328 aminoácidos. Tnt1 também contém as regiões não codificantes adjacentes às LTRs: PBS na extremidade 5', e PPT na extremidade 3', correspondendo aos sítios importantes no processo de transposição de retroelementos com LTRs (Grandbastien *et al.*, 1989).

Através de estudos de variabilidade de sequências parciais de Tnt1 em *Nicotiana tabacum*, analisando tanto sequências genômicas quanto sequências de cDNA, Casacuberta *et al.* (1995) demonstraram que o elemento não representava uma única sequência, e sim diferentes populações de sequências. Além de analisar e observar a diferença na variabilidade das regiões do estudo (Figura 3), os autores puderam constatar que havia uma maior diversidade de cópias genômicas quando comparadas com as cópias de cDNA. Ainda, análises de expressão demonstraram que as cópias de cDNA diferiam dependendo do órgão ou tecido da planta em que eram expressas. A grande diversidade de cópias genômicas sugere que apenas algumas dessas são expressas. Também, a população de cópias genômicas apresentou grande variação na porção U3 da região 3'-LTR, contrastando com as populações de cópias expressas, as quais apresentaram maior uniformidade nessa região.

Em 1998, Vernhettes *et al.* estudaram a evolução das sequências de Tnt1 em sete espécies de *Nicotiana*. Os autores puderam observar que a região regulatória U3 das diferentes cópias de Tnt1 divergiu dando origem a três clados, classificando as sequências como Tnt1A, Tnt1B, ou Tnt1C. Ainda, os autores sugeriram que essa variabilidade em uma região de regulação da transcrição poderia ser influenciada por questões ambientais, e

permitiria a evolução de diferentes mecanismos regulatórios, melhorando a coexistência do elemento com o genoma em que habita.

Tnt1 pode ser ativado por diversos estresses bióticos e abióticos, incluindo agentes oxidantes, infecção por patógenos e elicitores microbianos (Pouteau *et al.*, 1991). Grandbastien *et al.* (1997), estudando a expressão induzida do elemento por diferentes patógenos, puderam verificar que a região U3 da LTR contém os principais componentes *cis*- regulatórios para a ativação da transcrição em associação com as respostas de defesa da planta.

Com base nas informações quanto aos diferentes grupos e diferentes tipos de estresse que podem induzir a replicação de Tnt1, Beguiristain *et al.* (2001) investigaram a expressão dos três grupos diferentes em espécies de *Nicotiana*. Nesse estudo foi possível observar que Tnt1A, B, e C estão presentes nas espécies estudadas, sugerindo que esses elementos são ativos, porém diferem em abundância. Também foi observado que quatro dos tipos de estresse testados induzem a expressão de cada tipo de elemento com intensidade diferente, demonstrando que cada grupo difere quanto ao padrão de expressão, e isso é devido aos diferentes elementos de indução do promotor presente na região U3 da 3'-LTR. Essa variabilidade na expressão pode ter consequências na evolução dos elementos, de forma que a exposição frequente a um tipo de estresse pode acarretar na amplificação de um dos tipos em particular, e isso pode variar de uma espécie para outra. Ainda, os autores constataram que os diferentes grupos são expressos em diferentes tecidos da planta, e que as sequências diferem dentro de um mesmo grupo de elementos.

Estudos relacionados com inserções de Tnt1 em *Nicotiana tabacum* e seus parentais (Le *et al.*, 2007) demonstraram uma possível preferência do elemento em se inserir próximo ou dentro de regiões codificadoras. Isso pode ocorrer devido a maior facilidade de se integrar em regiões eucromáticas de transcrição frequente. Os autores puderam constatar que tabaco apresenta maior número de inserções próximas ou dentro de genes que seus parentais, de forma que essas inserções podem ser mais bem toleradas devido à poliploidia da espécie.

De posse das informações de preferência de inserção do elemento, Hernández-Pinzón *et al.* (2009) investigaram a eficiência da terminação da transcrição de Tnt1 e sua habilidade de estender a transcrição para regiões adjacentes, o que pode influenciar na

regulação de genes próximos. Assim, os autores demonstraram que os transcritos de Tnt1 apresentam uma alta variabilidade de sítios de poliadenilação, originando populações de transcritos de diversos tamanhos, poucos terminando no sítio principal de terminação. Os autores também puderam verificar a capacidade que o elemento apresenta em transcrever regiões genômicas adjacentes, podendo gerar transcritos que se sobrepõem e se anelam formando moléculas de RNA de fita dupla, as quais podem ser processadas e utilizadas nas rotas de silenciamento de genes. Nesse trabalho também foi observado que sequências que apresentam grande similaridade na região U3 da 3'-LTR podem diferir quanto ao tamanho dos transcritos gerados, e que isso pode estar relacionado às deleções presentes nessas sequências. Assim, esse trabalho ainda levanta a hipótese de que a região U3 não está apenas associada à ativação da transcrição do elemento, mas também em sua terminação.

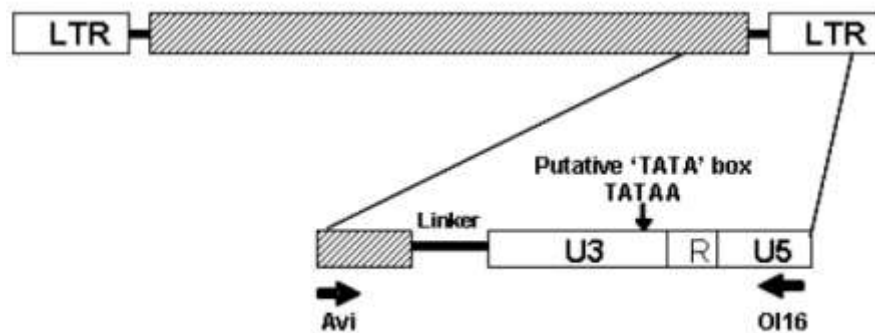


Figura 3: Representação esquemática da região de estudo de diversidade do elemento Tnt1 (Manetti *et al.*, 2007).

### 1.6.1. Elementos Similares a Tnt1 em Solanaceae

Elementos similares a Tnt1 já foram detectados em pimenta (Tam *et al.*, 2005), berinjela (Grandbastien *et al.*, 2005), tomate, batata, e petúnias-de-jardim (Grandbastien *et al.*, 1989). Porém, ainda são poucos os estudos de relacionamento entre esses elementos e o genoma no qual estão inseridos.

Em 2001, Araújo *et al.* investigaram a diversidade das sequências de Retrolyc1 (Costa *et al.*, 2000), elemento isolado de *Lycopersicon peruvianum* (atualmente no gênero *Solanum*), que apresenta similaridade a Tnt1, com exceção da porção U3 da LTR. Nesse trabalho foi detectada a existência de duas subfamílias: Retrolyc1A e Retrolyc1B, e a

existência de motivos repetitivos na U3 de Retrolyc1A, similares àqueles encontrados nas cópias ativas de Tnt1, sugerindo, então, a existência de cópias de Retrolyc1A ativas, a qual foi observada através da análise da expressão do elemento.

Manetti *et al.* (2007), analisando a diversidade das sequências similares a Tnt1 no gênero *Solanun*, puderam verificar a existência de três grupos do elemento denominado Retrosol. Ao comparar as sequências com aquelas de Tnt1 e Retrolyc1, os autores puderam observar que não há compartilhamento de sequências entre os gêneros, sugerindo que esses elementos surgiram de forma independente na família Solanaceae.

Em 2009, Tam *et al.* estudando as inserções de elementos similares a Tnt1 e T135 em espécies do gênero *Capsicum* puderam verificar que o polimorfismo de inserções é bastante útil no estudo da diversidade genética e do relacionamento entre as espécies do gênero.

## **1.7. O Gênero *Petunia***

*Petunia* Juss. (Solanaceae) é um gênero subtropical, endêmico da América do Sul, que apresenta 14 espécies (Stehmann *et al.*, 2009), a maioria delas encontradas no Sul e Sudeste do Brasil. Conhecidas popularmente pelas petúnias de jardim, híbrido artificial obtido no início do século XIX a partir do cruzamento entre as espécies *P. integrifolia* (Hook) Schinz & Tell e *P. axillaris* (Lam.) Britton, Sterns & Poggenb, este grupo de espécies apresenta uma longa história de cruzamentos artificiais. A obtenção de híbridos demonstra a falta de barreiras genéticas entre as espécies (Watanabe *et al.*, 1996).

As espécies de *Petunia* são subarborescentes ou herbáceas, decumbentes ou procumbentes, com raízes axiais. São plantas, em geral, anuais, monóicas, predominantemente heliófilas e podem ser encontradas nos mais variados tipos de solo. São habituais em campos, afloramentos rochosos e capoeiras de zonas entre o campo e a mata. O caule é cilíndrico e herbáceo, esverdeado ou purpúreo, com folhas sésseis ou subsésseis e flores vistosas, infundibuliformes em grande parte das espécies. Apresentam cálice fendido até próximo à base; a corola possui prefloração imbricada, podendo ser actinomorfa ou zigomorfa, de coloração predominantemente púrpura e, menos frequentemente, vermelha ou branca. Apresentam cinco estames, com anteras amarelas ou violáceas (Stehmann *et al.*, 2009).

O gênero apresenta história evolutiva recente, e sua taxonomia é complexa. Em 2006, Kulcheski *et al.* apresentaram a filogenia do gênero baseada em marcadores platiniais, mitocondriais, e nucleares. Nesse trabalho é possível observar a grande similaridade entre as espécies, que se reflete no baixo suporte dos clados formados. Ainda assim, três agrupamentos podem ser observados, bem como a proximidade com o gênero *Calibrachoa* La Llave & Lex., considerado seu grupo irmão.

Em 2007, Chen *et al.* publicou uma filogenia do gênero a partir de *exons* de um gene nuclear, e *introns* de um gene plastidial. Porém, foi observada uma separação do gênero em dois clados: um agrupando as espécies que ocupam principalmente o oeste e o sul da distribuição do gênero, e outro agrupando as espécies que ocorrem no nordeste da distribuição do gênero (Figura 4).

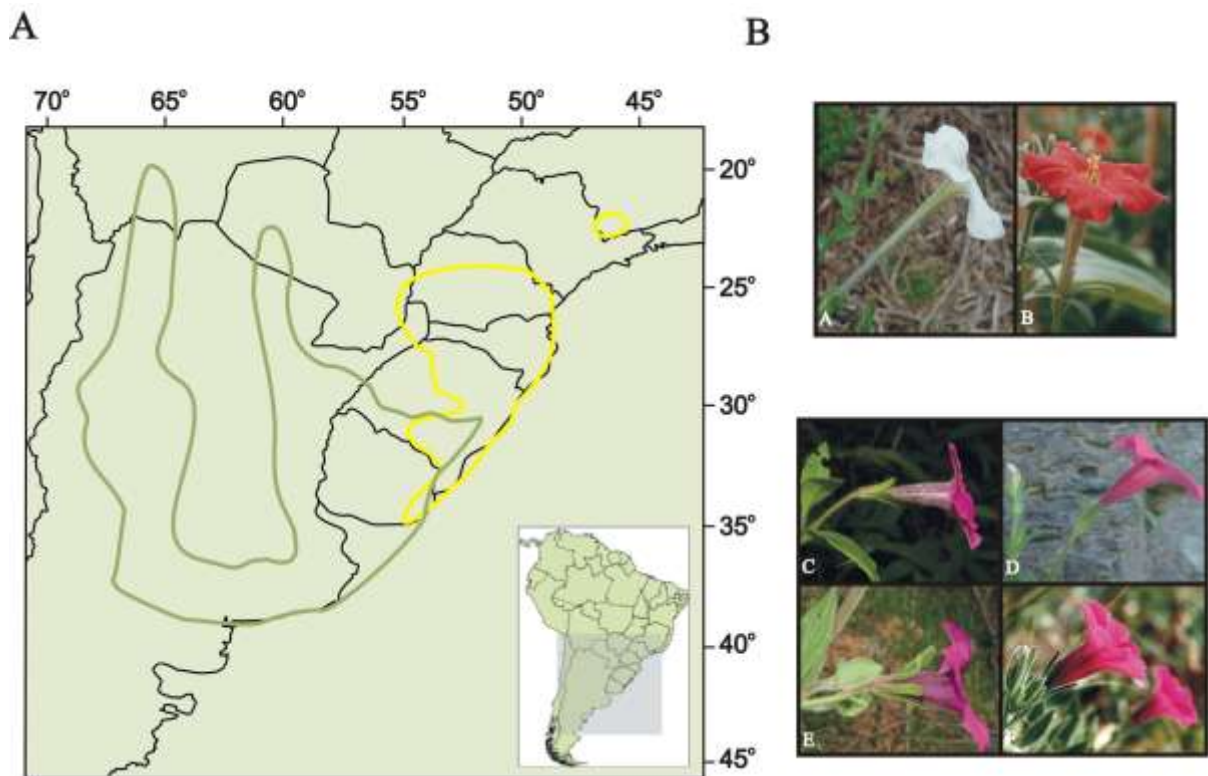


Figura 4: A- Distribuição do gênero *Petunia*. O contorno verde refere-se ao oeste da distribuição, e o contorno amarelo refere-se ao leste da distribuição. B- *P. axillaris* (A) e *P. exserta* (B), representantes do clado I; *P. mantiqueirensis* (C), *P. saxicola* (D), *P. bajeensis* (E), e *P. reitzii* (F), representantes do clado II (adaptado de Stehmann *et al.*, 2009).



### 1.7.1. Retrotransposons com LTR em *Petunia*

Poucos são os estudos de diversidade de retrotransposons com LTR no gênero *Petunia*. Através da análise filogenética dos domínios RT e RNaseH de *P. hybrida* V26, *P. axillaris* subesp. *axillaris* Ando, *P. axillaris* subesp. *parodii* Ando, *P. inflata* R.E.Fr., e *Calibrachoa parviflora* (Juss.) Wijsman, Richert-Pöggeler & Schwarzacher (2009) puderam constatar a presença de dois grupos nos genomas estudados: um grupo subdividido em representantes de dois dos três gêneros da família Ty1/Copia (*Hemivirus* ainda não foi detectado em plantas), e outro subdividido em representantes dos três gêneros de Ty3/Gypsy. Porém, nenhum subgrupo apresentou sequências exclusivas de apenas um genoma, ou de algum fenótipo (por exemplo, flores brancas vs. flores púrpuras), o que difere de diversos grupos de plantas nos quais os retroelementos são capazes de refletir a filogenia das espécies. Elementos de transposição são considerados por muitos como os componentes mais antigos dos genomas, de forma que sua história poderia levar a uma composição diferente de elementos entre as espécies. Essa heterogeneidade de sequências mantida nas espécies estudadas, corrobora a idéia de um processo de especiação recente (Kulcheski *et al.*, 2006), com unidades taxonômicas recém isoladas, permitindo cruzamentos interespecíficos com uma consequente troca de retroelementos.

## **OBJETIVOS**

## Objetivo geral

O objetivo geral desse trabalho é caracterizar a diversidade dos retrotransposons com LTR do clado Tork da superfamília *Copia*, relacionados ao elemento Tnt1 de tabaco, em espécies nativas de *Petunia*, tendo como objetivos específicos:

- a) Obter sequências similares a Tnt1 em espécies silvestres de *Petunia*;
- b) Caracterizar a diversidade das sequências obtidas;
- c) Analisar o relacionamento das sequências entre si; e
- d) Comparar as sequências obtidas para o gênero *Petunia* com as demais sequências deste e similares a este elemento disponíveis na literatura.

**Novel transposable elements in Solanaceae: evolutionary relationship  
among Tnt1-related sequences in wild *Petunia* species**

Novel transposable elements in Solanaceae: evolutionary relationship among  
Tnt1-related sequences in wild *Petunia* species

Raquel A. Kriedt<sup>1</sup>, Jéferson N. Fregonezi<sup>1</sup>, Guilherme M.Q. Cruz<sup>2</sup>, Sandro L. Bonatto<sup>3</sup>,  
Loreta B. Freitas<sup>1</sup>.

<sup>1</sup>Molecular Evolution Laboratory, Department of Genetics, Universidade Federal do Rio Grande do Sul, Bento Gonçalves 9500, PoBox 15053, 91501-970, Porto Alegre, RS, Brazil.

<sup>2</sup>GaTE Lab, Department of Botany, Universidade de São Paulo, Rua do Matão 277, São Paulo, 05508-090 SP, Brazil.

<sup>3</sup>Genomic and Molecular Biology Laboratory, Pontifícia Universidade Católica do Rio Grande do Sul, Ipiranga 6681, 90610-001, Porto Alegre, RS, Brazil.

Short title: Tnt1-related elements in *Petunia*

Correspondence author: Loreta B. Freitas, Phone 55 51 3308-6731, Fax: 55 51 3308-9823;  
e-mail: [loreta.freitas@ufrgs.br](mailto:loreta.freitas@ufrgs.br)

## Abstract

Transposable elements (TEs) are widespread among eukaryotic genomes. TEs diversity and abundance are highly variable among species, and may reflect a specific relationship between species and TEs in its genome. There are usually many TEs families in a single genome, and amplification of a given family can have an impact in another family's dynamics. LTR retrotransposons are very abundant in flowering plants, with Tnt1 being one of most well known plant LTR retrotransposons, and one of the few retroelements with demonstrated transpositional activity. Tnt1 was first characterized in tobacco, but Tnt1-related sequences have been reported in other Solanaceae genera. In this work we investigated the Tnt1-related sequences profile among species of the *Petunia*, *Calibrachoa* and *Fabiana* genera through genomic amplification and cloning of Tnt1-related element's partial sequence. Sequences analysis revealed very diverse sequences, not as closely related to Tnt1 as previously thought. Sequences classification gave rise to ten possible TEs families, all of them as part of the Tork clade of *Copia* superfamily. Sequences did not follow species phylogeny, however, they were not homogeneously distributed. One family comprises only sequences of five of the 21 taxa studied in this work, which inhabit dryer areas in *Petunia* geographic distribution. We were not able to associate the remaining families with any pattern. These finding corroborated to previous suggestions of relationship among these species, as well as an early association of Tnt1-related elements in the Solanaceae's evolution.

## Introduction

Transposable elements (TEs) are characterized as DNA segments capable of mediating their own movement and insertion on different regions of the genome [1]. These elements are divided in two major classes, according to the presence or absence of an RNA transposition intermediate: class I TEs transpose via an RNA intermediate, while class II transpose directly through DNA, and contains two subclasses that differ accordingly to the number of DNA strands that are cut during transposition [2].

TEs are a major component of angiosperm DNA [3], and retroelements may constitute over 90% of TE fraction in some plant species [4; 5]. Retrotransposons are separated into five orders [2], LTR (long terminal repeat) retrotransposons being the predominant order in plants [2; 6]. LTR retrotransposons present long terminal repeat in both ends, in direct orientation. These LTRs are subdivided in unique 3'RNA (U3), repeat RNA (R), and unique 5'RNA (U5) regions, and contain promoters and terminators associated with element's transcription [7]. Also, LTRs have been reported as the faster evolving regions of retrotransposons [8; 9].

LTR retrotransposons have two major genes (*gag* and *pol*) which encode a number of proteins. The *gag* gene encodes a capsid-like protein and the *pol* gene encodes a polyprotein that is responsible for protease, reverse transcriptase (RT), ribonuclease (RNase H), and integrase activities, which are required for replicative transposition of the element [7]. By copying themselves into new locations, retroelements are able to increase the genome size of their host [10; 11; 12], being very abundant in flowering plants, representing up to 70% of angiosperms nuclear genome [2; 6; 13]. LTR retrotransposons are classified into five superfamilies: Ty1/*Copia* (*Pseudoviridae*), Ty3/*Gypsy* (*Metaviridae*), *Bel/ Pao*, retroviruses and *ERV* (endogenous retroviruses). Up to now,

members of the three last superfamilies have been detected only in metazoan hosts, while Ty1/*Copia* and Ty3/*Gypsy* elements have been reported in all eukaryotic lineages [2].

Tnt1 is one of the best known plant LTR retrotransposons, being one of the few retroelements for which transpositional activity has been demonstrated [14; 15]. Tnt1 is a member of the Tork clade of Ty1/*Copia* superfamily [16], represented in high copy number in the *Nicotiana tabacum* genome [14]. There are at least three different Tnt1 elements in several *Nicotiana* species: Tnt1A, Tnt1B, and Tnt1C [17; 18; 19], and studies were able to detect not only that the Tnt1 element can be activated by biotic and abiotic stress [20], but that different types of stress can activate each subfamily in a different way, due to variability of the U3 regions, which contain different stress-inducible promoters [21]. Tnt1-related elements have been identified by membrane hybridizations and PCR approach in other Solanaceae species. Sequences were identified in garden petunias [14], tomato [22; 23], eggplant [15], potato [24], and pepper species [25].

TEs distribution in a given genome is neither homogeneous nor random [1]. Accumulation and loss of LTR retrotransposons can contribute to interspecific variation in nuclear DNA content, helping to explain how closely related species can differ in genome sizes despite sharing a similar evolutionary history [26]. There may be many TE families in a single genome, and the amplification of a particular family can have an impact in another family's dynamics [1]. Differences in retrotransposon expression can affect an element's evolutionary history, and frequent exposition to a specific stress may result in amplification of a certain type of element, which can lead to different retrotransposons profile among species [21]. In this study, we investigated the Tnt1-related sequences profile among species of the *Petunia* genus through genomic amplification of Tnt1-related element's partial sequence. We were able to detect ten possible families of Tork elements



in *Petunia* species and its close relative *Calibrachoa*, as well as a differential distribution of retrotransposon families among species.

## **Results**

### **Sequence diversity and classification**

The presence of Tnt1-related sequences among *Petunia*, *Calibrachoa*, and *Fabiana* species (referred to as petunias) was verified by amplification and cloning of fragments that comprised the terminal portion of ribonuclease H (RNaseH) domain, as well as the Linker DNA (polypurine tract), and the 3'LTR (U3, R, and part of the U5 region) of the elements. The number of cloned sequences from each species varied from 5 to 20, resulting in a total of 148 non-redundant fragments. Sequences varied from 250-450 bp long. Overall, sequences resulted in a 477 bp alignment. On account of none of these species having well-characterized genomes, we do not rely on complete genomic information. For this reason, we could not determine if the fragments amplified are representative of the total sequence diversity. Nevertheless, this seems to be an appropriate manner of accessing Tnt1-related sequence diversity [19; 21; 22; 24].

Sequence MegaBlast was performed to evaluate the relatedness of sequences to Tnt1-like elements available on GenBank. According to attempts on establishing a universal classification system for TEs [2], classification under the same family relies on at least 80% identity across 80% of either an elements' coding region or its terminal repeats, with a minimum of 80 bp. None of the retrieved sequences followed that parameter when compared to Tnt1, Retrosol, or Retrolyc1 sequences. Therefore, even though there is great similarity between our sequences group and these Solanaceae LTR *Copia* elements group previously described, especially so because primers used to retrieve these sequences of

*Petunia* were those designed for *Solanum* species, Blast results strongly suggests that we are dealing with different families.

The nucleotide diversity among sequences was  $\pi = 0.167 \pm 0.005$ . As described for other Tnt1-related sequences, there was significant heterogeneity in the U3 region of the 3'LTR, therefore, we performed a Blast analysis for all sequences against each other. Those that were able to fit the 80-80-80 rule were considered as the same family, which gave rise to ten possible families, one with only *Calibrachoa* sequences. When submitting the partial RNaseH sequences to Gypsy Database (GYDB) [16] core search, the closest hits were those of Tnt1, Tork, and Forf elements, respectively. This indicates that our sequences represent, as Tnt1, the Tork clade of *Copia* elements.

According to current classification system [2], the suited name for these new sequences are RLC\_peTork1-9 (standing for Retroelement LTR *Copia* Tork of *Petunia*, families one through nine) and RLC\_cpTork1, for that one family were only represented by sequences retrieved of *Calibrachoa* species so far. Also, as indels within sequences of a same group could characterize subfamilies, we may as well be able to detect a few clusters within some of the families (for example see Figure S1 in Supplementary Material). Families'  $\pi$  values are presented in Table 1. RLC\_peTork3 was the less diverse family, with  $\pi = 0.018$ , and RLC\_peTork9 was shown to be the most diverse family, with  $\pi = 0.215$ . However, RLC\_peTork9 is represented by only two sequences, although of the same species.

Because Retrosol studies have shown that a few sequences have no stop codon [24], and frequent transcriptional read through of Tnt1 retrotransposon have been reported [27], we investigated a possible lack of stop codons in RNaseH domain of *Petunia* sequences. Even though RLC\_cpTork1, and RLC\_peTork1, 8 and 9 have a different three nucleotide

(TAG) sequence comparing to RLC\_peTork2 to 6 (TAA), all sequences presented a stop codon in the same place as Tnt1's putative stop codon.

### **Relationship among petunias' sequences**

Sequences' relationship was accessed by Neighbor-Joining analysis of full-length fragments. Sequences from one species can be found in more than one RLC family, therefore not following species' phylogeny. Relationship among all sequences is shown in Figure 1, and two main groups can be detected, followed by internal clades of different bootstrap values. Group 2 accounts more than 65% of overall sequences. Table 1 shows species and families which sequences belong to. RLC\_peTork1 comprises sequences of species closer related to each other than to other *Petunia* species. These taxa inhabit dryer areas of *Petunia*'s distribution range [28] (for RLC\_peTork1 distribution see Figure S2 in the Supplementary Material). RLC\_peTork5, 7, 8 and 9, are represented by sequences of purple flowered *Petunia* species, which is also the case of RLC\_peTork3 and 4, except for the sharing of *Calibrachoa* and *Fabiana* sequences, which comprises purple, white, and pinkish flowers. RLC\_peTork2 and 6 mix sequences of both red and purple *Petunia* flowers, although in very different proportions (only one sequence of red flower taxa in each family).

Given the well established role of U3 region on element's transcription and group determination, we performed relationship analysis excluding this region of the alignment. As shown in Figure 2, relationship among sequences, and especially among groups, were not able to maintain the previous monophyly of families.

Network relationship analysis has been designed to investigate closely related sequences. This approach allows us to visualize multifurcation and reticulations, as well as the ancestral nodes that are not solved when analyzing a phylogenetic tree [30]. Median-

Joining network was able to recover the two main groups cited above. The most frequent sequence is named master copy, or  $\alpha$ , and occupies a central position in the network, having many connections to closely related sequences [30]. As shown in Figure 3, there may be one master copy in Group 2 ( $\alpha_1$ ), accounting for almost 44% of all petunias sequences.

### **Relationship among all Tnt1-related sequences**

When combining our data with 110 sequences of already described Tnt1-related elements, three main clades were formed. One comprising RLC\_peTork3-7 families, another with RLC\_peTork1 and 2 families, and the third one comprising RLC\_peTork8 and 9, as well as RLC\_cpTork1, and Tnt1, Retrolyc1, and Retrosol sequences, although bootstrap values supporting this clade was low (Figure 4). Analysis excluding the U3 region resulted in two main clades: one with all ingroup sequences, and another with all Tnt1-related elements (for relationship analysis excluding the U3 region see Figure S3 in the Supplementary Material). As we placed our sequences with remaining Tnt1-like elements in a Median-Joining network analysis of sequences' coding region, both petunias groups appear to be closer to previously described elements than to each other. Nonetheless, both groups seem to be equally divergent of any other Tnt1-related element, though there are Tnt1 sequences within petunias' groups (Figure 4). There are no RNaseH sequences shared among Tnt1, Retrolyc1, Retrosol, and petunias sequences.

### **Discussion**

LTR retrotransposons are ubiquitous elements present in almost all analyzed genomes. In the era of genome sequencing projects, we are surrounded by information on TE composition of different genomes. Being able to compare entire genomes made the discovery of TEs sharing among different genera possible. Tnt1 was first identified in

tobacco and detected in other Solanaceae species by probe hybridization afterward [14]. Since then, Tnt1-related elements have been identified and studied in wild *Nicotiana* species [17], tomato [22; 23] and potato [24], and their wild relatives, being widespread in solanaceous plant species. Here we report novel Tnt1-related families in three genera of wild species: *Petunia*, *Calibrachoa*, and *Fabiana*.

The pool of sequences recovered in our research is remarkably diverse, giving rise to ten novel TEs families present in several species. Our data point to a complex evolutionary history among TEs families, which may include expansions, losses and maybe even transfers due to hybridization possibilities among species. Groups of sequences with short branch lengths and low values of nucleotide diversity are consistent with predictions for retrotransposon recent proliferation, once proliferative elements generate copies that are identical at the time of insertion, acquiring mutations independently afterward [26]. The recent proliferation of these TEs families may be related to *Petunia*'s ongoing speciation process. As colonization of new habitat occurred, taxa may have been exposed to different types of stress. Environmental stress has been implicated as a causal agent of LTR retrotransposon derepression and proliferation in plants [26]. Thus, it seems possible that the different habitats and associated environmental stresses experienced by emerging species could be associated with these different TEs proliferations.

All sequences belonging to RLC\_peTork1 family are, of plants that inhabit the Western territory of *Petunia*'s range of distribution, which is dryer than the remaining parts of its range (*P. axillaris* subsp. *parodii* and *P. occidentalis*). Also *P. exserta*, which can only be found in small "caves" formed in sandstones of the Serra do Sudeste (Rio Grande do Sul, Brazilian state), *P. secreta*, which inhabits sandstone's tops in the same area, and

*P. axillaris* subs. *axillaris*, which can be found in Southern Brazil, along the coast of Uruguay, and Northeast Argentina [28]. These findings are in agreement with previous phylogenetic reconstruction in which these taxa were placed together by *Hfl* gene analysis, as well as plastid *trnK* intron [31]. Furthermore, this clade contains only the self-compatible species of *Petunia* genus [28], including the natural hybrid between *P. axillaris* vs. *P. exserta* [29]. These taxa are the most different, morphologic wise, when compared to the remaining taxa. They are very closely related, and can all be found in rocky substrate with little humidity. Correlation between water deprivation and retrotransposon copy number has been reported for *BARE-1* elements of barley. In such case, the copy number of *BARE-1* in *Hordeum spontaneum* increased with aridity across its range, suggesting that *BARE-1* expression and propagation may be stress induced, and arid condition could be responsible for a lower rate of loss of integrated elements [32]. Taking into account that Tnt1 elements from tobacco are differentially expressed depending on specific stress conditions [21], even though there are no expression studies for these *Petunia* elements, our hypothesis is that plants exposed to water deprivation, or most likely their ancestral, may have preferentially amplified one RLC\_peTork family.

As mentioned before, the amplification of a particular family can have an impact in another family's dynamics [1]. In this manner, if the ancestral of the five *Petunia* taxa represented by RLC\_peTork1 presented a family of elements which insertion had some advantage under arid conditions, this family may have been over amplified in comparison to other families. Once differences in retrotransposon expression may affect TE evolutionary history, which can lead to different retrotransposons profile among species [21], it is possible that RLC\_peTork elements distribution among *Petunia* was influenced by the genera evolutionary history. Even though *Petunia* is placed in a basal position of Solanaceae phylogenetic tree [33], its species present a very recent evolutionary history, in

which species' border are not well established, and its relationship to its sister groups are still being solved [34]. The genetic proximity of *Petunia* and *Calibrachoa* species was also reflected in our tree. With exception of RLC\_cpTork1, which comprises only *Calibrachoa* sequences, *Calibrachoa* and *Fabiana* sequences were placed among *Petunia*'s ones.

Despite the fact that RLC\_peTork1 family is represented by five very closely related taxa, one sequence retrieved from *P. axillaris* vs. *P. exserta*, and one sequence from *P. exserta* were placed in different families (RLC\_peTork6 and 2, respectively), indicating that these TE lineages are ancient and predate the divergence of species. This pattern can be found throughout the tree (Figure 1), and it can also be detected genus wise, once the sequence RLC\_peTork3.2 is shared among *C. excellens*, *C. pygmaea*, *F. imbricata*, and *P. scheideana*. The observed sequence plasticity of these tnt1-like elements within Solanaceae species may account for adaptation and preservation of these elements within distinct host genomes after invasion by an ancestral element early in Solanaceae's evolution [24].

Diversity of the U3 region was determinant on Tnt1 elements group division, giving rise to Tnt1A, B, and C [19]. The same occurs within our sequences, in which the U3 region plays an important role on distinguishing among families, even though U5 region also showed to be very diverse among groups. When the U3 region was excluded, relationship among sequences, and especially among groups, did not maintain the previous monophyly of families (Figure 2). Families that were able to maintain their identity could only do so for its U5 sequence profile. However, when excluding the U5 portion of the aligned sequences, all families were able to maintain their monophyly, but relationships among families were not the same (for relationship analysis excluding the U5 region see Figure S4 in the Supplementary Material).

Network analyses allow us to better visualize relationship among closely related sequences. The partly star-like relationship of petunias' sequences in the network, as well as the proportion of elements that are not directly connected to the central master copy, suggest that several members may contribute to expansion of these elements in *Petunia*. The presence of a master copy ( $\alpha 1$ ) in Group 1 suggests that this sequence may represent an active element, and that the adjacent copies may be active as well [30]. Copies that emerge from  $\alpha 1$  may also contribute to elements' expansion, for it only differs of the master copy by one point mutation. In the way Group 2 network relationships are depicted with not as strong star-like topology as Group 1, it is plausible that many family members are capable of amplification [30].

The present composition of elements in species genomes emerged of an adaptative response of the ancestral elements' population to different host [15]. The different history of these RLC families may be representative of past selective amplifications of ancestral copies that could have influenced species adaptation to their current habitat. These families represent either advantageous or non-deleterious elements, considering that non-transcribed copies would probably have been rapidly inactivated and lost. No species correlation among RLC families retrieved in this study was found, but there seem to be a bias of a group of species and its Tnt1-related sequences, in accordance to species relatedness described in nuclear exon and chloroplast intron phylogenetic analysis [30]. Also, the Tnt1-like sequence profile among petunias included in this work suggests different amplification and/or loss of elements in different species (or groups of species). Although we cannot definitively exclude PCR or other artifacts as having artificially contributed to observed levels of sequence diversity, the amount of sequences retrieved is unlikely to bias our results. Analysis of copy number variation of different families among species, as well as complete element characterization, may help to understand Tnt1-related



elements evolutionary history among the studied species. Expression studies of these families may be able to correlate the presence of RLC\_peTork1 and a possible advantage for those taxa that live under arid conditions. Such association would be of special interest for those taxa that live under arid conditions. Such association would be of special interest for plant breeding programs, since garden petunias are one of the world's most important ornamental plants, with seed trading generating millions of dollars every year. It would be very interesting to be able to strongly suggest a correspondence between quantity and quality of elements and host ecology, especially with great potential for enhancing cultivated species utilizing native *Petunia* as a source of genetic variability and agronomic features [35].

## **Materials and methods**

### **DNA extraction and amplification**

Total DNA was extracted from young dried leaves, with cetyltrimethyl ammonium bromide (CTAB) protocol [36]. One specimen of most wild *Petunia* taxa, including the natural hybrid between *P. axillaris* vs. *P. exserta*, as well as one *Calibrachoa parviflora*, *C. pygmaea*, *C. excellens*, and one *Fabiana imbricata* specimens (Table 2), was submitted to PCR reactions as follows: 0.2  $\mu$ M of each primer [24], 1.5 mM MgCl<sup>+2</sup>, 0.2 mM dNTPs, 1x Taq buffer and 0.5 Unit of Taq polymerase were mixed in a total reaction volume of 10  $\mu$ l. The PCR cycle conditions were: one cycle of 94°C for 5 min, followed by 35 cycles of 94°C for 45 sec, 51°C for 45 sec, 72°C for 60 sec; and finally one extension cycle of 72°C for 10 min. Fragments are composed of the end portion of ribonuclease H (RNAseH) domain of the element, as well as the Linker DNA (polypurine tract), and the 3'LTR (U3, R, and part of the U5 region). The PCR products were checked by horizontal electrophoresis in 1% agarose gel, stained with GelRed (Biotium).

## **Cloning and sequencing of amplified fragments**

Each reaction was cloned into pGen-T-Easy vector (Promega) and transformed into chemically competent *Escherichia coli* XL1 blue following manufactures' instructions. Cloned fragments were retrieved using SP6 and T7 primers according to manufactures' instructions, and purified with PEG20% [29]. Sequencing reactions were performed using ET terminators kit (GE Healthcare Biosciences, Pittsburgh, Pennsylvania, USA) on MegaBACE 1000 automatic machine (GE Healthcare Biosciences) following manufacturer's instructions. Redundant sequences from the same specimen were excluded (for GenBank numbers see Table S1 in the Supplementary Material).

## **Sequence analysis**

Sequence alignments were performed with ClustalW [37] with subsequent manual adjustment when necessary. To evaluate sequences relationships, Neighbor-Joining trees with 1000 bootstrap replications were performed with Mega 5 [38], based on Kimura 2-parameter distance model. Average nucleotide diversity ( $\pi$ ) was calculated using DnaSP version 5.00.03 [39]. To better understand our sequences relationship with previous Tnt1-like elements, representatives of all three Tnt1 groups (n = 33), all three Retrosol groups (n = 52), as well as both Retrolyc1 groups (n = 23) were included in our analysis. Also, to evaluate the importance of the U3 region on group assigning, analysis were performed with and without this region portion in the alignment.

As multifurcating network analyses better represent complex evolutionary relationships among closely related sequences [40], Median-Joining networks were constructed for the coding region (RNAse H) of the retrieved sequences, using the software Network version 4.510 [41] with equal weights to all variable sites, to evaluate overall relationship, as well as existence of master copies in our ingroup.

## References

1. Hua-Van A, Le Rouzic A, Boutin T, Filée J, Capy P (2011) The struggle for life of the genome's selfish architects. *Biol Direct* 6: 19-48.
2. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, *et al.* (2007) A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* 8: 973-982.
3. Tenaillon MI, Hollister JD, Gaut BS (2010) A triptych of the evolution of plant transposable elements. *Trends Plant Sci* 15: 471-478.
4. Mao L, Wood TC, Yu Y, Budiman MA, Tomkins J, *et al.* (2000) Rice transposable elements: A survey of 73,000 sequence-tagged-connectors. *Genome Res* 10: 982-990.
5. Wei F, Stein JC, Liang C, Zhang J, Fulton RS, *et al.* (2009) Detailed analysis of a contiguous 22-Mb region of the maize genome. *PLoS Genet* 5: e1000728.
6. Kalendar R, Flavell AJ, Ellis THN, Sjakste T, Moisy C, *et al.* (2011) Analysis of plant diversity with retrotransposon-based molecular markers. *Heredity* 106: 520-530.
7. Kumar A, Bennetzen JL (1999) Plant Retrotransposon. *Annu Rev of Geneti* 33: 479-532.
8. Kalmykova AI, Gvozdev VA (2004) Selective expansion of the newly evolved genomic variants of retrotransposon *1731* in the *Drosophila* genomes. *Mol Biol Evo* 21:2281-2289.
9. Vicient CM, Kalendar R, Schulman AH (2005) Variability, recombination, and mosaic evolution of the barley *BARE-1* retrotransposon. *J Mol Evol* 61: 275-291.
10. Boeke JD, Corces VG (1989) Transcription and reverse transcription of retrotransposons. *Annu Rev Microbiol* 43: 403-434.
11. Wilhelm M, Wilhelm FX (2001) Reverse transcription of retroviruses and LTR retrotransposons. *Cell Mol Life Sci* 58: 1246-1262.
12. Zedek F, Smerda J, Smarda P, Bures P (2010) Correlated evolution of LTR retrotransposons and genome size in the genus *Eleocharis*. *BMC Plant Biol* 10: 265-275.
13. Bennetzen JL (2005) Transposable elements, gene creation and genome rearrangement in flowering plants. *Curr Opin Genet Dev* 15: 621-627.
14. Grandbastien MA, Spielmann A, Caboche M (1989) Tnt1, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. *Nature* 337: 376-380.
15. Grandbastien MA, Audeon C, Bonnivard E, Casacuberta JM, Chalhoub B, *et al.* (2005) Stress activation and genomic impact of Tnt1 retrotransposons in Solanaceae. *Cytogenet Genome Res* 110: 229-241.

16. Llorens C, Futami R, Covelli L, Dominguez-Escriba L, Viu JM, *et al*, (2011) The Gypsy Database (GyDB) of mobile genetic elements: Release 2.0. *Nucleic Acids Res* 39 (Database issue): D70-D74.
17. Hirochika H, Otsuki H, Yoshikawa M, Otsuki Y, Sugimoto K, *et al* (1996) Autonomous transposition of the tobacco retrotransposon Tto1 in rice. *Plant Cell* 8: 725-734
18. Casacuberta JM, Vernhettes S, Grandbastien MA (1995) Sequence variability within the tobacco retrotransposon Tnt1 population. *EMBO J* 14: 2670-2678.
19. Vernhettes S, Grandbastien MA, Casacuberta JM (1998) The evolutionary analysis of the Tnt1 retrotransposon in *Nicotiana* species reveals the high variability of its regulatory sequences. *Mol Biol Evol* 15: 827-836.
20. Pouteau S, Huttner E, Grandbastien MA, Caboche M (1991) Specific expression of the tobacco Tnt1 retrotransposon in protoplasts. *EMBO J*. 10: 1911-1918.
21. Beguiristain T, Grandbastien MA, Puigdomenech P, Casacuberta JM (2001) Three Tnt1 subfamilies show different stress-associated patterns of expression in tobacco. Consequences for retrotransposon control and evolution in plants. *Plant Physiol* 127: 212-221.
22. Costa APP, Scortecci KC, Hashimoto RY, Araujo PG, Grandbastien MA, *et al*. (1999) Retrolyc1-1, a member of the Tnt1 retrotransposon super-family in the *Lycopersicon peruvianum* genome. *Genetica* 107: 65-72.
23. Araujo PG, Casacuberts JM, Costa APP, Hashimoto RY, Grandbastien MA, *et al*. (2001) Retrolyc1 subfamilies defined by different U3 LTR regulatory regions in the *Lycopersicon* genus. *Mol Genet Genomics* 266: 35-41.
24. Manetti ME, Rossi M, Costa APP, Clausen AM, Van Sluys MA (2007) Radiation of the Tnt1 retrotransposon superfamily in three Solanaceae genera. *BMC Evol Biol* 7:34-46.
25. Tam SM, Lefebvree V, Palloix A, Sage-Palloix AM, Mhiri C, *et al*. (2009) LTR-retrotransposons Tnt1 and T135 markers reveal genetic diversity and evolutionary relationships of domesticated peppers. *Theor Appl Genet* 119: 973-989.
26. Kawakami T, Strakosh SC, Zhen Y, Ungerer MC (2010) Different scales of Ty1/Copia-like retrotransposon proliferation in the genomes of three diploid hybrid sunflower species. *Heredity* 104: 341-350.
27. Hernández-Pinzón I, Jesús E, Santiago N, Casacuberta JM (2009) The frequent transcriptional readthrough of the tobacco Tnt1 retrotransposon and its possible implications for the Control of Resistance Genes. *J Mol Evol* 68: 269-278.
28. Tsukamoto T, Ando T, Kokubun H, Watanabe H, Tanaka R, *et al*. (1998) Differentiation in the status of self-incompatibility among all natural taxa of *Petunia* (Solanaceae). *Acta Phytotaxon Geobot* 49: 115-133.

29. Lorenz-Lemke AP, Mäder G, Muschner VC, Stehmann JR, Bonatto SL, *et al.* (2006) Diversity and natural hybridization in a highly endemic species of *Petunia* (Solanaceae): A molecular and ecological analysis. *Mol Ecol* 15: 4487-4497.
30. Cordaux R, Hedges DJ, Batzer MA (2004) Retrotransposition of *Alu* elements: how many sources? *Trends Genet* 20: 464-467.
31. Chen S, Matsubara K, Omori T, Kokubun H, Kodama H, *et al.* (2007) Phylogenetic analysis of the genus *Petunia* (Solanaceae) based on the sequence of the Hf1 gene. *J Plant Res* 120: 385-397.
32. Kalendar R, Tanskanen J, Immonen S, Nevo E, Schulman AH (2000) Genome evolution of wild barley (*Hordeum spontaneum*) by *BARE-1* retrotransposon dynamics in response to sharp microclimatic diversity. *Proc Natl Acad Sci USA* 97: 6603-6607.
33. Olmstead RG, Bohs L, Migid HA, Santiago-Valentin E, Garcia VF, *et al.* (2008) A molecular phylogeny of the Solanaceae. *Taxon* 57:1159-1181.
34. Kulcheski FR, Muschner VC, Lorenz-Lemke AP, Stehmann JR, Bonatto SL, *et al.* (2006) Molecular phylogenetic analysis of *Petunia* Juss. (Solanaceae). *Genetica* 126: 3-14.
35. Gerats T, Vandenbusshe M (2005) A model system for comparative research: *Petunia*. *Trends Plant Sci* 10: 251-256.
36. Roy A, Frascaria N, MacKay J, Bousquet J (1992) Segregating random amplified polymorphic DNAs (RAPDs) in *Betula alleghaniensis*. *Theor Appl Genet* 85: 173-180.
37. Thompson JD, Higgins DG, Gibson TJ (1994) Clustal-W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4673-4680.
38. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, *et al.* (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol* 28: 2731-2739.
39. Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19: 2496-2497.
40. Posada D, Crandall KA (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14: 817-818.
41. Bandelt HJ, Forster P, Sykes BC, Richards MB (1995). Mitochondrial portraits of human-populations using median networks. *Genetics* 141: 743-753.

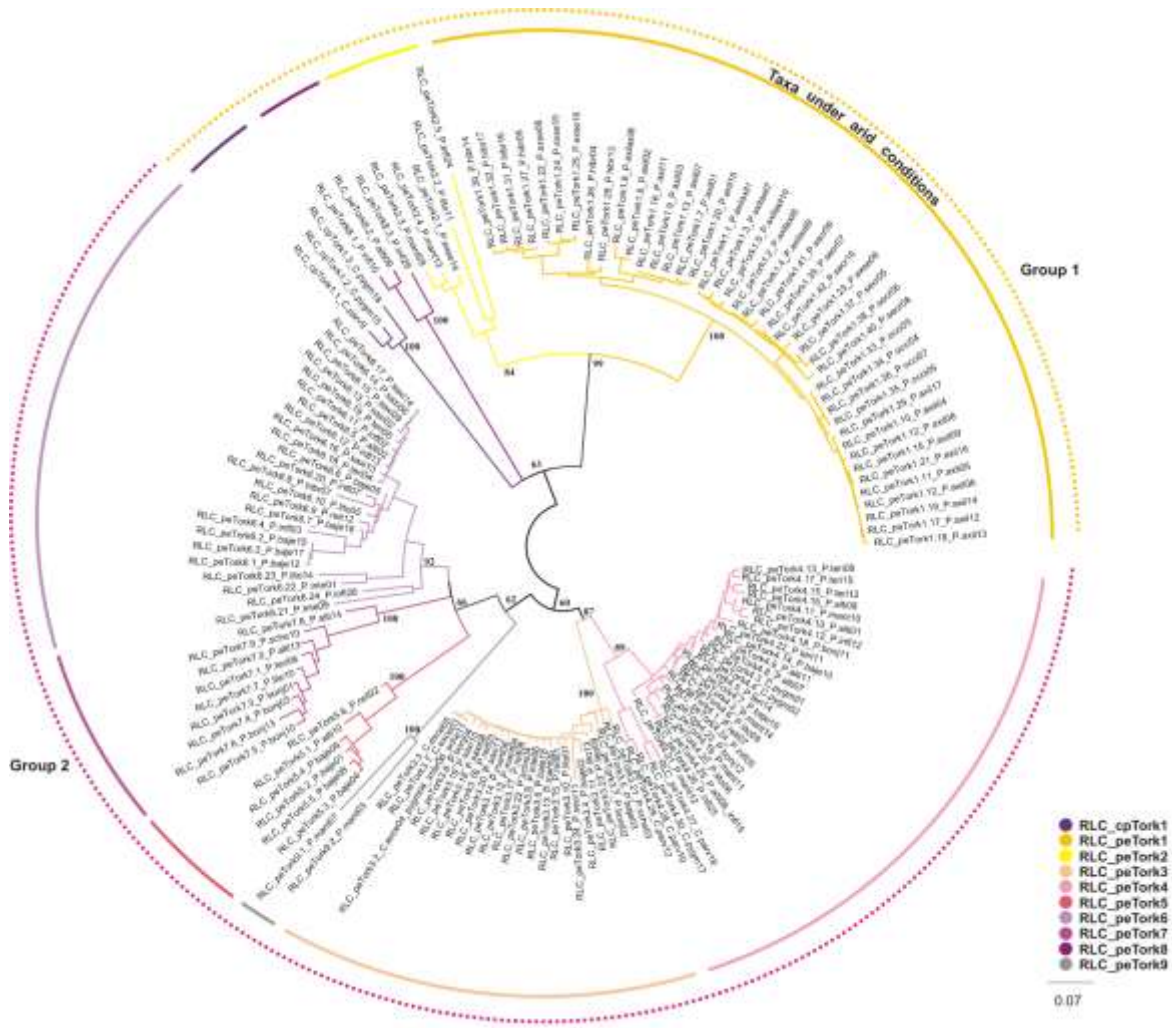
## Figure legends

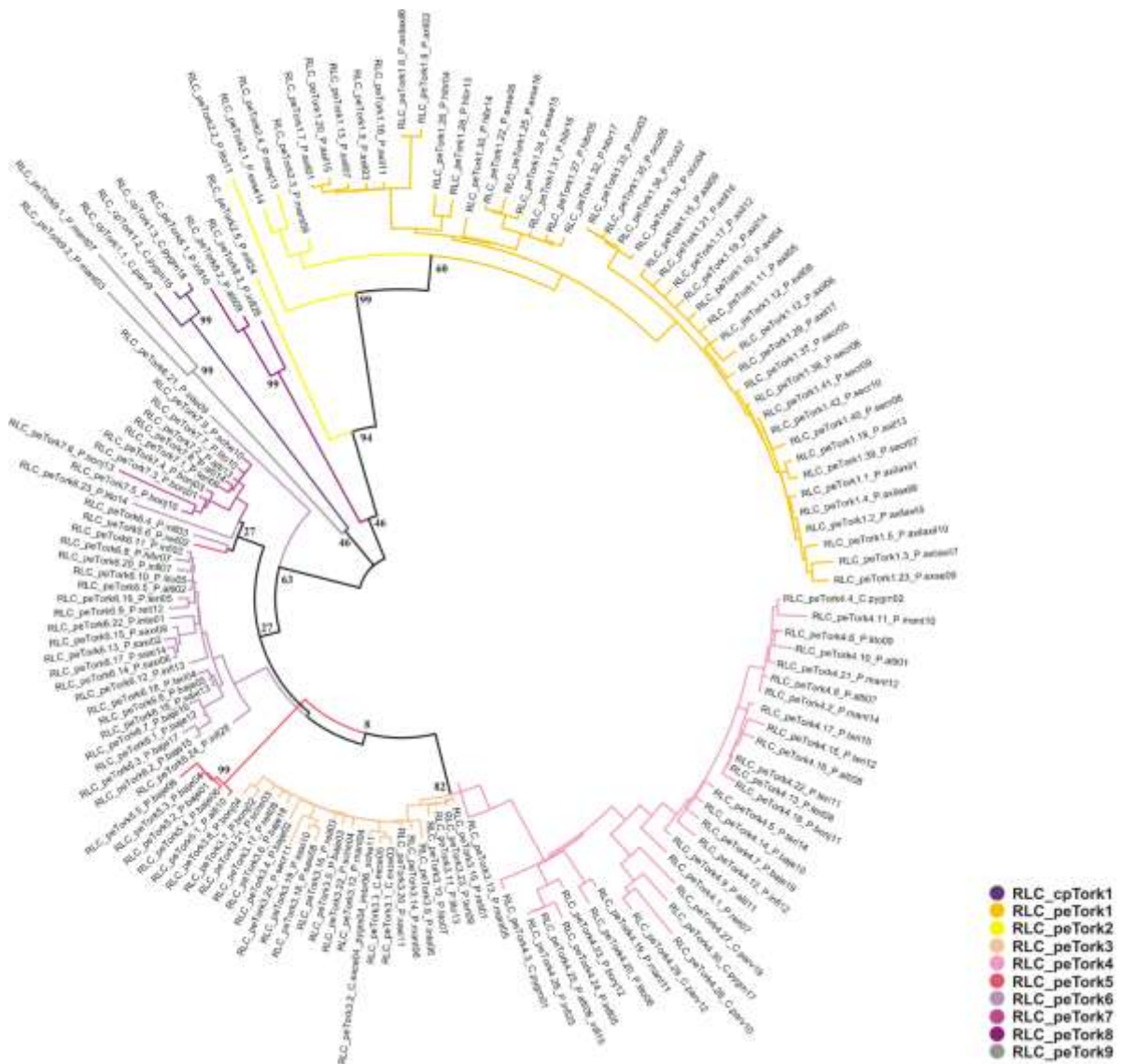
Figure 1: Neighbor-Joining tree of petunias' sequences relationship. Families' are indicated by different colors; Groups are represented by dotted lines. Clade of sequences belonging to species that live under dryer conditions is indicated, along with main bootstrap values near to nodes.

Figure 2: Neighbor-Joining tree of petunias' relationship excluding U3 region of aligned sequences. Families' are indicated by different colors.

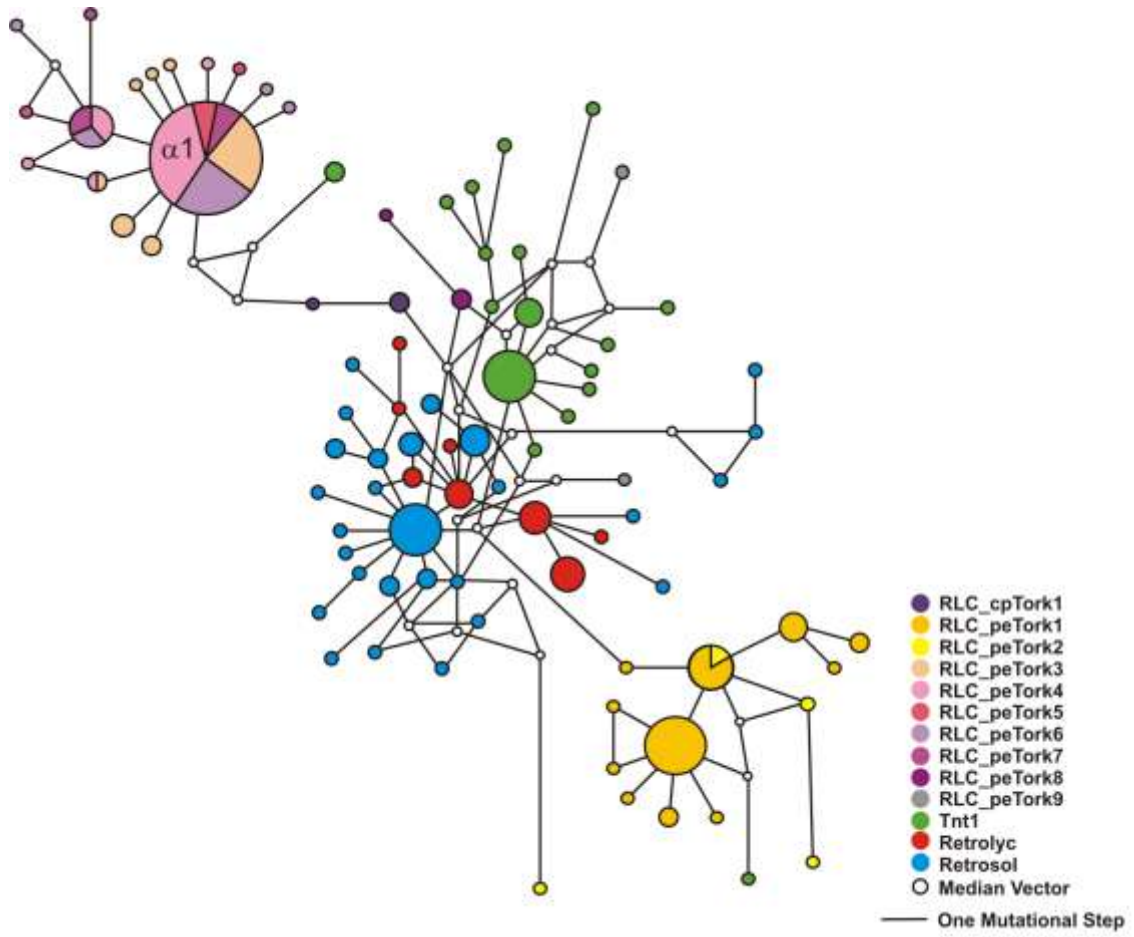
Figure 3: Median-Joining network of petunias' sequences and Tnt1, Retrolyc1, and Retrosol representatives. Petunias' families' and remaining TEs are indicated by colors. Master copy is indicated by  $\alpha 1$  symbol.

Figure 4: Neighbor-Joining tree of petunias' sequences and Tnt1, Retrolyc1, and Retrosol representatives. Petunias' families and remaining TEs are indicated by different colors.











**Table 1:** Number of sequences per taxa, per RLC family, and nucleotide diversity for all RLC families.

	cpTork1	peTork1	peTork2	peTork3	peTork4	peTork5	peTork6	peTork7	peTork8	peTork9	Sequences per taxa
<i>P. altiplana</i>					5	1	1	2	1		10
<i>P. axillaris axillaris</i>		6									6
<i>P. axillaris parodii</i>		16									16
<i>P. bajeensis</i>				3	2	4	5				14
<i>P. bonjardinensis</i>				2	2			4			8
<i>P. exserta</i>		4	1								5
<i>P. exserta vs. P. axillaris</i>		6					1				7
<i>P. inflata</i>			1		4		6		2		13
<i>P. integrifolia integrifolia</i>				1			3				4
<i>P. integrifolia depauperata</i>			1	2	2		2	1			8
<i>P. mantiqueirensis</i>			2	3	4					2	11
<i>P. occidentalis</i>		4									4
<i>P. reitzii</i>				3	1	1	1				6
<i>P. saxicola</i>				3			5				8
<i>P. scheideana</i>				3				1			4
<i>P. secreta</i>		6		1							7
<i>P. interior</i>				1	5		2	1			9
<i>C. excellens</i>				3							3
<i>C. parviflora</i>	1				3						4
<i>C. pygmaea</i>	2			1	2						5
<i>F. imbricata</i>				1							1
Sequences per family*	3	42	5	24	30	6	24	9	3	2	
Total unique sequences											148
Nucleotide diversity ( $\pi$ )	0.087 $\pm 0.026$	0.043 $\pm 0.003$	0.141 $\pm 0.039$	0.018 $\pm 0.003$	0.060 $\pm 0.008$	0.023 $\pm 0.012$	0.065 $\pm 0.008$	0.032 $\pm 0.009$	0.076 $\pm 0.024$	0.215 $\pm 0.107$	

\*numbers may not add up because more than one species can share a same sequence.

**Table 2:** Site collection information of each taxa used in this study.

<b>Taxa</b>	<b>Sample Localities</b>	<b>Voucher</b>	<b>GPS</b>
<i>P. altiplana</i>	Jaquirana, RS, Brazil	ICN 164586	28°88'47"S 50°35'78"W
<i>P. axillaris</i> subsp. <i>axillaris</i>	Minas do Camaquã, RS, Brazil	BHCB 140431	30°12'45"S 59°29'26"W
<i>P. axillaris</i> subsp. <i>parodii</i>	Corrientes, Argentina	BHCB 116997	30°12'45"S 59°29'26"W
<i>P. bajeensis</i>	Bagé, RS, Brazil	BHCB 102128	31°24'05"S 54°08'06"W
<i>P. bonjardinensis</i>	Bom Jardim da Serra, SC, Brazil	BHCB 8116947	28°33'69"S 49°62'47"W
<i>P. exserta</i>	Caçapava do Sul, RS, Brazil	BHCB 117010	30°50'18"S 53°30'38"W
<i>P. exserta</i> vs <i>P. axillaris</i>	Minas do Camaquã, RS, Brazil	BHCB 76031	30°53'29"S 53°23'26"W
<i>P. inflata</i>	Santo Cristo, RS, Brazil	BHCB 14462	27°50'16"S 54°38'03"W
<i>P. integrifolia</i> subsp. <i>integrifolia</i>	Santana do Livramento, Rs, Brazil	BHCB 114610	30°50'40"S 55°26'45"W
<i>P. integrifolia</i> subsp. <i>depauperata</i>	Capão da Canoa, RS, Brazil	BHCB 104853	29°19'48"S 49°42'56"W
<i>P. mantiqueirensis</i>	Camanducaia, MG, Brazil	BHCB 78265	22°39'32"S 45°51'20"W
<i>P. occidentalis</i> *	Urundel, Salta, Argentina, 1988	NA	NA
<i>P. reitzii</i>	Bom Retiro, SC, Brazil	BHCB 38655	25°15'27"S 47°40'33"W
<i>P. saxicola</i>	Otaçílio Costa, SC, Brazil	NA	27°20'45"S 49°30'56"W
<i>P. scheideana</i>	Biturana, PR, Brazil	BHCB 96617	26°15'57"S 51°54'22"W
<i>P. secreta</i>	Caçapava do Sul, RS, Brazil	BHCB 76027	30°53'56"S 53°55'29"W
<i>P. interior</i>	Xaxim, SC, Brasil	NA	26°57'08"S 52°30'42"W
<i>C. excellens</i>	Santa Cecília, SC, Brazil	BHCB 151094	27°03'32"S 50°39'23"W
<i>C. parviflora</i>	Quintão, RS, Brazil	NA	30°20'08"S 50°15'58"W
<i>C. pygmaea</i>	Quaraí, RS, Brazil	BHCB 102112	30°24'394"S 56°24'06"W
<i>F. imbricata</i> **	Quail Bot Gard, 23/10/1988.	UCSB81342	NA

\*herbarium material donated by J. R. Stehmann

\*\*DNA sample donated by R. G. Olmstead;

NA – not available; Brazilian states abbreviations: RS – Rio Grande do Sul; SC – Santa Catarina

## Supporting Information

Figure S1: Example of indels in a family's alignment.

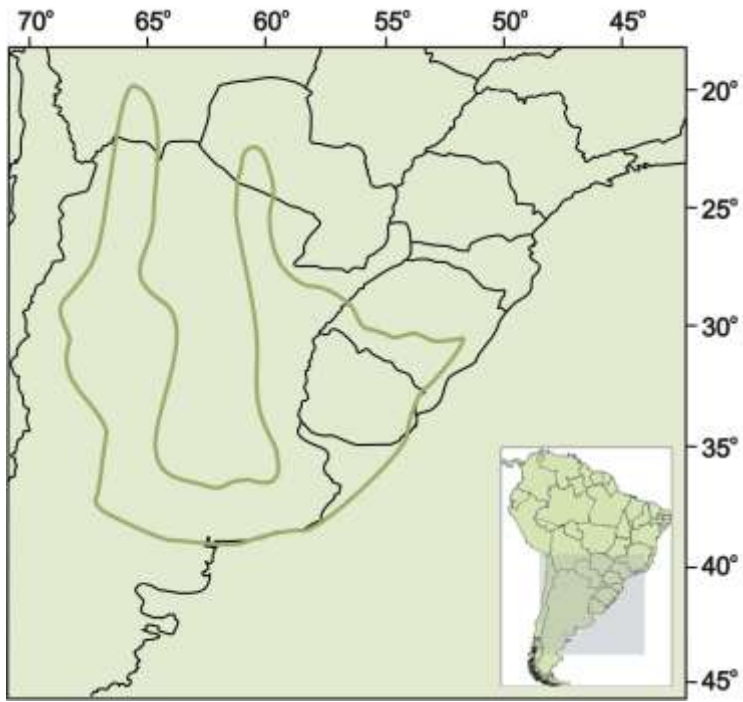
Figure S2: Distribution of RLC\_peTork1 *taxa*.

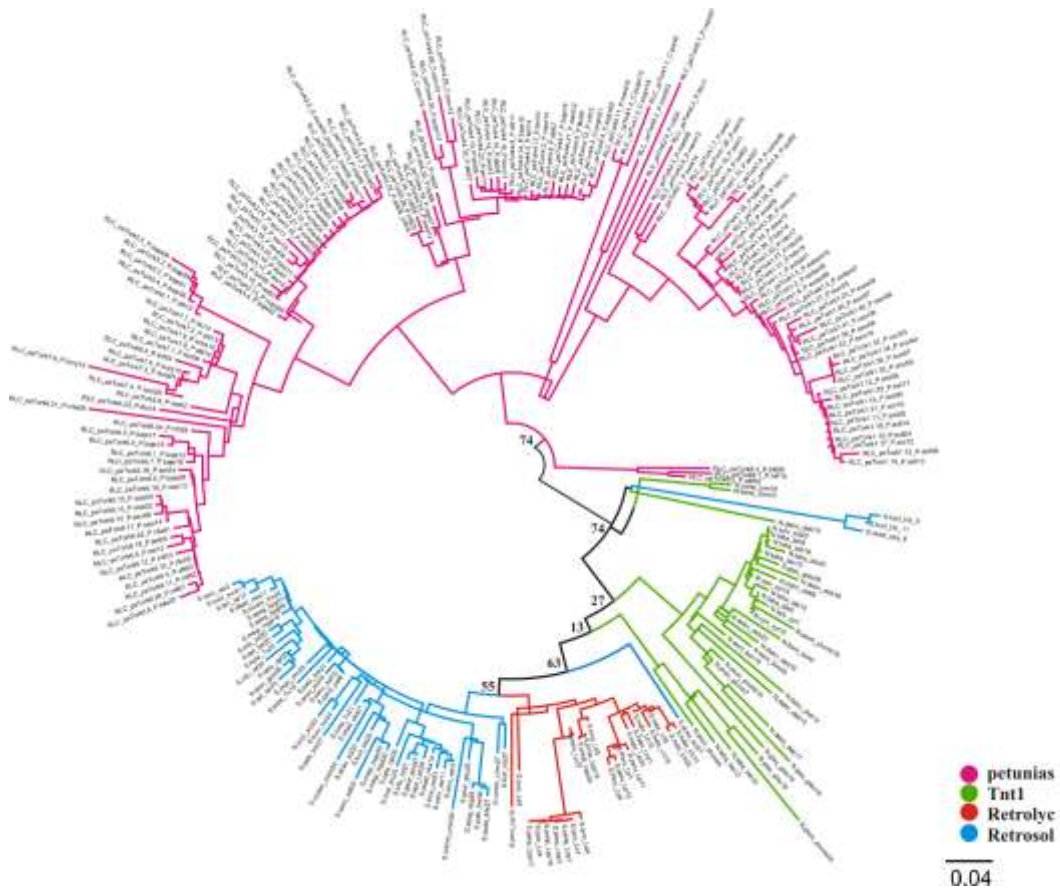
Figure S3: Neighbor-Joining tree of petunias' sequences and Tnt1, Retrolyc1, and Retrosol representatives excluding the U3 region of the alignment. Petunias' and remaining TEs are indicated by different colors.

Figure S4: Neighbor-Joining tree of petunias' sequences and excluding the U5 region of the alignment. Petunias' families are indicated by different colors.

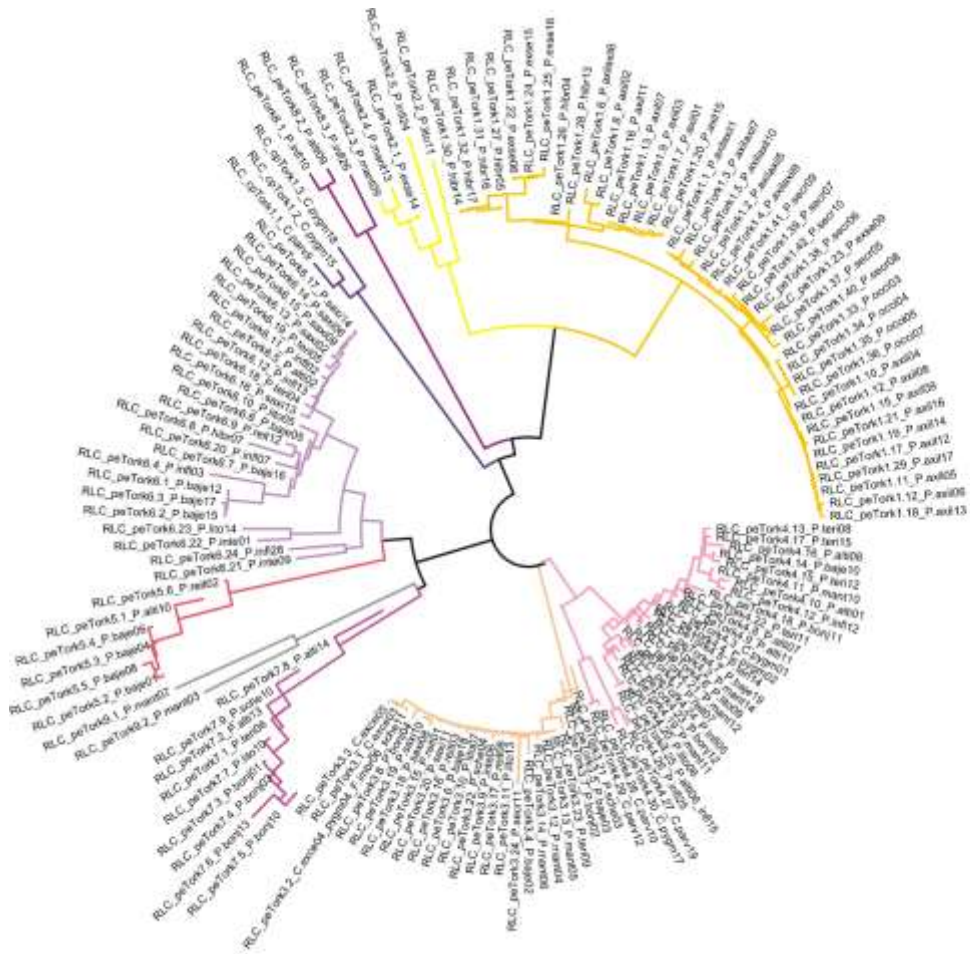
	110	120	130	140	150	160	170	180	190	200	210	220				
<i>RLC_pe Tork3.1_C.exce03/1-332</i>	TGTAAG	TGGATAGGCTCTCCAT	AAAA	CAAGGGCC	TTGTTTT	CATG	AAAA	CAAGGAAA	---	GTGGATAGGCTCTCCACAAAA	CAAGGGCC	TTTTTTTT	CCACAAAAAA	GGC		
<i>RLC_pe Tork3.2_C.exce04_pyygt04_imbr06_sche11/1-333</i>	TGTAAG	TGGATAGGCTCTCCAT	AAAA	CAAGGGCC	TTGTTTT	CATG	AAAA	CAAGGAAA	---	GTGGATAGGCTCTCCACAAAA	CAAGGGCC	TTTTTTTT	CCACAAAAAA	GGC		
<i>RLC_pe Tork3.3_C.exce05/1-332</i>	TGTAAG	TGGATAGGCTCTCCAT	AAAA	CAAGGGCC	TTGTTTT	CATG	AAAA	CAAGGAAA	---	GTGGATAGGCTCTCCACAAAA	CAAGGGCC	TTTTTTTT	CCACAAAAAA	GGC		
<i>RLC_pe Tork3.6_P.baje18/1-335</i>	TGTAAG	TGGATAGGCTCTCCAC	AAAA	CAAGGGCC	TTGTTTT	CATG	AAAA	CAAGGAAA	---	GTGGATAGGCTCTCCACAAAA	CAAGGGCC	TTTTTTTT	CCACAAAAAA	GGC		
<i>RLC_pe Tork3.8_P.bonj04/1-337</i>	TGTAAG	TGGATAGGCTCTCCAC	AAAA	CAAGGGCC	TTGTTTT	CATG	AAAA	CAAGGAAA	---	GTGGATAGGCTCTCCACAAAA	CAAGGGCC	TTTTTTTT	CCACAAAAAA	GGC		
<i>RLC_pe Tork3.9_P.inte06/1-331</i>	TGTAAG	TGGATAGGCTCTCCAC	AAAA	CAAGGGCC	TTGTTTT	CATG	AAAA	CAAGGAAA	---	GTGGATAGGCTCTCCACAAAA	CAAGGGCC	TTTTTTTT	CCACAAAAAA	GGC		
<i>RLC_pe Tork3.10_P.lito07/1-334</i>	TGTAAG	TGGATAGGCTCTCCAC	AAAA	CAAGGGCC	TTGTTTT	CATG	AAAA	CAAGGAAA	---	GTGGATAGGCTCTCCACAAAA	CAAGGGCC	TTTTTTTT	CCACAAAAAA	GGC		
<i>RLC_pe Tork3.11_P.lito13/1-328</i>	TGTAAG	TGGATAGGCTCTCCAC	AAAA	CAAGGGCC	TTGTTTT	---	CATG	AAAA	---	GTGGATAGGCTCTCCACAAAA	CAAGGGCC	TTTTTTTT	CCACAAAAAA	GGC		
<i>RLC_pe Tork3.12_P.mant04/1-334</i>	TGTAAG	TGGATAGGCTCTCCAC	AAAA	CAAGGGCC	TTGTTTT	CATG	AAAA	CAAGGAAA	---	ATGGATAGGCTCTCCACAAAA	CAAGGGCC	TTTTTTTT	CCACAAAAAA	GGC		
<i>RLC_pe Tork3.13_P.mant05/1-309</i>	TGTAAG	TGGATAGGCTCTCCAC	AAAA	CAAGGGCC	TTGTTTT	CATG	AAAA	CAAGGAAA	---	GTGGATAGGCTCTCCACAAAA	CAAGGGCC	TTTTTTTT	CCACAAAAAA	GGC		
<i>RLC_pe Tork3.14_P.mant06/1-335</i>	TGTAAG	TGGATAGGCTCTCCAC	AAAA	CAAGGGCC	TTGTTTT	CATG	AAAA	CAAGGAAA	---	GTGGATAGGCTCTCCACAAAA	CAAGGGCC	TTTTTTTT	CCACAAAAAA	GGC		
<i>RLC_pe Tork3.15_P.reit01/1-335</i>	TGTAAG	TGGATAGGCTCTCCAC	AAAA	CAAGGGCC	TTGTTTT	CATG	AAAA	CAAGGAAA	---	GTGGATAGGCTCTCCACAAAA	CAAGGGCC	TTTTTTTT	CCACAAAAAA	GGC		
<i>RLC_pe Tork3.16_P.reit03/1-334</i>	GGTAAG	TGGATAGGCTCTCCAC	AAAA	CAAGGGCC	TTGTTTT	CATG	AAAA	CAAGGAAA	---	GTGGATAGGCTCTCCACAAAA	CAAGGGCC	TTTTTTTT	CCACAAAAAA	GGC		
<i>RLC_pe Tork3.17_P.reit09/1-331</i>	TGTAAG	TGGATAGGCTCTCCAC	AAAA	CAAGGGCC	TTGTTTT	---	TGAAAA	CAAGGAAA	---	GTGGATAGGCTCTCCACAAAA	CAAGGGCC	TTTTTTTT	CCACAAAAAA	GGC		
<i>RLC_pe Tork3.18_P.saxi08/1-334</i>	TGTAAG	TGGATAGGCTCTCCAC	AAAA	CAAGGGCC	TTGTTTT	CATG	AAAA	CAAGGATA	---	GTGGATAGGCTCTCCACAAAA	CAAGGGCC	TTTTTTTT	CCACAAAAAA	GGC		
<i>RLC_pe Tork3.19_P.saxi10/1-334</i>	TGTAAG	TGGATAGGCTCTCCAC	AAAA	CAAGGGCC	TTGTTTT	CATG	AAAA	CAAGGATA	---	GTGGATAGGCTCTCCACAAAA	CAAGGGCC	TTTTTTTT	CCACAAAAAA	GGC		
<i>RLC_pe Tork3.20_P.saxi11/1-336</i>	TGTAAG	TGGATAGGCTCTCCAC	AAAA	CAAGGGCC	TTGTTTT	CATG	AAAA	CAAGGAAA	---	GTGGATAGGCTCTCCACAAAA	CAAGGGCC	TTTTTTTT	CCACAAAAAA	GGC		
<i>RLC_pe Tork3.22_P.sche04/1-334</i>	TGTAAG	TGGATAGGCTCTCCAC	AAAA	CAAGGGCC	TTGTTTT	CATG	AAAA	CAAGGAAA	---	GTGGATAGGCTCTCCACAAAA	CAAGGGCC	TTTTTTTT	CCACAAAAAA	GGC		
<i>RLC_pe Tork3.23_P.ten09/1-340</i>	TGTAAG	TGGATAGGCTCTCCAC	AAAA	CAAGGGCC	TTGTTTT	CATG	AAAA	CAAGGAAA	AAA	GTGGATAGGCTCTCCACAAAA	CAAGGGCC	TTTTTTTT	CCACAAAAAA	GGC		
<i>RLC_pe Tork3.24_P.secr11/1-337</i>	TGTAAG	TGGATAGGCTCTCCAC	AAAA	CAAGGGCC	TTGTTTT	ATTG	TAA	---	---	GTGGATAGGC	TCCACAAAA	CAAGGGCTTTGTTT	CC	TAAAAACAA	GGGC	TTATTTTT
<i>RLC_pe Tork3.4_P.baje02/1-334</i>	TGTAAG	TGGATAGGCTCTCCAC	AAAA	CAAGGGCC	TTGTTTT	CCAC	AAAA	CAAGGAAA	---	GTGGATAGGCTCTCCACAAAA	CAAGGGCC	TTTTTTTT	CCACAAAAAA	GGC		
<i>RLC_pe Tork3.5_P.baje03/1-281</i>	TGTAAG	TGGATAGGCTCTCCAC	AAAA	CAAGGGCC	TTGTTT	---	---	---	---	---	TCCACAAAA	AAA	GGCTTTG	---	TCC	---
<i>RLC_pe Tork3.7_P.bonj02/1-282</i>	TGTAAG	TGGATAGGCTCTCCAC	AAAA	CAAGGGCC	TTGTTT	---	---	---	---	---	TCCACAAAA	AAA	GGCTTTG	---	TCC	---
<i>RLC_pe Tork3.21_P.sche03/1-282</i>	TGTAAG	TGGATAGGCTCTCCAC	AAAA	CAAGGGCC	TTGTTT	---	---	---	---	---	TCCACAAAA	AAA	GGCTTTG	---	TCC	---











**Table S1:** GeneBank acessions numbers of petunia sequences

Os números de acesso das sequências serão disponibilizados apenas na publicação.

## **CONSIDERAÇÕES FINAIS E PERSPECTIVAS**

Os retrotransposons com LTR são a ordem predominante em plantas, chegando a representar 70% do genoma nuclear de algumas angiospermas (Kalendar *et al.*, 2011). Porém, poucos são conhecidos no gênero *Petunia*. Até o ano de 2009, apenas dois elementos do tipo *Gypsy*, e três do tipo *Copia* foram relatados no genoma de *Petunia* (Richert-Pöggeler & Schwarzacher, 2009). Sendo os retrotransposons geradores de diversidade genética, e também ferramentas para detecção de mudanças genômicas associadas à sua atividade (Kalendar *et al.*, 2011), é surpreendente que mesmo com a importância das petúnias de jardim, ainda sejam escassos os estudos relacionados a retrotransposons nesse gênero.

Este trabalho teve como objetivo a obtenção de retrotransposons com LTR do clado Tork da superfamília *Copia*, relacionados ao elemento Tnt1 de tabaco, em espécies nativas de *Petunia*, *Calibrachoa*, e *Fabiana Ruiz & Pav.*, gêneros proximamente relacionados da família Solanaceae (Olmstead *et al.*, 2009). As diferentes sequências obtidas possibilitaram o estudo do relacionamento e diversidade dessas sequências, e abrem portas para a utilização desses elementos em estudos de diversidade nos gêneros em questão, assim como em outros gêneros relacionados.

Estudos filogenéticos e de biodiversidade são essenciais para a preservação das espécies. Marcadores moleculares baseados em elementos transponíveis são especialmente adequados para estudar as relações filogenéticas e a diversidade genética entre e dentro das espécies. Atualmente existem seis métodos que utilizam marcadores moleculares baseados em retrotransposons (SSAP, IRAP, REMAP, RBIP, TAM, e RUP). Retrotransposons podem ser usados na análise quantitativa da diversificação dos genomas, como no caso em que o diferente número de cópias de *BARE-1* entre cultivares de cevada proporcionou evidências da atividade do elemento e sua influência no genoma (Soleimani *et al.*, 2006). Como cada evento de transposição gera polimorfismos de inserção que podem ser identificados através de técnicas moleculares, a presença ou ausência de um elemento transponível em um dado local do genoma pode ser usada como um marcador molecular para a investigação da diversidade e filogenia de um grupo (Grzebelus, 2006). Tam *et al.* (2009), demonstraram a utilidade de polimorfismos de inserção de Tnt1 e T135 na investigação da diversidade genética e relacionamento inter e intra-específico em espécies de pimenta. A descoberta de novos elementos transponíveis em *Petunia* gera a possibilidade de utilizá-los como marcadores moleculares que possam contribuir na delimitação filogenética das espécies do gênero, assim como em gêneros próximos.

Como demonstrado na literatura (Beguiristain *et al.*, 2001; Kalendar *et al.*, 2000; Pouteau *et al.*, 1991) e também sugerido nesse trabalho, é possível a ocorrência de um determinado tipo de elemento e adaptações a certas condições de estresse. Através de estudos de variação no número de cópias de um elemento em diferentes espécies, assim como a expressão e localização dos elementos nas espécies, a descoberta de uma forte associação entre a presença de certos elementos em espécies de *Petunia* adaptadas a condições específicas, seria de grande valia para os programas de melhoramento genético de espécies de interesse comercial, uma vez que as variedades de petúnias de jardim são provenientes de cruzamentos interespecíficos artificiais (originalmente entre duas espécies nativas do sul do Brasil) e têm grande importância no comércio mundial de plantas ornamentais. Além dos conhecidos híbridos comerciais de petúnias, *Calibrachoa* também é parte importante no comércio de plantas ornamentais, apresentando flores mais delicadas que aquelas de petúnias. Embora algumas espécies tenham baixa quantidade de sequências isoladas, é possível que a distribuição desigual de sequências nos clados do Grupo 2 seja fruto de amplificação ou perda desigual das diferentes famílias de elementos encontrados nesse trabalho. Esta amplificação ou perda pode ser resultante de deriva ou mecanismos de seleção. Porém, a relação de elementos presentes em *taxa* que habitam condições de baixa umidade pode ser mais que apenas coincidência.

*Petunia* é um gênero de história evolutiva recente (Kulcheski *et al.*, 2006; Chen *et al.*, 2007), embora esteja posicionado de forma basal na filogenia da família Solanaceae (Olmstead *et al.*, 2009). O resultado desse trabalho, demonstrando uma possível amplificação diferencial de elementos nos diferentes grupos de espécies, abre a possibilidade para estudos quanto à importância desses elementos na história evolutiva dos gêneros em questão. É interessante que a grande maioria das espécies de flores púrpuras, tanto de *Petunia* quanto *Calibrachoa*, encontram-se agrupadas em um mesmo clado, enquanto as sequências de espécies de flores brancas, rosas, e vermelhas, encontram-se preferencialmente em outro. Esse resultado pode ser fruto de um cenário passado em que o ancestral das diferentes espécies apresentava certa uniformidade genômica quanto à quantidade e qualidade dos elementos transponíveis identificados nesse estudo. Porém, diferenças de habitats e nichos podem ter acarretado em ampliações ou perdas diferenciais desses elementos nas espécies nascentes, resultando no perfil encontrado nas espécies atualmente.

Com a identificação de novos elementos em *Petunia*, relacionados ao retrotransposon Tnt1 de tabaco, e a possibilidade de haver associação entre elementos e condições ambientais das espécies, tem-se como perspectiva o aprofundamento do estudo desses elementos para maior conhecimento e utilização como ferramenta em investigações quanto ao relacionamento interespecífico nos gêneros utilizados nesse trabalho, através da análise de polimorfismos de inserção no genoma das espécies.

A diferença qualitativa de elementos entre as espécies autocompatíveis e as autoincompatíveis levanta questões quanto aos diferentes mecanismos de repressão de elementos no genoma, de forma que a fertilização cruzada permitiria que os elementos se espalhassem pelo genoma (Wright & Finnegan, 2001). Assim, a presença da grande maioria das sequências das espécies autocompatíveis em apenas um clado, gera a possibilidade de melhor investigar o relacionamento do sistema de cruzamento e elementos transponíveis. Dessa forma, para observar uma possível atividade dos elementos, para melhor compreender uma associação destes com condições hídricas, e para investigar a relação dos elementos nas espécies autocompatíveis, são necessários estudos que avaliem o número de cópias e a expressão das diferentes famílias aqui identificadas.

**REFERÊNCIAS BIBLIOGRÁFICAS**  
(capítulos 1 e 4)

- Arillonnet S & Wessler SR (1998) Extreme structural heterogeneity among the members of a maize retrotransposon family. *Genetics* 150: 1245-1256.
- Araujo PG, Casacuberts JM, Costa APP, Hashimoto RY, Grandbastien MA & Van Sluys MA (2001) Retrolyc1 subfamilies defined by different U3 LTR regulatory regions in the *Lycopersicon* genus. *Mol Genet Genomics* 266: 35-41.
- Baucom RS, Estill JC, Leebens-Mack J & Bennetzen JL (2009) Natural Selection on gene function drives the evolution of LTR retrotransposons in the rice genome. *Genome Res* 19: 243-254.
- Beguiristain T, Grandbastien M-A, Puigdomenech P & Casacuberta JM (2001) Three Tnt1 subfamilies show different stress-associated patterns of expression in tobacco. Consequences for retrotransposon control and evolution in plants. *Plant Physiol* 127: 212-221.
- Capy P, Bazin C, Higuete D & Langin T (1998) Dynamics and evolution of transposable elements. Landes Bioscience, Austin, Texas, 197 pp.
- Casacuberta JM, Vernhettes S & Grandbastien MA (1995) Sequence variability within the tobacco retrotransposon Tnt1 population. *EMBO J* 4: 2670-2678.
- Chain PS, Carniel E, Larimer FW, Lamerdin J, Stoutland PO, Regala WM, Georgescu AM, Vergez LM, Land ML, Motin VL, *et al*, (2004) Insights into the evolution of *Yersinia pestis* through whole-genome comparison with *Yersinia pseudotuberculosis*. *Proc Natl Acad Sci USA* 101: 13826-13831.
- Chen S, Matsubara K, Omori T, Kokubun H, Kodama H, Watanabe H, Hashimoto G, Marchesi E, Bullrich L & Ando T (2007) Phylogenetic analysis of the genus *Petunia* (Solanaceae) based on the sequence of the *Hfl* gene. *J. Plant Res.* 120: 385-397.
- Clegg MT & Durbin ML (2000) Flower color variation: a model for the experimental study of evolution. *Proc Natl Acad Sci USA* 97: 7016-7023.
- Costa APP, Scortecci K, Hashimoto RY, Araujo PG, Grandbastien MA & Van Sluys MA (1999) Retrolyc-1, a member of the Tnt1 retrotransposon super-family in the *Lycopersicon peruvianum* genome. *Genetica* 197: 65-72.
- Delprat A, Negre B, Puig M & Ruiz A (2009) The transposon Galileo generates natural chromosomal inversions in *Drosophila* by ectopic recombination. *PLoS One* 4: e7883.



- Finnegan DJ (1989) Eukaryotic transposable elements and genome evolution. *Trends Genet* 5: 103-107.
- Gerats AG, Huits H, Vrijlandt E, Marana C, Souer E & Beld M (1990) Molecular characterization of a nonautonomous transposable element (dTph1) of petunia. *Plant Cell* 2: 1121–1128.
- Gerats T (2009) Identification and Exploitation of Petunia Transposable elements: A Brief History. In: Gerats, T & Strommer J (ed). *Petunia: evolutionary, developmental and physiological genetics*. Springer, New York. pp 365-379.
- Grandbastien MA, Lucas H, Morel JB, Mhiri C, Vernhettes S & Casacuberta JM (1997) The expression of the tobacco Tnt1 retrotransposon is linked to plant defense responses. *Genetica* 100: 241-252.
- Grandbastien MA, Spielman A a& Caboche M (1989) Tnt1, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. *Nature* 337: 376-380.
- Grzebelus D (2006) Transposon insertion polymorphism as a new source of molecular markers. *J. Fruit Ornam. Plant Res.* 14: 21-29.
- Haas BJ, Kamoun S, Zody MC, Jiang RH, Handsaker RE, Cano LM, Grabherr M, Kodira CD, Raffaele S, Torto-Alalibo T, *et al*, (2009) Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* 461: 393-398.
- Hernández-Pinzón I, Jesús E, Santiago N & Casacuberta JM (2009) The frequent transcriptional readthrough of the tobacco Tnt1 retrotransposon and its possible implications for the control of resistance genes. *J Mol Evol* 68: 269-278.
- Hirochika H, Otsuki H, Yoshikawa M, Otsuki Y, Sugimoto K & Takeda S (1996) Autonomous transposition of the tobacco retrotransposon Tto1 in rice. *Plant Cell* 8: 725-734.
- Hua-Van A, Le Rouzic A, Boutin T, Filée J & Capy P (2011) The struggle for life of the genome's selfish architects. *Biol Direct* 6: 19-48.
- Inukai T & Sano Y (2002) Sequence rearrangement in the AT-rich minisatellite of the novel rice transposable element Basho. *Genome* 45: 493-502.
- Kalendar R, Flavell AJ, Ellis THN, Sjakste T, Moisy C & Schlman AH (2011) Analysis of plant diversity with retrotransposon-based molecular markers. *Heredity* 106: 520-530.
- Kidwell MG & Lisch D (1997) Transposable elements as sources of variation in animals and plants. *Proc Natl Acad Sci USA* 94: 7704-7711.
- Kidwell MG & Lisch DR (2001) Perspective: transposable elements, parasitic DNA, and genome evolution. *Evolution Int J Org Evolution* 55: 1-24.

- Kidwell MG (2002) Transposable elements and the evolution of genome size in eukaryotes. *Genetica* 115: 49-63.
- Kulcheski FR, Muschner VC, Lorenz-Lemke AP, Stehmann JR, Bonatto SL, Salzano FM & Freitas LB (2006) Molecular phylogenetic analysis of *Petunia* Juss. (Solanaceae). *Genetica* 126: 3-14.
- Kumar A & Bennetzen JL (1999) Plant Retrotransposon. *Annu Rev Genet* 33: 479-532.
- Le QH, Melayah D, Bonnivard E, Petit M & Grandbastien MA (2007) Distribution dynamics of the Tnt1 retrotransposon in tobacco. *Mol Genet Genomics* 278: 639-651.
- Llorens C, Munoz-Pomer A, Bernard L, Botella H & Moya A (2009) Network dynamics of eukaryotic LTR retroelements beyond phylogenetic trees. *Biol Direct* 4: 41-72.
- Loreto ELS, Carareto CM & Capy P (2008) Revisiting horizontal transfer of transposable elements in *Drosophila*. *Heredity* 100: 545-554.
- Lukens LN & Zhan S (2007) The plant genome's methylation status and response to stress: implications for plant improvement. *Curr Opin Plant Biol* 10: 317-322.
- Manetti ME, Rossi M, Costa APP, Clausen AM, Van Sluys MA (2007) Radiation of the Tnt1 retrotransposon superfamily in three Solanaceae genera. *BMC Evol Biol* 7:34-46.
- Morgante M, Brunner S, Pea G, Fengler K, Zuccolo A & Rafalski A (2005) Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. *Nat Genet* 37: 997-1002.
- Nuzhdin SV (1999) Sure facts, speculations, and open questions about the evolution of transposable element copy number. *Genetica* 107: 129-137.
- Parkhill J, Sebaihia M, Preston A, Murphy LD, Thomson N, Harris DE, Holden MT, Churcher CM, Bentley SD, Mungall KL, *et al.* (2003) Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. *Nat Genet* 35: 32-40.
- Pouteau S, Huttner E, Grandbastien MA & Caboche M (1991) Specific expression of the tobacco Tnt1 retrotransposon in protoplasts. *EMBO J.* 10: 1911-8.
- Raskina O, Barber JC, Nevo E & Belyayev A (2008) Repetitive DNA and chromosomal rearrangements: speciation-related events in plant genomes. *Cytogenet Genome Res* 120: 351-357.
- Richert-Pöggeler KR & Schwarzacher T (2009) Impact of retroelements in shaping the *Petunia* genome. In: Gerats, T & Strommer J (ed). *Petunia: evolutionary, developmental and physiological genetics*. Springer, New York. pp 343-360.

- SanMiguel P, Tikhonov A, Jin Y-K, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, & Bennetzen JL (1996) Nested retrotransposons in the intergenic regions of the maize genome. *Science* 274: 765-768.
- SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y & Bennetzen JL (1998) The paleontology of intergene retrotransposons of maize. *Nat Genet* 20: 43-45.
- Soleimani VD, Baum BR & Johnson DA (2006) Quantification of the retrotransposon BARE1 reveals the dynamic nature of the barley genome. *Genome* 49: 389–396.
- Stehmann JR, Lorenz-Lemke AP, Freitas, LB & Semir, J (2009) The genus *Petunia*. In: Gerats, T & Strommer J. (ed). *Petunia* evolutionary, developmental and physiological genetics. Springer, New York. pp 1-28.
- Tam SM, Mhiri C, Vogelaar A, Kerkveld M, Pearce S & Grandbastien MA (2005) Comparative analyses of genetic diversities within tomato and pepper collections detected by retrotransposon-based SSAP, AFLP and SSR. *Theor Appl Genet* 110: 819–831.
- Tam SM, Lefebvre V, Palloix A, Sage-Palloix AM, Mhiri C & Grandbastien MA (2009) LTR-retrotransposons Tnt1 and T135 markers reveal genetic diversity and evolutionary relationships of domesticated peppers. *Theor Appl Genet* 119: 973-989.
- Ullu E & Tschudi C (1984) *Alu* sequences are processed 7SL RNA genes. *Nature* 312: 171- 172.
- Venner S, Feschotte C & Biemont C (2009) Dynamics of transposable elements: towards a community ecology of the genome. *Trends Genet* 25: 317-323.
- Vernhettes S, Grandbastien MA & Casacuberta JM (1998) The Evolutionary Analysis of the Tnt1 Retrotransposon in *Nicotiana* Species Reveals the High Variability of its Regulatory Sequences. *Mol Biol Evol* 15: 827-836.
- Vicient CM, Kalendar R & Schulman AH (2005) Variability, recombination, and mosaic evolution of the barley *BARE-1* retrotransposon. *J Mol Evol* 61:275-291.
- Vitte C & Panaud O (2005) LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model. *Cytogenet Genome Res* 110: 91-107.
- Wang W, Zheng HK, Fan CZ, Li J, Shi JJ, Cai Z, Zhang G, Liu D, Zhang J, Vang S, *et al*, (2006) High rate of chimeric gene organization by retroposition in plant genomes. *Plant Cell* 18: 1791-1802.
- Watanabe H, Ando T, Iida S, Suzuki A, Buto K, Tsukamoto T, Hashimoto G & Marchesi E (1996) Cross compatibility of *Petunia* cultivars and *P. axillaris* with native taxa of *Petunia* in relation to their chromosome number. *J Jap Soc Hortic Sci* 65: 625-634.

- Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, *et al*, (2007) A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* 8: 973-982.
- Wright S & Finnegan D (2001) Genome evolution: sex and transposable elements. *Curr Biol* 11: R296-R299.
- Xiao H, Jiang N, Schaffner E, Stockinger EJ & van der Knaap E (2008) A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science* 319: 1527-1530.
- Xiao W, Su Y, Sakamoto W, & Sodmergen (2007). Isolation and characterization of Ty1/*Copia*-like retrotransposons in mung bean (*Vigna radiata*). *J Plant Res* 120: 323-328.
- Zedek F, Smerda J, Smarda P, Bures P (2010) Correlated evolution of LTR retrotransposons and genome size in the genus *Eleocharis*. *BMC Plant Biol* 10: 265-275.
- Zhang J, Yu C, Pullettikurti V, Lamb J, Danilova T, Weber DF, Birchler J & Peterson T (2009) Alternative Ac/Ds transposition induces major chromosomal rearrangements in maize. *Genes Dev* 23: 755-765.
- Zufall RA & Rausher MD (2003) The genetic basis of a flower color polymorphism in the common morning glory (*Ipomoea purpurea*). *J Hered* 94: 442-448.